Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.38754

Federated Reinforcement Learning for Traffic Signal Control: Aggregation Strategies under Intersection-Level Heterogeneity

Jikai Tang

Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China
2020jd237t@sjtu.edu.cn

Abstract. Urban traffic signal control must adapt to demand that varies across intersections, time of day, and network topology, and centralized reinforcement learning becomes impractical when every intersection must transmit raw observations to a single learner. Federated reinforcement learning (FRL) relaxes this constraint by training local controllers and exchanging only model updates, but the evidence supporting it is dispersed across incompatible experimental settings. This article reviews ten recent studies of federated and distributed reinforcement learning for traffic signal control, together with a public benchmark for the underlying control task, and organizes them through a common comparison framework covering traffic-control setting, learning paradigm, reinforcement learning algorithm, federation topology, aggregation strategy, data heterogeneity, efficiency, privacy and security, evaluation design, and deployment readiness. The review separates federated methods from distributed and multi-agent reinforcement learning, which coordinate through peer consensus or local interaction rather than a federation server, and it traces a shift from uniform global aggregation toward partial-layer sharing, clustered and hierarchical federation, and personalized combinations of global and local models. The synthesized results indicate that model-level collaboration can accelerate convergence and reduce delay and queue length in several calibrated networks, yet the gains are conditional: fixed-time control remains competitive under light or balanced demand, shared parameters weaken on networks with irregular intersection geometry, and improvements are seldom accompanied by repeated trials, uncertainty estimates, or adversarial privacy evaluation. Progress therefore depends on task-specific aggregation design and on evaluation protocols that hold the traffic scenario constant while the federation configuration is varied.

Keywords: Reinforcement learning, federated reinforcement learning, urban traffic signal control

1. Introduction

Urban traffic congestion is a serious problem. It becomes worse when cities expand and travel demand grows. Conventional traffic signal control methods, such as fixed-time and rule-based strategies, cannot handle traffic conditions that change. Reinforcement Learning (RL) can be a

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.38754

useful method here. It is worth noting that traffic signal controllers can learn adaptive policies through interaction with traffic environment. However, conventional RL-based traffic signal control can face difficulties when this paper applies it to large road networks with many intersections. Centralized approaches require traffic data to be transmitted to central server, which can create communication and computation problems and also bring scalability and privacy concerns and separate agents cannot benefit from knowledge acquired at other intersections. These limitations motivate the use of federated reinforcement learning, which allows intersections to collaborate without transferring raw traffic data to a central server.

Federated Reinforcement Learning (FRL) uses decision-making methods from RL and distributed training from federated learning. In FRL-based traffic signal control system, each intersection can train a local agent using traffic data collected at that location and they share model updates for aggregation so that it is not required to transfer raw traffic data and agents in different locations can learn from each other. In fact, FedLight is an early work that applies this idea to multi-intersection traffic signal control.

These studies show that FRL-based traffic signal control is developing. It is moving from uniform global model aggregation toward scalable architectures that can work across different domains and can be personalized for specific users according to local traffic patterns, which shows the field is making progress and can support more complex traffic scenarios in the future.

Section 2 examines how federated reinforcement learning has been developed for traffic signal control, focusing on system architecture, reinforcement learning algorithms, federated aggregation strategies, the handling of non-IID traffic conditions, communication efficiency, privacy protection, and experimental evaluation. It also compares the assumptions, methods, limitations, and other characteristics of representative studies to identify current research gaps. Section 3 discusses the requirements for deploying FRL-based traffic control systems in large-scale heterogeneous urban networks and outlines directions for future work.

2. Methodology

2.1. Preliminaries of federated reinforcement learning

Federated reinforcement learning combines RL, which optimizes sequential decisions through interaction, with federated learning (FL), which coordinates training across distributed clients. In traffic signal control, an intersection or a group of intersections is modeled as an agent operating in a Markov decision process or a partially observable Markov game. Local observations commonly describe queues, occupancy, speed, waiting time, and signal status; actions select or extend feasible phases; and rewards approximate network objectives such as delay, pressure, queues, or travel time. Each client generates trajectories locally, updates a value function or actor-critic policy, and periodically transmits parameters or gradients to an aggregator. The returned model then initializes the next local-learning interval [1-4]. Raw observations therefore remain local, but the effectiveness of the workflow still depends on the compatibility of state and action spaces, the aggregation interval, the treatment of non-IID traffic, and the information that can be inferred from exchanged updates.

2.2. Federated value based and actor-critic learning

FedLight established an early multi-intersection implementation of federated Advantage Actor-Critic learning [1]. The released code represents each intersection by lane-pressure values and the current

DOI: 10.54254/2755-2721/2026.38754

phase, uses the negative absolute intersection pressure as the local reward, and selects among admissible phases. Actor and critic gradients are averaged across agents every five decision updates and applied to all local networks; after each episode, the implementation also copies the highest-reward agent weights to the other agents. The supplied CityFlow configurations contain synthetic networks with 4, 9, and 16 controlled intersections and a 12-intersection real-road network, with 3,600 simulation steps per episode and 200 episodes. These details make the model-sharing mechanism more reproducible than the paper-level description alone, but they also show that FedLight assumes synchronous participation, compatible network dimensions, and reliable global exchange [1]. Fu and Di subsequently evaluated an A3C-based federated workflow on a calibrated New York City road-network simulation, extending the evidence from regular grids to a realistic urban topology while retaining periodic server coordination [5].

Value-based FRL addresses the same coordination problem through discrete action values. Bao et al. use DQN controllers and separate the neural network into globally shared feature layers and locally retained output layers [2]. A zero-padded, intersection-independent state matrix aligns inputs with different numbers of approaches and lanes, while each client preserves its own phase-dependent action output. The cloud averages only the shared feature-extraction parameters after fixed local-learning intervals; once federation converges, the global layers are frozen and a new intersection fine-tunes its local output layers. This design differs from the distributed DRL comparison of Liu and Ding, which exchanges model information through consensus over a peer graph rather than through a federated server [6]. Server aggregation and peer consensus can both share knowledge without transferring replay samples, but they introduce different assumptions about synchronization, connectivity, and failure recovery.

2.3. Heterogeneity aware aggregation and personalization

Partial aggregation provides a concrete response to architectural heterogeneity. Bao et al. train DQNs with three fully connected feature layers of 64, 128, and 256 units, while the intersection-specific output layer remains local; clients communicate after every 20 training episodes [2]. The reward combines halted-vehicle queues with the waiting time of the lead vehicle in each incoming lane. Experiments use an eight-signal Cologne network for training and a Monaco network containing eight four-way and six three-way intersections for transfer evaluation. The reported tests vary demand across 1,200, 1,800, and 2,400 vehicles per hour and compare independent DQN, fine-tuned DQN, the federated model before local adaptation, and the proposed fine-tuned model. The results support transfer of shared features in these SUMO networks, but they do not establish that the same layer boundary is optimal for other state encodings or control architectures [2].

Li et al. instead federate PPO actor-critic models across independent single-intersection environments with different traffic-flow distributions [3]. Each local state is a 16-dimensional vector containing queue length and lead-vehicle waiting time for eight incoming lanes; the four actions correspond to signal phases, and the reward is the change in network-average waiting time between decisions. The server applies a soft FedAvg update, while experiments vary the local update interval, the retained weight of the preceding global model, and performance-based client weights. The selected configuration uses ten local updates between communication rounds, a global-model retention rate of 0.1, and flexible weights derived from positive local rewards. Under the authors' convergence definition, the final model is reported to converge 47.69% faster than individually trained PPO and 45.35% faster than aggregated PPO, while reducing waiting time by an average of 27.34% across 24 test patterns [3]. However, fixed timing remains competitive or superior under

DOI: 10.54254/2755-2721/2026.38754

low-saturation and approximately balanced flows, indicating that federation does not uniformly improve every operating regime.

Hierarchical FRL limits averaging among dissimilar intersections. Fu et al. formulate local control as a partially observable Markov game and use A2C agents whose states include lane occupancy, queue length, average speed, and phase ratios; a binary action either retains the current phase or advances through a safe predefined sequence [4]. FedClusterLight applies K-means to local model parameters and performs FedAvg within dynamically updated clusters. FedFomoLight instead calculates a personalized weighted combination of other actor-critic models according to local objective improvement and model distance. Evaluation covers 3 by 3 and 5 by 5 synthetic grids and a SUMO network near Columbia University calibrated with TomTom and manually counted traffic, using centralized RL, decentralized RL, FedAvg, fixed-time, actuated, and multi-agent baselines. FedFomoLight is strongest on the calibrated New York network and under light demand on the 5 by 5 grid, whereas centralized RL remains strongest in some smaller or heavily loaded synthetic cases [4]. These results support adaptive grouping rather than a universal hierarchy. Edge-assisted FedProx and density-based weighting address a related form of local-update drift, but should not be treated as equivalent to cluster or model-level personalization [7].

2.4. Communication and real time architectures

Recent work embeds federation within wider communication architectures. RT-FedFlow combines federated multi-agent reinforcement learning, local coordination, a congestion-oriented reward mechanism, and an emergency-vehicle pre-emption module; its SUMO evaluation reports queue length, intersection delay, inference time, fuel consumption, and emergency-clearance outcomes [8]. FedDRL-SDN uses a two-level topology: each junction trains locally, each Ryu-based software-defined networking controller aggregates the junctions in a geographic or administrative subgraph, and a cloud server performs global aggregation across controllers [9]. Each subgraph contains nine junctions, and the paper applies full periodic aggregation to the complete exchanged state-action value structure. These methods make controller placement and aggregation scope part of the learning design rather than treating the communication network as a transparent channel.

In FedDRL-SDN, the local state is a vector of queue lengths for eight incoming lanes, the action selects one of six signal phases, and the reward is the negative waiting time [9]. Although the method is framed as DQN, the exchanged object is described as a Q-table of state-action values. Junction-level values are decoded and merged at an SDN controller, re-encoded for cloud aggregation, and returned to local replay memories for retraining. Quantization reduces the representation from 32-bit values to an 8-bit binary form. Tests use SUMO, Ryu, PyTorch, 3 by 3, 9 by 9, and 12 by 12 grids, and arrival rates from 1,200 to 1,800 vehicles per hour. Relative to DRL and DRL-SDN, the reported waiting-time reductions range from 1.38% to 78.96%, whereas queue-length reductions range from 0.26% to 4.90%, depending on the grid and comparator. Quantization leaves waiting time and queue length nearly unchanged but produces 3% higher mean-square error while reducing simulated transfer latency and bandwidth in the 3 by 3 case [9]. Peer-to-peer consensus [6], edge-assisted aggregation with FedProx and differential privacy [7], and VANET-based adaptive federation [10] represent different placements of coordination and should not be treated as interchangeable implementations of the same architecture.

DOI: 10.54254/2755-2721/2026.38754

2.5. Evaluation protocols and benchmark comparability

Evaluation in this literature follows a shared, though informal, protocol. All reviewed studies report simulated rather than field results, and the larger evaluations rely on established microscopic simulators, with CityFlow used for the synthetic multi-intersection network [1] and SUMO used for the calibrated urban and grid networks [2, 4, 8, 9]. Reported metrics also overlap substantially, concentrating on average waiting time and average queue length, and adding delay, travel time, fuel consumption, inference time, and communication volume in individual studies [2-4, 8, 9, 11]. Baselines are largely shared in kind as well, with conventional fixed-time or actuated control as the reference condition, and with independent or decentralized agents, centralized reinforcement learning, plain FedAvg, and study-specific federated variants as learning comparators [2-4, 9, 11]. What is not shared is the specification of the federation problem itself: the number of clients, the demand levels, the training budget, the local update interval, the repetition policy, and the treatment of randomness differ across studies, and uncertainty is rarely reported [3, 4, 9]. RESCO offers the closest available benchmark for the underlying control task, combining a common SUMO interface, calibrated Cologne and Ingolstadt scenarios, single-signal, corridor, and regional task scales, and fixed-time, max-pressure, Greedy, IDQN, IPPO, MPLight, and FMA2C controllers evaluated over five random seeds with one-standard-deviation bands [11]. Because it is not a federated benchmark, it standardizes the traffic side of FRL evaluation while leaving the federation side, namely client partitions, non-IID demand construction, aggregation schedules, and communication accounting, to be defined.

3. Discussion

3.1. Challenges

The challenge specific to traffic signal control is that data heterogeneity is imposed by the road network rather than produced by a statistical partition of a shared dataset. An intersection observes only its own approaches, and its reward depends on the phase decisions of its neighbours, so local data distributions differ according to geometry, turning demand, and signal coordination, and the locally optimal policy changes as neighbouring controllers continue to learn. Because the client boundary itself, whether an intersection, a corridor, or a subnetwork, is a design decision without an established convention, the same aggregation rule can be evaluated on formulations that are not comparable. The reviewed studies also place their federation designs on very different task definitions. Bao et al. train on eight Cologne signals and transfer to fourteen Monaco intersections using DQN and three demand levels [2]; Li et al. federate PPO across four independent single-intersection environments with 24 arrival-rate combinations and 50 repeated evaluations [3]; Fu et al. compare A2C-based methods on 3 by 3 and 5 by 5 grids and a calibrated New York network [4]; and FedDRL-SDN evaluates increasingly large regular grids under randomly generated demand while reporting neither random seeds nor uncertainty intervals [9]. RESCO additionally shows that a favourable table entry can come from the best training episode rather than from final policy behaviour [11]. As a result, the effect of an aggregation strategy remains entangled with the traffic setting in which it is measured, and no reviewed study isolates the contribution of the federation design from the scenario itself.

The full-text evidence also shows that heterogeneity does not uniformly favor federation. Partial DQN aggregation improves transfer by keeping action-specific layers local [2]. Federated PPO is stable across diverse demand patterns, yet fixed timing performs better when traffic is uniformly

DOI: 10.54254/2755-2721/2026.38754

light, and aggregated PPO can attain slightly lower waiting time when centralized data access is allowed [3]. In the hierarchical study, FedFomoLight outperforms FedAvg and centralized RL on the calibrated New York network, but centralized RL remains strongest on the 3 by 3 grid and under heavy demand on the 5 by 5 grid [4]. RESCO similarly finds that shared control parameters can become unstable when irregular intersections impose dissimilar control tasks, whereas independent learners can remain competitive on realistic networks [11]. Personalization, clustering, and partial sharing should therefore be selected according to the source of heterogeneity and evaluated at each intersection, because an improved network mean can conceal negative transfer.

Privacy, communication, and reproducibility claims require further qualification. Keeping trajectories local reduces direct data disclosure, but FedLight, partial DQN aggregation, federated PPO, hierarchical FRL, and FedDRL-SDN do not evaluate reconstruction, membership inference, poisoning, or Byzantine attacks [1-4, 9]. FedDRL-SDN encodes 32-bit values into an 8-bit representation and reports lower simulated upload and download latency and bandwidth, but compression is reversible and is not a formal confidentiality mechanism. Its quantized variant also has 3% higher mean-square error than the uncompressed variant [9]. The PPO study estimates 51.07 KB per transmission at an assumed link rate, while the hierarchical study reports analytically estimated communication bytes rather than measurements on a deployed network [3, 4]. The FedLight repository further exposes synchronous gradient sharing, end-of-episode best-agent copying, and inconsistent labels for its 3 by 3 and 4 by 4 datasets [1]. These implementation details and the absence of common random-seed and uncertainty reporting limit reproducibility across the literature [9, 11].

3.2. Future prospects

Future work should therefore treat the federation configuration as part of the traffic-control problem rather than as a generic machine-learning choice, and two questions follow from the reviewed evidence. The first concerns client definition. Intersections, corridors, and subnetworks induce different degrees of non-IID data, and the reviewed studies adopt all three without comparing them under a fixed traffic scenario, so controlled experiments that hold the network, demand, and training budget constant while varying only the client partition and the aggregation rule would separate the effect of federation from the effect of scenario [1-5, 9]. The second concerns coupling between neighbouring intersections. Partial-layer sharing and dynamic clustering both restrict averaging, and the transfer from Cologne to Monaco, the clustered actor-critic controllers, and the instability of shared parameters at irregular intersections in RESCO suggest that the appropriate restriction depends on network geometry rather than on demand alone [2, 4, 11]. A benchmark suited to this field would therefore require calibrated networks with variable intersection geometry, explicit client partitions, a defined construction of non-IID traffic, and communication accounting per control decision, so that a federation design can be evaluated independently of the simulator, reward function, and training budget. For each configuration, means, dispersion, final-policy performance, and failure cases should be reported for delay, queues, throughput, travel time, fairness, communication volume, wall-clock training, and decision latency [11].

Aggregation research should treat traffic heterogeneity and unreliable communication jointly. A useful design could combine partial-layer sharing with dynamic clustering while allowing asynchronous or stale-update-aware aggregation when intersections have unequal compute or intermittent links [2, 4]. FedDRL-SDN explicitly identifies controller failure, cloud-link disruption, synchronization delay, and poor sensor data as risks; its fallback is fixed-time operation when communication fails [9]. Future experiments should therefore inject packet loss, controller outages,

DOI: 10.54254/2755-2721/2026.38754

stale updates, sensor corruption, and heterogeneous processing delays, then measure degradation and recovery time rather than assuming continuous connectivity. Client selection and weighting should reflect demand coverage, update quality, staleness, and local performance. Edge, SDN, and VANET architectures provide candidate testbeds for these experiments [7, 9, 10].

Privacy and deployment readiness require explicit threat and safety models. Differential privacy should be evaluated against control degradation and communication cost and combined with secure aggregation, poisoning detection, and Byzantine-robust aggregation under stated attacker capabilities [7]. Quantization should be evaluated as a compression mechanism unless a separate cryptographic or statistical privacy guarantee is demonstrated [9]. Traffic-control tests should enforce minimum greens, yellow and clearance intervals, pedestrian service, emergency priority, and safe fallback policies. Validation should progress from calibrated simulation to controller- or hardware-in-the-loop testing, shadow operation on live feeds, and limited field pilots with human oversight. FedDRL-SDN itself identifies a hardware testbed as future work, confirming that its SUMO and Ryu results are not field evidence [9]. Cross-city transfer among New York, Cologne, Monaco, and Douala would provide stronger evidence than separate simulator improvements [2, 4, 5, 10].

4. Conclusion

The problem that motivated federated reinforcement learning for traffic signal control, coordinating many intersections without centralizing their observations, remains only partly resolved. This article compared ten recent studies of federated and distributed reinforcement learning for traffic signal control through a shared framework of traffic-control setting, learning paradigm, algorithm, federation topology, aggregation strategy, heterogeneity, efficiency, privacy, evaluation design, and deployment readiness, and its main contribution is a structured account of how aggregation choices respond to intersection-level heterogeneity, together with an explicit separation of federated learning from distributed and multi-agent alternatives. The reviewed evidence indicates that model-level collaboration can accelerate convergence and reduce delay and queue length, particularly when shared and local layers are separated or when aggregation is restricted to similar intersections. Those gains are conditional and simulation-bound, however: fixed-time control remains competitive under light demand, shared parameters degrade on irregular networks, and few studies report repeated trials, uncertainty, or adversarial privacy analysis. Future work should therefore treat client definition, the source of heterogeneity, and the communication and privacy assumptions as design variables to be evaluated jointly, and should progress from calibrated simulation toward controller-in-the-loop testing and limited field validation before deployment claims are made.

References

[1] Ye, Y., Zhao, W., Wei, T., Hu, S., & Chen, M. (2021). FedLight: Federated reinforcement learning for autonomous multi-intersection traffic signal control. *Proceedings of the 58th ACM/IEEE Design Automation Conference (DAC)*, 847–852.

[2] Bao, J., Wu, C., Lin, Y., Zhong, L., Chen, X., & Yin, R. (2023). A scalable approach to optimize traffic signal control with federated reinforcement learning. *Scientific Reports, 13*, 19184.

[3] Li, M., Pan, X., Liu, C., & Li, Z. (2025). Federated deep reinforcement learning-based urban traffic signal optimal control. *Scientific Reports, 15*, 11724.

[4] Fu, Y., Zhong, L., Li, Z., & Di, X. (2026). Federated hierarchical reinforcement learning for adaptive traffic signal control. *Transportation Research Part C: Emerging Technologies, 192*, 105894.

[5] Fu, Y., & Di, X. (2023). Federated reinforcement learning for adaptive traffic signal control: A case study in New York City. *Proceedings of the 26th IEEE International Conference on Intelligent Transportation Systems (ITSC)*,

DOI: 10.54254/2755-2721/2026.38754

5738–5743.

[6] Liu, B., & Ding, Z. (2022). A distributed deep reinforcement learning method for traffic light control. *Neurocomputing, 490*, 390–399.

[7] Dayoub, A. J., & Suleiman, E. S. (2026). Federated reinforcement learning for intelligent traffic signal control: A privacy-preserving approach with edge-assisted aggregation. *Proceedings of the 8th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE)*, 1–5.

[8] Akbar, A., Ullah, S. S., Malik, A., & Qaisar, S. M. (2025). RT-FedFlow: An efficient framework for real-time traffic signal optimization using federated multi-agent reinforcement learning. *Engineering Applications of Artificial Intelligence, 161*, 112147.

[9] Kumari, S., Ranga, V., & Meena, J. (2026). FedDRL-SDN: A federated deep reinforcement learning framework with software-defined networking for traffic signal control systems. *Ain Shams Engineering Journal, 17*(10), 104383.

[10] Pidy Pidy, L. T., Ngossaha, J. M., & Tsakou, S. B. (2026). Optimizing traffic light control in developing cities using adaptive federated reinforcement learning in vehicular ad hoc networks. *International Journal of Intelligent Transportation Systems Research*, Advance online publication.

[11] Ault, J., & Sharon, G. (2021). Reinforcement learning benchmarks for traffic signal control. *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, 1.