

Article

SA-MAGPPO: Security-Aware Multi-Agent DRL for Edge-Assisted Public Transit Systems in Low-Altitude Intelligent Transportation Environments

Muhammad Mustafa ¹, Ehzaz Mustafa ^{2,*} , Sardar Khaliq uz Zaman ²  and Faisal Rehman ² 

¹ Foothill Transit, West Covina, CA 91790, USA; mmustafa@foothilltransit.org

² Department of Computer Science, Comsats University Islamabad, Abbottabad Campus, Abbottabad 22060, Pakistan; skhaleeq@cuiatd.edu.pk (S.K.u.Z.); frehman@cuiatd.edu.pk (F.R.)

* Correspondence: aizazmustafa@cuiatd.edu.pk

Abstract

Public transit systems play an important role in reducing traffic congestion, energy consumption, and greenhouse gas emissions in smart urban transportation environments. However, large-scale transit operations still suffer from inefficient routing, scheduling, and resource management decisions under highly dynamic traffic and passenger demand conditions. Furthermore, existing intelligent transportation approaches often ignore communication unreliability, anomalous transportation observations, computation offloading overhead, and network congestion in low-altitude intelligent transportation systems. To address these challenges, this paper proposes a secure and resilient UAV-assisted edge-enabled transit offloading framework based on Security-Aware Multi-Agent Proximal Policy Optimization (SA-MAGPPO). In the proposed framework, UAVs operate as low-altitude communication assistants that enhance V2I and V2V connectivity. These UAVs provide aerial traffic observations and reduce communication congestion in dense urban transportation networks. The proposed framework jointly optimizes transit routing, charging scheduling, fleet management, and computation offloading decisions between onboard units and edge servers. Unlike conventional MAGPPO, the proposed framework augments the agent state representation with communication reliability and anomaly information and incorporates security-aware policy learning to improve robustness against unreliable traffic observations, communication disruptions, and unstable network conditions. Extensive simulations under multiple operational scenarios demonstrate that the proposed framework consistently outperforms RP, GS, RBH, SPPO, and conventional MAGPPO approaches in terms of energy, operational efficiency, offloading reliability, and robustness against communication anomalies.

Keywords: edge computing; public transit; computation offloading; proximal policy optimization; intelligent transportation systems



Academic Editor: Paolo Bellavista

Received: 15 June 2026

Revised: 25 July 2026

Accepted: 29 July 2026

Published: 31 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Public transit systems play an important role in reducing traffic congestion, energy consumption, and greenhouse gas emissions in modern urban transportation environments [1]. Large-scale fleet operations improve transportation accessibility and reduce dependence on private vehicles, which ultimately contributes to environmentally sustainable mobility. Moreover, efficient optimization of public transit systems can significantly reduce CO₂

emissions, which remain a major concern in climate change and urban air pollution. However, transit fleet operations still face significant challenges such as fluctuating passenger demand, traffic variations, complex route planning, charging coordination, and dynamic fleet resource management [2,3]. In addition, the increasing integration of edge-enabled intelligent transportation infrastructure and low-altitude communication support introduces new challenges related to communication reliability, network congestion, and anomalous transportation observations in dynamic urban environments.

These complex tasks, such as route planning, passenger demand prediction, fleet scheduling, and real-time operational optimization, may not be efficiently processed on the onboard units of public fleets due to limited computational resources [4]. To overcome these challenges, edge resources may be utilized to provide storage and computational support to vehicles in closer proximity [5,6]. The edge servers integrated with roadside infrastructure provide on-demand services such as intelligent route optimization, traffic analytics, and real-time passenger management using Vehicle-to-Infrastructure (V2I) communication links. Moreover, Vehicle-to-Vehicle (V2V) communication links enable real-time information exchange among transit vehicles for congestion avoidance and traffic-aware coordination [7,8]. Recently, low-altitude intelligent transportation support using Unmanned Aerial Vehicles (UAVs) has also emerged as an effective solution to improve communication coverage, enhance traffic monitoring, and reduce communication congestion in dense transportation environments. Figure 1 shows the visual representation of public transit fleets with edge resources and UAV-assisted communication support. However, traditional scheduling methods, such as rule-based and greedy approaches, fail to adapt to highly dynamic transportation environments with varying traffic conditions, communication instability, and fluctuating passenger demands. This results in poor fleet resource utilization, increased passenger waiting times, communication delays, and higher energy consumption and emissions.

Artificial Intelligence (AI) offers effective approaches to address the above challenges through adaptive and data-driven decision-making in real time [9]. In particular, deep reinforcement learning (DRL) enables intelligent agents to learn optimal policies through continuous interaction with dynamic environments [10,11]. DRL can effectively adapt to stochastic transportation conditions, varying communication reliability, and changing passenger mobility patterns in real time. Furthermore, DRL enables multiple intelligent agents, such as buses, edge controllers, and scheduling entities, to cooperatively optimize routing, path planning, charging schedules, fleet dispatching, and computation offloading decisions based on both historical patterns and real-time environmental observations [12].

Motivated by the above discussion and the limitations of conventional rule-based, greedy optimization approaches and traditional DRL methods, this paper proposes a Security-Aware Multi-Agent Proximal Policy Optimization (SA-MAGPPO) framework for resilient and environmentally sustainable public transit operations. An example of the considered public transit network with V2I communication and UAV-assisted low-altitude connectivity is illustrated in Figure 1. The proposed framework integrates demand forecasting, adaptive transit control, and computation offloading decisions within a unified DRL framework. Specifically, UAVs assist the transportation infrastructure by improving communication reliability, reducing network congestion, and providing aerial traffic observations for real-time fleet coordination. The proposed SA-MAGPPO framework jointly optimizes routing, charging scheduling, fleet management, and computation offloading decisions between onboard units and edge servers. In addition, anomaly-aware and reliability-aware learning mechanisms are incorporated to improve robustness against unreliable transportation observations and communication disruptions. Unlike conventional MAGPPO, where agents primarily optimize decisions based on traffic and operational

observations, the proposed SA-MAGPPO further augments each agent's observation space with communication reliability and anomaly-related information obtained from the transportation environment. Consequently, the learned policy is conditioned not only on traffic dynamics and passenger demand but also on network reliability and communication quality. This enables cooperative agents to proactively avoid unreliable communication conditions while jointly optimizing routing, scheduling, charging, and computation offloading decisions, thereby improving the robustness of fleet operations under dynamic urban environments.

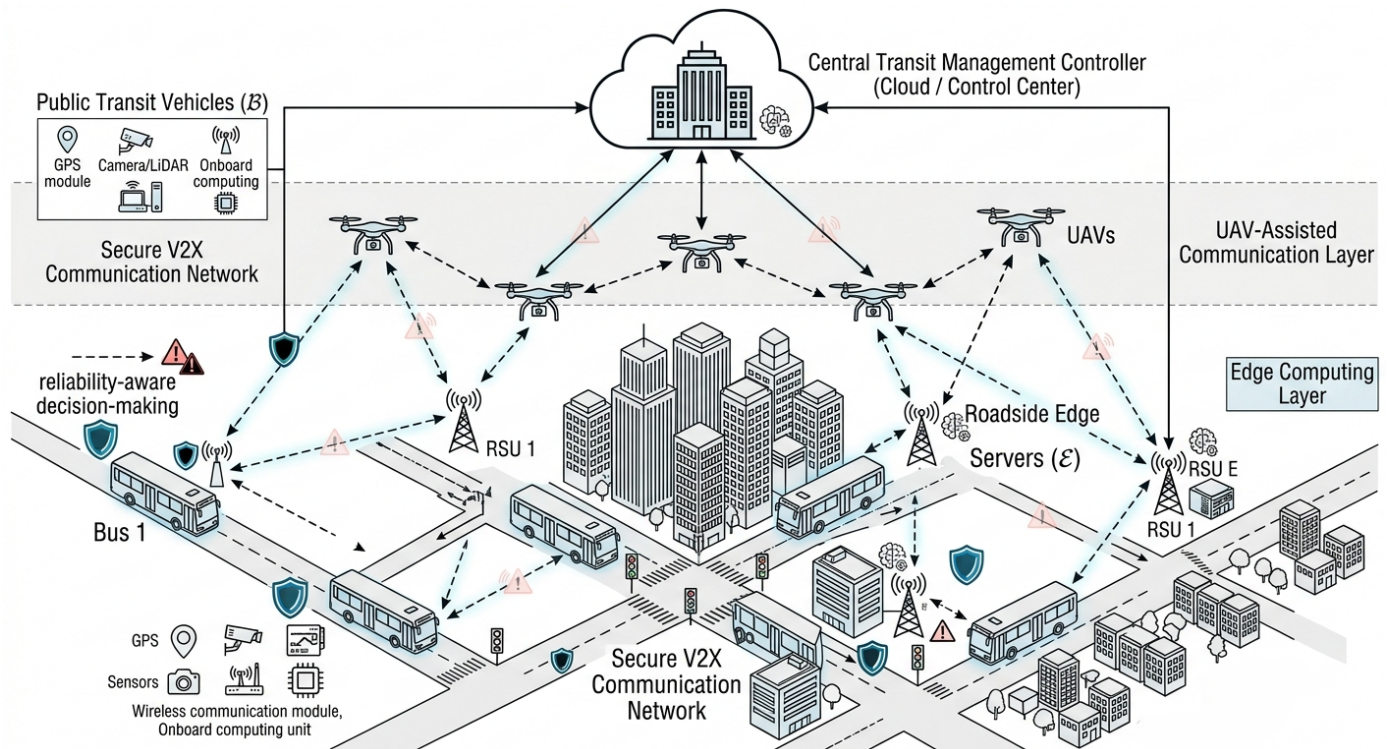


Figure 1. System model of public transit fleets with edge resources.

Our objective is to minimize operational cost, passenger delay, energy consumption, computation latency, and CO₂ emissions while maintaining resilient transportation operations. Simulation results demonstrate that the proposed framework effectively outperforms conventional RP, GS, RBH, SPPO, and standard MAGPPO approaches under varying operational scenarios. The proposed SA-MAGPPO framework achieves lower passenger waiting times, improved communication reliability, reduced operational cost, and better environmental sustainability under dynamic and anomaly-aware transportation environments. The following are the key contributions of this work.

- This research proposes a secure UAV-assisted edge-enabled public transit architecture that supports intelligent fleet operations, low-latency computation offloading, and reliable V2I and V2V communication in dynamic transportation environments.
- A novel Security-Aware Multi-Agent Proximal Policy Optimization (SA-MAGPPO) framework is developed by extending conventional MAGPPO with an augmented state representation that incorporates communication reliability and anomaly information, enabling security-aware joint optimization of routing, scheduling, charging, and computation offloading decisions between onboard units and edge servers.
- A reliability-aware and anomaly-aware learning mechanism is incorporated to improve robustness against unreliable transportation observations, communication instability, and anomalous network conditions in low-altitude intelligent transportation systems.

- The coupled spatio-temporal interactions are modeled among traffic conditions, passenger demand, communication reliability, edge server load, and battery-constrained fleet operations for more adaptive and energy-efficient transportation control.
- The extensive simulations under diverse traffic and operational conditions evaluate the effectiveness of the proposed framework. Simulation results demonstrate significant improvements in operational cost, energy consumption, communication reliability, and CO₂ emissions compared with conventional RP, GS, RBH, SPPO, and MAGPPO approaches.

The organization of this article is as follows. Section 2 briefly discusses recent studies related to public transit fleet management using AI-centered techniques. Section 3 discusses the system model, followed by the problem formulation in Section 4. Section 5 discusses the Multi-Agent Proximal Policy Optimization framework. Section 6 presents simulation results, and Section 7 concludes the article.

2. Related Work

AI has emerged as a promising technology in optimizing public transit networks and fleet services while merging both real-time data analytics and sustainability goals. In [13], the authors presented a DRL-based approach to investigate the problem of headway deviation in a public transport system and to increase the regularity and resiliency of the services in urban environments. A data-driven approach is used to model cumulative disturbances to increase system resilience. The authors of [14] discussed the applications of AI in route optimization and traffic management, as well as in electrifying the fleet to improve fuel efficiency and to cut emissions costs. Urban logistics can be utilized in the optimization of transit fleets, especially by combining AI and autonomous systems. Mohsen in [15] provided an AI-based system where autonomous vehicles with Internet of Things (IoT) infrastructure allow a better usage scheme to enhance traffic capacity with intelligent route planning. The authors in [16] developed an extension by implementing adaptive routing with deep Q-learning and achieved lower fuel consumption, shorter route lengths, and significantly reduced emissions.

In [17], the authors introduced an algorithm tackling various features of urban transport, such as the optimization of the public transport system, the deployment of autonomous vehicles, e-parking, and safety improvement. In [18], Tsolakis et al. investigated the environmental usage of automated guided vehicles in container ports, which shows significant emission reductions. In [19], the authors investigated a carbon footprint management framework with AI for observing and decreasing emissions across global supply chains.

Data-driven approaches further enhance the optimization of mixed-fleet public transportation. In [20], the authors designed a framework to predict route-level energy consumption for electric and internal combustion vehicles by incorporating vehicle-level, traffic, and weather information. The authors of [21] developed a novel framework to manage energy demand and charging infrastructure to facilitate electric mobility, based on floating car data, and freely available sources for simulation-based decision-making. These predictive modeling solutions are fundamental in implementing optimization strategies that are not only viable in operational terms but also in environmental sustainability during fleet decarbonization processes.

This research [22] proposes a GraphSAGE-based GNN architecture integrated with multi-agent deep deterministic policy gradient (MADDPG) for resource allocation-based task offloading in V2X networks. A high-speed vehicular environment is modeled using a GNN, accounting for dynamic topology and interference awareness to enable joint action decision-making. The framework achieved better performance in task completion ratio,

latency, and energy overhead than DQN and traditional MADDPG approaches, especially in high-load traffic.

This study presents a vehicular offloading framework based on task generation using a traffic density model with a congestion coefficient relying on road occupancy and vehicle speed [23]. Gaussian-based density refinement is used for task offloading modelled as an MDP and solved using a DDPG approach to minimize latency and energy consumption. The proposed framework can adapt traffic offloading decisions to varying vehicular traffic density. The results indicated that the method provided better QoE with reduced latency and enhanced resource management than existing schemes.

Deep Adaptive Assignment Framework (DAAF) is proposed in this paper [24], which is a deadline-aware fuzzy logic-based task offloading technique for V2V networks. Considering multiple parameters with task deadlines enables adaptation of offloading decisions from uncertainty through fuzzy logic. Extensive results show that DAAF reduces the average completion time by 12%, achieves a 90% task completion ratio, and improves execution time when compared to traditional fuzzy techniques and baseline methods.

The authors of [25] propose a dependency-aware and energy-efficient scheduling (DESR) mechanism for offloading tasks in VEC scenarios. To address the problem, the task dependency is presented with directed acyclic graphs (DAGs), and the delay and energy consumption minimization problem of task offloading is formulated as an NP-hard MINLP problem. The decomposition and interior point-based method is exploited to jointly optimize the task offloading decisions and resource allocation. Simulation results show that DESR outperforms state-of-the-art methods in terms of system delay and energy consumption.

In [26], the authors used AI to enhance smart cities' environmental management by improving knowledge levels and enabling smarter decision-making. This enhances resource usage patterns and environmental governance. In [27], the researchers proposed a data-driven approach by determining the multi-perspective parameters in service economies. They provided a framework for embedding sustainability in multidimensional systems and concluded that the data-driven approaches with AI are critical for optimizing the efficiency of the public transport fleet, passenger satisfaction, and environmental responsibility.

In the context of security and privacy, the authors of [28] investigated a security-aware offloading of IoT applications to be processed by the mobile edge servers. The aim was to minimize computation overheads measured by the task processing delay and energy consumption of Internet-of-Things devices. The researchers integrated dynamic voltage scaling technology with task offloading and AES cryptography. The simulation results validated the effectiveness of the proposed work. Similarly, the authors of [29] investigated the application execution on edge servers using dynamic voltage scaling and AES encryption for offloading and spectrum resource allocation. In [30], the researchers provided a cooperative jamming-based scheme. This scheme involves a UAV jammer that transmits jamming signals to interfere with eavesdroppers during devices' computation offloading processes. They used a multi-agent twin delayed deep deterministic policy gradient (TD3) to solve the non-linear programming problem. The simulation results showed that the proposed methods surpassed the baseline in terms of task completion delay. Recent advances in edge-enabled intelligent transportation systems have increasingly focused on integrating multi-agent deep reinforcement learning, UAV-assisted edge computing, and communication-aware resource management to improve the adaptability and scalability of large-scale transportation networks [31]. Moreover, recent studies have emphasized robust computation offloading and resilient decision-making under dynamic network conditions, highlighting the importance of communication reliability and trustworthy transportation observations for intelligent edge services [32].

The above studies, taken together, provide a broad perspective on the role of AI and other schemes in urban sustainability, including public transport, freight logistics, industrial operations, and urban planning. AI, combined with IoT, predictive analytics, and data-driven simulation, can solve significant urban issues such as traffic congestion, gas emissions, and energy demands. However, the above methods overlook the security anomaly and optimization of CO₂ emissions and operational costs in low-altitude intelligent transportation systems. To this end, we propose SA-MAGPPO to optimize the abovementioned parameters in such highly dynamic networks. Table 1 provides a comparison of the proposed work with recent studies based on multiple parameters. From the table, it is evident that existing works focus on isolated aspects such as routing, energy prediction, or task offloading, etc. These do not jointly model passenger demand and battery-constrained operations under dynamic spatio-temporal environments. To this end, the proposed framework integrates all these factors into a unified AI-driven optimization approach.

Table 1. Comparison of proposed SA-MAGPPO-based framework with existing studies.

Ref.	Contribution	Fleet Op.	AI	Emission Reduction	Spatio-Temporal	Battery-Aware	Passenger Demands	Security and Privacy
[13]	DRL-based headway control for transit regularity	✓	✓	✗	✗	✗	✗	✗
[14]	AI applications in routing, traffic, electrification (survey)	✓	✓	✗	✗	✗	✗	✗
[15]	AI + IoT-based routing for traffic efficiency	✓	✓	✗	✗	✗	✗	✗
[16]	DRL-based adaptive routing reducing fuel/emissions	✓	✓	✓	✗	✗	✗	✗
[17]	AI-driven urban transport system optimization framework	✓	✓	✗	✗	✗	✗	✗
[18]	Emission reduction using AGVs in port operations	✗	✗	✓	✗	✗	✗	✗
[19]	AI-based carbon footprint monitoring framework	✗	✓	✗	✗	✗	✗	✗
[20]	Data-driven energy consumption prediction for fleets	✓	✓	✗	✗	✗	✗	✗
[21]	EV charging and energy demand management framework	✓	✓	✗	✗	✓	✗	✗
[22]	GNN + MADDPG for V2X task offloading	✗	✓	✗	✓	✗	✗	✗
[26,27]	AI-driven sustainability frameworks for smart cities	✓	✓	✗	✓	✗	✓	✗
[28]	Security-aware IoT offloading with AES encryption and dynamic voltage scaling for energy-efficient edge computing	✗	✓	✗	✗	✗	✗	✓
[29]	Secure edge execution with AES encryption and dynamic voltage scaling for resource allocation	✗	✓	✗	✗	✗	✗	✓
[30]	UAV-assisted secure offloading using cooperative jamming and multi-agent TD3 for eavesdropping protection	✗	✓	✗	✗	✗	✗	✓
Proposed	Multi-agent security-aware AI optimization for routing, scheduling, charging, and resilient transit control	✓	✓	✓	✓	✓	✓	✓

3. System Model

We consider a secure UAV-assisted edge-enabled public transit system deployed in a smart urban environment. This consists of a fleet of public transit vehicles, roadside edge servers, UAV-assisted communication units, and a centralized transit management

controller. The system supports intelligent transportation services through real-time communication, distributed edge analytics, and AI-driven operational control. In our system, $\mathcal{B} = \{1, 2, \dots, B\}$ denote the set of transit vehicles, $\mathcal{E} = \{1, 2, \dots, E\}$ denote the set of edge servers deployed along the road infrastructure, and $\mathcal{U} = \{1, 2, \dots, U\}$ represent the set of UAVs operating as aerial communication and monitoring assistants. Each bus is equipped with sensing modules, GPS units, wireless communication interfaces, and on-board computation resources capable of executing lightweight AI tasks. The edge servers provide low-latency computation support for computationally intensive transit analytics and decision-making tasks. The UAVs operate as aerial relay and monitoring nodes to enhance V2V and V2I communication reliability in highly congested urban environments. UAVs assist in extending communication coverage, and reducing wireless congestion. The considered system operates in a dynamic transportation environment where passenger demand, traffic congestion, communication quality, and operational conditions continuously vary over time. Furthermore, due to the reliance on real-time transportation information, the system may experience unreliable observations caused by communication failures, corrupted traffic reports, or malicious data injection attempts.

3.1. Transit Network Model

The urban transit network is represented by a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ denotes the set of transit nodes, including bus stops and stations, while $\mathcal{L}$ represents the set of road links connecting the nodes. Each vehicle $b \in \mathcal{B}$ traverses a predefined or dynamically adjusted route composed of a subset of links $\mathcal{L}_b \subseteq \mathcal{L}$. The traffic density on link $l \in \mathcal{L}$ at time slot t is denoted by $\rho_l(t)$, while the average vehicle speed is represented by $v_l(t)$. The passenger demand at node $i \in \mathcal{N}$ evolves dynamically according to temporal mobility patterns and is represented as $d_i(t)$. The demand variation depends on traffic conditions, historical mobility trends, and real-time passenger flow observations. The predicted passenger demand is continuously updated. To further improve congestion awareness, UAVs periodically monitor traffic conditions and collect aerial transportation observations. The UAV-assisted traffic monitoring information is represented as

$$\Omega_u(t) = \{\rho_u(t), v_u(t), \psi_u(t)\},$$

where $\rho_u(t)$ denotes the observed traffic density, $v_u(t)$ represents average traffic flow velocity, and $\psi_u(t)$ denotes congestion intensity captured by UAV u at time slot t . The collected aerial observations are forwarded to nearby edge servers for congestion-aware fleet coordination. The travel time of bus b on link l is expressed as

$$T_{b,l}(t) = \frac{L_l}{v_l(t)},$$

where L_l denotes the physical length of link l .

3.2. Communication and Reliability Model

The buses communicate with nearby edge servers through Vehicle-to-Infrastructure (V2I) wireless communication links. In congested or unreliable communication regions, UAVs assist as aerial relay nodes to improve communication stability and extend wireless coverage. The achievable uplink transmission rate between bus b and edge server e at time slot t is given by

$$r_{b,e}(t) = W \log_2 \left(1 + \frac{P_b h_{b,e}(t)}{\sigma^2 + I_{b,e}(t)} \right),$$

where W is the channel bandwidth, P_b is the transmission power of bus b , $h_{b,e}(t)$ denotes the wireless channel gain, σ^2 represents the noise power, and $I_{b,e}(t)$ denotes the interference power. For UAV-assisted communication, the achievable transmission rate between bus b and UAV u is expressed as

$$r_{b,u}(t) = W_u \log_2 \left(1 + \frac{P_b h_{b,u}(t)}{\sigma^2 + I_{b,u}(t)} \right),$$

where W_u denotes the UAV communication bandwidth and $h_{b,u}(t)$ represents the air-to-ground wireless channel gain. The UAV forwards the received information to the associated edge server through UAV-to-Infrastructure (U2I) communication links. Due to the dynamic transportation environment, communication links may suffer from instability, packet corruption, or unreliable traffic information. Therefore, we define a communication reliability score $\chi_{b,e}(t) \in [0, 1]$ where larger values indicate more reliable communication conditions. Similarly, the reliability of UAV-assisted communication links is represented by $\chi_{b,u}(t) \in [0, 1]$. In addition, an anomaly score $a_b(t) \in [0, 1]$ is introduced to quantify the abnormality level of the received transportation observations. Let $\mathbf{o}_b(t)$ denote the received observation vector of bus b , including traffic density, vehicle speed, and communication measurements, while $\hat{\mathbf{o}}_b(t)$ denotes its predicted normal value obtained from historical observations. The observation deviation is computed as

$$e_b(t) = \|\mathbf{o}_b(t) - \hat{\mathbf{o}}_b(t)\|_2.$$

The corresponding anomaly score is normalized as

$$a_b(t) = \min \left(1, \frac{e_b(t)}{\varepsilon} \right),$$

where ε is a normalization threshold representing the maximum acceptable observation deviation under normal operating conditions. Larger values of $a_b(t)$ indicate a higher likelihood that the received transportation information is unreliable due to sensing failures, communication errors, or malicious data manipulation attempts. The computed anomaly score is incorporated into the security-aware state representation and reward formulation of the proposed SA-MAGPPO framework.

3.3. Edge Computation Model

Each bus periodically generates a computation task

$$\tau_b(t) = \{S_b(t), C_b(t)\},$$

where $S_b(t)$ denotes the task input data size and $C_b(t)$ represents the required CPU cycles for task execution. The generated tasks may include passenger demand prediction, route optimization, fleet scheduling, and congestion-aware control decisions. Each task can either be processed locally at the onboard unit or offloaded to an edge server for remote execution. The binary offloading decision is represented as

$$x_b(t) = \begin{cases} 0, & \text{local execution,} \\ 1, & \text{edge offloading.} \end{cases}$$

For local processing, the execution delay is given by

$$\delta_b^{\text{loc}}(t) = \frac{C_b(t)}{f_b^{\text{loc}}},$$

where f_b^{loc} denotes the onboard computational capability. For direct edge execution, the uplink transmission delay is expressed as

$$\delta_b^{\text{tx}}(t) = \frac{S_b(t)}{r_{b,e}(t)}.$$

When UAV-assisted communication is utilized, the transmission delay becomes

$$\delta_b^{\text{uav}}(t) = \frac{S_b(t)}{r_{b,u}(t)} + \frac{S_b(t)}{r_{u,e}(t)},$$

where $r_{u,e}(t)$ denotes the transmission rate between UAV u and edge server e . The edge execution delay is given by

$$\delta_b^{\text{edge}}(t) = \frac{C_b(t)}{f_e(t)},$$

where $f_e(t)$ denotes the computational resources allocated by edge server e . Then, the total edge-assisted execution delay becomes

$$\delta_b^{\text{tot}}(t) = \delta_b^{\text{tx}}(t) + \delta_b^{\text{edge}}(t).$$

For UAV-assisted offloading, the total delay is expressed as

$$\delta_b^{\text{tot-uav}}(t) = \delta_b^{\text{uav}}(t) + \delta_b^{\text{edge}}(t).$$

3.4. Energy and Emission Model

The operational energy consumption of bus b depends on vehicle movement, onboard computation, and wireless communication activities. The communication energy consumption is expressed as $E_b^{\text{tx}}(t) = P_b \delta_b^{\text{tx}}(t)$. For UAV-assisted communication, the corresponding transmission energy consumption is given by $E_b^{\text{uav}}(t) = P_b \delta_b^{\text{uav}}(t)$. The local computation energy consumption is modeled as

$$E_b^{\text{loc}}(t) = \kappa \left(f_b^{\text{loc}} \right)^2 C_b(t),$$

where κ denotes the effective switched capacitance coefficient. The total energy consumption is then given by

$$E_b^{\text{total}}(t) = (1 - x_b(t)) E_b^{\text{loc}}(t) + x_b(t) (E_b^{\text{tx}}(t) + E_b^{\text{uav}}(t)).$$

The corresponding CO₂ emissions are modeled as $E_b^{\text{CO}_2}(t) = \eta_b E_b^{\text{total}}(t)$, where η_b denotes the emission factor associated with the vehicle type.

4. Problem Formulation

The proposed framework jointly optimizes transit operation control and edge-assisted computation decisions in order to minimize environmental impact, operational energy consumption, passenger delay, and unreliable transportation decisions in dynamic urban environments. At each time slot t , each bus agent determines operational control actions, task offloading decisions, charging and scheduling strategies, adaptive routing policies,

and UAV-assisted communication selection. The optimization problem is formulated over a finite operational horizon T . The overall system cost is defined as

$$\min_{\pi, x_b(t), a_b(t)} \mathcal{J} = \sum_{t=0}^T \sum_{b=1}^B \left(\alpha E_b^{\text{CO}_2}(t) + \beta E_b^{\text{total}}(t) + \gamma D_b(t) - \delta U_b(t) + \zeta a_b^{\text{anom}}(t) - \mu \chi_{b,u}(t) \right), \quad (1)$$

where $E_b^{\text{CO}_2}(t)$ denotes carbon emissions, $E_b^{\text{total}}(t)$ represents total energy consumption, $D_b(t)$ denotes passenger service delay, $U_b(t)$ represents passenger utility satisfaction, $a_b^{\text{anom}}(t)$ denotes the anomaly risk score associated with unreliable transportation observations, and $\chi_{b,u}(t)$ denotes the UAV-assisted communication reliability factor. In this equation, $\alpha, \beta, \gamma, \delta, \zeta$, and μ are non-negative weighting coefficients. The anomaly-aware penalty term enables the proposed framework to avoid unreliable routing and scheduling decisions caused by corrupted or abnormal transportation information. Furthermore, the UAV-assisted communication reliability term promotes stable aerial-ground communication links under highly congested transportation conditions. The operational control action of bus b at time slot t is represented as

$$a_b(t) = \{r_b(t), h_b(t), c_b(t), v_b(t), y_b(t)\},$$

where $r_b(t)$ denotes routing action, $h_b(t)$ denotes headway adjustment, $c_b(t)$ denotes charging decision, $v_b(t)$ denotes speed adaptation, and $y_b(t) \in \{0, 1\}$ denotes UAV-assisted communication selection. For edge-assisted execution, the total task completion delay includes uplink transmission delay, edge computation delay, and downlink feedback delay. Then, the latency constraint is expressed as

$$x_b(t) \left(\frac{S_b(t)}{r_{b,e}(t) \chi_{b,e}(t)} + \frac{C_b(t)}{f_e^{\text{edge}}(t)} + \frac{R_b(t)}{r_{e,b}(t)} \right) \leq \delta_b^{\text{max}}(t), \quad (2)$$

where $\chi_{b,e}(t) \in (0, 1]$ denotes the communication reliability factor, $R_b(t)$ denotes the returned control response size, and $\delta_b^{\text{max}}(t)$ denotes the maximum tolerable latency. When UAV-assisted communication is selected, the delay constraint becomes

$$y_b(t) \left(\frac{S_b(t)}{r_{b,u}(t) \chi_{b,u}(t)} + \frac{S_b(t)}{r_{u,e}(t)} + \frac{C_b(t)}{f_e^{\text{edge}}(t)} \right) \leq \delta_b^{\text{max}}(t). \quad (3)$$

For local execution, the onboard computation delay must satisfy

$$(1 - x_b(t)) \frac{C_b(t)}{f_b^{\text{loc}}} \leq \delta_b^{\text{max}}(t). \quad (4)$$

As each edge server has finite computational resources. Therefore, the aggregate computation demand allocated to edge server e must satisfy

$$\sum_{b \in \mathcal{B}_e} x_b(t) C_b(t) \leq F_e^{\text{cap}}, \quad (5)$$

where F_e^{cap} denotes the maximum computation capacity of edge server e . Due to limited wireless spectrum resources, the aggregate communication demand sharing the same wireless channel must satisfy

$$\sum_{b \in \mathcal{B}_c} \frac{S_b(t)x_b(t)}{\delta_b^{\text{max}}(t)} \leq B_c^{\text{phy}}, \quad (6)$$

where B_c^{phy} denotes the available physical wireless bandwidth. Similarly, the UAV-assisted aerial communication resources are constrained by

$$\sum_{b \in \mathcal{B}_u} \frac{S_b(t)y_b(t)}{\delta_b^{\text{max}}(t)} \leq B_u^{\text{phy}}, \quad (7)$$

where B_u^{phy} denotes the available UAV communication bandwidth. The battery state evolution of electric bus b is modeled as

$$\text{SOC}_b(t+1) = \text{SOC}_b(t) - \frac{E_b^{\text{drive}}(t)}{C_b^{\text{bat}}} + \frac{E_b^{\text{chg}}(t)}{C_b^{\text{bat}}}, \quad (8)$$

where $\text{SOC}_b(t)$ denotes the battery state-of-charge, C_b^{bat} denotes battery capacity, $E_b^{\text{drive}}(t)$ denotes driving energy consumption, and $E_b^{\text{chg}}(t)$ denotes charging energy. Then the battery operation must satisfy $\text{SOC}_{\min} \leq \text{SOC}_b(t) \leq \text{SOC}_{\max}$, $\forall b, t$. Each charging station supports a limited number of simultaneous charging vehicles. Then,

$$\sum_{b: \text{at } s} z_{b,s}(t) \leq Q_s, \quad z_{b,s}(t) \in \{0, 1\}, \quad (9)$$

where $z_{b,s}(t)$ denotes charging assignment and Q_s denotes charging station capacity. The routing actions must satisfy the road network topology constraints. If bus b moves from node i to node j , then $(i, j) \in \mathcal{L}$. The formulated optimization problem is computationally intractable because it jointly optimizes routing, fleet scheduling, computation offloading, and edge resource allocation under dynamic communication and mobility constraints. The binary offloading decisions, coupled with non-linear communication delay, energy consumption, and reliability constraints, result in a mixed-integer non-convex optimization problem. Moreover, computation offloading and task scheduling have been shown to be NP-hard in mobile and vehicular edge computing systems. Since the proposed formulation generalizes these NP-hard subproblems while introducing additional routing and communication reliability constraints, the overall optimization problem is also NP-hard. Therefore, obtaining the global optimum through exhaustive search becomes computationally infeasible for large-scale transit systems, motivating the use of a multi-agent deep reinforcement learning approach.

4.1. Problem Complexity Analysis

The formulated optimization problem jointly considers discrete offloading decisions, dynamic routing optimization, charging scheduling, UAV-assisted communication selection, and the anomaly aware reliability, etc. The coupling between communication, computation, transportation, aerial communication assistance, and energy constraints results in a high-dimensional non-convex mixed-integer optimization problem. Furthermore, the dynamic evolution of traffic conditions, UAV-assisted communication reliability, and passenger demand make real-time optimization computationally intractable using conventional optimization techniques. To this end, we employ an SA-MAGPPO-based DRL

framework (Figure 2) to learn adaptive and scalable decision policies through continuous interaction with the transportation environment.

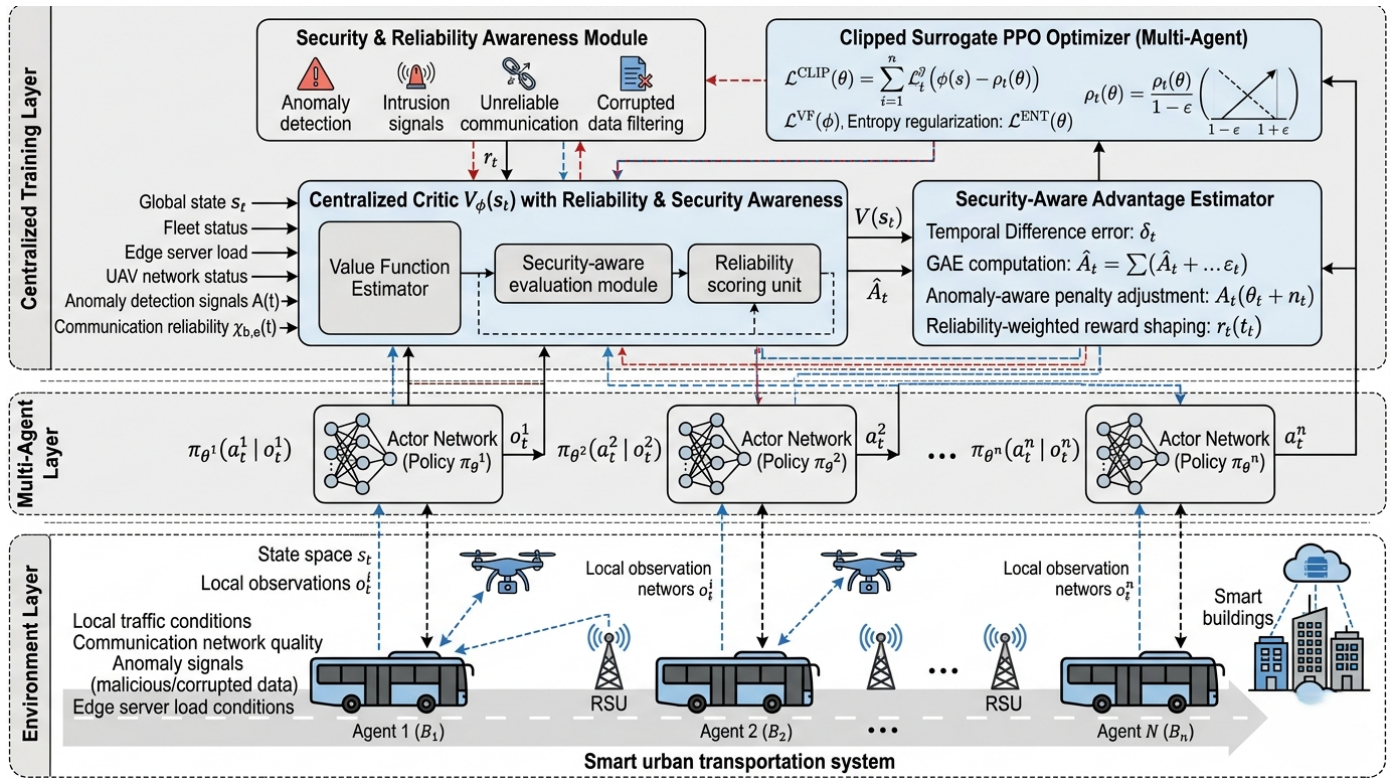


Figure 2. Architecture of proposed SA-MAGPPO framework.

4.2. MDP Formulation

We model the proposed problem as a Markov Decision Process (MDP) $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma)$, where the system evolves over discrete time slots $t \in \{0, 1, \dots, T\}$. The system state $s_t \in \mathcal{S}$ captures spatio-temporal fleet dynamics, edge and network conditions, UAV-assisted communication conditions, and reliability-aware environmental observations. It is defined as:

$$s_t = \{s_t^{\text{fleet}}, s_t^{\text{mob}}, s_t^{\text{net}}, s_t^{\text{edge}}, s_t^{\text{sec}}, s_t^{\text{uav}}\}$$

where state is given $s_t^{\text{fleet}} = \{d_b(t), l_b(t), \tau_b(t), \text{SOC}_b(t)\}_{b=1}^B$. The demand and mobility state is presented as $s_t^{\text{mob}} = \hat{D}(\cdot, t)$. The network state is $s_t^{\text{net}} = \rho_{\text{net}}(t)$, the edge computing state is $s_t^{\text{edge}} = \rho_{\text{srv}}(t)$, the security-aware reliability state is $s_t^{\text{sec}} = \{a_b(t), \chi_{b,e}(t)\}_{b=1}^B$, and the UAV-assisted communication state is given as $s_t^{\text{uav}} = \{\chi_{b,u}(t), \Omega_u(t)\}_{u=1}^U$ where $a_b(t) \in [0, 1]$ denotes anomaly level in transportation observations, $\chi_{b,e}(t) \in (0, 1]$ denotes communication reliability between bus b and edge server e , $\chi_{b,u}(t) \in (0, 1]$ represents UAV-assisted communication reliability, and $\Omega_u(t)$ denotes UAV-observed aerial traffic and congestion information. The action space $a_t \in \mathcal{A}$ is the joint control action that consists of four coupled decisions

$$a_t = \{a_t^{\text{mob}}, a_t^{\text{comp}}, a_t^{\text{energy}}, a_t^{\text{uav}}\}$$

where mobility control is given as $a_t^{\text{mob}} = \{r_b(t), v_b(t)\}_{b=1}^B$ for route selection and speed. The computation control is given as $a_t^{\text{comp}} = \{x_b(t)\}_{b=1}^B$ among local execution and edge offloading. The energy control is $a_t^{\text{energy}} = \{c_b(t)\}_{b=1}^B$ related to charging and UAV-assisted communication control is represented as $a_t^{\text{uav}} = \{y_b(t)\}_{b=1}^B$ where $y_b(t) \in \{0, 1\}$ determines whether UAV-assisted communication is utilized. The system transition probability $\mathcal{P}$ is

governed by stochastic interactions among traffic flow, passenger arrivals, battery dynamics, wireless channel variations, UAV-assisted communication conditions, and computation offloading outcomes. Specifically, the state evolution satisfies $s_{t+1} \sim \mathcal{P}(s_{t+1} | s_t, a_t)$ where transitions are influenced by time-varying traffic density, passenger demand fluctuations, battery consumption and charging processes, edge server load dynamics, communication reliability $\chi_{b,e}(t)$, UAV communication reliability $\chi_{b,u}(t)$, anomaly disturbances $a_b(t)$ affecting observation quality, and UAV-assisted aerial congestion observations. The offloading process is feasible only if:

$$\delta_b^{\text{tx}}(t) + \delta_b^{\text{comp}}(t) \leq \Delta_{\max}$$

under reliable communication conditions. Similarly, for UAV-assisted transmission,

$$\delta_b^{\text{uav}}(t) + \delta_b^{\text{comp}}(t) \leq \Delta_{\max}.$$

We define a resilience-aware multi-objective reward function that jointly captures sustainability, service quality, communication reliability, and system resilience:

$$r_t = -\alpha E^{\text{CO}_2}(t) - \beta E^{\text{energy}}(t) - \gamma D(t) + \eta U(t) - \xi A(t) + \mu \chi_u(t)$$

where $E^{\text{CO}_2}(t)$ represents the total emissions, $E^{\text{energy}}(t)$ denotes energy consumption, $D(t)$ denotes passenger delay, $U(t)$ is the service utility, $A(t)$ is the system anomaly risk, and $\chi_u(t)$ is the UAV-assisted communication reliability reward. Here, $\alpha, \beta, \gamma, \eta, \xi, \mu > 0$ are weighting factors. The weighting coefficients are empirically selected to achieve a balanced trade-off between energy consumption, passenger delay, operational cost, communication reliability, and anomaly robustness. During preliminary experiments, the coefficients were adjusted to obtain stable policy convergence while preventing any single optimization objective from dominating the learning process. The anomaly penalty $A(t)$ ensures that the learned policy avoids unreliable or potentially corrupted transportation states caused by communication failures or malicious data injections. Moreover, the UAV-assisted communication reward improves aerial-ground connectivity reliability and reduces communication instability in congested transportation environments.

5. Security-Aware Multi-Agent Proximal Policy Optimization (SA-MAGPPO)

To solve the formulated high-dimensional, stochastic, and reliability-aware optimization problem, we propose the SA-MAGPPO framework. The framework enables distributed decision-making across transit agents while maintaining centralized stability during training. Each agent $i \in \mathcal{N}$ (representing a bus or control unit) learns a stochastic policy $\pi_{\theta^i}(a_t^i | o_t^i)$, where o_t^i denotes the local observation derived from the global system state. The objective is to maximize the expected discounted return:

$$J(\theta) = \mathbb{E}_{\pi_{\theta}} \left[\sum_{t=0}^T \gamma^t r_t \right],$$

where r_t is the system reward defined in the previous section. Unlike conventional MAGPPO, SA-MAGPPO incorporates anomaly-aware and reliability-aware feedback into the advantage function. The temporal-difference error is defined as:

$$\delta_t = r_t + \gamma V_{\phi}(s_{t+1}) - V_{\phi}(s_t),$$

where the reward r_t explicitly includes anomaly penalty $A(t)$ and communication reliability effects $\chi_{b,e}(t)$. The generalized advantage estimate is computed as:

$$\hat{A}_t = \sum_{l=0}^{T-t} (\gamma\lambda)^l \delta_{t+l}.$$

The formulation ensures that trajectories influenced by unreliable or anomalous states receive lower advantage values. This guides the policy toward secure and stable operational regions. The policy update follows the PPO clipped objective extended to multi-agent settings, which ensures stable policy updates under highly dynamic and uncertain transit environments and given as

$$L^{\text{CLIP}}(\theta) = \mathbb{E}_t [\min(\rho_t(\theta)\hat{A}_t, \text{clip}(\rho_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)],$$

where:

$$\rho_t(\theta) = \frac{\pi_\theta(a_t|o_t)}{\pi_{\theta_{\text{old}}}(a_t|o_t)}.$$

A centralized value function $V_\phi(s_t)$ is used to stabilize training by aggregating global system information. This includes fleet state, edge load, network congestion, anomaly level, and communication reliability. The critic is trained by minimizing:

$$L^{\text{VF}}(\phi) = \mathbb{E}_t [(V_\phi(s_t) - \hat{R}_t)^2],$$

where the return $\hat{R}_t$ is computed as:

$$\hat{R}_t = \sum_{l=0}^{T-t} \gamma^l r_{t+l}.$$

To encourage exploration in uncertain and dynamic transportation environments, an entropy regularization term is added:

$$L^{\text{ENT}}(\theta) = \mathbb{E}_t [-\pi_\theta(a_t|o_t) \log \pi_\theta(a_t|o_t)].$$

The final SA-MAGPPO optimization objective is given by:

$$\max_{\theta} \mathbb{E}_t [L^{\text{CLIP}}(\theta) - c_1 L^{\text{VF}}(\phi) + c_2 L^{\text{ENT}}(\theta)].$$

During training, the centralized critic leverages full system observability. This includes anomaly and reliability signals to stabilize learning. During execution, each agent operates in a decentralized manner using only local observations o_t^i , which enables scalable deployment across large-scale transit fleets. This centralized training–decentralized execution ensures robustness, scalability, and adaptability in real-world intelligent transportation environments. This is specifically characterized by dynamic traffic conditions, uncertain communication quality, and potential data anomalies. Algorithm 1 provides the complete pseudocode of the proposed algorithm, and Figure 3 shows the multi-agent neural network architecture.

Algorithm 1 Security-Aware Multi-Agent Proximal Policy Optimization (SA-MAGPPO)

Require: Transit agents $\mathcal{N}$, policy parameters θ , critic parameters ϕ , learning rate α , discount factor γ , GAE parameter λ , clipping parameter ϵ

Ensure: Optimized policy parameters θ

- 1: Initialize actor π_θ and centralized critic V_ϕ
- 2: **for** each training episode **do**
- 3: Reset the transit environment
- 4: **for** each time step t **do**
- 5: **for** each agent $i \in \mathcal{N}$ **do**
- 6: Observe local state o_t^i
- 7: Select action $a_t^i \sim \pi_\theta(a_t^i | o_t^i)$
- 8: **end for**
- 9: Execute joint action $\mathbf{a}_t$
- 10: Receive reward r_t , next state s_{t+1} , anomaly score $A(t)$, and communication reliability $\chi(t)$
- 11: Store transition $(s_t, \mathbf{o}_t, \mathbf{a}_t, r_t, s_{t+1})$
- 12: **end for**
- 13: Compute temporal-difference error

$$\delta_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)$$

- 14: Compute generalized advantage estimate $\hat{A}_t$
- 15: Compute discounted return $\hat{R}_t$
- 16: Update actor by maximizing the clipped PPO objective

$$L^{\text{CLIP}}(\theta)$$

- 17: Update centralized critic by minimizing

$$L^{\text{VF}}(\phi)$$

- 18: Apply entropy regularization

$$L^{\text{ENT}}(\theta)$$

- 19: Update parameters (θ, ϕ)

- 20: **end for**

- 21: **return** Optimized decentralized policies π_θ

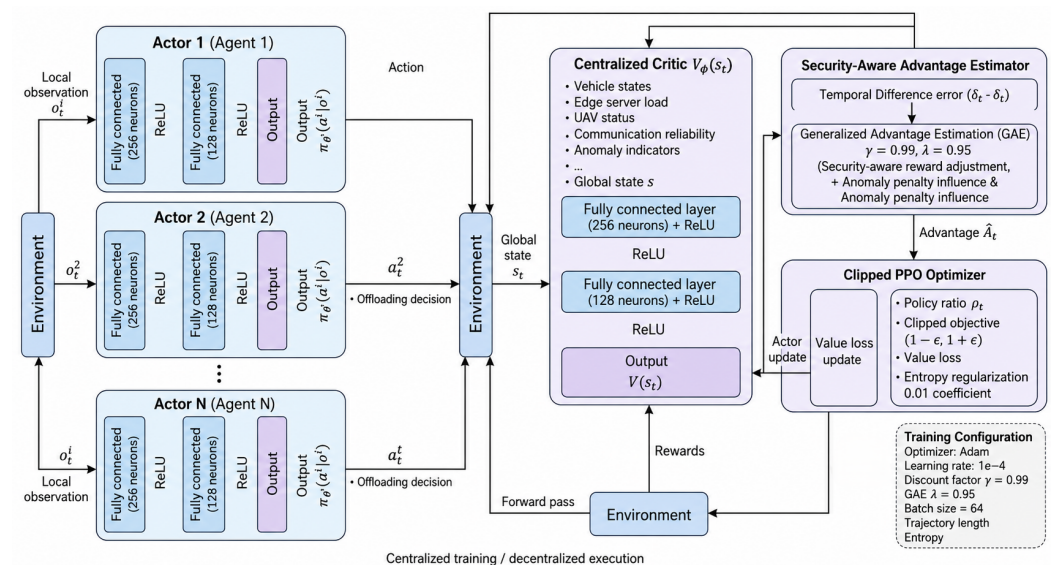


Figure 3. Multi-agent Neural Network Architecture.

6. Simulation Results

We evaluate the performance of the proposed SA-MAGPPO framework under diverse urban transit operational conditions. The simulations are conducted across four realistic time-dependent scenarios, including morning peak, midday off-peak, evening peak, and night off-peak. These scenarios capture temporal variations in passenger demand, traffic congestion, communication reliability, and energy consumption. In the morning peak (7:00–9:00 AM), the system experiences high passenger arrival rates, dense traffic conditions, and increased energy usage. The midday off-peak period (11:00 AM–2:00 PM) represents lower demand and moderate network congestion. The evening peak (5:00–7:00 PM) reflects high return-trip demand with increased routing complexity and energy consumption. Finally, the night off-peak period (10:00 PM–12:00 AM) has reduced traffic and minimal passenger load.

6.1. Simulation Environment Setup

The simulation environment is implemented in Python 3.11 using a custom-built fleet control platform that models urban transit operations as a dynamic edge-assisted system. The environment incorporates stochastic passenger arrivals, time-varying traffic density, and heterogeneous wireless channel conditions. NetworkX is used for modeling UAVs. In addition to standard operational dynamics, we introduce a reliability-aware communication layer to reflect real-world network instability in intelligent transportation systems. Specifically, wireless links are subject to random fading, packet loss, and intermittent congestion, which directly affects the communication reliability factor $\chi_{b,e}(t)$. To further evaluate robustness, we incorporate an adversarial disturbance model where a fraction of traffic demand signals are corrupted. The delayed or noisy state observations are introduced, and occasional anomaly injections also simulate abnormal transportation reports. The communication reliability score $\chi_{b,e}(t)$ is dynamically updated according to the instantaneous wireless channel quality, packet loss, and interference conditions, resulting in values within the range $[0, 1]$, where larger values indicate more reliable communication links. Similarly, the anomaly score $a_b(t)$ is computed as the normalized deviation between the received transportation observation and its predicted normal value, as defined in the Communication and Reliability Model. Both scores evolve over time with changing network conditions and are incorporated into the security-aware state representation and reward computation during training. All experiments are conducted on a workstation equipped with an Intel Core i9 processor, 32 GB RAM, and an NVIDIA RTX 4090 GPU. Unless otherwise stated, each experiment is repeated over five independent runs with different random seeds, and the reported results correspond to the average performance. The agents are trained for 2000 episodes, where each episode consists of 200 simulation time slots. The number of transit vehicles varies from 10 to 50, the number of edge servers ranges from 1 to 5, the number of UAVs ranges from 2 to 10, and the communication reliability and anomaly levels are varied according to the evaluated scenarios. The proposed SA-MAGPPO is compared against the following baseline methods:

- Greedy scheme (GS): selects locally optimal routing and offloading decisions without long-term optimization.
- Rule-based heuristic (RBH): uses predefined thresholds for charging, routing, and task offloading.
- Random policy (RP): randomly selects actions without learning.
- Single-agent PPO (SPPO): uses a single-agent PPO method to optimize the learning process.
- Multi-agent PPO (MAGPPO): uses the multi-agent PPO method to optimize the learning process.

The selected baseline methods represent different categories of existing transportation optimization approaches. RP, GS, and RBH correspond to conventional heuristic and rule-based methods, SPPO represents a single-agent DRL approach, while MAGPPO serves as the most relevant multi-agent DRL baseline for evaluating the effectiveness of the proposed security-aware extensions. Moreover, recent robust and security-aware multi-agent DRL methods are designed for different optimization objectives and application scenarios, making direct comparisons less appropriate. The SA-MAGPPO framework adopts an actor–critic architecture with centralized training and decentralized execution. The policy and value networks consist of two fully connected hidden layers with 256 and 128 neurons, respectively, using ReLU activation functions. The training process utilizes the Adam optimizer with a learning rate of 10^{-4} , discount factor $\gamma = 0.99$, and Generalized Advantage Estimation parameter $\lambda = 0.95$. Each update is performed using mini-batches of size 64 with trajectory length 2048. An entropy coefficient of 0.01 is used to encourage exploration in uncertain environments. All results are averaged over multiple independent simulation runs to ensure statistical significance. In each run, the system is initialized with different random seeds affecting passenger demand distributions, traffic congestion patterns, and communication channel realizations. To ensure fairness, all baseline methods are evaluated under identical environmental conditions. This includes the same demand patterns, network loads, and energy pricing models.

6.2. Results and Discussion

The convergence behavior in Figure 4 demonstrates that the proposed SA-MAGPPO achieves faster and more stable learning compared to all baseline methods. Random, greedy, and rule-based approaches remain nearly constant over training episodes. This suggests a lack of learning capability. SPPO shows gradual improvement but suffers from slower convergence due to a lack of coordination among agents. MAGPPO converges faster than SPPO, which confirms the benefits of multi-agent learning. However, it exhibits fluctuations under dynamic conditions. Compared to all these, SA-MAGPPO achieves both fast convergence and stable performance, due to the integration of reliability-aware and anomaly-aware feedback in the learning process.

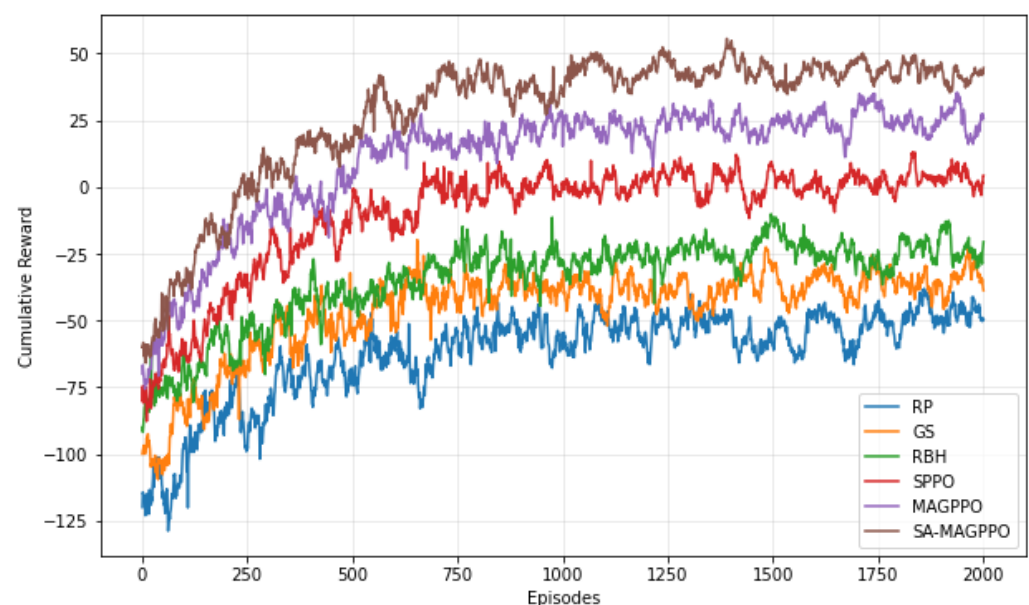


Figure 4. Performance of cumulative reward convergence.

First, we compare the performance of energy over varying parameters, including the number of vehicles, number of UAVs, MEC servers, and task arrival rates in Figure 5. The results in the four subfigures show the energy consumption trends of all methods under different system conditions. In all cases, energy consumption increases with system load. However, SA-MAGPPO consistently achieves the lowest energy usage compared to all baselines. This demonstrates the benefit of security-aware and reliability-driven decision-making. As the number of vehicles increases, SA-MAGPPO scales more efficiently by reducing unnecessary communication and offloading overhead. With more UAVs and edge servers, energy consumption decreases overall due to improved offloading opportunities. Here, SA-MAGPPO uses these resources more effectively than other methods. Under higher task arrival rates, all methods experience increased energy demand, but SA-MAGPPO maintains a clear advantage. This is due to the adaptation to dynamic workload conditions. Overall, the results confirm that SA-MAGPPO provides superior energy efficiency and scalability in complex environments. SA-MAGPPO also consistently achieves the lowest operational cost across all four scenarios, as shown in Figure 6. The reduction is more significant during morning and evening peak hours, where traffic congestion and passenger demand are high. Traditional, greedy, and rule-based methods result in higher costs due to short-sighted decision-making and a lack of global coordination. SPPO improves performance by learning a policy over time, and MAGPPO further reduces cost through coordinated multi-agent decisions. However, SA-MAGPPO outperforms all baselines by avoiding unreliable network conditions and suboptimal routing decisions. This leads to better long-term cost efficiency.

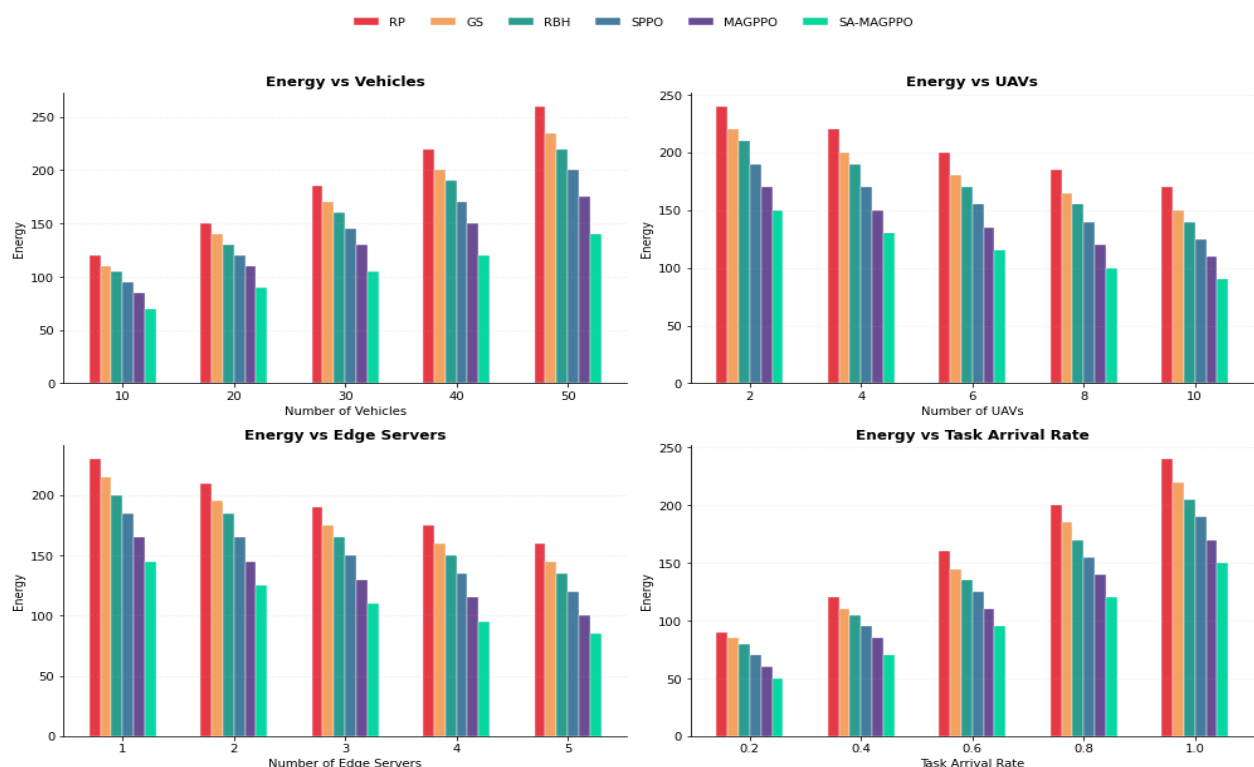


Figure 5. Energy consumption over varying parameters including number of vehicles, number of UAVs, MEC servers, and task arrival rates.

In Figure 7, we compare the performance of energy consumption. The energy consumption results show a consistent reduction achieved by SA-MAGPPO across all scenarios. During peak hours, the improvement is more evident due to optimized routing and reduced redundant vehicle movement. Greedy and rule-based methods consume more

energy because they rely on fixed or locally optimal decisions without considering the global system state. SPPO reduces energy consumption moderately, and MAGPPO further improves efficiency through cooperation among agents. SA-MAGPPO achieves the best performance by jointly considering mobility, computation, and energy constraints under dynamic network conditions. Similarly, SA-MAGPPO results in the lowest CO₂ emissions across all scenarios as presented in Figure 8. This reduction is mainly due to improved routing efficiency, reduced idle time, and better alignment between demand and supply. Greedy and rule-based approaches produce higher emissions because of inefficient fleet utilization. SPPO reduces emissions by learning adaptive policies. MAGPPO further improves coordination among agents. However, SA-MAGPPO achieves additional gains by avoiding decisions influenced by unreliable or anomalous system states, thereby ensuring environmentally efficient operation. Meanwhile, passenger delay is significantly reduced by SA-MAGPPO at all times, as presented in Figure 9, particularly during peak demand periods. Greedy and rule-based methods show the highest delays due to their inability to adapt to dynamic demand variations. SPPO improves delay by learning adaptive policies, and MAGPPO reduces it further through multi-agent coordination. SA-MAGPPO achieves the lowest delay by integrating demand prediction, reliable communication awareness, and efficient offloading decisions, which improves responsiveness to real-time system changes.

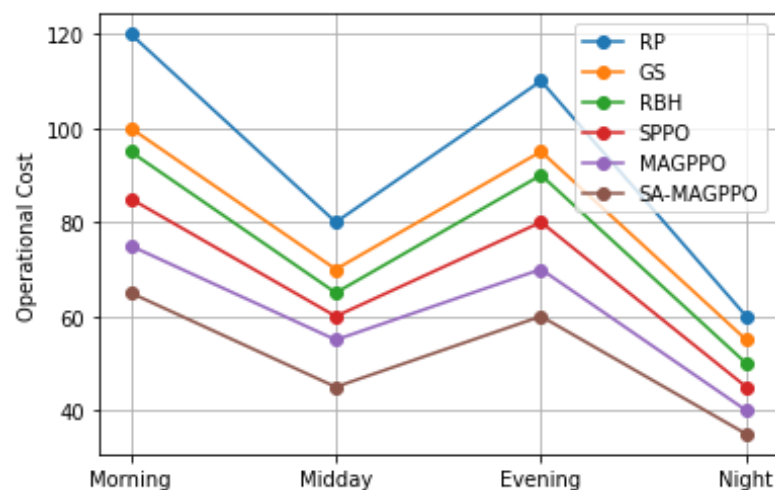


Figure 6. Operational cost over different times.

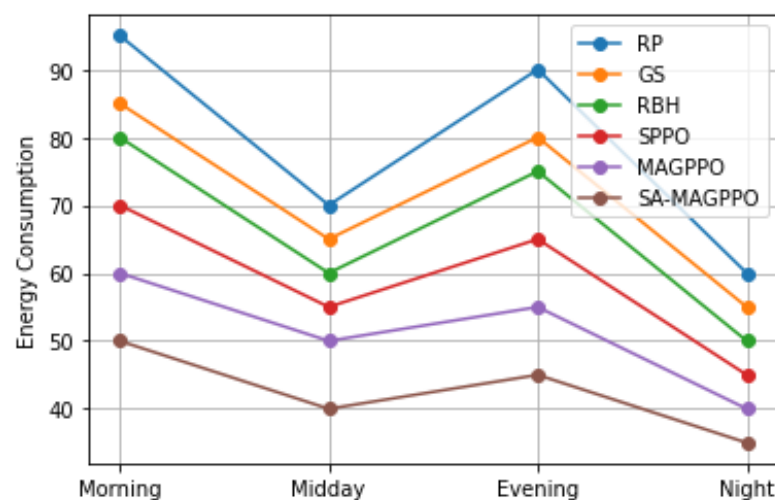


Figure 7. Energy consumption over different times.

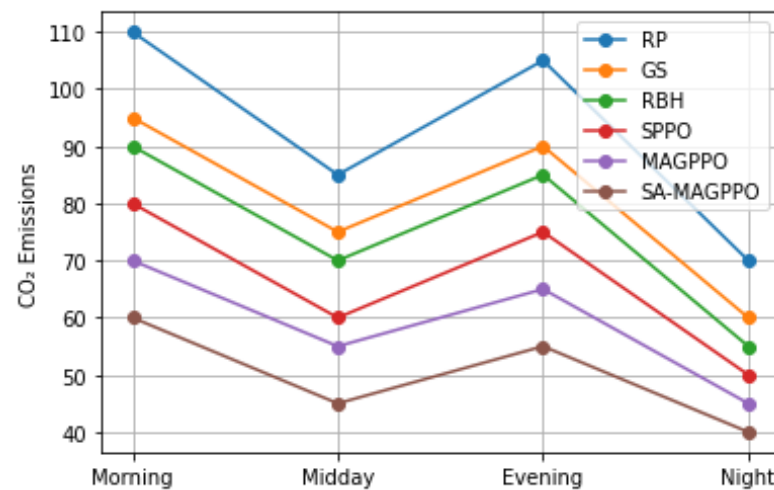


Figure 8. Emission passenger delay over different times.

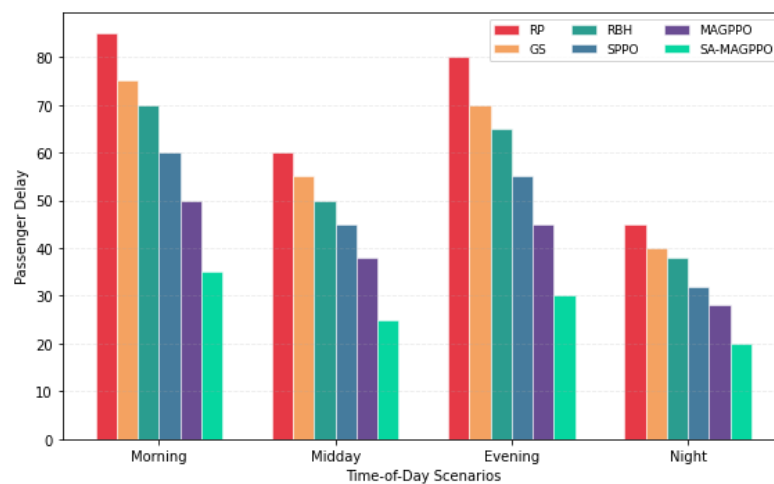


Figure 9. Passenger delay over different times.

The robustness evaluation in Figure 10 shows that SA-MAGPPO maintains stable performance even under increasing anomaly injection rates. In contrast, greedy and rule-based methods degrade rapidly due to their dependence on raw system inputs. SPPO shows moderate sensitivity to anomalies, and MAGPPO performs better due to cooperative decision-making. However, SA-MAGPPO demonstrates the highest robustness because it explicitly incorporates anomaly-aware and reliability-aware signals in both the state representation and advantage estimation. This reduces the impact of corrupted or unreliable information on policy decisions. We also computed the robustness results based on attack levels as shown in Figure 11. When the attack level increases from 0% to 40%, all baseline methods exhibit a gradual decline in performance. However, the rate of degradation varies significantly across methods. RP shows the highest sensitivity to disturbances, with a steep performance drop as it relies entirely on unstructured decision-making. Greedy and rule-based approaches also degrade rapidly due to their dependence on instantaneous and unreliable system observations without long-term adaptation. SPPO demonstrates improved stability compared to classical methods. However, they still experience noticeable performance loss under increasing anomaly levels. This is due to its lack of explicit modeling of communication reliability and corrupted state information. MAGPPO provides further improvement by reducing degradation compared to SPPO. The multi-agent coordination helps mitigate some effects of dynamic disturbances, but performance still declines under high anomaly rates because the learning process does not

explicitly account for reliability variations. In contrast, SA-MAGPPO achieves the most stable performance across all attack levels. The degradation curve remains significantly flatter compared to all baselines, which shows strong resilience to anomalies and communication failures. This improvement is due to reliability-aware state representation and anomaly-sensitive learning. This reduces the influence of corrupted or unreliable inputs during decision-making. As a result, SA-MAGPPO maintains consistent performance even under high disturbance conditions, demonstrating superior robustness in dynamic and uncertain transit environments.

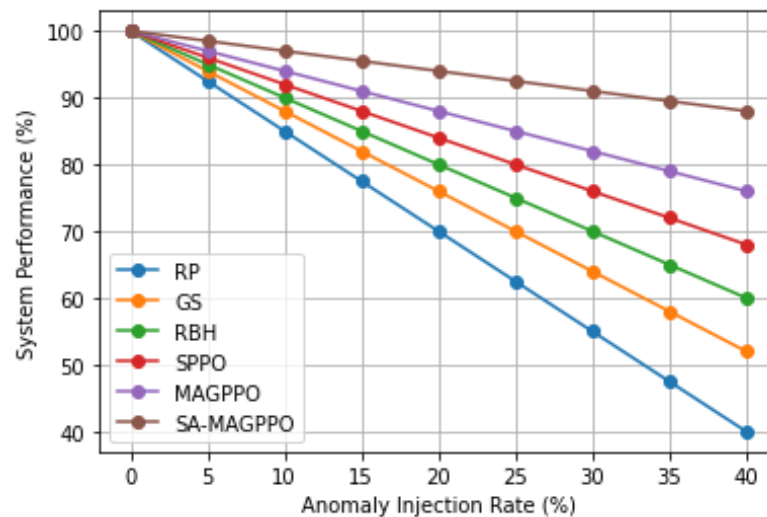


Figure 10. System Performance under Increasing Anomaly Injection Rates.

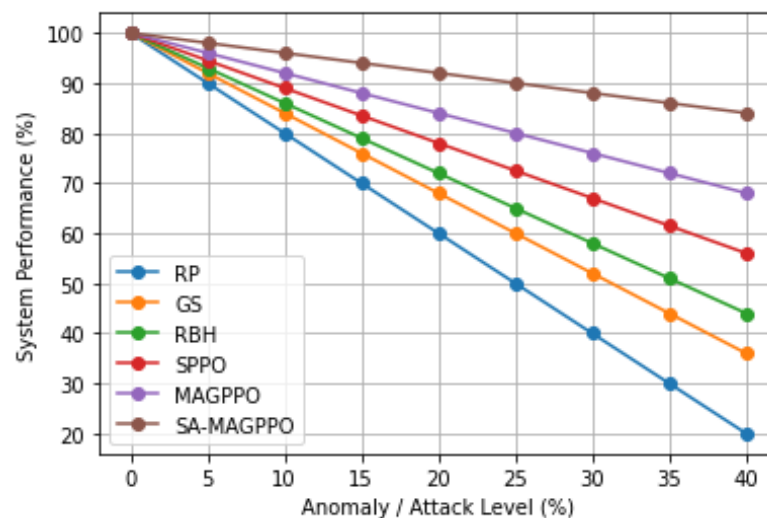


Figure 11. System Performance under Increasing Attack Levels.

Figure 12 presents the ablation study of the proposed SA-MAGPPO framework under different anomaly levels. The results show that both reliability-aware and anomaly-aware mechanisms individually improve the performance of the conventional MAGPPO. The complete SA-MAGPPO consistently achieves the highest system performance across all attack levels, demonstrating that the two mechanisms provide complementary benefits and jointly enhance the robustness of transit operations. Overall, compared with the strongest baseline (MAGPPO), the proposed SA-MAGPPO achieves an average reduction of approximately 16% in energy consumption across different simulation scenarios. Furthermore, under increasing anomaly and attack levels, SA-MAGPPO improves the average system

performance by approximately 21%, demonstrating its superior robustness and resilience against unreliable communication and anomalous transportation observations.

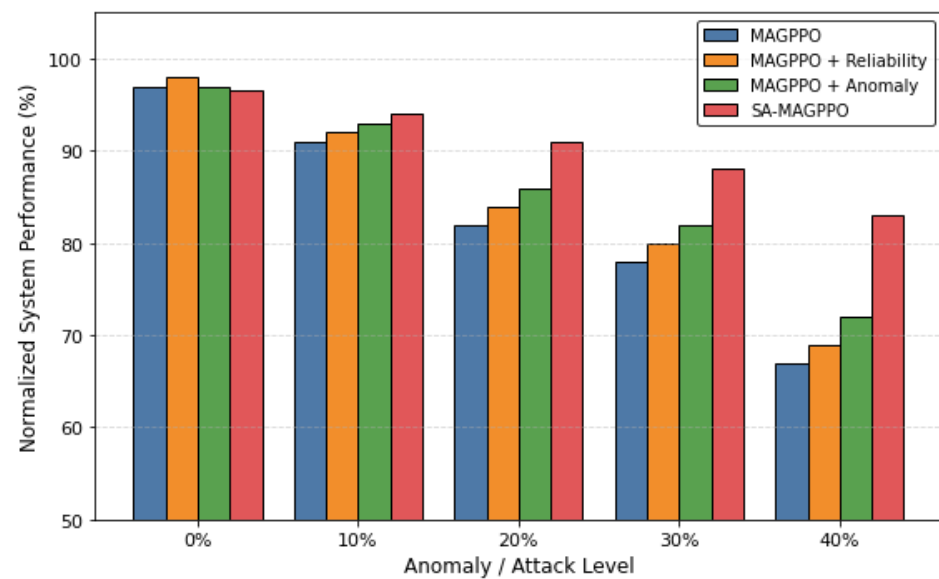


Figure 12. Ablation analysis.

7. Conclusions

This paper proposed a secure and resilient UAV-assisted edge-enabled framework for optimizing public transit operations using the Security-Aware Multi-Agent Proximal Policy Optimization (SA-MAGPPO) algorithm. By incorporating key factors such as passenger demand, traffic congestion, battery state-of-charge, communication reliability, anomaly-aware observations, and computation offloading, the proposed framework effectively reduces energy consumption and CO₂ emissions while maintaining reliable transportation services in dynamic urban environments. Furthermore, UAV-assisted low-altitude communication support improves V2I and V2V connectivity and reduces communication congestion during high-demand operational conditions. The simulation results demonstrate that the proposed SA-MAGPPO framework consistently outperforms conventional RP, GS, RBH, SPPO, and standard MAGPPO approaches, particularly under congested and unreliable network conditions, due to its coordinated multi-agent decision-making and reliability-aware learning capability. In particular, the proposed framework achieves significant reductions in operational cost and energy consumption while maintaining stable performance under anomalous and communication-disrupted environments. Overall, the proposed framework highlights the potential of integrating edge computing, UAV-assisted low-altitude communication support, and multi-agent deep reinforcement learning for secure, sustainable, and intelligent transportation systems.

Future work may focus investigating hybrid methods, including LLM reasoning, to improve the reasoning in offloading decisions. Moreover, the conditional value at risk may be incorporated in the reward section to minimize the probability of catastrophic failures. Finally, the applications of offloading in terms of metaverse applications can also be explored to enhance the applicability of the proposed framework. We will also consider integrating real-world urban road network datasets to enable geographic visualization of traffic congestion, including heat map-based analysis of regional traffic conditions and congestion mitigation.

Author Contributions: M.M.: Conceptualization, Methodology, Validation, Writing—Review and Editing. E.M.: Conceptualization, Methodology, Software, Writing—Original Draft. S.K.u.Z.: Data Curation, Validation, Formal analysis. F.R.: Supervision, Project administration, and Resources. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors acknowledge the use of OpenAI's ChatGPT and Gemini to assist with language refinement and enhancement of figure quality. All scientific content, technical contributions, and final manuscript revisions were verified and approved by the authors.

Conflicts of Interest: Author Muhammad Mustafa was employed by the company Foothill Transit. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Khameneh, R.; Barker, K.; Ramirez-Marquez, J. A hybrid machine learning and simulation framework for modeling and understanding disinformation-induced disruptions in public transit systems. *Reliab. Eng. Syst. Saf.* **2025**, *255*, 110656. [\[CrossRef\]](#)
2. Passman, D.; O'Hara, S.; Plummer, Y. Understanding the role of public transportation in supporting the care economy in Washington, DC, USA. *Sustainability* **2024**, *16*, 1288. [\[CrossRef\]](#)
3. Kapatsila, B.; Grisé, E.; Crumley, M.; El-Geneidy, A. Empirical analysis of battery-electric bus transit operations in Portland, OR, USA. *Transp. Res. Part D Transp. Environ.* **2024**, *128*, 104120. [\[CrossRef\]](#)
4. Mustafa, E.; Shuja, J.; Rehman, F.; Riaz, A.; Maray, M.; Bilal, M.; Khan, M. Deep Neural Networks meet computation offloading in mobile edge networks: Applications, taxonomy, and open issues. *J. Netw. Comput. Appl.* **2024**, *226*, 103886. [\[CrossRef\]](#)
5. Umer, A.; Ali, M.; Jehangiri, A. Fault tolerant and mobility-aware task offloading and scheduling model for IoT logistics. *J. Supercomput.* **2025**, *81*, 1203. [\[CrossRef\]](#)
6. Mustafa, E.; Shuja, J.; Rehman, F.; Namoun, A.; Ali, M.; Alourani, A. Deep Reinforcement Learning with Dueling DQN for Partial Computation Offloading and Resource Allocation in Mobile Edge Computing. *IEEE Access* **2025**, *13*, 94319–94335. [\[CrossRef\]](#)
7. Khan, A.; Rahman, S.; Ullah, F.; Khattak, M.; Bait-Suwailam, M.; El Sayed, H. V2I-VTL: IoT-Enabled adaptive traffic light controller and emission reduction at intersection. *PeerJ Comput. Sci.* **2024**, *10*, e2507. [\[CrossRef\]](#)
8. Naeem, M.; Chaudhary, S.; Meng, Y. Road to efficiency: V2v enabled intelligent transportation system. *Electronics* **2024**, *13*, 2673. [\[CrossRef\]](#)
9. Lai, B.; Wen, J.; Kang, J.; Du, H.; Nie, J.; Yi, C.; Kim, D.; Xie, S. Resource-efficient generative mobile edge networks in 6G era: Fundamentals, framework and case study. *IEEE Wirel. Commun.* **2024**, *31*, 66–74. [\[CrossRef\]](#)
10. Du, H.; Li, Z.; Niyato, D.; Kang, J.; Xiong, Z.; Shen, X.; Kim, D. Enabling AI-generated content services in wireless edge networks. *IEEE Wirel. Commun.* **2024**, *31*, 226–234. [\[CrossRef\]](#)
11. Yang, N.; Chen, S.; Zhang, H.; Berry, R. Beyond the edge: An advanced exploration of reinforcement learning for mobile edge computing, its applications, and future research trajectories. *IEEE Commun. Surv. Tutor.* **2024**, *27*, 546–594.
12. Khani, M.; Sadr, M.; Jamali, S. Deep reinforcement learning-based resource allocation in multi-access edge computing. *Concurr. Comput. Pract. Exp.* **2024**, *36*, e7995. [\[CrossRef\]](#)
13. Lukic Vujadinovic, V.; Damjanovic, A.; Cacic, A.; Petkovic, D.R.; Prelevic, M.; Pantovic, V.; Stojanovic, M.; Vidojevic, D.; Vranjes, D.; Bodolo, I. AI-driven approach for enhancing sustainability in urban public transportation. *Sustainability* **2024**, *16*, 7763. [\[CrossRef\]](#)
14. Sarkar, S.R.; Tahura, S. AI-Enhanced Public Transportation Systems for Urban Sustainability. In *AI and Emerging Technologies for Emergency Response and Smart Cities*; IGI Global Scientific Publishing: Hershey, PA, USA, 2025; pp. 97–116.
15. Mohsen, B.M. AI-driven optimization of urban logistics in smart cities: Integrating autonomous vehicles and IoT for efficient delivery systems. *Sustainability* **2024**, *16*, 11265. [\[CrossRef\]](#)
16. Thuraka, B. AI-Driven Adaptive Route Optimization for Sustainable Urban Logistics and Supply Chain Management. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **2021**, *7*, 667–684.
17. Samaei, S.R. A comprehensive algorithm for AI-driven transportation improvements in urban areas. In Proceedings of the 13th International Engineering Conference on Advanced Research in Science and Technology, Bandar-e Gaz, Iran, 18 January 2024. Available online: <https://civilica.com/doc/1930041> (accessed on 24 July 2026).

18. Tsolakis, N.; Zissis, D.; Papaefthimiou, S.; Korfiatis, N. Towards AI driven environmental sustainability: An application of automated logistics in container port terminals. *Int. J. Prod. Res.* **2022**, *60*, 4508–4528. [\[CrossRef\]](#)
19. Huang, R.; Mao, S. Carbon footprint management in global supply chains: A data-driven approach utilizing artificial intelligence algorithms. *IEEE Access* **2024**, *12*, 89957–89967. [\[CrossRef\]](#)
20. Ayman, A.; Wilbur, M.; Sivagnanam, A.; Pugliese, P.; Dubey, A.; Laszka, A. Data-driven prediction of route-level energy use for mixed-vehicle transit fleets. In *Proceedings of the 2020 IEEE International Conference on Smart Computing (SMARTCOMP)*; IEEE: Piscataway, NJ, USA, 2020; pp. 41–48.
21. Nigro, M.; Ferrara, M.; De Vincentis, R.; Liberto, C.; Valenti, G. Data driven approaches for sustainable development of E-mobility in urban areas. *Energies* **2021**, *14*, 3949. [\[CrossRef\]](#)
22. Zhu, Y.; Zhao, X.; Zhang, X. A Secure GNN-MADDPG Framework with Combinatorial Action Optimization for Task Offloading in Vehicular Networks. *Veh. Commun.* **2026**, *58*, 101000. [\[CrossRef\]](#)
23. Fan, J.; Xun, Y.; Wang, X.; Yang, H. Task offloading in vehicular edge computing based on traffic density-driven task generation. *Ad Hoc Netw.* **2026**, *186*, 104203. [\[CrossRef\]](#)
24. Kothari, V.; Tambe, S.; Franklin, A. Deadline-Aware Adaptive Fuzzification for Task Offloading in Vehicle-to-Vehicle Networks. In *Proceedings of the 2026 18th International Conference on COMMunication Systems and NETWORKS (COMSNETS)*; IEEE: Piscataway, NJ, USA, 2026; pp. 1150–1154.
25. Ni, Z.; Chen, H.; Xue, H.; Lin, K.; Chen, N.; Liu, W.; Yu, J. Dependency-Aware and Energy Efficient Task Offloading in Vehicular Edge Computing Systems. *IEEE Trans. Veh. Technol.* **2026**, *75*, 11634–11646. [\[CrossRef\]](#)
26. Al-Raei, M. The smart future for sustainable development: Artificial intelligence solutions for sustainable urbanization. *Sustain. Dev.* **2025**, *33*, 508–517. [\[CrossRef\]](#)
27. Alahmari, N.; Mehmood, R.; Alzahrani, A.; Yigitcanlar, T.; Corchado, J.M. Autonomous and Sustainable Service Economies: Data-Driven Optimization of Design and Operations through Discovery of Multi-perspective Parameters. *Sustainability* **2023**, *15*, 16003. [\[CrossRef\]](#)
28. Wang, Y.; Zhang, W.; Deng, H.; Li, X. Efficient resource allocation for security-aware task offloading in MEC system using DVS. *Electronics* **2022**, *11*, 3032. [\[CrossRef\]](#)
29. Li, X.; Wei, B.; Yang, X.; Tolba, A.; Zeng, Z.; Alfarraj, O. A Spectrum Allocation and Security-Sensitive Task Offloading Algorithm in MEC Using DVS. *Comput. Mater. Contin.* **2025**, *85*, 3437. [\[CrossRef\]](#)
30. Wu, M.; Wu, H.; Lu, W.; Guo, L.; Lee, I.; Jamalipour, A. Security-aware designs of multi-UAV deployment, task offloading and service placement in edge computing networks. *IEEE Trans. Mob. Comput.* **2025**, *24*, 11046–11060. [\[CrossRef\]](#)
31. Yin, P.; Liang, W.; Wen, J.; Kang, J.; Chen, J.; Niyato, D. Multi-agent DRL for multi-objective twin migration routing with workload prediction in 6G-enabled IoV. *IEEE Trans. Mob. Comput.* **2026**, *25*, 12222–12236. [\[CrossRef\]](#)
32. Xiao, Z.; Chen, Y.; Li, T.; Tian, S.; Min, G.; Tang, Z.; Li, K.; Li, K. Toward Secure Task Offloading in Vehicular Edge Computing Networks. *IEEE Trans. Dependable Secur. Comput.* **2026**, Early access. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.