



City Research Online

City St George's, University of London

Citation: Georgiades, M., Ioannou, I., Hou, Y., Islam, H., Ayaz, F. & Ho, K-H. (2026). Enhanced Vehicular Edge Offloading through SLA-Based Deep Reinforcement Learning and Risk-Aware Bandits. IEEE Transactions on Network and Service Management,

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/38473/>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Enhanced Vehicular Edge Offloading through SLA-Based Deep Reinforcement Learning and Risk-Aware Bandits

Michael Georgiades¹, Iacovos Ioannou², Yun Hou³, Hasan Islam⁴, Ferheen Ayaz⁵ and Kin-Hon Ho⁶

Abstract—This paper investigates service-level agreement (SLA)-aware computation offloading in vehicular edge computing using two complementary online learners. SRM-FA-UCB combines staleness-aware transport exploration, contextual service-time prediction, empirical SLA screening, and selected-arm feedback. SLA-MA-PPO couples a state-dependent SLA action mask with a two-hidden-layer actor-critic, generalized advantage estimation, and cooperative KPI shaping. We evaluate both methods in an extended EdgeCloudSim environment with deterministic SUMO urban-grid mobility, V2X radio, congestion and handover effects, 200–1000 vehicles, and 10 paired replications per load. Among the six primary schemes, SRM-FA-UCB attains the lowest mean SLA violation, service time, and modeled energy with near-unity fairness. Additional controlled attribution with two simple screened heuristics and a screened FA-UCB control shows that the large improvement over conventional FA-UCB is primarily associated with the shared SLA-aware prediction and admissibility-screening stage; no independent endpoint advantage of the more complex post-screen SRM selector is established after Holm correction. SLA-MA-PPO significantly reduces SLA violations, mean service time, and modeled energy relative to PPO-GAE, while ablations identify action masking as its critical component. A controlled change-point experiment does not establish a transient benefit from the restless term. Code, configurations, seeds, and analysis scripts are provided through a public EdgeCloudSim research fork.

Index Terms—Vehicular Edge Computing, Task Offloading, Service Level Agreements, Multi-Armed Bandits, Deep Reinforcement Learning.

I. INTRODUCTION

Connected and autonomous vehicles (CAVs) have evolved into computing-intensive platforms that combine cameras, LiDAR, radar, positioning, communications, and increasingly complex AI workloads. Katare et al. surveyed autonomous-driving Edge AI and emphasized that sensing, computation, wireless data exchange, and energy consumption needed to be

considered jointly in connected-vehicle services [1]. Arthurs et al. likewise showed that connected-vehicle and intelligent-transportation applications faced latency and bandwidth constraints that motivated moving computation from distant cloud infrastructure toward the network edge [2]. Lu and Shi further framed connected vehicles as mobile computing platforms and reported that connected/autonomous vehicles could generate data volumes as high as 19 TB per hour, intensifying the pressure on communication and computing resources [3].

Field and application studies further reinforced this systems perspective. Berger et al. reported more than 4.5 TB of data collected from cameras, GNSS-IMU sensors, and CAN buses during a year-long logistics-truck deployment [4]. Xia et al. identified onboard computation and communication as limiting factors for autonomous-driving services and used roadside edge devices for real-time lane scheduling around crossroads [5]. Naren et al. surveyed computation-resource allocation in IoT-enabled VEC and highlighted the recurring interaction among QoS, power consumption, network stability, and limited onboard resources [6]. Pang et al. formulated multi-area vehicular MEC scheduling as a joint optimization of communication delay, location privacy, and energy consumption and solved it using Double DQN [7]. At the wider connected-car ecosystem level, Pourrahim discussed the growing dependence of vehicles on standardized cellular connectivity and the resulting coordination across IoT automotive value chains [8].

Collectively, these peer-reviewed studies highlighted the role of Vehicular Edge Computing (VEC) in bringing computation closer to vehicles while also exposing offloading decisions to mobility, wireless variation, server load, application deadlines, and energy constraints. These characteristics motivate the adaptive learning-based orchestration considered in this work, since static or purely heuristic policies can struggle to maintain service-level objectives under changing conditions.

End-to-end offloading delay combines access transmission, compute/queueing, backhaul, and return-path effects. Excess delay can violate application deadlines, particularly in safety-critical scenarios, while congestion and mobility make the risk of a given path time varying. Effective orchestration therefore requires online feedback that links each selected action to its realized service outcome.

In this paper, we capitalize on application-specific SLA feedback and online service observations to improve offloading decisions under dynamic vehicular conditions. Based on this principle, two complementary SLA-aware offloading schemes are proposed: SRM-FA-UCB and SLA-MA-PPO, with the aim of reducing deadline violations while balancing service

Michael Georgiades is with the Department of Computer Science, Intelligent Systems Laboratory, Neapolis University Pafos, 8042 Pafos, Cyprus (e-mail: m.georgiades@nup.ac.cy).

Iacovos Ioannou is with the Department of Computer Science, Intelligent Systems Laboratory, Neapolis University Pafos, 8042 Pafos, Cyprus; CYENS Centre of Excellence, Nicosia, Cyprus; and the University of Cyprus (UCY), Nicosia, Cyprus (e-mail: iakovos.ioannou@nup.ac.cy).

Yun Hou is with the Department of Computer Science, Hang Seng University of Hong Kong, Hong Kong SAR, China. (e-mail: aileenhou@hsu.edu.hk).

Hasan Islam is with the Department of Computer Science, East West University, Dhaka, Bangladesh (e-mail: hasan.mahmood@ewubd.edu).

Ferheen Ayaz is with the Department of Computer Science, City St. George's, University of London, London (e-mail: ferheen.ayaz@city.ac.uk).

Kin-Hon Ho is with the Department of Business Administration, Hong Kong Shue Yan University, Hong Kong SAR, China. (e-mail: khho@hksyu.edu).

This work was partly supported by the Faculty Development Scheme (FDS) from the Research Grants Council of Hong Kong under Research Grant No. UGC/FDS14/E07/25.

delay, energy consumption, fairness, and adaptability across heterogeneous network conditions.

The main contributions of the paper are summarized as follows:

- We develop two SLA-aware offloading schemes: SRM-FA-UCB, which combines restless transport optimism, contextual prediction, empirical risk scoring, and SLA admissibility screening; and SLA-MA-PPO, which combines a two-hidden-layer actor-critic with a state-dependent SLA action mask and cooperative KPI shaping.
- We align the mathematical description with the implemented feedback semantics. Outcome-dependent bandit state is updated only for the selected arm, while the risk score and action mask are treated as empirical control mechanisms with explicit fallback rather than calibrated probabilistic safety guarantees.
- We evaluate six algorithms over 200–1000 vehicles using deterministic SUMO mobility, V2X radio, congestion and handover effects, 10 paired replications per load, confidence intervals, component ablations, and a controlled workload change. The analysis covers SLA violations, service-time tails, energy, fairness, placement, utilization, and the behavior of the Hybrid switching mechanism.
- We add a matched controlled-attribution study for SRM-FA-UCB using two simple post-screen heuristics and a screened FA-UCB rule that retain the same complete-service predictor, empirical risk computation, admissibility thresholds, fallback rule, selected-arm feedback semantics, workload realization, and paired seeds. Safe-set cardinality is logged at every decision so that the contribution of SLA-aware prediction/screening can be separated from the complexity of the post-screen selector.

The remainder of this paper is organized as follows. Section II reviews related work. Section III defines the system model and formulates the problem. Section IV presents the two proposed SLA-aware offloading schemes. Section V provides experimental results, followed by a detailed discussion in Section VI. Finally, Section VII concludes the paper and outlines future research directions.

II. RELATED WORK

This section categorizes and discusses the task offloading schemes proposed in Mobile Edge Computing (MEC) and Vehicular Edge Computing (VEC) literature, emphasizing the fundamental learning and optimization mechanisms behind each approach.

A. Vehicular Edge Computing Context for SLA-Aware Offloading

Prior work established the broader vehicular-edge context in which the offloading problem considered in this paper arises. Arthurs et al. [2] surveyed edge-cloud computing for intelligent transportation systems and connected vehicles, showing that moving computation closer to vehicles can alleviate latency and bandwidth limitations, while mobility,

infrastructure heterogeneity, and location-aware resource selection complicate orchestration. Naren et al. [6] reviewed computation-resource allocation in VEC and classified existing approaches according to their optimization objectives, algorithmic methods, and enabling technologies, with recurring emphasis on QoS, energy consumption, network stability, and limited onboard resources. From an autonomous-driving perspective, Katare et al. [1] reviewed approximate Edge AI and highlighted the need to jointly consider sensing, computation, communication, model complexity, and energy efficiency. Lu and Shi [3] introduced the broader Vehicle Computing vision, in which connected vehicles act not only as transportation platforms but also as mobile computing entities that interact with nearby vehicles, edge servers, IoT devices, and remote clouds. These studies collectively motivate the heterogeneous vehicle–edge–cloud decision space adopted in our work and the need for adaptive orchestration under time-varying service conditions.

Several application-oriented studies further illustrated the operational constraints that motivate SLA-aware offloading. Xia et al. [5] proposed an edge-assisted lane-scheduling framework for autonomous vehicles at crossroads, demonstrating the importance of low-latency edge processing for time-sensitive vehicular decisions. Pang et al. [7] formulated collaborative computation scheduling in a multi-area, multi-user, multi-MEC environment and jointly considered communication delay, vehicle-location privacy, and energy consumption using a Double-DQN-based solution. Berger et al. [4] provided complementary field evidence of the data-management burden in connected vehicles by reporting more than 4.5 TB of remotely collected camera, GNSS–IMU, and CAN-bus data from a logistics truck over one year. At the wider connected-car ecosystem level, Pourrahim [8] examined the increasing dependence of connected vehicles on standardized cellular connectivity and the associated data-driven value chain.

These works established the importance of edge proximity, adaptive resource allocation, mobility awareness, communication efficiency, and energy-conscious operation in connected-vehicle systems. However, they did not address the specific problem studied here: online vehicular offloading in which application-specific service-level deadlines are incorporated directly into action admissibility, empirical risk tracking, and learning feedback. Our work builds on this broader VEC literature by focusing on SLA-aware decision making under mobility, congestion, heterogeneous execution paths, and time-varying service conditions.

B. Upper Confidence Bound Approach

In [9], Wang and Zhang propose an Adaptive Task Offloading Algorithm (ATOA) that switches between ϵ -greedy and UCB1 depending on short-term traffic stability. Stability is estimated from the task-size variance v_t over a recent window D ; if $v_t < a$, ATOA exploits with ϵ -greedy, otherwise it uses UCB1 for safer exploration. This dynamic adaptation to “tidal” traffic minimizes per-slot latency across users and achieves faster convergence compared with fixed ϵ -greedy or UCB1. Simulations (with $I = 6$ users, $J = 2$ servers) demonstrate

reduced latency and computational complexity during stable periods.

Wu et al. [10] address MEC task offloading in a collocated 5G/Wi-Fi heterogeneous network by introducing a Fronthaul-Aware UCB (FA-UCB) algorithm. The model integrates multiple radio access technologies and explicitly exploits fronthaul state information to improve the accuracy of delay estimates. The algorithm maintains sublinear regret with respect to the optimal benchmark, and its convergence properties make it suitable for dynamic multi-RAT environments with asymmetric network state information.

C. Bandit-Based Approaches

Li et al. [11] propose a batched framework called Batch-based Multi-user Server Elimination (BMSE) to enhance scalability in large MEC systems. BMSE operates at both user (BMSE-UL) and system (BMSE-SL) levels, applying successive elimination principles to reduce the candidate action space. The dual-layer batching design lowers selection frequency and maintains sublinear regret, significantly improving latency and efficiency under dense user deployments.

Tran and Kim [12] present a Thompson-Sampling-based Bandit Learning for Online Task Offloading (TS-BLOTO) algorithm in dynamic fog computing systems. Each fog node maintains a Beta posterior over binary success outcomes, defined by whether task latency meets a predefined threshold. Posterior sampling ensures adaptive exploration, while exponential forgetting enables quick adaptation to environmental changes. Simulation results confirm that TS-BLOTO reduces average delay compared with ϵ -greedy and UCB baselines.

Hossain et al. [13] develop a microservice-enabled vehicular offloading strategy (MLCO) that integrates input-aware UCB exploration. The exploration term increases for smaller tasks and decreases for larger ones, balancing exploration-exploitation across heterogeneous workloads. Evaluations in highway-scale vehicular scenarios demonstrate reduced failure rate and improved average service time compared with heuristic baselines.

D. Machine Learning and Deep Learning-Based Approaches

Sonmez et al. [14] introduce a two-stage machine learning-based workload orchestrator for VEC. The first stage uses classification models (Naive Bayes, SVM, and MLP) to predict offloading success, while the second stage employs regression to estimate service time. Using EdgeCloudSim simulations, the MLP classifier provided the best overall accuracy, and the hybrid framework outperformed random, moving-average, and game-theoretic baselines in reducing failures and latency.

Yamansavascular et al. [15] propose *DeepEdge*, a Double Deep Q-Network (DDQN)-based orchestrator that adapts to stochastic edge conditions. The model is trained to meet task deadlines under mobility and variable loads, achieving superior performance in task completion rate and resource utilization compared with four classical orchestrators (network-, utilization-, hybrid-, and fuzzy-based).

Mustafa et al. [16] present a Proximal Policy Optimization (PPO)-based deep reinforcement learning scheme for computation offloading in dynamic vehicular networks. The PPO

framework stabilizes updates through surrogate clipping and uses Generalized Advantage Estimation (GAE) for efficient long-term learning. It achieves significant reductions in task delay (up to 19.48%) and dropped task ratio (up to 44%) compared with DQN, DDQN, DRQN, and DA-TODDPG baselines, demonstrating robustness under stringent deadlines and mobility constraints.

E. Limitations of the State of the Art

Despite the diversity of methods above, several critical limitations persist across existing MEC and VEC task offloading schemes:

- **Limited SLA awareness:** Most bandit and DRL methods optimize latency or energy but lack explicit service-level agreement (SLA) feedback loops to regulate deadline or reliability violations dynamically.
- **Absence of unified risk modeling:** Although UCB-based schemes quantify uncertainty via confidence bounds, they do not couple these bounds with real-time risk metrics such as violation probability or transport delay variance.
- **Weak adaptability under non-stationary conditions:** Fixed or batched learning policies (e.g., BMSE, ATOA) may lag behind abrupt changes in vehicular density or wireless congestion.
- **Lack of cooperative or global feedback:** Most bandit algorithms operate independently per node, ignoring inter-vehicle dependencies or shared congestion information available in multi-agent environments.
- **Insufficient integration of SLA screening or fairness mechanisms:** Deep reinforcement learning models such as DDQN and PPO often optimize average reward without explicit deadline-aware action filtering, fairness constraints, or SLA-driven penalties in the learning objective.

These gaps motivate SLA-aware offloading schemes that combine empirical risk tracking, deadline-aware action filtering, and adaptive learning under dynamic vehicular context. The two schemes studied here implement these elements through complementary bandit and policy-optimization mechanisms.

Table I summarizes how representative VEC/MEC offloading methods relate to the above limitations and highlights the specific gaps addressed by the two proposed schemes.

III. SYSTEM MODEL AND PROBLEM FORMULATION

Figure 1 combines the urban mobility setting with the action space used by the orchestrator. Each task is generated by a vehicle and is assigned to one of three execution paths: RSU-hosted edge processing, cloud execution through the RSU/WAN path, or cloud execution through the cellular path. The corresponding radio, congestion, and handover effects enter the service-delay estimates used by the decision logic.

A. Notation Summary

For convenience, the main symbols used in Sections III–IV are listed in Table II.

TABLE I
POSITIONING OF REPRESENTATIVE VEC/MEC OFFLOADING METHODS AND LIMITATIONS ADDRESSED BY THE PROPOSED SCHEMES.

Offloading method	Learning principle	Main strength	Relation to proposed schemes
ATOA [9]	Adaptive ϵ -greedy / UCB	Switches exploration mode according to traffic stability	Lacks explicit SLA-risk modeling, deadline-aware action screening, and global violation feedback
FA-UCB [10]	Fronthaul-aware UCB	Uses radio/fronthaul state to improve delay estimates	Does not incorporate deadline-margin screening, SLA pressure, or cooperative vehicle feedback
BMSE [11]	Batched server elimination	Reduces selection overhead in multi-user MEC	Can react slowly under abrupt vehicular density and congestion changes
TS-BLOTO [12]	Thompson sampling with forgetting	Adapts online to dynamic fog conditions	Uses success probability but does not integrate multi-KPI SLA penalties or tail-risk control
MLCO [13]	Input-aware UCB	Accounts for workload-size variation in VEC	Does not combine transport risk, SLA action masking, and cooperative SLA regulation
DeepEdge [15]	DDQN-based DRL	Learns task orchestration under stochastic edge conditions	Optimizes learned reward without explicit SLA action masking or adaptive violation pressure
PPO offloading [16]	PPO with GAE	Improves task delay and dropped-task ratio in vehicular networks	Does not jointly combine SRM-style admissibility screening, load-aware SLA shaping, and bandit-DRL complementarity
SRM-FA-UCB	Restless/contextual bandit with SLA screening	Combines transport staleness, contextual delay prediction, empirical SLA risk, and selected-arm feedback	Proposed bandit scheme; matched screened controls isolate prediction/screening from post-screen selection, while ablations isolate risk, context, and restless terms
SLA-MA-PPO	Masked multi-agent PPO with GAE	Uses the same SLA telemetry with nonlinear function approximation and global KPI shaping	Proposed DRL scheme; architecture, mask, and global-KPI ablations are evaluated separately

TABLE II
NOTATION SUMMARY

Symbol	Description
t	Decision epoch ($t = 1, 2, \dots$)
$\mathcal{T}_t = (L_t, A_t)$	Task; L_t load/size and A_t application class
$\mathcal{A}, a_t$	Action set and selected action (edge / RSU-cloud / cellular-cloud)
$\mathbf{x}_t$	Observed task/context vector
$S_t(a)$	End-to-end service time under action a (s)
$r_t^{\text{res}}, B, \epsilon_t^{\text{res}}$	Winsorized selected-arm residual observation, its bound ($B = 3$ s), and centered residual innovation
$D_t^{\text{acc}}(a), D_t^{\text{ret}}(a)$	Access and return-path delays (s)
$C_t^{\text{cmp}}(a)$	Compute time under a (s)
$E_t(a)$	Total energy for action a (J)
τ_t	Application deadline / SLA threshold (s)
$b_t(a)$	SLA band: 0 (on-time), 1 (mild), 2 (severe)
$Y_t(a), \ell_t(a)$	Delay ratio $S_t(a)/L_t$ and per-round loss
$T_t(a), U_t^{\text{ctx}}(a)$	Staleness-aware residual-service/path and contextual indices; RLS state ($\theta_t(a), P_t(a)$)
$\hat{s}_t(a)$	Full-service admissibility proxy
$R_t(a), \lambda_t$	Empirical risk score and global SLA-pressure state
$\mathcal{S}_t, M_t^i(a)$	SRM SLA-admissibility set and PPO binary action mask
R_T	Diagnostic cumulative regret versus per-round oracle

B. Entities and Decision Epochs

We consider mobile clients generating computational tasks, an edge infrastructure, and a remote cloud. Decisions occur at

discrete epochs $t = 1, 2, \dots$ upon task arrival. The action set is

$$\mathcal{A} = \{0, 1, 2\},$$

where $a = 0$ denotes execution at the *edge*, and $a = 1, 2$ denote execution at the *cloud* via two alternative *paths/modes*.

C. Task Model

At epoch t , a task is $\mathcal{T}_t = (L_t, A_t)$ with load $L_t > 0$ and application class $A_t \in \{0, 1, 2\}$. An optional deadline/SLA is $\tau_t > 0$.

D. Delay Decomposition

For action $a \in \mathcal{A}$, the end-to-end service time decomposes as

$$S_t(a) = D_t^{\text{acc}}(a) + C_t^{\text{cmp}}(a) + D_t^{\text{ret}}(a), \quad (1)$$

where $D_t^{\text{acc}}(a)$ and $D_t^{\text{ret}}(a)$ denote the uplink (access) and downlink (return) transmission delays, respectively. The compute delay $C_t^{\text{cmp}}(a)$ depends on the selected execution site as follows:

$$C_t^{\text{cmp}}(0) = \frac{L_t}{f_t^{\text{edge}}}, \quad C_t^{\text{cmp}}(1) = C_t^{\text{cmp}}(2) = \frac{L_t}{f_t^{\text{cloud}}}. \quad (2)$$



Fig. 1. Urban-grid vehicular edge-computing environment and the three execution paths implemented in the simulations: (1) vehicle $\rightarrow$ RSU/edge server, (2) vehicle $\rightarrow$ RSU $\rightarrow$ cloud, and (3) vehicle $\rightarrow$ cellular (LTE/5G) $\rightarrow$ cloud. Edge compute is co-located with the RSU; vehicles generate tasks but local on-vehicle execution is not an orchestration action.

Here, L_t represents the required number of CPU cycles to process the task at time t , while f_t^{edge} and f_t^{cloud} denote the available computational capacities (in CPU cycles per second) of the edge server and the remote cloud, respectively. The edge CPU rate f_t^{edge} typically varies with vehicular load and server utilization, whereas the cloud capacity f_t^{cloud} is assumed to be higher and relatively stable but subject to longer backhaul delay. We also use the shorthand $D_t^{\text{path}}(a) = D_t^{\text{acc}}(a) + D_t^{\text{ret}}(a)$ for total network delay.

E. Observations and Context

The decision-maker observes a context vector $\mathbf{x}_t$; a compact choice is

$$\mathbf{x}_t = [1, \tilde{L}_t, \mathbf{1}\{A_t = 0\}, \mathbf{1}\{A_t = 1\}, \mathbf{1}\{A_t = 2\}]^\top, \quad (3)$$

where $\tilde{L}_t$ is a normalization of L_t .

F. Stochastic Assumptions

Raw wireless, queueing, scheduling, and end-to-end service times may be nonstationary and can exhibit heavy-tailed behavior. We therefore do not infer bounded queue length or waiting time from finite link/server resources, and we do not impose a sub-Gaussian model on the raw service-time process $S_t(a)$.

Let $\mathcal{F}_{t-1}$ denote the filtration generated by observations available before the selected-arm outcome at epoch t . For the selected action a_t , the SRM implementation uses the bounded transformed residual

$$\begin{aligned} r_t^{\text{res}} &= \min\left\{B, \max\left\{0, S_t(a_t) - D_t^{\text{path}}(a_t)\right\}\right\}, \\ \varepsilon_t^{\text{res}} &= r_t^{\text{res}} - \mathbb{E}[r_t^{\text{res}} | \mathcal{F}_{t-1}], \quad B = 3 \text{ s}. \end{aligned} \quad (4)$$

Thus $0 \leq r_t^{\text{res}} \leq B$ by construction. Consequently, the centered residual innovation is conditionally bounded and admits the standard conditional variance and moment-generating-function bounds

$$\begin{aligned} \mathbb{E}[(\varepsilon_t^{\text{res}})^2 | \mathcal{F}_{t-1}] &\leq \frac{B^2}{4}, \\ \mathbb{E}[e^{\lambda \varepsilon_t^{\text{res}}} | \mathcal{F}_{t-1}] &\leq \exp\left(\frac{\lambda^2 B^2}{8}\right), \quad \forall \lambda \in \mathbb{R}. \end{aligned} \quad (5)$$

These bounds arise from the explicit winsorization of the transformed selected-arm residual, not from an assumption that the underlying queueing process is bounded. They are used only to motivate the concentration-style restless exploration term. They do not imply queue stability, a bound on raw waiting time, a sub-Gaussian raw service-time distribution, or a probabilistic SLA guarantee for the complete screened policy. Temporal dependence and drift in the conditional mean of the underlying service process remain permitted.

G. Energy Model

Per-action energy decomposes as

$$E_t(a) = E_t^{\text{tx}}(a) + E_t^{\text{rx}}(a) + E_t^{\text{cpu}}(a), \quad (6)$$

with generic proxies

$$\begin{aligned} E_t^{\text{tx}}(a) &= \bar{P}_{\text{tx}}(a) D_t^{\text{acc}}(a), & E_t^{\text{rx}}(a) &= \bar{P}_{\text{rx}}(a) D_t^{\text{ret}}(a), \\ E_t^{\text{cpu}}(a) &= \bar{P}_{\text{cpu}}(a) C_t^{\text{cmp}}(a). \end{aligned} \quad (7)$$

The reported joule values are simulator-derived modeled-energy proxies rather than physical power-meter measurements. Specifically, the implementation applies fixed power coefficients to modeled transmit/receive durations and CPU service time. Accordingly, the evaluation reports modeled energy per recorded outcome; this quantity is used only as a simulation-based comparative metric.

H. SLA and Risk Signals

We define the SLA indicator

$$\mathbb{I}_t^{\text{SLA}}(a) = \mathbf{1}\{S_t(a) > \tau_t\}. \quad (8)$$

and the three-band SLA state

$$b_t(a) = \begin{cases} 0, & S_t(a) \leq \tau_t, \\ 1, & \tau_t < S_t(a) \leq 1.5\tau_t, \\ 2, & S_t(a) > 1.5\tau_t. \end{cases} \quad (9)$$

I. Losses

We use the delay-normalized loss

$$Y_t(a) = \frac{S_t(a)}{L_t}. \quad (10)$$

Representative objectives (each fits within column width) are

$$\ell_t^{\text{lat}}(a) = Y_t(a), \quad (11)$$

$$\ell_t^{\text{lat+en}}(a) = Y_t(a) + \beta \frac{E_t(a)}{L_t}, \quad \beta \geq 0, \quad (12)$$

$$\ell_t^{\text{SLA}}(a) = Y_t(a) + \alpha g(b_t(a)), \quad \alpha \geq 0, \quad (13)$$

with $g(0) = 0 < g(1) < g(2)$.

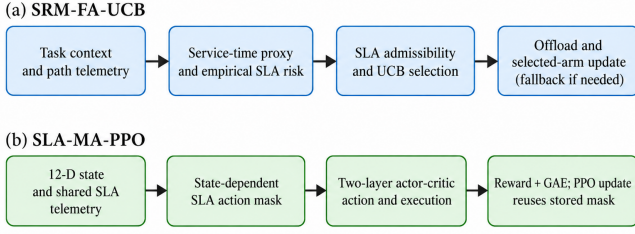


Fig. 2. Compact decision pipelines for the proposed schemes: (a) SRM-FA-UCB and (b) SLA-MA-PPO.

J. Objective and Service-Level Targets

Given a policy π selecting $a_t \in \mathcal{A}$ based on $(\mathbf{x}_t, \text{history})$, the optimization objective is

$$\min_{\pi} \mathbb{E} \left[\sum_{t=1}^T \ell_t(a_t) \right]. \quad (14)$$

A service-level reliability target can be written as

$$\frac{1}{T} \sum_{t=1}^T \Pr(S_t(a_t) \leq \tau_t) \geq 1 - \delta. \quad (15)$$

For diagnostic comparison, regret against a per-round oracle is defined as

$$R_T = \sum_{t=1}^T \ell_t(a_t) - \sum_{t=1}^T \min_{a \in \mathcal{A}} \ell_t(a). \quad (16)$$

Equations (15) and (16) define design targets and diagnostic benchmarks; the complete implemented policies are not claimed to satisfy a calibrated probabilistic safety bound or a closed-form no-regret guarantee.

IV. PROPOSED METHODS: RESTLESS SLA-AWARE BANDIT AND SLA-DRIVEN MULTI-AGENT PPO

We present two complementary schemes for online vehicular offloading under service-level agreements (SLAs): (i) a restless, transport-aware contextual bandit with an empirical SLA admissibility screen, and (ii) an SLA-driven multi-agent PPO with a state-dependent action mask and cooperative KPI shaping. The shared risk signals are operational scores rather than calibrated deadline-miss probabilities, and the screening/masking stages are guardrails rather than hard per-step safety guarantees. The corresponding decision and update loops are given in Algorithms 1 and 2.

A. Scheme 1: SLA-Screened Restless Access-UCB (SRM-FA-UCB)

1) *Design Goals and Intuition:* SRM-FA-UCB combines staleness-aware transport exploration, contextual delay prediction, an empirical SLA-risk/admissibility screen, and selected-arm-only outcome updates. Figure 2(a) summarizes the operational sequence, while Algorithm 1 gives the complete one-pass update loop.

Notation: At round t , choose $a_t \in \mathcal{A}$ (EDGE/RSU/GSM). Let $d_t^{\text{path}}(a)$ denote the currently measured network-path component available before selection, $\hat{b}_t(a)$ the most recent selected-arm residual-service observation for a , $x_t \in \mathbb{R}^d$ the task context, τ_t the SLA for service time, and L_t the task length.

Measured path vs. residual service. The implementation separates the current network-path telemetry from the residual part of realized service time. Specifically, $\hat{b}_t(a)$ is not a fronthaul/backhaul delay estimate: after arm a is selected, its residual observation is formed from total service time minus the measured network-path component, so it represents the remaining compute/queueing and other unmodeled service contribution. Only selected-arm realized outcomes update this residual state; current path-delay telemetry may be observed for all candidate actions before the decision.

2) *Restless Transport Optimism and Contextual Lower Bound:* Let $t_0(a)$ be the last time a was selected and $\{A(n)\}$ a nonincreasing schedule. We define the restless radius

$$r_t^{\text{rm}}(a) = \sqrt{\Xi \log t \sum_{n=t_0(a)+1}^t A(n)^2}. \quad (17)$$

The radius is used as a staleness-sensitive exploration bonus that grows as an arm remains unsampled. Its concentration-style motivation applies to the bounded selected-arm residual innovation in Section III, rather than to raw queueing or end-to-end service time. Because the complete decision rule also contains contextual prediction, empirical risk, normalization, screening, and fallback, this term is not interpreted as an algorithm-level safety, queue-stability, or regret guarantee.

The staleness-aware residual-service/path index (smaller is better) is

$$T_t(a) = d_t^{\text{path}}(a) + (\hat{b}_t(a) - r_t^{\text{rm}}(a)). \quad (18)$$

Thus, $T_t(a)$ combines the currently measured network-path component with an optimistic version of the selected-arm residual-service state. The residual term is shifted downward when an arm becomes stale; it is not itself a fronthaul/backhaul measurement and $T_t(a)$ is not a calibrated confidence bound for complete end-to-end service time. When no residual history exists, the implementation uses the warm-start service proxy.

For context, a per-arm LinUCB/RLS state $(\theta_t(a), P_t(a))$ yields a lower bound on normalized delay,

$$U_t^{\text{ctx}}(a) = x_t^\top \theta_t(a) - \alpha \sqrt{x_t^\top P_t(a) x_t}. \quad (19)$$

The RLS updates follow standard RLS recursions and the corresponding elliptical-potential confidence control; operationally, $(\theta_t(a), P_t(a))$ predicts $y_t = S_t/L_t$ from x_t , and subtracting the confidence term yields an optimistic lower bound whose width is set by α .

3) *SLA Signals, Risk, and Admissibility Set:* Let $\bar{b}_t(a) \in [0, 1]$ be an EWMA of a triple-zone SLA band (green/yellow/red $\mapsto$ 0/0.5/1), $\bar{v}_t(a) \in [0, 1]$ an EWMA of deadline violations, and $\hat{p}_t^{\text{succ}}(a) \in (0, 1)$ a discounted on-time score. A global multiplier $\lambda_t \geq 0$ tracks overall SLA pressure. The admissibility proxy $\hat{s}_t(a)$ is the current

measured network-path component plus the learned selected-arm residual-service component defined consistently with Eq. (24); before arm-specific residual observations exist, the analytical compute estimate supplies the warm start. This complete-service proxy is intentionally distinct from the optimistic index $T_t(a)$. With weights $w_b, w_g, w_d, w_\tau \geq 0$, define

$$R_t(a) = w_b \bar{b}_t(a) + w_g \lambda_t \bar{v}_t(a) + w_d (1 - \hat{p}_t^{\text{succ}}(a)) + w_\tau \mathbf{1}\{\hat{s}_t(a) > \rho_\tau \tau_t\} \quad (20)$$

$R_t(a)$ is an empirical composite risk score. The SLA-band EWMA, violation EWMA, and discounted success memory are not calibrated probabilities; the final indicator supplies an explicit penalty when the full-service proxy exceeds the tightened deadline margin.

The SLA-admissibility set for risk screen $\eta \in [0, 1]$ is

$$\mathcal{S}_t = \{a \in \mathcal{A} : R_t(a) \leq \eta \wedge \hat{s}_t(a) \leq \rho_\tau \tau_t\}. \quad (21)$$

Membership in $\mathcal{S}_t$ denotes admissibility under the implemented deterministic guardrail, not a bound on deadline-exceedance probability. When no arm passes the screen, the algorithm falls back to the full action set and the fallback is explicitly counted.

4) *Score and Selection*: With $w_T, w_U, w_R \geq 0$ and $w_T + w_U + w_R = 1$, let $\mathcal{N}_t[z(a)]$ denote the per-decision min-max normalization of component z across the three candidate arms, i.e., $\mathcal{N}_t[z(a)] = (z(a) - z_t^{\min}) / \max\{z_t^{\max} - z_t^{\min}, \epsilon\}$ with a small $\epsilon > 0$. The implemented weighted score is therefore

$$J_t(a) = w_T \mathcal{N}_t[T_t(a)] + w_U \mathcal{N}_t[U_t^{\text{ctx}}(a)] + w_R \mathcal{N}_t[R_t(a)], \quad (22)$$

and

$$a_t \in \begin{cases} \arg \min_{a \in \mathcal{S}_t} J_t(a), & \mathcal{S}_t \neq \emptyset, \\ \arg \min_{a \in \mathcal{A}} J_t(a), & \text{otherwise.} \end{cases} \quad (23)$$

Scoring and normalization. The evaluated configuration uses fixed decision weights $w_T = 0.50$, $w_U = 0.30$, and $w_R = 0.20$. Equation (22) now states the same per-decision across-arm normalization used by the evaluated implementation, so heterogeneous raw scales of the residual-service/path, contextual, and risk terms do not determine the weighted decision score. No vehicle-load-dependent retuning is applied.

5) *Selected-Arm Online Updates and Complexity*: After executing a_t , SRM-FA-UCB updates outcome-dependent state only for that arm. Let $d_t^{\text{path}}(a_t)$ denote the measured network-path component and define the bounded residual-service observation

$$r_t^{\text{res}} = \min\{3, \max\{0, S_t(a_t) - d_t^{\text{path}}(a_t)\}\} s. \quad (24)$$

Equation (24) makes the implementation semantics explicit: the learned residual is the portion of realized end-to-end service time left after subtracting the measured network-path component. It therefore captures compute/queueing and any remaining unmodeled service overhead; it is not described as a standalone fronthaul/backhaul delay. This is the same

Algorithm 1 SRM-FA-UCB decision/update loop

Input: Per-arm transport/context/risk state and SLA target τ_t

- 1 **for** decision epoch t **do**
- 2 Compute $T_t(a)$, $U_t^{\text{ctx}}(a)$, and $R_t(a)$ for all $a \in \mathcal{A}$;
- 3 Form SLA-admissible set $\mathcal{S}_t$ and use $\mathcal{A}$ only if $\mathcal{S}_t = \emptyset$;
- 4 Select $a_t = \arg \min J_t(a)$ over the active set and execute offloading;
- 5 Observe service/energy/SLA outcome;
- 6 Update the selected-arm residual-service observation, EWMA/dual/DTS statistics, and selected-arm RLS state;

3-s winsorized transformed quantity used in the stochastic discussion of Eqs. (4)–(5); boundedness is therefore introduced algorithmically before the residual-state update and is not asserted for the underlying queueing delay or raw service time. The residual mean is updated incrementally, the selected-arm pull count and last-pull epoch are advanced, and the SLA-band, violation-EWMA, discounted on-time, and contextual states receive the realized outcome. The contextual predictor uses the standard selected-arm RLS recursions

$$P^+ = P - \frac{P x_t x_t^\top P}{1 + x_t^\top P x_t}, \quad (25)$$

$$\theta^+ = \theta + P^+ x_t (S_t/L_t - x_t^\top \theta). \quad (26)$$

The global SLA-pressure state is updated from realized system telemetry. Quantities available before selection, such as current path-delay estimates, are used as context but are not entered as unselected-arm service outcomes. Thus the implementation retains bandit feedback semantics. All scalar updates are $O(1)$ and the RLS step is $O(d^2)$. The fixed SRM weights and screening parameters are given in Table III; no load-specific retuning is used.

6) *Analytical Scope and Implementation Semantics*: The confidence terms in SRM-FA-UCB are motivated by standard contextual concentration arguments under bounded contexts and by concentration of the explicitly winsorized selected-arm residual innovation defined in Eqs. (4)–(5). No corresponding tail assumption is imposed on raw queueing or end-to-end service time. The complete implemented policy also includes a composite empirical risk score, adaptive global pressure, deterministic screening, normalization, and fallback. We therefore do not claim queue stability, a calibrated probabilistic safety guarantee, or a closed-form no-regret guarantee for the complete algorithm. The analytical components motivate the estimator and update structure; their practical contribution is assessed through paired ablations and load-dependent experiments.

B. Scheme II: SLA-Driven Multi-Agent PPO with Action Mask (SLA-MA-PPO)

We develop a DRL counterpart that (i) applies the same SLA-oriented admissibility logic through a state-dependent action mask, (ii) optimizes a cooperative reward shaped by global failure, tail-delay, and utilization KPIs, and (iii) retains per-agent delay, energy, and deadline terms. The MLP actor/critic provides nonlinear function approximation for richer shared and local context. The training loop and masked action selection are given in Algorithm 2.

High-level picture. Figure 2(b) summarizes the PPO sequence. The current SLA mask is stored with each transition and reused in the PPO probability ratio; empty-mask fallback is recorded. Rewards combine local delay/energy/deadline terms with shared failure, utilization, and tail-delay shaping.

1) *Multi-Agent Setting and KPIs:* There are N_t concurrent vehicles indexed by i . Each agent observes s_t^i , selects $a_t^i \in \mathcal{A}$ with policy π_θ , and receives reward r_t^i . Shared KPIs:

$$\text{FailPct}_t, \text{VMUtil}_t, \text{p99Over}_t, \ell_t \in [0, 1].$$

Let SLA targets be FailPct^* , VMUtil^* , τ^{abs} , and a relative factor κ_τ with $\tau_t^{\text{rel}} = \kappa_\tau \hat{s}_t^{\text{est}}$.

KPI semantics. FailPct_t aggregates deadline misses; VMUtil_t captures infrastructure saturation; p99Over_t reflects tail-latency risk. The load proxy ℓ_t (e.g., normalized active-vehicle count) shapes the early bonus to bias learning toward robust behavior in low-to-moderate load.

SLA action mask. With per-arm $R_t^i(a)$ and the same full-service admissibility proxy $\hat{s}_t^i(a)$ used by SRM, define

$$M_t^i(a) = \mathbf{1}\{R_t^i(a) \leq \eta \wedge \hat{s}_t^i(a) \leq \rho_\tau \tau_t^i\}. \quad (27)$$

The mask applies the same deterministic SLA guardrail as SRM; membership is an admissibility decision, not a probabilistic safety statement. The resulting masked policy is

$$\tilde{\pi}_\theta(a|s_t^i) = \frac{\pi_\theta(a|s_t^i) M_t^i(a)}{\sum_{a'} \pi_\theta(a'|s_t^i) M_t^i(a')}, \quad (28)$$

with fallback to π_θ if the denominator is zero. Policy gradients are evaluated on the stored admissible support, with no backpropagation through mask construction.

Mask rationale and implementation. Equation (27) mirrors the bandit admissibility screen: actions exceeding the risk threshold or the margin-tightened deadline are masked. In practice, their logits are set to $-\infty$ (or a large negative) without backpropagating through mask construction, so only currently admissible actions receive probability mass when the set is nonempty. If all actions are masked, the implementation reverts to the unmasked policy to prevent deadlock and records the fallback event.

2) *Per-Agent Reward with SLA Terms:* For chosen a_t^i , let S_t^i be service (s), L_t^i task length, $\hat{s}_t^{\text{est}}$ an estimate (for τ_t^{rel}), and E_t^i the energy (J). Define local ratios

$$\text{DelayRatio}_t^i = \frac{S_t^i}{L_t^i}, \quad \text{EnergyRatio}_t^i = \frac{\max\{0, E_t^i\}}{L_t^i}.$$

With $\beta_E, \beta_\tau > 0$, the local penalty is

$$\mathcal{L}_t^{\text{local}} = \text{DelayRatio}_t^i + \beta_E \text{EnergyRatio}_t^i + \beta_\tau \left(\mathbf{1}\{S_t^i > \rho_\tau \tau_t^{\text{rel}}\} + \mathbf{1}\{S_t^i > \tau^{\text{abs}}\} \right) \quad (29)$$

The cooperative global shaping penalizes KPI exceedances:

$$\begin{aligned} \mathcal{L}_t^{\text{glob}} = & w_F [\text{FailPct}_t - \text{FailPct}^*]_+ \\ & + w_U [\text{VMUtil}_t - \text{VMUtil}^*]_+ \\ & + w_{99} \text{p99Over}_t. \end{aligned} \quad (30)$$

Favor early and low-load phases via

$$\mathcal{B}_t^{\text{early}} = \gamma_{\text{early}} (0.5 - \ell_t)_+. \quad (31)$$

With $\beta_{\text{risk}} > 0$, the SRM risk penalty is $\mathcal{L}_t^{\text{risk}} = \beta_{\text{risk}} R_t^i(a_t^i)$, and the final reward is

$$r_t^i = -\mathcal{L}_t^{\text{local}} - \mathcal{L}_t^{\text{glob}} - \beta_{\text{risk}} R_t^i(a_t^i) + \mathcal{B}_t^{\text{early}}. \quad (32)$$

The local term penalizes delay, energy, and deadline misses; the global term adds shared pressure when failure, utilization, or tail-delay KPIs exceed their targets; the early-load term changes reward shaping when the load proxy is low; and the empirical-risk term exposes the PPO agent to the same SLA telemetry used by SRM. These terms define the implemented reward and are evaluated empirically rather than interpreted as guarantees that the KPI targets are always met.

3) *PPO Objective with Action Mask and GAE:* Let A_t^i be advantages from $\text{GAE}(\gamma, \lambda)$; with $\rho_t^i = \tilde{\pi}_\theta(a_t^i|s_t^i)/\tilde{\pi}_{\theta_{\text{old}}}(a_t^i|s_t^i)$ and entropy weight β_{ent} ,

$$\begin{aligned} \mathcal{J}_{\text{PPO}}(\theta) = & \mathbb{E} \left[\min(\rho_t^i A_t^i, \text{clip}(\rho_t^i, 1 - \epsilon, 1 + \epsilon) A_t^i) \right] \\ & + \beta_{\text{ent}} \mathbb{E} [\mathcal{H}(\tilde{\pi}_\theta(\cdot|s_t^i))]. \end{aligned} \quad (33)$$

The critic minimizes

$$\mathcal{L}_V(\phi) = \frac{1}{2} \mathbb{E} [(V_\phi(s_t^i) - R_t^{i, \text{return}})^2], \quad (34)$$

and the total objective (maximize) is

$$\mathcal{J}(\theta, \phi) = \mathcal{J}_{\text{PPO}}(\theta) - c_V \mathcal{L}_V(\phi). \quad (35)$$

The PPO ratio is evaluated with the exact selection-time mask stored in each transition, so the old and updated policies are compared on the same admissible support. Gradients therefore redistribute probability mass only within that stored support, while entropy on $\tilde{\pi}_\theta$ promotes exploration within it. Stability tricks apply: normalize advantages, clip value updates (or use a clipped value loss), gradient-norm clipping, and KL early stopping.

4) *SLA as a Soft Constraint via Lagrangian Pressure:* Define $g_t = [\text{FailPct}_t - \text{FailPct}^*, \text{VMUtil}_t - \text{VMUtil}^*, \text{p99Over}_t]_+$ and update a nonnegative multiplier by dual ascent

$$\lambda_{t+1} = [\lambda_t + \eta_\lambda \mathbf{1}^\top g_t]_+, \quad (36)$$

or absorb λ_t into (30) via adaptive weights $w_F, w_U, w_{99} \propto \lambda_t$ as a practical primal-dual pressure mechanism. We use this construction as optimization guidance rather than a claim that the non-stationary multi-agent process satisfies the KPI targets asymptotically.

5) *Complexity and Scalability:* Actor-critic updates are $O(|\mathcal{B}| |\mathcal{A}| d_\pi)$ per epoch for batch $\mathcal{B}$ and policy dimension d_π ; the mask adds $O(|\mathcal{A}|)$ per agent for per-action transport/risk evaluation. Compared with the one-pass bandit update, PPO incurs higher optimization cost but supports nonlinear function approximation and shared KPI shaping. The low-load reward term is an implementation choice and is not interpreted as a complexity or convergence guarantee.

6) *Convergence and feasibility scope:* SLA-MA-PPO optimizes the standard clipped PPO surrogate on the stored masked support. Standard PPO/GAE does not provide unconditional monotonic improvement, per-step SLA safety, or vanishing violations in this non-stationary multi-agent setting. The mask is state- and time-dependent and can be bypassed only by

Algorithm 2 SLA-MA-PPO masked rollout/update loop

Input: Shared actor/critic, SLA action mask, KPI targets, rollout size N

- 1 **repeat**
- 2 Observe global KPIs and each agent state;
- 3 Compute per-action transport/risk values and SLA action mask;
- 4 Sample from the masked policy (fallback if all actions are masked);
- 5 Execute, form local/global SLA-aware reward, and store transition;
- 6 **until** buffer contains N transitions;
- 7 Compute GAE advantages; optimize clipped PPO and value losses for E epochs;
- 8 Optionally update projected KPI pressure λ ;

the explicit empty-mask fallback. Accordingly, the manuscript treats the mask and KPI-pressure terms as implemented control mechanisms whose contribution is evaluated empirically rather than as sources of hard guarantees.

Implementation summary. The PPO implementation uses normalized observations, GAE, clipped policy/value updates, gradient clipping, and the same online SLA/risk telemetry used by SRM. The action mask is applied at rollout and inference; if every action is masked, the unmasked policy is used as a deadlock-prevention fallback, the event is recorded, and the reward retains the SLA penalty.

C. Connections and Complementarity

The schemes share SLA telemetry and guardrail logic but differ in optimization style: SRM-FA-UCB uses one-pass per-arm learning, whereas SLA-MA-PPO uses function approximation, GAE, and cooperative KPI shaping. Their matched ablations are evaluated under the same campaign settings.

V. RESULTS

A. Experimental Protocol, Verification, and Reproducibility

We use EdgeCloudSim [17] as the discrete-event foundation for task generation, edge/cloud execution, networking, and orchestration, while extending its modular vehicular interfaces for the conditions required by this study. In particular, deterministic SUMO urban-grid trajectories, an explicit RSU/cellular overlay, V2X radio-path effects, load-dependent congestion, handover interruption, application-specific SLA deadlines, deterministic seed streams, and per-task service tracing are integrated around the native task lifecycle. Where applicable, the scenario follows 3GPP references: TR 37.885 for the V2X urban grid [18], TR 38.901 for the 500-m, three-sector UMa macro geometry [19], and TS 38.331 for Event-A3 handover/TTT [20]; Table III distinguishes these from our evaluated parameter choices. This design retains EdgeCloudSim's reproducible compute/offloading core while exposing both learners to mobility- and path-dependent service outcomes rather than fixed access-delay assumptions.

Six schemes are evaluated at 200, 400, 600, 800, and 1000 vehicles with 10 independent paired replications per load. The root seed is deterministically expanded into environment, workload, and policy streams; within each replication, all policies reuse the same SUMO trace, application mix, infrastructure configuration, and environment/workload realization. Curves report raw unsmoothed replication means with

Student- t 95% confidence intervals, and 1000-vehicle same-family comparisons use paired tests with Holm correction. The harness archives the active configuration, science-critical source files, scripts, and checksums; no replication is removed as an outlier and the minimum task-outcome coverage is 99.86%. For contribution attribution, we additionally evaluate a predicted-complete-service heuristic, a minimum-risk heuristic, and a screened FA-UCB selector under the same SRM predictor/screening formulation and the same five loads and 10 paired seeds. Together with conventional FA-UCB, full SRM-FA-UCB, full SLA-MA-PPO, and its no-mask control, the merged attribution package contains 350 runs (5 loads $\times$ 7 policies $\times$ 10 paired replications). Only the post-screen decision rule differs among the four screened SRM variants; predictor, empirical-risk computation, admissibility thresholds, fallback semantics, and selected-arm update rules are held fixed. The implementation and reproducibility artifacts are maintained in a public research fork of the official EdgeCloudSim repository at <https://github.com/mgeorgiades/EdgeCloudSim-VEC-Extensions>. The fork preserves upstream attribution and includes the simulator extensions, frozen configurations, deterministic seed logic, experiment harness, and analysis scripts used in this study.

Outcome coverage is defined as $N_{\text{out}}/N_{\text{dec}}$. A decision without a terminal completion/failure outcome at the simulation horizon is retained separately as *unfinished-at-end*; across all 300 primary runs, $N_{\text{dec}} = N_{\text{out}} + N_{\text{unfinished}}$ and $N_{\text{out}} = N_{\text{completed}} + N_{\text{failed}}$, so the summary accounting contains no residual unexplained task category. The minimum single-run coverage over the five loads is 99.8974% (ATO), 99.8626% (BMSE), 99.9052% (FA-UCB), 99.9193% (SRM-FA-UCB), 99.9088% (PPO-GAE), and 99.9103% (SLA-MA-PPO); hence the previously quoted 99.86% minimum is the BMSE minimum, not the SRM-FA-UCB coverage. Reported SLA-violation and failure rates use N_{dec} as denominator. For active vehicle i , Jain fairness uses $x_i = N_i^{\text{on-time}}/N_i^{\text{decisions}}$ and $J = (\sum_i x_i)^2 / (n_a \sum_i x_i^2)$, where n_a is the number of vehicles with at least one decision. The implementation increments $N_i^{\text{decisions}}$ at decision time, so unfinished decisions remain represented in the fairness denominator.

B. Load-Dependent Service Performance

Figure 3 shows a consistent separation in the central service metrics. Among the six primary schemes shown in this figure, SRM-FA-UCB maintains the lowest SLA-violation level throughout the sweep (0.039% at 200 vehicles and 0.026% at 1000) together with the lowest mean service time and modeled energy; the additional matched attribution controls are reported separately below. At 1000 vehicles, SLA-MA-PPO records 0.139% SLA violations versus 3.736% for PPO-GAE, mean service time of 0.138 s versus 0.153 s, and mean modeled energy of 0.603 J versus 0.697 J. Its P95 interval is wider at some loads, so the data do not support a uniform tail-latency advantage. BMSE degrades most strongly as density increases, reaching 21.858% SLA violations at 1000 vehicles.

At 1000 vehicles, the conservative unfinished-task sensitivity raises SRM-FA-UCB from 0.02581% reported SLA viola-

TABLE III
SIMULATION AND LEARNING SETTINGS.

Item	Evaluated setting	Item	Evaluated setting
Simulator/actions	EdgeCloudSim + V2X extensions; RSU/edge, RSU-WAN-cloud, cellular-cloud.	Time/load	300 s with 60 s warm-up; 200, 400, 600, 800, 1000 vehicles.
RSU topology	16 RSUs on the 3GPP TR 37.885 urban grid; 433 m $\times$ 250 m unit; 1299 m $\times$ 750 m area [18].	Applications	Traffic/danger/infotainment = 30/35/35%; task lengths 3000/10000/20000; Poisson interarrival 3/5/15 s.
Cellular overlay	7 modeled UMa sites; 500 m inter-site distance; 3 sectors/site [19].	SLA deadlines	0.20/0.30/1.00 s for traffic/danger/infotainment.
Mobility	Deterministic SUMO urban-grid trace; 0.1 s sampling; same trace paired across policies.	Runs/seeds	10 paired replications per load; root seed 20260825 $\rightarrow$ environment/workload/policy streams.
V2I wireless	5.9 GHz, 20 MHz; RSU transmit power 23 dBm; V2X radio/overlay enabled.	Statistics	Raw unsmoothed means; Student- t 95% CIs; paired tests at 1000 vehicles; Holm correction.
Cellular wireless	4.0 GHz, 40 MHz; cellular access modeled separately from RSU access.	PPO	12-D/3 actions; MLP 12 $\rightarrow$ 64 $\rightarrow$ 64 $\rightarrow$ 3 actor, 12 $\rightarrow$ 64 $\rightarrow$ 64 $\rightarrow$ 1 critic; buffer 256; 4 epochs; batch 32; $\gamma = .99$, GAE $\lambda = .95$, clip .20, LR 3×10^{-3} .
Backhaul/congestion	RSU-cloud 8 ms; inter-RSU 1 ms; snapshot equal-share congestion on realized paths.	SRM-FA-UCB	Transport/context/risk = .50/.30/.20; risk threshold .16; $\rho_\tau = .95$; $\Xi = 1.01$, $A_0 = .050$ s, decay .99.
Handover	3GPP Event-A3 [20]: +3 dB offset, 1 dB hysteresis, 160 ms TTT; 50 ms modeled interruption.	Public artifact	Public EdgeCloudSim research fork; repository URL given in Section V.

TABLE IV

HIGH-LOAD COMPARISON AT 1000 VEHICLES (MEANS OVER 10 PAIRED REPLICATIONS). “WORST SLA” IS A CONSERVATIVE SENSITIVITY IN WHICH EVERY UNFINISHED-AT-END TASK IS COUNTED AS AN SLA VIOLATION; “COV.” IS THE MEAN RECORDED-OUTCOME COVERAGE. ENERGY VALUES ARE MODELED PROXIES, NOT PHYSICAL MEASUREMENTS.

Scheme	SLA viol. (%)	Worst SLA (%)	Cov. (%)	Fail. (%)	Mean svc. (s)	P95 svc. (s)	Modeled E (J)	Jain fair.
ATOA	4.3428	4.4032	99.9396	0.0004	0.1580	0.3257	0.7226	0.99562
BMSE	21.8576	21.9501	99.9075	1.9614	0.2391	0.6494	1.0861	0.89967
FA-UCB	4.0613	4.1185	99.9428	0.0002	0.1550	0.3185	0.7070	0.99613
SRM-FA-UCB	0.0258	0.0757	99.9501	0	0.1289	0.3121	0.5409	0.99997
PPO-GAE	3.7356	3.7946	99.9410	0	0.1526	0.3167	0.6968	0.99669
SLA-MA-PPO	0.1394	0.1918	99.9476	0	0.1378	0.3899	0.6027	0.99992

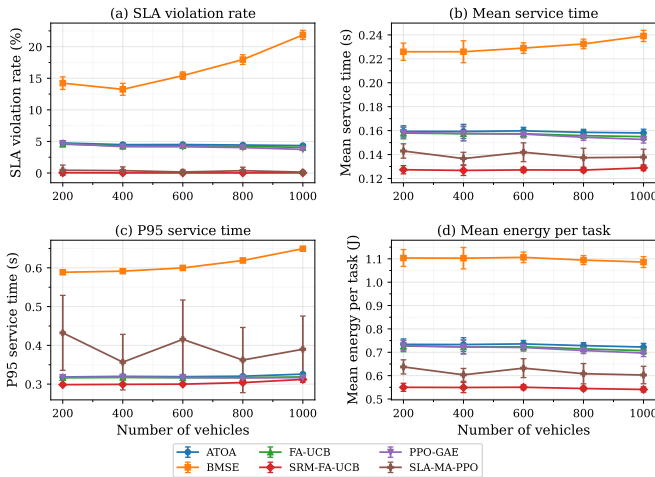


Fig. 3. Service-level performance versus vehicle density: (a) SLA violation rate, (b) mean service time, (c) P95 service time, and (d) mean modeled energy per task. Each point is the mean of 10 paired replications; error bars are 95% Student- t confidence intervals.

tions to 0.07568%, while FA-UCB changes from 4.06131% to 4.11849%. SLA-MA-PPO changes from 0.13944% to 0.19179%, compared with 3.73565% to 3.79464% for PPO-GAE. Thus, even assigning an SLA violation to every unfinished-at-end task does not explain the low reported violation rates of the proposed schemes.

At 1000 vehicles, paired comparisons against the closest same-family baselines remain significant after Holm correction across the five main endpoints. Relative to FA-UCB, SRM-FA-UCB lowers SLA violations by 4.035 percentage points (95% CI $[-4.186, -3.885]$), mean service time by 0.0261 s,

P95 service time by 0.00646 s, and mean modeled energy by 0.166 J, while increasing Jain fairness by 0.00384 (all Holm-adjusted $p < 10^{-12}$). Relative to PPO-GAE, SLA-MA-PPO lowers SLA violations by 3.596 percentage points (Holm $p = 4.37 \times 10^{-10}$), mean service time by 0.0148 s ($p = 9.24 \times 10^{-4}$), and modeled energy by 0.0941 J ($p = 6.47 \times 10^{-4}$), while increasing fairness by 0.00323 ($p = 1.65 \times 10^{-12}$). The P95 difference is not significant after correction ($p = 0.168$).

C. Placement, Resource Use, and Cross-Metric Trade-offs

Figure 4 complements the load-sweep curves with placement and resource-use information. The six policies reach distinct operating points through materially different use of edge, RSU-cloud, and cellular-cloud execution paths. At 1000 vehicles, SRM-FA-UCB occupies the lowest-modeled-energy/lowest-mean-delay region, while SLA-MA-PPO is the next closest learned policy. Across the observed utilization range, both proposed schemes maintain substantially lower SLA-violation levels than their closest same-family baselines. Because utilization is itself influenced by the orchestration policy, these panels are interpreted as operating-point associations rather than causal effects of utilization on SLA performance.

D. Replication-Level Distributions at High Load

Figure 5 makes between-run variation explicit rather than relying only on aggregate means. The violin/point representation shows all ten observed SLA outcomes for each scheme and reveals the markedly lower replication-level violation regime of SRM-FA-UCB and SLA-MA-PPO. The empirical cumulative curves provide a second, distribution-free view of replication-level mean service time at the highest evaluated

TABLE V
COMPLETE QUANTITATIVE COMPONENT ABLATIONS AT 1000 VEHICLES (MEANS OVER 10 PAIRED REPLICATIONS).

Variant	SLA viol. (%)	Mean svc. (s)	P95 svc. (s)	Modeled E (J)	Jain
SRM-FA-UCB family					
SRM-FA-UCB	0.02581	0.12888	0.31207	0.54091	0.999967
SRM no-risk	0.03024	0.12883	0.31178	0.54088	0.999966
SRM no-context	0.02337	0.12880	0.31188	0.54101	0.999968
SRM no-restless	0.02037	0.12889	0.31211	0.54057	0.999968
SLA-MA-PPO family					
SLA-MA-PPO (MLP)	0.13944	0.13785	0.38991	0.60274	0.999924
PPO linear	0.03048	0.13771	0.33236	0.60174	0.999958
PPO no-global-KPI	0.10330	0.13755	0.38948	0.60051	0.999940
PPO no-mask	4.39841	0.15664	0.38737	0.71727	0.994208

Means are computed over the 10 paired replications for each 1000-vehicle ablation profile. Within each family, the environment and workload realization are shared across paired variants. Statistical interpretation in the text uses paired tests with Holm correction over the reported endpoints.

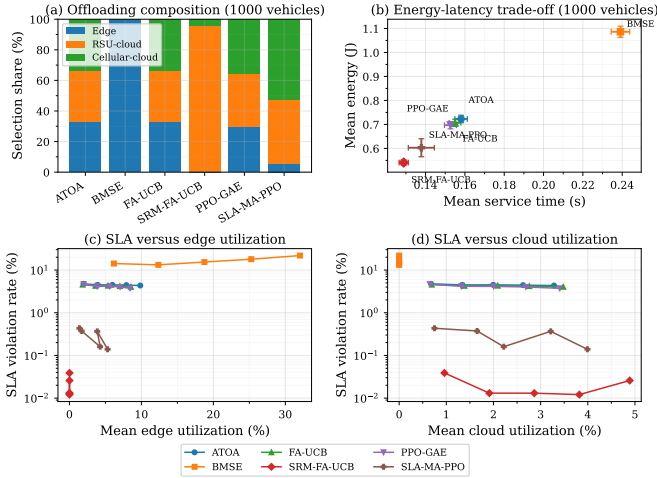


Fig. 4. System-level operating behavior of the six evaluated schemes: (a) edge/RSU-cloud/cellular-cloud selection shares at 1000 vehicles, (b) mean modeled energy versus mean service time at 1000 vehicles, (c) SLA violation rate versus mean edge utilization, and (d) SLA violation rate versus mean cloud utilization. The utilization panels connect each scheme's SLA behavior to the operating points induced across the five evaluated vehicle loads. Error bars in the energy-latency panel denote 95% Student-*t* confidence intervals.

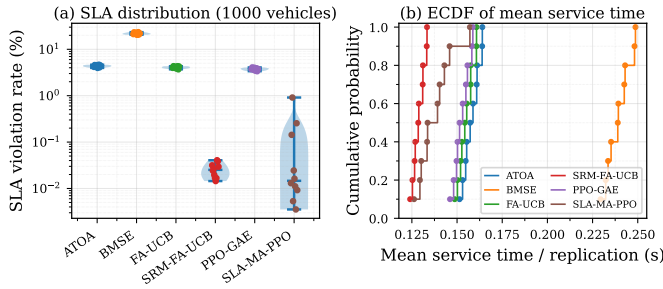


Fig. 5. Replication-level distributions at 1000 vehicles: (a) SLA-violation distribution across the 10 independent replications, with all observed replication values superimposed, and (b) empirical cumulative distribution of the replication-level mean service time. These panels show between-run variation directly and do not represent task-level latency distributions.

load. These views complement the confidence intervals in Fig. 3 and the exact high-load values in Table IV.

E. Component Analysis and Dynamic Conditions

Table V reports the complete steady-regime numerical ablations that were previously summarized only partially. For

SRM-FA-UCB, removing the risk term or the contextual term does not establish an independent endpoint advantage after Holm correction. In contrast, removing the restless term reduces the mean SLA-violation rate from 0.02581% to 0.02037% (paired difference -0.00544 percentage points, 95% CI $[-0.00754, -0.00334]$, family-level Holm $p = 0.00359$) and reduces modeled energy by 0.000342 J (95% CI $[-0.000475, -0.000209]$, family-level Holm $p = 0.00359$). Mean service time, P95 service time, and fairness do not show a corresponding significant change. Thus, the evaluated steady regime exposes a small but statistically detectable exploration cost for the restless term rather than a steady-state endpoint benefit.

For SLA-MA-PPO, neither the linear-vs.-MLP comparison nor removal of the global-KPI term establishes a significant main-endpoint advantage for the full MLP configuration. The action mask is the consequential component: removing it increases SLA violations from 0.13944% to 4.39841%, mean service time from 0.13785 to 0.15664 s, and modeled energy from 0.60274 to 0.71727 J; these three effects remain significant under family-level Holm correction ($p = 0.0129$, 0.00488, and 0.00200, respectively). The P95 and fairness differences do not remain significant under the same family-level correction. These results attribute the principal PPO endpoint gain to SLA-aware action filtering rather than to nonlinear depth or the global-KPI term alone.

The dynamic-condition checks are reported separately because they use different operating points. At 400 vehicles, the Hybrid dynamic and static variants obtain SLA-violation rates of 0.02116% and 0.03502%, respectively, but none of the five paired endpoint differences is significant after Holm correction. Dynamic switching occurs in 2/10 runs versus 10/10 for the static-switch comparator (exact paired $p = 0.0078125$), demonstrating a behavioral switching difference rather than an endpoint-performance advantage. At 1000 vehicles, removing the Hybrid warm start increases the mean SLA-violation rate from 0.04719% to 0.36282%, but none of the five paired endpoints reaches significance after Holm correction. Finally, after the controlled arrival-rate doubling at 180 s, $\Delta_{SLA} = +1.70$ percentage points (95% CI $[-1.55, 4.95]$, $p = 0.267$); hence the present change-point experiment does not establish a transient SLA advantage for the restless term.

1) *Controlled Attribution of Prediction/Screening and Post-Screen Selection:* To isolate the contribution of the more

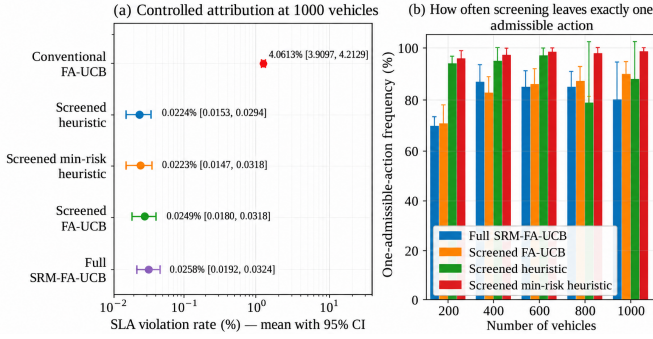


Fig. 6. Controlled contribution attribution and screening restrictiveness. (a) SLA-violation means with 95% Student- t confidence intervals at 1000 vehicles for conventional FA-UCB, two simple screened heuristics, screened FA-UCB, and full SRM-FA-UCB; the horizontal axis is logarithmic. (b) Replication-mean frequency with which the common SRM admissibility screen leaves exactly one action, reported across 200–1000 vehicles for full SRM-FA-UCB, screened FA-UCB, the predicted-service heuristic, and the minimum-risk heuristic; error bars denote 95% Student- t confidence intervals over the 10 paired replications. Panel (a) isolates the large gain associated with introducing the common complete-service predictor and SLA-aware screening machinery, whereas panel (b) quantifies how strongly that screen constrains the downstream selector across load. Conventional FA-UCB is omitted from panel (b) because it does not implement the SRM admissibility screen and therefore has no corresponding $|\mathcal{A}_{\text{safe}}|$ cardinality.

complex SRM learning components, three matched controls are evaluated in addition to conventional FA-UCB and full SRM-FA-UCB. The *predicted-service heuristic* selects the admissible arm with the smallest predicted complete service time, the *minimum-risk heuristic* selects the admissible arm with the smallest empirical risk score, and *screened FA-UCB* applies an FA-UCB-style post-screen rule. All four screened SRM variants use the same complete-service predictor, risk computation, admissibility test, thresholds, fallback behavior, selected-arm feedback semantics, scenario realizations, and paired seeds; only the post-screen selector is changed. Different selectors can nevertheless produce different realized safe-set trajectories because their selected-arm observations update subsequent predictor/risk state.

Figure 6(a) shows that the large improvement relative to conventional FA-UCB appears once the matched full-service predictor and SLA-aware admissibility screen are introduced. At 1000 vehicles, conventional FA-UCB minus screened FA-UCB gives +4.036 percentage points in SLA violations, +0.02605 s in mean service time, +0.2470 s in P99 service time, and +0.1662 J in mean modeled energy; all six reported paired endpoints, including fairness, remain significant after Holm correction ($p_{\text{Holm}} < 10^{-12}$). In contrast, pairwise comparisons among the predicted-service heuristic, minimum-risk heuristic, screened FA-UCB, and full SRM-FA-UCB do not establish an endpoint advantage for the more complex post-screen selector after Holm correction. The dominant empirical gain is therefore associated with the shared complete-service prediction and SLA-aware screening machinery rather than with post-screen selector complexity alone.

Figure 6(b) addresses the reviewer’s question of how often screening leaves only one admissible action. For full SRM-FA-UCB, the replication-mean one-action frequencies are 69.69%, 86.88%, 84.94%, 84.53%, and 80.57% at 200, 400, 600,

800, and 1000 vehicles, respectively. The controlled selectors exhibit method-dependent trajectories despite sharing the same screening formulation because each selected action generates different subsequent selected-arm observations and hence different predictor/risk states. In particular, the minimum-risk heuristic is almost always reduced to one admissible action at high load, whereas full SRM-FA-UCB retains multiple admissible alternatives in roughly one fifth of 1000-vehicle decisions. Separately, the full SLA-MA-PPO mask leaves exactly one action in only about 0.005% of 1000-vehicle decisions, so its substantial no-mask degradation cannot be explained by the mask merely forcing a single choice. These diagnostics support a narrower and more defensible attribution: the SLA-aware prediction/screening stage is the dominant empirical mechanism, while the full transport/context/risk score remains an interpretable rule for choosing among the alternatives that survive screening without a demonstrated independent endpoint gain in the evaluated regime.

VI. DISCUSSION

Across the evaluated loads, application-specific SLA information before execution, followed by learning from realized service outcomes, improves both learning families relative to their closest same-family baselines. At 1000 vehicles, SRM-FA-UCB reduces FA-UCB SLA violations from 4.061% to 0.026%, mean service time from 0.155 to 0.129 s, and modeled energy from 0.707 to 0.541 J, while Jain fairness rises from 0.99613 to 0.99997. SLA-MA-PPO similarly reduces PPO-GAE violations from 3.736% to 0.139%, service time from 0.153 to 0.138 s, and modeled energy from 0.697 to 0.603 J, while fairness rises from 0.99669 to 0.99992.

Across 200–1000 vehicles, low violation rates persist, indicating that the benefit is not confined to one operating point. The new matched attribution controls sharpen the earlier interpretation: conventional FA-UCB is decisively worse than screened FA-UCB, whereas no screened heuristic/FA-UCB/full-SRM comparison establishes an independent endpoint advantage for the more complex post-screen selector after Holm correction. The evidence therefore supports the complete-service predictor plus SLA-aware admissibility screen as the dominant empirical SRM mechanism, with realized selected-arm feedback adapting the state that drives subsequent decisions. No claim is made beyond the evaluated topology and load range.

For PPO, removing the SLA mask raises violations from 0.139% to 4.398% and increases service time and modeled energy, identifying action filtering as a central contributor even though the full PPO mask leaves only one admissible action in approximately 0.005% of 1000-vehicle decisions. For SRM-FA-UCB, the same diagnostic shows a much stronger action-set restriction (80.714% one-action decisions at 1000 vehicles). The controlled results consequently separate two effects that were previously conflated: the SLA-aware predictor/screening machinery produces the dominant endpoint improvement relative to conventional FA-UCB, while the transport/context/risk post-screen score provides a structured, interpretable selector without a demonstrated independent

endpoint gain in this regime. Wider SLA-MA-PPO P95 uncertainty at some loads also precludes uniform tail-latency dominance. Thus, the evidence supports two complementary SLA-aware schemes rather than universal dominance of one learning family.

The complete ablation profiles further narrow the component-level interpretation. Within SRM-FA-UCB, neither risk nor context removal yields a statistically established main-endpoint difference, whereas removing restless exploration slightly but significantly improves steady-regime SLA violation and modeled energy. Within SLA-MA-PPO, neither nonlinear MLP representation nor the global-KPI term shows an independent endpoint advantage, while removing the action mask causes large degradations in SLA violation, mean service time, and modeled energy. The Hybrid experiments similarly distinguish adaptation behavior from endpoint improvement: switching incidence changes significantly, but the measured SLA, delay, energy, and fairness endpoints do not. Accordingly, the manuscript treats these components according to the evidence actually observed rather than attributing the overall gains uniformly to every element of the proposed designs.

EdgeCloudSim [17] provides the compute, load, network, and orchestration core; Table III documents the added standard-referenced SUMO/V2X geometry, congestion, handover, seeded replications, and tracing. These extensions expose both schemes to mobility- and resource-induced service variation. SRM-FA-UCB favors interpretable low-cost updates, whereas SLA-MA-PPO offers richer function approximation at higher training cost. Neither guardrail is a calibrated safety guarantee; site-specific propagation, heterogeneous hardware, signaling overhead, multidimensional security SLAs, and real-testbed validation remain future work.

VII. CONCLUSIONS AND FUTURE WORK

Embedding application-specific SLA information into offloading decisions improves deadline adherence while balancing service time, modeled energy, and fairness. SRM-FA-UCB provides lightweight, interpretable adaptation, whereas SLA-MA-PPO supports richer state spaces through nonlinear function approximation and shared KPI shaping. The controlled attribution study further shows that the dominant SRM gain over conventional FA-UCB is associated with the shared complete-service prediction and SLA-aware screening stage; no independent endpoint advantage of the more complex post-screen selector is established under the evaluated conditions. The complete component ablations identify SLA masking as the empirically consequential PPO component, do not establish independent endpoint gains for SRM risk/context or PPO depth/global-KPI terms, and show a small steady-regime cost for restless exploration; Hybrid switching changes behavior without a demonstrated endpoint advantage. The ablations therefore favor an evidence-based interpretation. Future work will study broader non-stationarity, heterogeneous infrastructure, signaling overhead, 6G-enabled V2X, multidimensional SLAs, and real-V2X or hardware-in-the-loop validation. In addition, the associated EdgeCloudSim extensions will be refined to improve modularity, portability, reproducibility, and reuse, thereby supporting future research in VEC.

REFERENCES

- [1] D. Kataré, D. Perino, J. Nurmi, M. Warnier, M. Janssen, and A. Y. Ding, "A survey on approximate edge AI for energy efficient autonomous driving services," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2714–2754, 2023.
- [2] P. Arthurs, L. Gillam, P. Krause, N. Wang, K. Halder, and A. Mouzakitis, "A taxonomy and survey of edge cloud computing for intelligent transportation systems and connected vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6206–6221, 2022.
- [3] S. Lu and W. Shi, "Vehicle computing: Vision and challenges," *Journal of Information and Intelligence*, vol. 1, no. 1, pp. 23–35, 2023.
- [4] C. Berger, A. Karsolia, F. Giaimo, and O. Benderius, "Over 60,000 km in a year: Remotely collecting large-volume high-quality data from a logistics truck," *SN Applied Sciences*, vol. 4, no. 10, p. 277, 2022.
- [5] C. Xia, X. Jin, L. Kong, C. Xu, and P. Zeng, "Lane scheduling around crossroads for edge computing based autonomous driving," *Journal of Systems Architecture*, vol. 95, pp. 1–8, 2019.
- [6] Naren, A. K. Gaurav, N. Sahu, A. P. Dash, G. S. S. Chalapathi, and V. Chamola, "A survey on computation resource allocation in IoT enabled vehicular edge computing," *Complex & Intelligent Systems*, vol. 8, no. 5, pp. 3683–3705, 2022.
- [7] M. Pang, L. Wang, and N. Fang, "A collaborative scheduling strategy for IoV computing resources considering location privacy protection in mobile edge computing environment," *Journal of Cloud Computing*, vol. 9, no. 1, p. 52, 2020.
- [8] M. Pourrahim, "A new approach to FRAND royalty determination in IoT value chains: Insights from connected cars," *IIC – International Review of Intellectual Property and Competition Law*, vol. 55, no. 7, pp. 1097–1128, 2024.
- [9] L. Wang and J. Zhang, "Adaptive multi-armed bandit learning for task offloading in mobile edge computing," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 5285–5289.
- [10] B. Wu, T. Chen, and X. Wang, "An MAB approach for MEC-centric task-offloading control in multi-RAT HetNets," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.
- [11] H. Li, L. Li, S. Tan, X. Zhong, and S. Zhang, "A multi-user effective computation offloading mechanism for MEC system: Batched multi-armed bandits approach," *IEEE Transactions on Network and Service Management*, 2024.
- [12] H. Tran-Dang and D.-S. Kim, "An efficient bandit learning based online task offloading in fog computing-enabled systems," in *IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2023, pp. 1–5.
- [13] M. D. Hossain, T. Sultana, S. Akhter, M. I. Hossain, G.-W. Lee, C. S. Hong, and E.-N. Huh, "Computation offloading strategy based on multi-armed bandit learning in microservice-enabled vehicular edge computing networks," in *2023 International Conference on Information Networking (ICOIN)*. IEEE, 2023, pp. 769–774.
- [14] C. Sonmez, C. Tunca, A. Ozgovde, and C. Ersoy, "Machine learning-based workload orchestrator for vehicular edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2239–2251, 2020.
- [15] B. Yamansavascular, A. C. Baktir, C. Sonmez, A. Ozgovde, and C. Ersoy, "DeepEdge: A deep reinforcement learning based task orchestrator for edge computing," *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 1, pp. 538–552, 2022.
- [16] E. Mustafa, J. Shuja, F. Rehman, A. Namoun, M. Bilal, and A. Iqbal, "Computation offloading in vehicular communications using PPO-based deep reinforcement learning," *The Journal of Supercomputing*, vol. 81, no. 4, pp. 1–24, 2025.
- [17] C. Sonmez, A. Ozgovde, and C. Ersoy, "EdgeCloudSim: An environment for performance evaluation of edge computing systems," *Transactions on Emerging Telecommunications Technologies*, vol. 29, no. 11, p. e3493, 2018.
- [18] 3rd Generation Partnership Project (3GPP), "Study on evaluation methodology of new vehicle-to-everything (V2X) use cases for LTE and NR," 3GPP, Technical Report TR 37.885 V15.3.0, Jun. 2019.
- [19] —, "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP, Technical Report TR 38.901, 2020.
- [20] —, "NR; radio resource control (RRC); protocol specification," 3GPP, Technical Specification TS 38.331, 2020.



Michael Georgiades (Member, IEEE) received the B.Eng. degree in communications and radio engineering (Hons.) from King's College London, the M.Sc. degree in telecommunications from University College London, the M.Sc. degree in machine learning and data science (Hons.) from Imperial College London, and the Ph.D. degree in wireless and mobile networks from the University of Surrey. He is an Assistant Professor at the Department of Computer Science, Neapolis University Pafos. He is also an Adjunct Professor at the Cyprus University

of Technology and at the Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh. He has also served as an Adjunct Professor at the Open University of Cyprus, a Visiting Professor at the National Technical University of Athens, and an External Mentor at Royal Holloway, University of London. His previous industry and research positions include Research and Development Manager at Primetel PLC and Research Fellow at the University of Surrey. He has participated in numerous EU-funded ICT research projects, is a patent holder, and has contributed to international publications and IETF standards. His research interests include the Internet of Things, applied machine learning, vehicular networks, network security, mobile and wireless communications, explainable AI, and federated learning.



Hasan Mahmood Aminul Islam is an Assistant Professor in the Department of Computer Science and Engineering (CSE) at East West University in Dhaka, Bangladesh. He earned a Doctor of Science degree in Technology (CSE) from the School of Science at Aalto University, Finland, in 2018. Previously, he worked as a Specialist in System-on-Chip (SoC) Software at Nokia headquarters in Espoo, Finland, where he developed Layer 1 Low (L1Low) software drivers for 5G technology from 2019 to 2022. During his doctoral studies, he played a key role in the EU-H2020 POINT Project, which focused on future internet architecture (ICN). For his master's thesis, he collaborated with the Nomadic Laboratory at L.M. Ericsson in Finland and worked closely with the IETF community on data channel protocol options for RTCWeb. His research interests include data communications and distributed systems, Internet protocols, IoT protocols, future Internet architecture, information-centric networking, and delay-tolerant networking.



Iacovos Ioannou (Senior Member, IEEE) received the B.Sc. and Ph.D. degrees in computer science from the University of Cyprus in 2005 and 2021, respectively, and the M.Sc. degree in security and defense from the Open University of Cyprus in 2016. He is currently an Assistant Professor of Computer Science with Neapolis University Pafos, a Senior Researcher with the CYENS Centre of Excellence, a Researcher with GÉANT Europe, and a Scientific Collaborator with the Networks Research Laboratory at the University of Cyprus. He also serves as an Infrastructure and Systems Architect with Areté Strategy AI. He previously held Assistant Professor positions with Philips University and European University Cyprus and industry positions with PrimeTel and the Cyprus Stock Exchange. He has authored or coauthored more than 130 publications, including over 35 peer-reviewed journal articles. His published work includes the book *Distributed Artificial Intelligence for 5G/6G Communications* (CRC Press, 2024). His research interests include 5G/6G and beyond wireless networks, distributed artificial intelligence, machine learning for network optimization and control, Internet of Things security, intrusion detection, federated learning, device-to-device communications, reconfigurable intelligent surfaces, localization, smart grid and mini-grid management, and edge/fog computing. He has contributed to European and national research projects, including ADROIT6G, METACITIES, GÉANT GN5-2, Cybersafe Cyprus, and the National CSIRT-CTIP System. He serves as an Associate Editor for *IEEE Networking Letters* and the *International Journal of Antennas and Propagation*. He has served as a Technical Program Committee Chair and Workshop Chair at international conferences, delivered several keynote presentations, and reviewed more than 800 manuscripts for major IEEE conferences and journals. His teaching experience spans undergraduate and postgraduate courses in computer science, communications, and cybersecurity, along with supervising B.Sc. and M.Sc. theses.

frastructure and Systems Architect with Areté Strategy AI. He previously held Assistant Professor positions with Philips University and European University Cyprus and industry positions with PrimeTel and the Cyprus Stock Exchange. He has authored or coauthored more than 130 publications, including over 35 peer-reviewed journal articles. His published work includes the book *Distributed Artificial Intelligence for 5G/6G Communications* (CRC Press, 2024). His research interests include 5G/6G and beyond wireless networks, distributed artificial intelligence, machine learning for network optimization and control, Internet of Things security, intrusion detection, federated learning, device-to-device communications, reconfigurable intelligent surfaces, localization, smart grid and mini-grid management, and edge/fog computing. He has contributed to European and national research projects, including ADROIT6G, METACITIES, GÉANT GN5-2, Cybersafe Cyprus, and the National CSIRT-CTIP System. He serves as an Associate Editor for *IEEE Networking Letters* and the *International Journal of Antennas and Propagation*. He has served as a Technical Program Committee Chair and Workshop Chair at international conferences, delivered several keynote presentations, and reviewed more than 800 manuscripts for major IEEE conferences and journals. His teaching experience spans undergraduate and postgraduate courses in computer science, communications, and cybersecurity, along with supervising B.Sc. and M.Sc. theses.



Yun Hou (M'05–SM'20) received her Ph.D. degree in Electrical and Electronic Engineering from Imperial College London, UK, in 2009. She is an Associate Professor in the Department of Computer Science at the Hang Seng University of Hong Kong. Before joining HSUHK, she held several teaching and research positions in Hong Kong SAR and the UK, and has been conducting applied research at the Hong Kong Applied Science and Technology Research Institute (ASTRI) as a Senior Lead Engineer for 3G, 4G, and 5G mobile communication

systems. She also has experience working at IBM and Alcatel-Lucent. Her main research interests include V2X, smart cities, 5G, network optimization, and machine learning.



Ferheen Ayaz is a Lecturer at City St. George's, University of London, UK. She received her Ph.D. degree from the University of Sussex, UK, in 2022 on the topic of blockchain security for vehicular communications. She completed her M.Sc. and B.Sc. degrees in electronic engineering from NED University of Engineering and Technology, Pakistan, in 2014 and 2011, respectively. She conducted post-doctoral research at the University of Sussex and the University of Glasgow. Her research interests include energy sustainability through 6G communications, vehicular networks, and security of AI. She is a member of the UK Telecom Innovation Network Academic Strategy Working Group and NetworkEurope. She is a recipient of the N2Women Fellowship and has been recognized as a Global Talent by the Royal Academy of Engineering, UK.



Kin-Hon Ho (M'03–SM'22) earned a BSc in Computer Studies with First Class Honors from City University of Hong Kong, a PGDip in Entrepreneurship from the University of Cambridge, an MSc(Eng) in Data Communications with Distinction from the University of Sheffield, an MSc in Business Analytics from the Hong Kong University of Science and Technology, an MSc in Data Analytics and Artificial Intelligence from Hong Kong Baptist University, and a PhD in Computer Networking from the University of Surrey. He has held various faculty and research

positions in Hong Kong SAR and the UK and is currently an Associate Professor and Associate Head in the Department of Business Administration at Hong Kong Shue Yan University. Dr. Ho has published over 60 papers in international journals, conferences, and book chapters. His research interests include computer networking and financial data analysis. Notably, he received the Best Paper Award at the IFIP/IEEE International Conference on Multimedia and Mobile Networks and Service Management (MMNS) in 2006 and was the runner-up for the Best Paper Award at the IFIP Networking Conference in 2008.