Reliability and Traffic Aware Resource Allocation for UAV-assisted Vehicular O-RAN

Hayla Nahom Abishu, *Member, IEEE,* Ahmed Badawy, *Member, IEEE,* Amr Mohamed, *Senior Member, IEEE,* and Carla Fabiana Chiasserini, *Fellow, IEEE*

***Abstract*—The rapid advancements of next-generation vehicular networks require intelligent, low-latency, and efficient resource management to support heterogeneous services. In this work, we propose a Traffic-aware Dynamic Resource Allocation (TADRA) architecture for UAV-assisted vehicular O-RAN to address the challenges of dynamic traffic conditions, infrastructure failures, and stringent quality of service (QoS) requirements. Due to the dynamic mobility and flexible deployment characteristics, UAV Open Radio Units (O-RUs) in the TADRA architecture support the terrestrial infrastructure under overload or failure conditions, dynamically extending coverage, balancing traffic loads, and restoring service to maintain uninterrupted QoS across diverse and heterogeneous traffic demands. Unlike existing static or single-layer solutions, our proposed TADRA integrates RAN Intelligent Controllers (RICs) with a Hierarchical Traffic-Aware Multi-Agent Twin-Delayed (TMT) algorithm to optimize the allocation of computation and radio resources. This joint optimization problem is NP-hard, highly dynamic, and coupled across agents, making TMT a tractable and adaptive alternative. This hierarchical framework performs traffic prioritization at the upper (application) layer and resource allocation at the lower (MAC) layer, facilitating adaptive decision-making under diverse vehicular traffic patterns. Numerical results demonstrate that our solution provides substantial gains over MATD3, MADDPG, and GA, achieving 17% lower latency, 10% higher throughput, 14% lower energy consumption, and 6.5% higher reliability.**

***Index Terms*—Dynamic resource allocation, open RAN, UAV, vehicular network.**

I. Introduction

With the advent of next-generation vehicular networks, ultra-reliable and low-latency communications (URLLC), massive machine-type communication (mMTC), and high-throughput enhanced mobile broadband (eMBB) have become increasingly important [1]. URLLCs support mission-critical functions, such as collision avoidance, by providing high reliability and ultra-low latency [2]. On the other hand, eMBB supports high-speed services such as infotainment and augmented reality, placing a greater emphasis on data rate rather than strict latency constraints. Meanwhile, mMTC supports a wide range of low-power devices, such as vehicle sensors, cameras, and LiDAR units—essential for environment monitoring in autonomous driving. These stringent and heterogeneous service requirements in autonomous driving and mobility-oriented applications have motivated the adoption of the Open Radio Access Network (O-RAN) architecture [1], [3]. This is because O-RAN uses a modular and disaggregated architecture that separates hardware and software components, thereby enabling multi-vendor interoperability through software-defined controls. To take advantage of this architecture, RAN Intelligent Controllers (RICs) are integrated into O-RAN, which provide AI-driven analytics, policy enforcement, and radio resource control. This integration includes both Near-Real-Time (Near-RT RIC) and Non-Real-Time (Non-RT RIC), enabling the O-RAN system to respond effectively to varying traffic conditions in real time [4], [5].

However, meeting these heterogeneous and stringent service requirements in vehicular O-RAN environments remains challenging. In particular, the existing resource allocation solutions based on the conventional static or heuristic, single-agent learning, and flat-structure multi-agent learning-based resource allocation algorithms often fail to adapt well in such complex and dynamic dense urban settings, where traffic load, user mobility, and service heterogeneity fluctuate rapidly, resulting in increased latency, packet loss, or degradation of service for critical URLLC traffic. Furthermore, in dynamic and congested traffic conditions or terrestrial O-RAN infrastructure failures, balancing these heterogeneous service requirements and efficiently allocating network resources is a complex and challenging task [6].

Motivated by the above-mentioned challenges and the advantage of the O-RAN architecture, this paper proposes a Traffic-aware Dynamic Resource Allocation (TADRA) architecture for a Uncrewed Aerial Vehicle (UAV)-assisted vehicular O-RAN system. The proposed framework leverages the intelligence and flexibility of the RIC architecture to dynamically prioritize and allocate radio and computational resources based on real-time network state and application-level QoS requirements. In this framework, we deploy UAVs as aerial Open Radio Units (O-RUs) to provide radio and computation resources in emergency scenarios. Due to the high-dimensional, dynamic, and multi-agent nature of autonomous driving systems, coupled with stringent real-time requirements for their services, heuristic and conventional learning-based algorithms struggle to achieve nearly optimal resource allocation decisions. Thus, by leveraging the Multi-Agent Twin Delayed Deep Deterministic Policy Gradient (MATD3) algorithm, we develop a Hierarchical Traffic-aware Multi-Agent Twin-Delayed (TMT) algorithm, which makes dynamic and optimal resource allocation decisions.

H. N. Abishu, A. Badawy, and A. Mohamed are with the Department of Computer Science and Engineering, Qatar University, Doha, Qatar. C. F. Chiasserini is with Politecnico di Torino, Torino, Italy, with CNIT, Parma, Italy, and CNR-IEIIT, Torino, Italy. Research reported in this publication was supported by the Qatar Research Development and Innovation Council ARG01-0525-230339. The content is solely the responsibility of the authors and does not necessarily represent the official views of Qatar Research Development and Innovation Council.

Existing works on UAV-assisted vehicular O-RAN address reliability, offloading, and resource allocation, but several challenges remain. The high-dimensional and time-varying nature of vehicular traffic, coupled with the dynamic topology and mobility of UAVs and vehicles, as well as the heterogeneous traffic types, leads to a highly complex and large-scale multi-agent resource allocation problem characterized by non-convexity and stringent real-time latency constraints. Despite recent progress, existing research remains limited in addressing these challenges through efficiently integrated UAV-assisted O-RAN frameworks that jointly optimize aerial and terrestrial resources. These gaps highlight the need for hierarchical MADRL-based resource allocation approaches that provide a unified, adaptive multi-objective framework with real-time adaptability under dynamic and failure-prone conditions, as well as traffic-aware architectures capable of maintaining QoS across heterogeneous vehicular services like the one proposed in this work. Our framework leverages xApps deployed in the Near-RT RIC to analyze traffic patterns, user mobility, and service priorities, while coordinating with the Non-RT RIC for long-term policy optimization. The proposed framework dynamically orchestrates radio and computation resources across aerial and terrestrial network infrastructures based on the traffic conditions to ensure the QoS requirements of each traffic class.

The main contributions of this work are summarized as follows.

- We propose a traffic-aware dynamic resource allocation framework for UAV-assisted vehicular O-RAN systems. In this framework, O-RUs are located both on ground and aerial networks, while Open Distributed Units (O-DUs) and Open Centralized Units (O-CUs) are deployed on ground or edge servers, allowing for seamless O-RU reassignment in response to ground unit overloads or failures. In the case of variable loads in the vehicular network environment, mobility-induced disruptions, and potential ground infrastructure failures, UAVs serve as mobile and intelligent O-RAN entities capable of filling coverage and computational gaps during disasters or congestion.
- We formulate the joint traffic prioritization and resource allocation decision-making problem in the TADRA architecture as a Multi-Objective Markov Decision Process (MOMDP) and then introduce a novel traffic-aware multi-agent learning approach, the TMT algorithm, that integrates hierarchical decision-making to manage heterogeneous vehicular traffic (URLLC, eMBB, mMTC) in a dynamic O-RAN environment.
- We integrate a lightweight learning technique, Reinforcement Learning (RL) with Linear Function Approximation, into the proposed TMT framework to predict traffic demand and classify flows into URLLC, eMBB, and mMTC based on real-time network conditions, mobility patterns, and application context. In the upper layer of the framework, we fully use TMT for traffic prioritization policy to prioritize URLLC, eMBB, and mMTC traffic types. This module is also responsible for handling complex and high-dimensional decision-making while coordinating multiple interacting agents (vehicles, ground and UAV O-RUs, and RIC) that have heterogeneous QoS requirements. In contrast, we use RL with linear function approximations in the lower layer to efficiently allocate radio and computation resources to prioritized traffic in real time.
- Through extensive experiments, we evaluate the performance of our proposed approach compared to state-of-the-art methods in terms of latency, throughput, reliability, and energy consumption per traffic class. Our results demonstrate that the proposed approach achieves 17% lower latency, 10% higher throughput, 6.5% higher reliability, and 14% lower energy consumption compared to the benchmark algorithms.

The rest of the paper is organized as follows. Section II reviews some relevant related work. Section III presents the proposed system model, while our optimization problem is formulated in Section IV. Section V discusses the proposed framework, which is the evaluated and benchmarked in Section VI. Finally, Section VII draws our conclusions and highlights directions for further research.

II. Related work

A large body of work has investigated resource allocation in O-RAN environments, aiming to enhance spectrum use and energy efficiency [7]. However, most of these studies rely on specific stationarity assumptions, show slow convergence of their optimization or learning algorithms, or impose significant computational complexity, thereby limiting their real-time application. In particular, the work in [8] presented a lightweight online learning framework that combines Thompson Sampling with probabilistic inference for energy-efficient and robust resource allocation decisions in virtualized base stations (BSs). [3] proposed, instead, a resource allocation threshold policy to balance the performance and energy consumption of virtualized BSs in an O-RAN environment. To achieve this, the authors introduced an online learning technique that operates under minimal assumptions and requires little knowledge of the environment. The proposed framework is designed to be capable of handling non-stationary and adversarial demands and conditions. Additionally, this study employs a meta-learning scheme that selects the algorithm that performs best based on the environment when other algorithms are available. The study in [9] explored a cell-free massive MIMO system in the virtualized O-RAN environment. This work attempts to minimize power consumption by jointly optimizing the radio, optical fronthaul, and virtualized cloud resources allocation. In this work, two schemes are compared: (1) a radio resource optimization scheme, which optimizes radio resources independently of the cloud, and (2) a local cloud coordination scheme, which performs orchestration only between radio units within the same cluster. To solve this problem more manageably for larger cell-free setups, the study proposed a concave-convex programming algorithm (CCP).

The research in [10] presented an RL-based resource allocation framework aimed at optimizing the distribution and processing of requests between the near-RT RIC and non-RT RIC, thereby improving request acceptance and service latency. To address the resource allocation problem, the authors applied a Double Deep Q Network (DDQN) with two variations: with a cache and without a cache. In this approach, each request type is intelligently assigned to the near-RT RIC or non-RT RIC resource to finish the request. Similarly, an RL-based two-layer network slicing architecture is presented in [1] for URLLC and eMBB, targeting latency reduction for IoT applications in a shared core and RAN in the O-RAN architecture. The architecture enables end-to-end network slicing across both the core and edge network domains, facilitating radio resource allocation based on eMBB and URLLC service requirements of end users. The formulated optimization problem is modeled as a Markov decision process (MDP) at both the high and low levels of the framework, and a hierarchical RL-based solution is presented to solve it. At a higher level, a Double Deep Q-network (DDQN)-based SDN controller allocates radio resources to gNBs (next-generation NodeB), while at the lower level, each gNodeB trained using a DDQN allocates its pre-allocated resources to its end-users. [11] presented a DRL, i.e., Proximal Policy Optimization (PPO) and a meta-learning-based resource allocation strategy tailored for 6G O-RAN environments, supporting heterogeneous services such as eMBB and URLLC. The proposed strategy optimizes network resource management to improve energy efficiency and reduce URLLC latency in highly dynamic channel and traffic conditions. The model uses meta-learning to swiftly adjust to fluctuating network conditions and more accurately estimate traffic distributions across multiple connections.

Furthermore, [6] proposed a multi-UAV-enabled O-RAN resource allocation architecture in which UAVs serve as aerial BSs to augment the terrestrial infrastructure during emergencies. The research formulated a joint UAV trajectory and task offloading optimization problem and presented a multi-agent RL (MARL) framework combined with online learning to achieve near-optimal solutions. Finally, the work in [4] presented a UAV-assisted O-RAN architecture in which UAVs complement terrestrial networks to extend wireless coverage and enhance energy efficiency. The authors formulated the joint optimization problem of O-RU assignment, the deployment of aerial O-RUs, and the allocation of O-RU resources as a non-convex optimization, maximizing energy efficiency. To solve this complex non-convex problem, a double-loop iterative algorithm is created. The outer loop applies the Dinkelbach method to handle the fractional objective function. In contrast, the inner loop applies a Block Coordinate Descent-based iterative optimization framework to efficiently solve the decoupled subproblems.

Progress beyond the state of the art. In summary, most of the existing studies focus on improving the allocation and orchestration of virtualized O-RAN resources in terrestrial infrastructures, with the aim of minimizing energy consumption and end-to-end latency, either in combination or independently. Such works have applied either heuristic algorithms or single-agent RL approaches to solve the formulated optimization problems, but often fail to perform well in complex and dynamic conditions. Further research efforts expand this focus to UAV-assisted O-RANs using multi-agent RL, focusing on issues such as user association, aerial O-RU deployment, UAV trajectory, or task offloading optimization, with the primary aim of increasing energy efficiency. Despite these advancements on multi-agent coordination using classical RL, they still struggle to provide the best or most adaptable solutions when the network is dynamic, high-dimensional, and complex. To the best of our knowledge, our study is the first to investigate hierarchical and traffic-aware resource allocation in UAV-assisted networks where O-RUs placed on UAVs serve as aerial extensions of the network, dynamically providing resources when terrestrial O-RUs become overloaded or fail due to disasters or interruptions. Also, the framework intelligently orchestrates both radio and computational resources based on the heterogeneous requirements of the URLLC, eMBB, and mMTC traffic classes, by leveraging traffic prioritization. The result is a traffic-aware hierarchical multi-agent approach that optimally trades off between latency, energy consumption, throughput, and reliability.

III. System Model

We consider vehicular users that generate heterogeneous traffic types, namely, URLLC, eMBB, and mMTC. The URLLC traffic is given by mission-critical safety messages, such as those enabling cooperative collision avoidance. The eMBB traffic is represented by infotainment and high-definition map downloads, while mMTC traffic includes vehicle diagnostics and environmental sensing traffic. As depicted in Fig. 1, besides vehicles, the system includes ground and aerial O-RAN components that work collaboratively to ensure resilient communications.

To ensure such resilient and adaptive vehicular connectivity, the three major O-RAN components, such as O-RU, O-DU, and O-CU, are deployed at different locations and interconnected through the standardized open interfaces. The O-RUs are deployed on both UAVs and ground MECs; they perform RF processing, digital front-end operations, and radio transmission and reception. Under normal conditions, the architecture relies primarily on the O-RUs deployed on terrestrial infrastructure. However, when the ground infrastructure fails or becomes excessively overloaded, aerial O-RU functionalities are enacted to extend connectivity and provide continuity to vehicular services. UAV O-RUs dynamically allocate resources based on the three aforementioned types of traffic (i.e., URLLC, eMBB, and mMTC). The O-DUs are located at ground MECs to handle real-time lower-layer operations, including MAC scheduling, channel coding/decoding, and modulation/demodulation. The Open Front-Haul interface connects O-DU to O-RU to efficiently transmit digitized RF data. The O-CU is deployed on the edge server and managed by cloud infrastructure, performing higher-level functions like mobility management, QoS enforcement, and global resource allocation. It also manages multiple O-DUs and their associated O-RUs (both ground and UAV).

To achieve this, the O-CUs connect with O-DUs through the F1 interface and seamlessly coordinate across multiple O-DUs. This hierarchical disaggregation of the O-RAN components optimally balances latency, energy consumption, throughput, and reliability of the traffic classes, URLLC, MMTC, and eMBB, while ensuring scalability. Thus, deploying O-DUs closer to vehicular users reduces end-to-end latency, whereas centralizing O-CUs enables efficient control-plane management and resource pooling. The RIC with a learning-driven optimization algorithm is built into this architecture to make it even more resilient and adaptable. The Near-RT RIC uses the E2 interface to optimize the network and prioritize traffic, while the Non-RT RIC uses the A1 interface to give policy guidance and training feedback. This integrated deployment ensures rapid service restoration, dynamic resource allocation, and reliable connectivity to meet the stringent service requirements of the three traffic classes under mission-critical conditions.

The RIC uses AI-driven near-real-time and non-real-time apps to control the activation of UAV O-RUs, the classification of traffic, and the distribution of resources. Through this intelligent control, prioritized routing of URLLC, eMBB, and mMTC traffic is offloaded to ground O-RUs in normal conditions and to UAV O-RUs in emergency situations. As a result, the proposed system maintains low-latency, reliable, and uninterrupted vehicular connectivity by adapting UAV deployment and processing capabilities in response to real-time network load and vehicular mobility patterns.

A. Communication Model

In our proposed UAV-assisted vehicular O-RAN, the ground segment consists of multiple vehicular users, denoted as $\mathcal{V} = \{1, 2, \ldots, V\}$, and multiple ground O-RUs, denoted as $\mathcal{G} = \{1, 2, \ldots, G\}$. These O-RUs are connected with the vehicles through wireless links to provide radio and computation resources to the vehicular users under normal conditions. In contrast, the aerial segment consists of multiple UAV O-RUs to extend coverage and provide radio and computation resources during ground infrastructure congestion, partial outages, or failures. Let the set of UAV O-RUs be defined as $\mathcal{U} = \{1, 2, \ldots, U\}$. The vehicular users are randomly distributed in the network coverage and generate traffic belonging to one of the traffic classes $\psi(v) \in \{URLLC, eMBB, mMTC\}$. To achieve the stringent requirements of these traffic classes, bandwidth and compute resources are dynamically allocated among these three distinct vehicular traffic classes, based on their specific QoS requirements. For ground O-RUs and vehicles, we use 2D location coordinates, while for UAV O-RUs, we use 3D coordinate spaces, defined as (x_g, y_g), (x_v, y_v), and (x_u, y_u, H), respectively. We define the distance between the vehicular user v and resource provider O-RU $n \in \mathcal{N} = \mathcal{G} \cup \mathcal{U}$ as:

$$d_{v,n} = \begin{cases} d_{v,g} = \sqrt{(x_g - x_v)^2 + (y_g - y_v)^2} \\ d_{v,u} = \sqrt{(x_u - x_v)^2 + (y_u - y_v)^2 + H^2}, \end{cases} \tag{1}$$

where $d_{v,g}$ is the distance between vehicle v and the ground O-RU g and $d_{v,u}$ is the distance between vehicle v and UAV O-RU u.

In our proposed framework, both terrestrial and aerial O-RUs use Orthogonal Frequency Division Multiple Access (OFDMA) is used to divide the frequency bandwidth into orthogonal subcarrier frequencies and dynamically assign them to vehicular users based on channel conditions, enabling fine-grained and interference-free allocation [12]. The framework performs hierarchical and traffic-aware resource allocation decisions based on instantaneous channel status information (CSI), user mobility patterns, and traffic priority indices to optimize subcarrier and computation resources in real time. URLLC traffic is then allocated the most reliable and low-latency channels, eMBB traffic is allocated high-throughput subcarriers when the channel quality indicators (CQIs) are favorable, and mMTC traffic is allocated opportunistically when there is low load [11]. Let the total bandwidth of the O-RU n (ground or UAV) be $\mathcal{B}_n$, divided into subcarriers $\mathcal{K}_n = \{1, 2, \ldots, K_n\}$. Each subcarrier k has a bandwidth $b_{n,k} = \frac{\mathcal{B}_n}{K_n}$ and each subcarrier $k \in \mathcal{K}_n$ of RU n (ground O-RU g or UAV O-RU u) can be assigned to at most one vehicle v at time t, which is expressed as:

$$\sum_{v \in \mathcal{V}} x_{v,n,k}(t) \leq 1, \forall n \in \{g, u\}, \forall k \in \mathcal{K}, \forall t \tag{2}$$

where $x_{v,n,k} \in \{x_{v,g,k}, x_{v,u,k}\} \in \{0, 1\}$ is a binary variable indicating whether the subcarrier k of the ground O-RU g or UAV O-RU u is allocated to vehicle v at time t. On the other hand, each vehicle $v \in \mathcal{V}$ can be assigned a subcarrier from either a ground O-RU or UAV O-RU, but not simultaneously from both. This constraint is given as:

$$\sum_{g \in \mathcal{G}} x_{v,g,k}(t) + \sum_{u \in \mathcal{U}} x_{v,u,k}(t) \leq 1, \forall v \in \mathcal{V}, \forall k \in \mathcal{K}, \forall t. \tag{3}$$

This approach enables exclusive subcarrier allocation and avoids simultaneous transmissions from multiple O-RUs to the same vehicle, ensuring orthogonality and interference-free operation across the vehicular network. The bandwidth allocated to vehicle v by the O-RU n over one or multiple subchannels is expressed as:

$$B_{v,n} = \sum_{k \in \mathcal{K}_n} x_{v,n,k} b_{n,k}, n \in \{g, u\}. \tag{4}$$

The total of bandwidth allocated to all vehicles by a ground or UAV O-RU must not exceed its available bandwidth $\mathcal{B}_n$.

$$\sum_{v \in \mathcal{V}} B_{v,n} = \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_n} x_{v,g,k} b_{n,k}(t) \leq \mathcal{B}_n, n \in \{g, u\}. \tag{5}$$

In this work, we consider two communication links at any given time t: vehicle to ground O-RU (V2G) and vehicle to UAV O-RU (V2U). Unlike the V2U link, which typically maintains a high probability of line-of-sight (LoS) connectivity owing to the elevated deployment and flexible positioning of UAVs, the terrestrial V2G link is more exposed to signal obstruction and degradation caused by static and dynamic obstacles such as buildings, trees, and surrounding vehicles. These environmental factors introduce shadowing, scattering, and multipath effects, which considerably degrade the channel quality and increase the probability of non-line-of-sight (NLoS) conditions. We use Rayleigh fading to model NLoS links, where the direct path is blocked and the received signal consists of multiple scattered components. In NLoS links, the

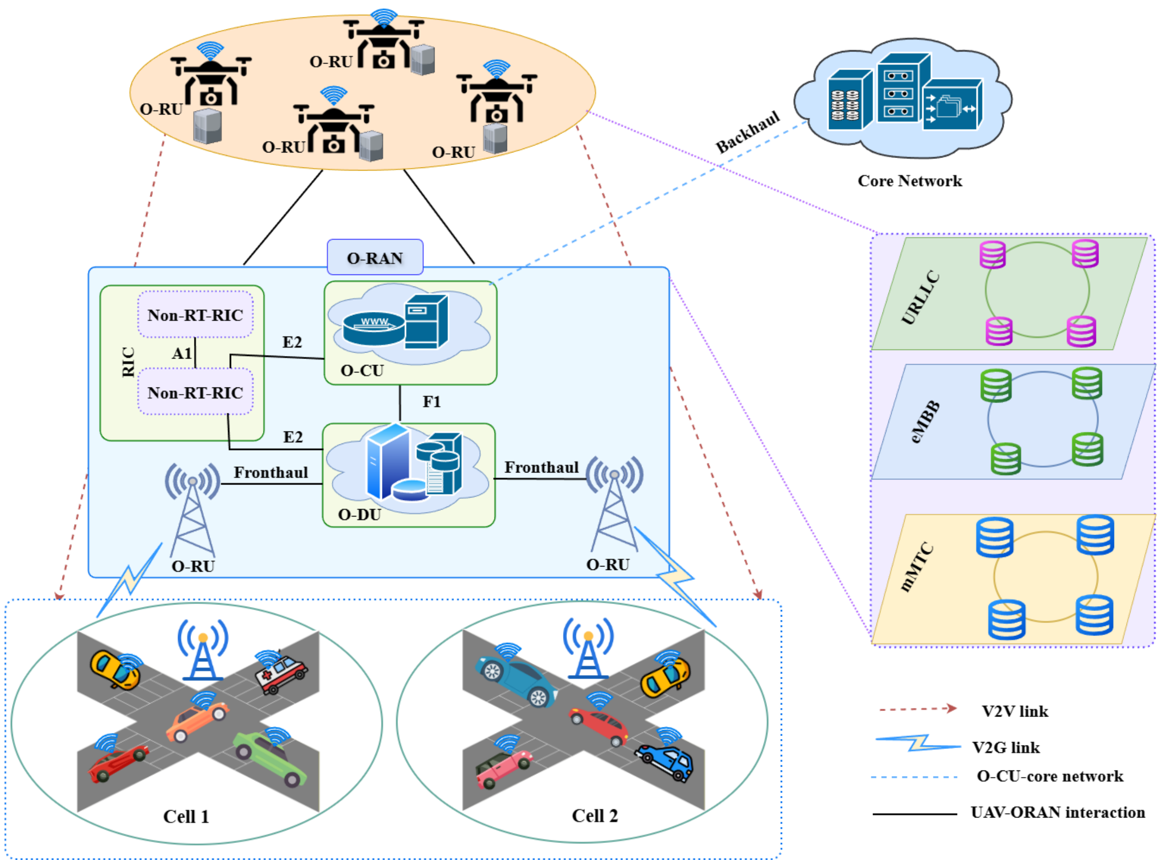

Fig. 1: System model for a traffic-aware dynamic resource allocation in UAV-assisted vehicular O-RAN.

large-scale fading follows the log-distance path loss model, while small-scale fading is Rayleigh-distributed, reflecting the lack of a dominant LoS path and presence of extensive multipath scattering [9]. Accordingly, the channel gain between vehicle v and the ground O-RU g over subchannel k at time t can be expressed as

$$G_{v,g,k}(t) = G_g \frac{\rho_0(f_c)}{[d_{v,g}(t)]^{\varrho}} \left|h^{\mathrm{NLoS}}_{v,g,k}(t)\right|^2, \tag{6}$$

where G_g denotes the antenna gain of the ground O-RU g, $\rho_0(f_c)$ is the reference channel power gain at a distance of 1 m for the carrier frequency f_c, $d_{v,g}(t)$ is the distance between vehicle v and the ground O-RU g, and ϱ is the path loss exponent. The term $|h^{NLoS}_{v,g}(t)|^2$ denotes the small-scale fading power gain, which reflects the fast fluctuations due to multipath propagation in dense urban settings. The instantaneous received power at vehicle v from the O-RU g on the subcarrier k is given as $P^{rx}_{v,g,k}(t) = p_{g,k} G_{v,g,k}(t)$, where $p_{g,k}$ is the transmit power the O-RU g assigns to subcarrier k.

The corresponding NLoS path loss for the V2G link, expressed in dB, is given by

$$L^{NLoS}_{v,g,k}(t) = 20\log_{10}\big(\frac{4\pi f_c d_{v,g}(t)}{c}\big) + 10\varrho\log_{10}\big(d_{v,g}(t)\big) + \chi^{NLoS}, \tag{7}$$

where χ^{NLoS} and ϱ denote the shadowing loss in dB and the path loss exponent. In this formulation, the first term corresponds to the free-space path loss, while the second term represents the additional attenuation induced by environmental obstacles, making this model suitable for realistic obstructed vehicular scenarios. The SINR at vehicle v receiving from the ground O-RU g over subchannel k at a time t is given as:

$$SINR_{v,g,k}(t) = \frac{p_{g,k} G_{v,g,k}(t)}{\sigma^2 + \sum_{g' \in \mathcal{G}, g' \neq g} p_{g',k} G_{v,g',k}(t)} \tag{8}$$

where $p_{g,k}$ denotes the transmission power of the ground O-RU, $G_{v,g,k}$ is the V2G channel gain, and σ^2 is the noise power at the receiver. In eqn. (8), the term $\sum_{g' \in \mathcal{G}, g' \neq g} p_{g',k} G_{v,g',k}$ is co-channel interference from all other transmitting ground O-RUs. The transmission rate between the vehicle v and the ground O-RU g on the subcarrier k is expressed using the Shannon formula as follows [4], [10].

$$R_{v,g,k}(t) = x_{v,g,k} b_{v,g,k} \log_2\big(1 + SINR_{v,g,k}(t)\big), \tag{9}$$

where $b_{v,g,k}$ denotes the bandwidth resource allocated by the ground O-RU g to vehicle v over the subscriber k at a time t. The rate of vehicle v (served by its associated O-RU g) is

$$R_{v,g}(t) = \sum_{g \in \mathcal{G}} \varsigma_{v,g} \sum_{k \in \mathcal{K}} R_{v,g,k}. \tag{10}$$

The V2U link is primarily a line of sight (LoS), where the large-scale fading component follows the free-space path loss model, while the small-scale fading is modeled by Rician fading, which captures both the deterministic LoS component and the scattered multipath effects [13]. Consequently, the instantaneous channel gain between vehicle v and UAV O-RU u over the subchannel k at time t is expressed as [14], [15]:

$$G_{v,u,k}(t) = \big(\frac{c}{4\pi d_{v,u} f_c}\big)^2 . g^a_{v,u,k} . |h^{LoS}_{v,u,k}(t)|^2 \quad (11)$$

where c denotes the speed of light, f_c is the carrier frequency, and $d_{v,u}$ represents the distance between vehicle v and UAV u. The term $|h^{LoS}_{v,u,k}(t)|^2$ is the small-scale power gain modelled as the Rician fading distribution between vehicle v and UAV O-RU u. While $g^a_{v,u,k}$ is the attenuation gain due to environmental effects, which depends on the distance and elevation angle between vehicle v and UAV O-RU u. The path loss for the LoS links is given as:

$$L^{LoS}_{v,u,k} = 20\log(4\pi f d_{v,u}/c) + \chi^{LoS} \quad (12)$$

where f, c, and χ^{LoS} denote the frequency, speed of light, and additional attenuation factor, respectively [6]. The SINR between the vehicle user v on the subcarrier k from the UAV O-RU u is given as follows [16]:

$$SINR_{v,u,k}(t) = \frac{p_{u,k}|G_{v,u,k}|^2}{\sigma^2 + \sum_{u'\neq u} P_{u'}|G_{v,u',k}|^2}, \quad (13)$$

where $p_{u,k}$ denotes the transmission power of the UAV O-RU u assigned to the subcarrier k and σ^2 is the noise power at the receiver. In eqn. (13), the term $\sum_{u\neq u'} P'_u|G_{v,u',k}|^2$ is co-channel interference from all other transmitting O-RUs $u' \neq u$. The transmission rate between the vehicle v and the UAV O-RU u on the subcarrier k is expressed using the Shannon formula as follows [4].

$$R_{v,u,k}(t) = b_{v,u,k}\log_2(1 + SINR_{v,u,k}(t)), \quad (14)$$

where $b_{v,u,k}$ is the bandwidth allocated to vehicle v from the UAV O-RU u over the subcarrier k at time t. The overall transmission rate between the vehicle v and the UAV O-RU u for offloading a task of size D_v via one or multiple allocated subcarriers is expressed as follows.

$$R_{v,u}(t) = \sum_{u\in\mathcal{U}} \varsigma_{v,u} \sum_{k\in\mathcal{K}} R_{v,u,k}. \quad (15)$$

To avoid ping-pong between the ground O-RU and the UAV O-RU under mobility and small-scale fading, we adopt a 3GPP-style hysteresis and time-to-trigger (TTT) rule. Let $\bar{\mathrm{SINR}}_{v,n}(t)$ denote a short-term filtered SINR (e.g., moving average over a window W):

$$\bar{\mathrm{SINR}}_{v,n}(t) \;=\; \frac{1}{W}\sum_{\tau=t-W+1}^{t} \mathrm{SINR}_{v,n}(\tau), \quad n\in\{g,u\}. \quad (16)$$

Define the SINR margin $\Delta_v(t) \triangleq \bar{\mathrm{SINR}}_{v,u}(t) - \bar{\mathrm{SINR}}_{v,g}(t)$. A handover from the ground O-RU to the UAV O-RU is triggered if

$$\Delta_v(\tau) > H \quad \forall\tau\in[t,\, t+\mathrm{TTT}-1], \quad (17)$$

$$\bar{\mathrm{SINR}}_{v,u}(t) \geq \mathrm{SINR}_{\min}, \quad \ell_u(t) < \ell_{\mathrm{th}}, \quad (18)$$

where $H > 0$ is the hysteresis margin, TTT is the time-to-trigger in slots, $\mathrm{SINR}_{\min}$ is a minimum quality threshold, and $\ell_u(t)$ is the UAV O-RU load with limit ℓ_{th}. Conversely, a return handover from the UAV O-RU to the ground O-RU occurs if

$$-\Delta_v(\tau) > H \quad \forall\tau\in[t,\, t+\mathrm{TTT}-1], \quad (19)$$

$$\bar{\mathrm{SINR}}_{v,g}(t) \geq \mathrm{SINR}_{\min}, \quad \ell_g(t) < \ell_{\mathrm{th}}. \quad (20)$$

We encode the serving-O-RU decision with binary variables $\varsigma_{v,g}(t), \varsigma_{v,u}(t) \in \{0,1\}$ such that

$$\varsigma_{v,g}(t) + \varsigma_{v,u}(t) = 1, \qquad \forall v\in\mathcal{V}, \quad (21)$$

and select the O-RU consistent with (17)–(20). Hence, the vehicle remains on its current O-RU unless the alternative O-RU offers a persistent advantage greater than H for at least TTT slots and meets the QoS/load constraints. This suppresses rapid oscillations while remaining responsive to sustained quality improvements.

To complement this adaptive association strategy, we define a dynamic priority mechanism that manages traffic-class- and state-aware resource allocation. We define a class-level priority weight $\omega_{\psi(v)}(t)$ for vehicle v based on its traffic type:

$$\omega_{\psi(v)}(t) = \begin{cases} \omega_{\psi(v,u)}, & \text{if} \quad \psi(v) = URLLC, \\ \omega_{\psi(v,e)}, & \text{if} \quad \psi(v) = eMBB, \\ \omega_{\psi(v,m)}, & \text{if} \quad \psi(v) = mMTC, \end{cases} \quad (22)$$

where $\omega_{\psi(v,u)} > \omega_{\psi(v,e)} > \omega_{\psi(v,m)}$ and $\omega_{\psi(v,u)} + \omega_{\psi(v,e)} + \omega_{\psi(v,m)} = 1$. This encodes the target service hierarchy. Then, we compute a vehicle-specific, normalized state factor $s_v(t)$ that aggregates latency-deadline urgency, energy urgency, and queue backlog via tunable coefficients. The per-vehicle priority is the normalized product of traffic class-based weight and dynamic state factor, expressed as, $\omega_v(t) = \frac{\omega_{\psi(v)} s_v(t)}{\sum_{v'\in\mathcal{V}} \omega_{\psi(v')} s'_v(t)}, \forall v' \neq v$, where $\omega_v(t) \in (0,1)$ and $\sum_v \omega_v(t) \leq 1$. This formulation ensures that the priority of each vehicle user v dynamically captures both the intrinsic importance of its traffic class and the real-time network conditions. By integrating these factors, the TMT framework achieves adaptive, traffic-aware, and fair resource allocation across heterogeneous vehicle users, effectively balancing performance objectives such as latency, throughput, energy consumption, and reliability. This dynamic priority design enables the system to adaptively coordinate radio and computing resources according to current network states and traffic demands. Once bandwidth is allocated, each vehicle v offloads its computation tasks to the associated computation server of O-RU n (i.e., ground or aerial), depending on link quality and resource availability. The uplink transmission time for this offloading process depends on the input data size D_v and the achievable uplink transmission rate $R_{v,n}$, between vehicle v and O-RU n, given by $T^{tx}_{v,n} = \frac{D_v}{R_{v,n}}$. The subsequent subsection details the computational resource allocation model.

B. Computation Model

In the proposed framework, each computing O-RU n on the ground or aerial network has a finite computational capacity $\mathcal{F}_n$, which directly influences the latency and energy consumption. Vehicles with latency-critical URLLC tasks must be allocated higher computing resources $f_{v,n}$ from the computing O-RU to minimize task execution latency $T^{exe}_{v,n}$. The allocation of computation resources must satisfy the following set of requirements to ensure efficient, reliable, and fair execution of tasks across all vehicles:

$$\sum_{v\in\mathcal{V}} x_{v,n,k} f_{v,n} \leq \mathcal{F}_n, f_{v,n} > 0. \tag{23}$$

Eq. (23) ensures that the total amount of computing resource allocated to all vehicles associated with serving O-RU n does not exceed the O-RU available processing capability.

To further capture the impact of computational resource allocation on end-to-end performance, the overall latency incorporates both the queuing delay and task execution time at the computing O-RU. To achieve the traffic prioritization objective, we use a priority-based queuing mechanism that aligns with the proposed traffic-aware allocation strategy. URLLC tasks are assigned the highest priority, followed by eMBB and mMTC, imposing that stringent latency requirements are consistently met. We consider that URLLC, eMBB, and mMTC traffic flows are assigned to separate priority queues. As widely assumed in the literature, we assume that the arrival process of each traffic class follows a Poisson distribution with rate $\lambda_{\psi(v)}$, and the service time distribution is exponential with mean service rate $\mu_{\psi(v)}$. Although URLLC traffic generated by IoT devices may not strictly follow a Poisson process, this is a fair assumption in vehicular environments where events such as collision avoidance messages or emergency alerts occur randomly and independently, thus making the memoryless property of the Poisson model hold. For this reason, the Poisson–exponential queuing framework has also been effectively adopted in prior works for modeling stochastic URLLC traffic arrivals in vehicular and 6G networks [1], [17]. However, it is worth mentioning that Poisson arrivals may not fully capture the burstiness, temporal correlation, and mobility-driven traffic dynamics observed in practical vehicular environments. In addition, the exponential service-time assumption may oversimplify heterogeneous processing and transmission characteristics in realistic O-RAN systems. Therefore, more advanced traffic and queuing models, such as Markov-modulated Poisson processes (MMPP), self-similar or trace-driven traffic models, and non-Markovian queuing systems (e.g., G/G/1 queues), are promising directions for future work to further enhance the realism, analytical depth, and practical robustness of the proposed framework.

The queuing system can then be modeled as an M/M/1 priority queuing system, where the queuing latency $T^{qu}_{\psi(v)}$ of traffic class $\psi(v)$ is given by [18]:

$$T^{qu}_{\psi(v),n} = \frac{1}{\mu_{\psi(v)} - \lambda^{\text{eff}}_{\psi(v)}},$$

where $\lambda^{\text{eff}}_{\psi(v)}$ is the arrival rate of the class $\psi(v)$ that can be adjusted based on the priority set such that higher-priority classes (e.g., URLLC) preempt or reduce the waiting time of lower-priority ones (e.g., eMBB and mMTC). In this way, we can establish a trade-off where URLLC achieves the minimum latency due to strict priority, eMBB experiences balanced latency and energy efficiency, and mMTC prioritizes energy savings at the cost of higher queuing delays. The task computation time at the computation O-RU n can be computed as:

$$T^{exe}_n = \frac{C_v}{f_{v,n}} = \frac{D_v \kappa_v}{f_{v,n}}, \tag{24}$$

where C_v is the CPU cycle requirement of vehicle v, with κ_v cycles/bit, and $C_v = D_v\kappa_v$. The total latency of a task is the sum of transmission, queuing, and computation delays. The total latency specific to the vehicle's traffic class for the vehicle v offloading its task to O-RU n is given by: $\mathcal{L}_{v,n} = T^{tx}_{v,n} + T^{qu}_{\psi(v),n} + T^{exe}_n$.

Similarly, we consider the energy consumed during end-to-end offloading and execution, which includes both the transmission energy of vehicle v and the computational energy consumed by the computing O-RU n, depending on the vehicle's traffic class. The energy consumed by vehicle v during the task transmission is given by: $E^{tx}_{v,n} = p^{tx}_{v,n} T^{tx}_{v,n} = p^{tx}_{v,n} \frac{D_v}{R_{v,n}}$, where $p^{tx}_{v,n}$ is the transmission power of vehicle v. Likewise, the energy consumed by the O-RU n during the execution of the task can be expressed as: $E^{exe}_{v,n} = p^{exe}_n T^{exe}_n$, where $p^{exe}_n = \xi_n f^3_{v,n}$, and ξ_n is a switched-capacitance coefficient for computing at O-RU n. This expression can be reformulated as: $E^{exe}_{v,n} = \xi_n f^3_{v,n} \frac{C_v}{f_{v,n}} = \xi_n f^2_{v,n} C_v$. We then derive the total energy consumption associated with task offloading and execution, defined as the sum of transmission energy at vehicle v and execution energy at O-RU n:

$$E_{v,n} = E^{tx}_{v,n} + E^{exe}_n. \tag{25}$$

We also calculate the system throughput, which is particularly relevant for eMBB traffic to meet the high throughput requirements in applications such as AR/VR. In our problem, we aim to optimize throughput while balancing resource allocation across different traffic classes. The throughput can be written as

$$\Gamma_{v,n} = \frac{D_v}{\mathcal{L}_{v,n}} \mathcal{R}_{v,n}, \tag{26}$$

where D_v is the data size of vehicle v, $\mathcal{L}_{v,n}$ is the total latency of the task, and $\mathcal{R}_{\text{suc}}$ is the reliability (success probability). Notably, $\Gamma_{v,n}$ captures the effective throughput that accounts for both data rate and reliability. For URLLC traffic, reliability quantifies deadline-constrained success via SINR or finite-blocklength error. Thus, we measure reliability in terms of success probability for a packet of class $\psi(v)$ offloaded by vehicle v to O-RU n, which is given by:

$$\mathcal{R}_{v,n} = \Pr\{SINR_{v,n} \geq SINR^{\min}_{v,n}\}, \tag{27}$$

where $SINR^{\min}_{v,n}$ is the class-specific SINR threshold (tighter for URLLC) required for correct decoding. The above expression represents the reliability of a single transmission attempt. For K repetitions, we get:

$$\mathcal{R}^K_{v,n} = 1 - \left(1 - \mathcal{R}_{v,n}\right)^K, \tag{28}$$

which signifies the reliability improvement when K independent repetitions are used.

IV. Problem Formulation

The proposed framework is designed to achieve a dynamic balance between the inherently system conflicting metrics, namely, latency, throughput, reliability, and energy consumption, based on the service requirements of different traffic classes. In URLLC traffic, minimizing latency ensures ultra-reliable and time-critical transmissions. In contrast, eMBB traffic demands a balanced trade-off between throughput efficiency and energy consumption, while mMTC traffic prioritizes energy conservation to sustain massive, low-power device connectivity. With this adaptive balancing mechanism, resource allocation strategies can be tailored to suit heterogeneous QoS demands of the vehicular O-RAN systems. We express such a long-term objective as a weighted sum of latency, energy consumption, throughput, and reliability, as follows:

$$\mathcal{O}(t) = \sum_{v\in\mathcal{V}}\sum_{n\in\mathcal{N}} x_{v,n}\Big[\theta_{\psi(v)}\mathcal{L}_{v,n} + \beta_{\psi(v)}E_{v,n} - \eta_{\psi(v)}\Gamma_{v,n} - \delta_{\psi(v)}\mathcal{R}_{v,n}\Big], \tag{29}$$

where θ, β, η, and δ are traffic-aware trade-off weights for latency, energy, throughput, and reliability, respectively. These weighted metrics can be summed per class and then normalized. These weight factors are used to balance the trade-off between performance metrics of each class of traffic while ensuring fair and consistent priority process between URLLC, eMBB, and mMTC traffic. Hence, $\theta+\beta+\eta+\delta = 1$ and $\theta,\beta,\eta,\delta = [0,1]$. The joint optimization problem is formulated to minimize the weighted sum function in (29) and written as:

$$\boldsymbol{P}: \quad \min_{x,\mathcal{F},\mathcal{B}} \mathbb{O}(t) \tag{30}$$

$$\begin{aligned}
\text{s.t. } & C1: \sum_{v\in\mathcal{V}} x_{v,n,k} = 1, x_{v,n,k}\in\{0,1\}, \forall v, n\\
& C2: \varsigma_{v,g}(t) + \varsigma_{v,u}(t) = 1, \varsigma_{v,g}, \varsigma_{v,u}\in\{0,1\}, \qquad \forall v\in\mathcal{V}\\
& C3: \sum_{v\in\mathcal{V}} B_{v,n} \le \mathcal{B}_n, B_{v,n} > 0, \forall v, n\\
& C4: \sum_{v\in V} x_{v,n,k} f_{v,n} \le \mathcal{F}_n, f_{v,n} > 0, \forall v, n\\
& C5: \sum_{k\in\mathcal{K}_n} p_{n,k} \le P_n^{max}, p_{n,k} > 0, \forall n, k\\
& C6: \mathcal{L}_{v,n} \le \mathcal{L}_{\psi(v)}^{max}, \forall v\in\mathcal{V}\\
& C7: Pr(\mathcal{L}_{v,n} \le \mathcal{L}_{URLLC}^{max}) \ge 1-\epsilon, \forall v: \psi(v) = URLLC\\
& C8: \mathcal{R}_{v,n} \ge \mathcal{R}_{\psi(v)}^{min}, \forall v\in\mathcal{V}\\
& C9: \Gamma_{v,n} \ge \Gamma_{\psi(v)}^{min}, \forall v\in\mathcal{V}\\
& C10: E_{v,n} \le E_{\psi(v)}^{max}, \forall v, n,
\end{aligned}$$

where ϵ is a very small violation probability, representing the acceptable risk that latency exceeds the target.

In the above problem, the constraint $C1$ makes use of a binary subscriber assignment variable, allowing each sub-channel from RU n to be allocated to at most one vehicle. $C2$ defines a binary association variable that enforces that each vehicle can only be associated with at most one O-RU at a time. In constraints $C3$ and $C4$, the bandwidth and computing resources allocated to all vehicles cannot exceed the capacity of each O-RU, ensuring a balanced utilization of resources. $C5$ ensures that the transmission power allocated cannot be greater than the defined maximum limit. For URLLC services that are critical to safety, $C6$ ensures that end-to-end latency stays within the prescribed delay limits for all traffic, while $C7$ and $C8$ ensure strict latency requirements and the probability of successful transmission and processing exceeds a reliability threshold for URLLC traffic. In addition, $C9$ ensures that the average throughput does not fall below the required level to meet eMBB's high data-rate requirements. Furthermore, $C10$ limits energy consumption per task within acceptable limits, thereby prolonging device lifetime and supporting sustainable operations in mMTC environments.

The formulated optimization is characterized as a mixed-integer nonconvex program (MINCP), where nonconvexity comes from the nonlinear coupling of latency, energy, throughput, and reliability functions with multiple decision variables, and integer variables represent the subscriber assignment and resource allocation decisions.

Property 1. *The formulated joint optimization problem* $\boldsymbol{P}$ *defined in (30) and subject to constraints* $C1$*–*$C10$ *is* ***NP-hard***.

Proof. The formulated optimization problem $\boldsymbol{P}$ is an MINCP involving both discrete and continuous decision variables. The nonconvexity arises from the nonlinear coupling among latency, energy consumption, throughput, and reliability functions. The binary variables $x_{v,n,k}\in\{0,1\}$ and $\varsigma_{v,n}\in\{0,1\}$ represent the discrete subcarrier assignment and vehicle-O-RU associations variables, while $\mathcal{B}_n$, $\mathcal{F}_n$, and $p_{n,k}$ denote continuous resource allocation variables. To prove NP-hardness, consider a simplified version of $\boldsymbol{P}$ containing only the subscriber assignment ($C1$), association ($C2$), and bandwidth capacity ($C3$) constraints. It can be reduced to a combinatorial assignment problem, which is proven to be NP-hard. This reduced problem is equivalent to the well-known *Generalized Assignment Problem (GAP)*, where each subscriber is assigned to one vehicle ($C1$) and each vehicle is associated with one O-RU ($C2$), subject to the bandwidth capacity of each O-RU ($C3$). The objective function $\mathbb{O}(t)$ can be expressed as the weighted sum of the assignment costs (or equivalently, the negative of profits in GAP). Since GAP is NP-hard, and $\boldsymbol{P}$ generalizes it by including additional nonlinear coupling constraints ($C4$–$C10$) and continuous variables, $\boldsymbol{P}$ is also NP-hard. Consequently, the overall joint optimization problem inherits this NP-hardness due to the presence of multiple interdependent integer and continuous variables. □

It follows that exploiting convex optimization techniques may lead to multiple local optima, and finding exact solutions for large-scale vehicular O-RAN systems is computationally challenging. To tackle this NP-hard optimization problem, the proposed TMT algorithm is employed, leveraging its hierarchical and traffic-aware design to achieve scalable, adaptive, and near-optimal decision-making in dynamic vehicular O-RAN environments.

V. The Proposed MADRL-Based Solution

In light of the complexity of the above problem, we introduce a low-complexity algorithmic framework that enables the solution of large problem instances in real time. We underline that traditional approaches to solve MINCP problems in real-time fail in our complex scenario with dynamic traffic arrivals, UAV mobility, and heterogeneous QoS requirements [19]. We thus adopt an MADRL-based solution that adapts to changing conditions and serves the goal of prioritizing traffic. To address the complexity of the joint optimization problem, we first formulate it as an MOMDP and, then, design the TMT resource allocation algorithm, presented in Fig.2.

A. MOMDP Formulation

We start by defining three types of agents that collaboratively operate within the hierarchical vehicular O-RAN.

1) Vehicle agents autonomously decide association, offloading, and resource demand decisions based on channel conditions, queue states, and traffic priorities.
2) O-RU agents, including terrestrial and aerial O-RUs, dynamically allocate radio and computing resources, balance load, and maintain service continuity.
3) The RIC agent serves as the central coordinator, observes the state of the global network, optimizes the distribution of resources, and ensures fair and efficient cooperation among all agents. Together, these agents form a hierarchical traffic-aware system that jointly optimizes throughput, latency, energy efficiency, and reliability across heterogeneous vehicular services.

We then define the MOMDP as $\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{P}, R, \gamma\}$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ is the state transition probability, R is the reward function, and γ is the discount factor [20]. We use centralized training with decentralized execution, where agents observe $o(t) \subseteq s(t)$. Specifically, the observation of agent i at time t is represented as a feature vector $o_i(t) \in \mathbb{R}^{d_i}$, where $d_i = |o_i(t)|$ denotes the observation dimension of agent i. When the agents make a joint action, the current state $s(t)$ transforms into a new state $s(t+1)$ and the state transition probability is defined as $\mathcal{P}\big(s(t+1)|s(t), a(t)\big)$.

State Space: The state space at time t is defined as $\mathcal{S}(t) = \{\mathcal{S}_V, \mathcal{S}_N, \mathcal{S}_{RIC}\}$, where $\mathcal{S}_V, \mathcal{S}_N$, and $\mathcal{S}_{RIC}$ denote the state spaces for vehicle, O-RU, and RIC agents. We define the vehicle state as $s_v(t) \in \mathcal{S}_V \subset \mathcal{S} = \{x_{v,n,k}, \varsigma_{v,n}, q_v(t), D_v, G_{v,n,k}(t), E_v(t), \bar{\rho}_v, \psi_v\}$, where $x_{v,n,k}$ is a binary subscriber assignment variable and $\varsigma_{v,n} = \{\varsigma_{v,g}, \varsigma_{v,u}\}$ denotes the binary association decision of vehicle v to either a ground or UAV O-RU. Here, $q_v(t)$ represents the current queue backlog at vehicle v, D_v denotes the data size of the task to be offloaded, $G_{v,n,k}$ signifies the channel gain between vehicle v and O-RU n on subcarrier k, $E_v(t)$ indicates the remaining energy of vehicle v, $\bar{\rho}_v$ denotes the position of vehicle v, and ψ_v specifies the traffic class of vehicle v. Similarly, we define the O-RU state as $s_n(t) \in \mathcal{S}_n \subset \mathcal{S}(t) = \{\mathcal{B}_n(t), \mathcal{F}_n(t), \bar{\rho}_u, \mathcal{V}_n, \mathcal{K}_n, l_n(t)\}$, where $\mathcal{B}_n(t), \mathcal{F}_n(t), \bar{\rho}_u, \mathcal{V}_n, \mathcal{K}_n$, and $l_n(t)$ denote the available bandwidth, computing resource, position of UAV, vehicle served, the set of subscribers, and the load at O-RU n, respectively. Moreover, the RIC state is defined as $s_r(t) \in \mathcal{S}_{RIC} \subset \mathcal{S} = \{\{\Gamma_{v,n}(t)\}_v, \omega_\psi, \{\mathcal{L}_{v,n}(t)\}_v, l(t), \bar{\mathrm{SINR}}_{v,n}, \Delta_v(t)\}$, where $\{\Gamma_{v,n}(t)\}_v, \omega_\psi, \{\mathcal{L}_{v,n}(t)\}_v, l(t), \bar{\mathrm{SINR}}_{v,n}$, and $\Delta_v(t)$ represent throughput per vehicle, global weights per traffic class, latency per vehicle, aggregated O-RU load, link quality measure (SINR), and handover control threshold, respectively.

Action Space: The joint action space is given by: $a(t) \in \mathcal{A}(t) = \{\mathcal{A}_V(t), \mathcal{A}_N(t), \mathcal{A}_{RIC}(t)\}$, where $\mathcal{A}_V(t), \mathcal{A}_N(t)$, and $\mathcal{A}_{RIC}(t)$ denote the action sets of the vehicle, O-RU, and RIC agents, respectively. Each vehicle agent v observes the current state and makes its action set $a_v(t) \in \mathcal{A}_v(t) = \{a_v^{\varsigma}(t), a_v^b(t), a_v^f(t), a_v^{QoS}(t)\}$. Within this action set, the agent selects its serving O-RU (ground or UAV) and makes the association decision $a_v^{\varsigma}(t)$ to establish the optimal service link for task execution. The vehicle agent also determines its bandwidth demand $a_v^b(t)$ and computational resource request $a_v^f(t)$, while dynamically setting the QoS threshold $a_v^{QoS}(t)$ to meet the stringent service requirements of each traffic class. Similarly, the action set of each O-RU agent is defined as $a_n(t) \in \mathcal{A}_N = \{a_n^k(t), a_n^b(t), a_n^f(t), a_n^p(t)\}$, where $a_n^k(t)$,$a_n^b(t)$,$a_n^f(t)$, and $a_n^p(t)$ denote the channel assignment, bandwidth allocation, computing resource allocation, and transmission power allocation decisions, respectively. Each O-RU agent n autonomously adjusts these parameters based on the observed state information to optimize resource utilization and maintain service continuity across heterogeneous vehicular environments. In addition, we define the action set of the RIC agent $a_r(t) \in \mathcal{A}_{RIC}(t) = \{a_r^\pi(t), a_r^r(t), a_r^\omega(t), a_r^h(t)\}$, where $a_r^\pi(t), a_r^r(t), a_r^\omega(t)$, and $a_r^h(t)$ denote the policy adaptation, resource orchestration, traffic prioritization, and handover management decisions, respectively.

Reward: We define a multi-objective reward function that combines four critical performance metrics. This reward function is directly aligned with the optimization problem objectives in (30), aiming to minimize latency and energy consumption while maximizing throughput and reliability. To capture this, the reward is expressed as follows:

$$R\big(s(t), a(t)\big) = \theta R^{la}(t) + \beta R^{en}(t) + \eta R^{th}(t) + \delta R^{re}(t)\}, \tag{31}$$

where $R^{la}(t), R^{en}(t), R^{th}(t)$, and $R^{re}(t)$ are the reward contributions from latency, energy consumption, throughput, and reliability. Therefore, $R\big(s(t), a(t)\big) = \min_{x,f,\mathcal{B}} \mathbb{O}(t)$. To solve this multi-objective optimization problem in the TMT framework, we aim to find the optimal policy π^* that maximizes the expected cumulative reward:

$$\pi^* = \arg\max_{\pi} \mathbb{E}\Big[\sum_{t=0}^{T} \gamma^t R\big(s(t), a(t)\big)\Big], \tag{32}$$

where $\gamma \in [0, 1]$ is the discount factor. Since $R\big(s(t), a(t)\big)$ is vector-valued, we applied a scalarization technique to convert this multi-objective reward into a single scalar signal suitable for learning. In this way, the proposed TMT framework effectively balances latency, energy efficiency, throughput, and reliability under highly dynamic vehicular conditions.

To support post-incident analysis and potential regulatory auditing, the proposed TMT framework maintains structured

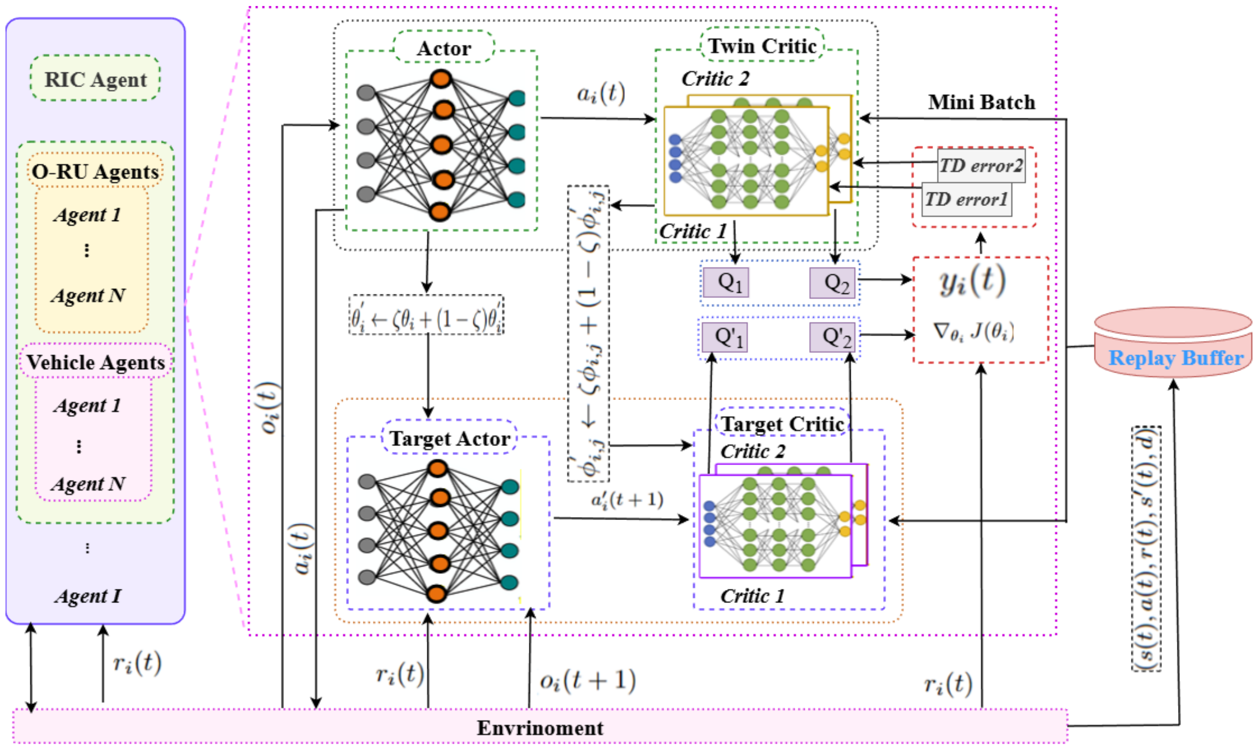

Fig. 2: Schematic representation and functional workflow of the proposed hierarchical TMT algorithm.

logs of key network and decision-making events within the RIC architecture. The recorded information includes network state observations such as traffic load, queue status, SINR, latency, throughput, energy consumption, and O-RU resource utilization, together with vehicle–O-RU association and handover decisions. In addition, the framework logs traffic prioritization actions, bandwidth and computation resource allocation decisions, policy adaptation updates, and corresponding reward evaluations generated by the TMT algorithm. Timestamped records of UAV O-RU activation and load balancing operations are also maintained to ensure traceability, transparency, and offline root-cause analysis under dynamic vehicular network conditions.

B. Hierarchical TMT Algorithm Design

To handle the dynamics and complexity in the formulation optimization problem, we present the TMT algorithm, which comprises an actor and two critic networks. TMT extends the classical MATD3 framework by incorporating traffic-awareness and hierarchical decision-making into the learning process. The algorithm leverages twin critics and delayed policy updates to mitigate overestimation bias and stabilize learning in dynamic vehicular O-RAN environments. Unlike benchmark methods such as MATD3 and Multi-Agent Deep Deterministic Policy Gradient (MADDPG) or flat DRL approaches, which struggle to handle the heterogeneous QoS demands of end devices due to their lack of traffic awareness, large action spaces, and vulnerability to overestimation bias, our proposed TMT framework introduces a hierarchical traffic-aware design. This framework allows critical services like URLLC to be prioritized while balancing eMBB and mMTC requirements to effectively address the multi-objective and traffic-sensitive nature of the formulated problem. Furthermore, the twin-critic mechanism mitigates value overestimation and enhances stability, leading to faster convergence, improved scalability, and robust adaptation in dynamic vehicular O-RAN environments.

Algorithm 1 TMT Algorithm for Dynamic Resource Allocation

```
Initialize actors π_{θ_i} and twin critics {Q_{i,1}, Q_{i,2}} and their weight parameters θ_i, φ_{i,1}, φ_{i,2}, respectively.
Initialize the target actor π', and target twin critics Q'_{i,1} and , Q'_{i,2} with their weight parameters θ'_i, φ'_{i,1}, φ'_{i,2}, respectively.
Initialize replay buffer D
for episode =1 to 3000 do
    Initialize state s(t)
    for env step t = 1, 2, ..., T do
        Observe the global state s and each local o_i
        Each agent selects a_i = π_{θ_i}(o_i) + exploration
        Execute joint action a, observe s', r(t)
        Store (s(t), a(t), r(t), s'(t), d) in D
        if time to update then
            Sample minibatch of B transitions from D
            Compute target actions as: a'_i(t+1) = π̄_{θ_i}(o'_i) + ε
            Compute targets as in (34)
            Update critics by minimizing the loss for j = 1, 2 as in (33)
            if mod(update_step, d_π) == 0 then
                For each agent i, update the policy as (35)
                Apply traffic-aware weighting: J_i^ψ = ωJ(θ_i)
                Soft-update all target networks with factor ζ as:
                φ'_{i,j} ← ζφ_{i,j} + (1 − ζ)φ'_{i,j},   θ'_i ← ζθ_i + (1 − ζ)θ'_i
            end if
        end if
    end for
end for
```

We apply the Multi-Objective Actor–Critic (MOAC) framework, integrated with TMT, which extends the standard actor–critic paradigm to handle multiple, and often conflicting, objectives, such as latency, energy consumption, throughput, and reliability, in vehicular O-RAN systems [21]. Within this framework, the actor network generates traffic-aware resource allocation and prioritization strategies, while the critic networks evaluate these policies against the weighted scalarized objective that captures the heterogeneous QoS requirements of URLLC, eMBB, and mMTC traffic. The hierarchical design of TMT further enhances MOAC by decomposing the decision process: the upper layer actor-critic pair manages traffic prioritization across classes, while the lower layer actor-critic agents refine continuous resource allocation (e.g., bandwidth, computation, and transmit power) for the selected traffic. This joint application ensures that the learning process remains scalable and adaptive, while effectively balancing multiple objectives in a dynamic and large-scale vehicular O-RAN environment. The reward function in (31) compresses multiple objectives into a traffic-aware scalar signal for learning. Let $i \in \mathcal{I} = \{\mathcal{V}, \mathcal{U}, \mathcal{G}, RIC\}$ denote the set of all agent types (vehicles ($\mathcal{V}$), UAV O-RUs ($\mathcal{U}$), ground O-RUs ($\mathcal{G}$), and RIC controller) and $j = \{1, 2\}$ be the twin critic index. Hence, the twin-critic loss in the actor-critic update process is expressed as [22]:

$$\mathcal{L}(\phi_{i,j}) = \mathbb{E}_{(o_i(t),a_i(t),r_i(t),o_i(t+1))}\Big[\big(Q_{\phi_{i,j}}(o_i(t), a_i(t)) - y_i(t)\big)^2\Big], \tag{33}$$

where $y_i(t)$ denotes the target Q-value, calculated using the minimum of the two target critic networks ($Q^{'}_{i,1}$ and $Q^{'}_{i,2}$) to effectively reduce overestimation bias and enhance the stability and accuracy of value estimation during training, expressed as:

$$y_i(t) = r_i(t) + \gamma \min_{j\in\{1,2\}} Q_{\phi^{'}_{i,j}}\big(o_i(t+1), \pi_{\theta^{'}_i}(o_i(t+1))\big), \tag{34}$$

with $r_i(t) \in R$ denoting the scalarized traffic-aware reward. Here, $\theta^{'}_i$ and $\phi^{'}_{i,j} (j = 1, 2)$ are weight parameters for the target actor and critic networks, respectively.

The deterministic policy gradient with a twin-critic applies to the traffic-aware scalarized objective. In line with the TD3 principle, TMT adopts a delayed policy update mechanism, where the actor network is updated once every d critic updates. This reduces overestimation bias and improves training stability. In the scenario considered, the system state changes rapidly due to the mobility of the vehicles and the UAVs, and the fluctuation of channel conditions. Thus, delaying policy updates enables the critic networks to estimate Q-values more accurately under non-stationary conditions before the policy is adjusted, thereby preventing instability caused by overly frequent actor updates. For the agent i, the delayed actor update is

$$\nabla_{\theta_i} J(\theta_i) = \mathbb{E}_{o_i}\Big[\nabla_{a_i} \min_{j\in\{1,2\}} Q_{\phi_{i,j}}\big(o_i(t), a_i(t)\big)\big|_{a=\pi_\theta(o_i(t))} \nabla_\theta \pi_\theta(o_i(t))\Big], \tag{35}$$

where $Q_{\phi_{i,1}}$ and $Q_{\phi_{i,2}}$ are the twin critics evaluating the scalarized reward, π_θ is the actor policy, and the expectation is over the joint observation o_t. The soft target updates of TMT is then given by:

$$\phi^{'}_{i,j} \leftarrow \zeta\phi_{i,j} + (1-\zeta)\phi^{'}_{i,j}, \quad \theta^{'}_i \leftarrow \zeta\theta_i + (1-\zeta)\theta^{'}_i, \tag{36}$$

where $\zeta \ll 1$ is the soft update coefficient. This update is performed periodically following the delayed policy update strategy of TMT. The step-by-step process of the proposed approach is systematically presented in **Algorithm 1**.

C. Complexity Analysis

The computational complexity and convergence behavior of the proposed hierarchical TMT algorithm are analyzed to assess its practicality in large-scale vehicular scenarios. Let N denote the number of agents, d the observation dimension, and $|\mathcal{A}|$ the action space size. The baseline multi-agent actor–critic framework incurs a per-update complexity of $\mathcal{O}(N{\cdot}d{\cdot}|\mathcal{A}|)$. The proposed TMT algorithm introduces additional components with moderate overhead. Specifically, the hierarchical policy structure yields a complexity of $\mathcal{O}(N \cdot (d_h + d_l) \cdot |\mathcal{A}|)$ due to decomposed decision spaces, while the traffic-aware weighting mechanism incurs only $\mathcal{O}(N)$ computational cost. The twin-delayed mechanism increases the value estimation complexity to $\mathcal{O}(2 \cdot N \cdot d)$, which remains linear in the number of agents. Furthermore, the inter-agent communication module introduces a worst-case complexity of $\mathcal{O}(N^2)$, although this is significantly reduced in practice through localized interactions. Overall, the total complexity of the TMT algorithm can be expressed as $\mathcal{O}(N \cdot d \cdot |\mathcal{A}| + N^2)$, which remains tractable for realistic system scales. Despite the additional computational overhead, the TMT algorithm demonstrates faster and more stable convergence compared to its ablated variants, achieving near-optimal performance in fewer training episodes. In contrast, removing key components such as traffic-aware weighting or the twin-critic mechanism leads to slower convergence, increased variance, and reduced adaptability under dynamic traffic conditions. These results highlight that the proposed TMT algorithm effectively balances computational efficiency with improved learning stability and scalability.

VI. Performance Evaluation

This section presents the experimental settings, benchmark algorithms, and numerical results, followed by a detailed analysis and discussion of the obtained results to demonstrate the performance of the proposed TMT algorithm under dynamic vehicular network conditions.

A. Experimental settings and benchmarks

Experimental settings. To evaluate the performance of the proposed TMT scheme, we conduct extensive simulations in a vehicular O-RAN environment. We consider a 1000×1000 m^2 urban area with $V = 40$ vehicles, $G = 3$ ground O-RUs, and $U = 3$ UAV O-RUs. UAV O-RU altitude is $H = 100$ m. The bandwidth per O-RU is $\mathcal{B}_n = 10$ MHz with $K = 20$ orthogonal subcarriers per O-RU. Noise power is $\sigma^2 = -104$ dBm, and transmission power is uniformly distributed in $[18, 30]$ dBm.

TABLE I: Parameter settings, including vehicular network and algorithm configurations

Parameter	Value
Area size	1000×1000 m^2
Number of vehicles	80
Ground O-RUs	4
UAV O-RUs	3
Bandwidth per O-RU	10 MHz
Number of subcarriers per O-RU	20
Noise power	-104 dBm
Max. transmission power	30 dBm
O-RU compute capacity	50 GHz (ground), 30 GHz (UAV)
Actor/Critic layers	[256, 256], ReLU activation
Learning rate	3×10^{-4}
Discount factor	0.99
Soft update	0.005
Replay buffer size	10^6
Batch size	256

Task sizes are uniformly distributed as $D_v \in [0.5, 2]$ KB for URLLC, $[500, 1500]$ KB for eMBB, and $[1, 10]$ KB for mMTC. The CPU cycles per bit κ_v are set to 500, 1000, and 200 cycles/bit, respectively. Tasks arrive according to a Poisson process with class-specific rates of $\lambda_u = 0.5$ tasks/s (URLLC), $\lambda_e = 0.2$ tasks/s (eMBB), and $\lambda_m = 0.3$ tasks/s (mMTC). Each O-RU is modeled as an M/M/1 priority queue with class priority URLLC $>$ eMBB $>$ mMTC. The SINR threshold $SINR_{\min}$ is 10 dB for URLLC, 0 dB for eMBB, and -5 dB for mMTC. Each ground O-RU provides $\mathcal{F}_n = 50$ GHz computation capacity, while UAV O-RUs offer 30 GHz. The switched-capacitance coefficient is $\xi_n = 10^{-28}$. Vehicles follow a random waypoint mobility model with speed ranges of [5,15] m/s (URLLC), [0,10] m/s (eMBB), and [0,5] m/s (mMTC).

We use a centralized training and decentralized execution paradigm. Each O-RU, vehicle, and RIC acts as an agent. The actor network has two hidden layers of size 256 with ReLU activations, and the critic uses a twin-critic architecture with the same layer sizes. The replay buffer size is 10^6, minibatch size is 256, and Adam optimizer is used with learning rate 3×10^{-4}. The discount factor is $\gamma = 0.99$, and the target networks are updated with $\zeta = 0.005$. The parameter settings are summarized in TABLE I.

Benchmarks. To validate the effectiveness of TMT, we compare its performance against MADDPG [23], MATD3 [22], [24], and a greedy algorithm (GA) [25], across multiple metrics, such as reward convergence, latency, energy consumption, throughput, and reliability. The details of the considered benchmarks are summarized as follows.

MATD3: An extension of the DDPG algorithm to multi-agent environments, incorporating twin critics to mitigate overestimation bias and enhance training stability.

MADDPG: A standard multi-agent actor–critic framework employing centralized training with decentralized execution, but lacking the twin-critic mechanism and delayed policy update strategy.

GA: A heuristic optimization method based on evolutionary operations, widely used for global search and performance comparison in resource allocation problems.

These benchmark algorithms are selected because they represent state-of-the-art reinforcement learning and heuristic optimization paradigms that are extensively utilized in dynamic resource management, control optimization, and multi-agent coordination within wireless and vehicular network environments. For a rigorous and fair comparison, all baseline models were systematically tuned under identical experimental conditions using a consistent and structured hyperparameter (e.g., learning rate, batch size, discount factor, and network architecture) optimization procedure within well-established ranges. Furthermore, all methods operate within the same environment, state–action spaces, reward formulation, and training budget across both terrestrial and non-terrestrial scenarios, ensuring strict experimental consistency. The reported results correspond to the best-performing configuration of each method, thereby providing a reliable and unbiased benchmark against the proposed TMT.

B. Numerical Results

TMT convergence. We first evaluate the performance of the TMT framework as a function of the number of training episodes, as shown in Fig. 3. Specifically, the effectiveness of TMT is evaluated over 3,000 episodes in optimizing critical performance metrics, namely, normalized reward, latency, throughput, and energy consumption. Fig. 3a shows the rapid increases in normalized reward during the initial training episodes and a convergence at a value around episode 500, indicating fast convergence and stable policy learning. Fig. 3b presents the end-to-end latency, which starts at a very high level, decreases sharply during early episodes, and stabilizes at a much lower level, illustrating the framework ability to effectively minimize communication and processing delays across the different traffic classes. This demonstrates that under our approach URLLC traffic experiences ultra-low latency for mission-critical services and eMBB traffic benefits from improved transmission efficiency, while mMTC traffic maintains reliable connectivity under massive device access. This shows that the framework not only reduces overall system latency but also adapts to heterogeneous QoS requirements of vehicular O-RAN systems.

Similarly, we evaluate the system throughput as the number of episodes increases. Fig. 3c shows a consistent throughout increase as learning proceeds, which demonstrates the capability of the proposed framework to optimize bandwidth allocation and maximize the data rate. This improves eMBB services provisioning, which demands high throughput, while also ensuring that URLLC traffic meets stringent data rate requirements and mMTC traffic achieves scalable, stable communication. On the other hand, energy consumption decreases sharply in the early episodes and converges to a low stable level, demonstrating efficient resource allocation, as shown in Fig. 3d. In UAV-assisted vehicular O-RAN, energy efficiency directly affects service sustainability. In UAV-assisted vehicular O-RAN, energy efficiency is pivotal for service sustainability, as limited UAV power directly determines operational lifetime and continuous service availability. The results show that our framework significantly reduces energy consumption while sustaining high throughput and low latency, thereby

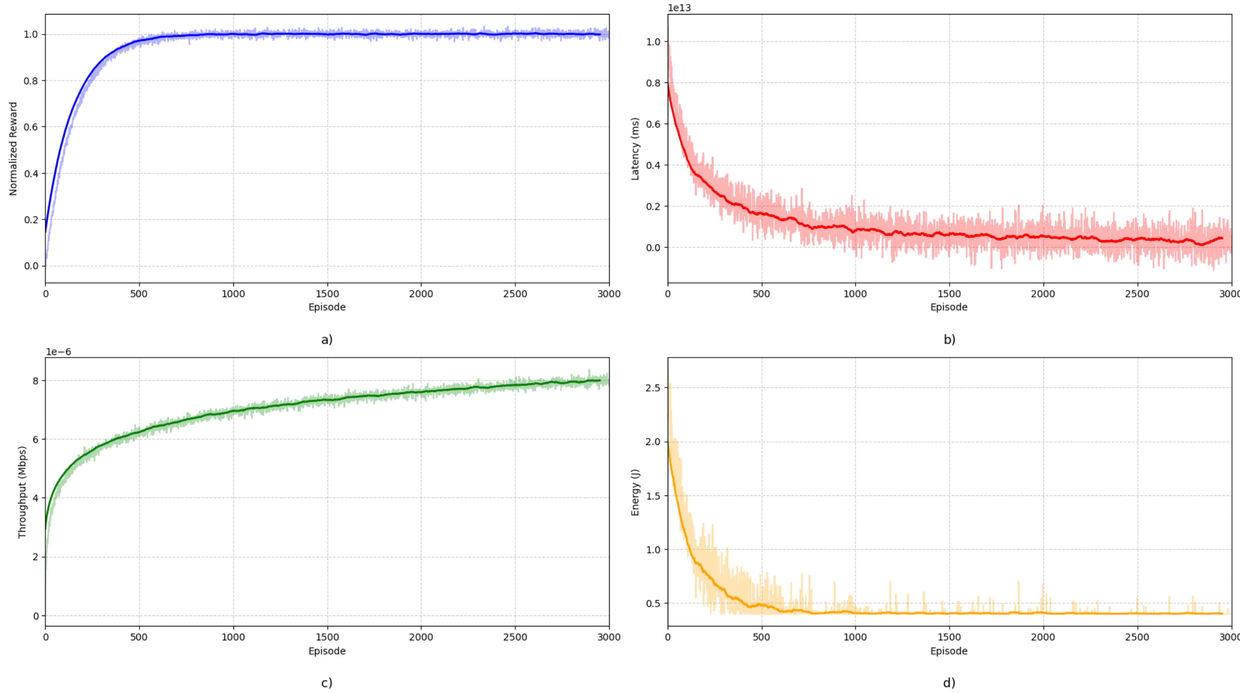

Fig. 3: Performance of TMT as a function of the number of episodes: a) Normalized reward; b) Latency; c) Throughput; d) Energy consumption.

ensuring reliable long-term support for URLLC, eMBB, and mMTC services. Collectively, these results show that the proposed approach achieves a favorable balance among reward maximization, latency reduction, throughput improvement, and energy efficiency, thereby meeting the stringent QoS requirements of vehicular O-RAN scenarios.

Ability to support different traffic classes. We present a comparative evaluation of energy consumption, latency, throughput, and success rate across URLLC, eMBB, and mMTC traffic classes in UAV-assisted vehicular O-RAN, as shown in Fig. 4. The results in Fig.4a demonstrate that eMBB traffic records the highest energy consumption due to its bandwidth-intensive nature, followed by URLLC, while mMTC shows the lowest energy consumption. Regarding the end-to-end latency across traffic types, URLLC traffic achieves the lowest latency, aligning with its strict QoS requirements for mission-critical services, as illustrated in Fig. 4b. eMBB traffic experiences moderate latency acceptable for high-data-rate applications, whereas mMTC traffic records the highest latency, reflecting its tolerance for delay in massive connectivity scenarios. Likewise, we evaluate the throughput performance of the three traffic classes, as shown in Fig. 4c. eMBB traffic attains the highest throughput to support data-intensive services, URLLC traffic sustains moderate throughput suitable for real-time applications, and mMTC traffic yields the lowest throughput. Finally, we assess the success rate for the three traffic classes in Fig. 4d. URLLC traffic achieves near-perfect reliability (99%), eMBB traffic achieves a strong success rate (96%) to support continuous service delivery, and mMTC traffic maintains a comparatively lower success rate (90%), primarily due to massive device access and interference in dense vehicular environments.

In summary, the results demonstrate that the proposed TMT efficiently adapts resource allocation to the heterogeneous QoS requirements of URLLC, eMBB, and mMTC (ensuring ultra-reliable and low-latency communication for URLLC, high-throughput service for eMBB, and scalable connectivity for mMTC), supporting diverse vehicular O-RAN traffic demands.

Normalized reward. We investigate how the normalized reward varies as the number of training episodes increases, for both the proposed and benchmark algorithms. Fig. 5 highlights that the proposed TMT framework converges faster and more stably than its alternatives, reaching steady performance in about 500 episodes. Compared to the benchmarks, TMT achieves normalized reward up to 12%, 17%, and 25% higher than MATD3, MADDPG, and GA, respectively. Such faster convergence of TMT is due to its hierarchical and traffic-aware architecture of the proposed framework, which improves policy coordination, accelerates learning, and ensures efficient trade-offs between exploration and exploitation across heterogeneous vehicular environments. In comparison, MATD3 converges more slowly than TMT but remains moderately stable as the training episodes increase. Furthermore, MADDPG exhibits a comparable final performance, but it takes longer to train because its policy updates are slow and its noise regularization is weak. Lastly, the GA has the slowest convergence and the highest levels of instability because it uses a heuristic search and is unable to learn from its mistakes.

Latency analysis. Fig. 6 presents the overall latency with respect to the number of training episodes for the proposed TMT and three benchmark algorithms, such as traditional MATD, MADDPG, and GA, across heterogeneous traffic

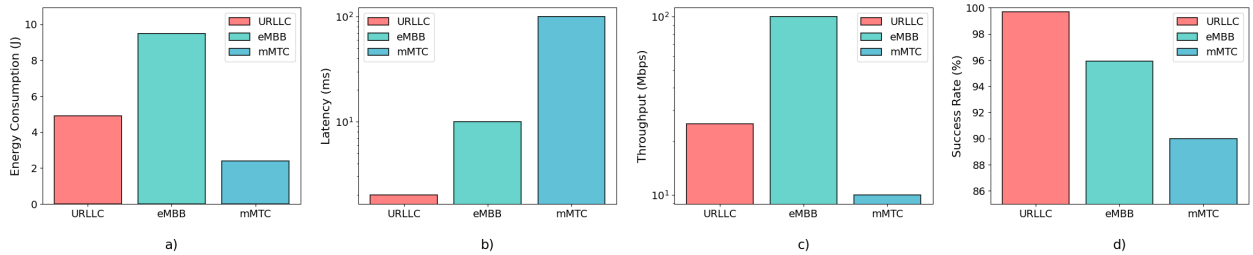

Fig. 4: Performance of the proposed TMT for the different traffic classes: a) Energy consumption; b) Latency; c) Throughput; d) Reliability.

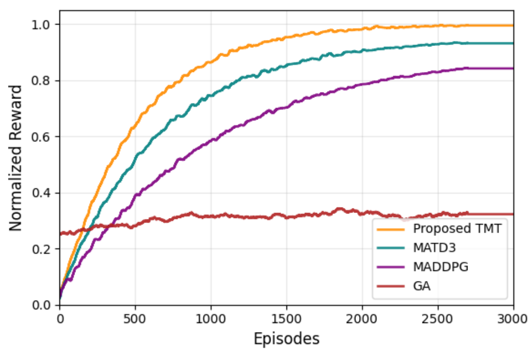

Fig. 5: Analysis of the convergence behavior and performance stability of all compared algorithms in terms of normalized reward over increasing training episodes.

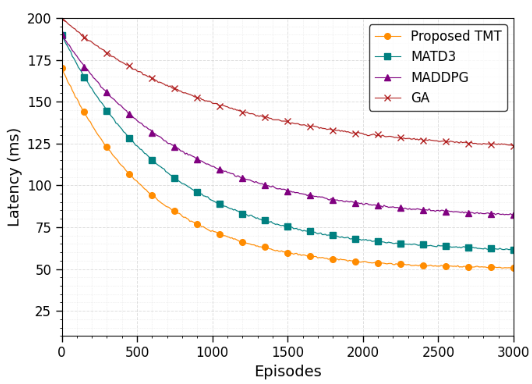

Fig. 6: Comparison of latency reduction achieved by the proposed TMT relative to its benchmarks across training episodes.

classes. The results show that, as the training process evolves, TMT consistently exhibits the lowest latency. This remarkable performed is achieved due to its hierarchical and traffic-aware policy structure that enables adaptive coordination between UAV and MEC O-RUs. As a result of its hierarchical and traffic-aware design, TMT enables efficient task offloading, rapid handover between UAV O-RUs, and efficient resource utilization, thereby enabling fast convergence and reducing delay.

Like TMT, MATD3 shows a steady decrease in latency but a higher latency than TMT because it requires longer training periods to stabilize, resulting in slower learning and less efficient adaptation to dynamic vehicular environments. The other learning-based approach, the MADDPG algorithm, has a moderate latency reduction compared to the TMT and MATD3 algorithms but converges more slowly because of its limited exploration and noise regularization efficiency. The GA consistently shows the highest latency compared to the learning-based algorithms and exhibits the slowest performance improvement due to its static optimization nature and lack of an adaptive learning technique to dynamically adjust its resource allocation strategies over time. Overall, the results demonstrate that the proposed TMT achieves better performance in latency reduction than the other methods, MATD3, MADDPG, and GA, by 25.4%, 33.5%, and 65.1%, respectively.

Throughput. Fig. 7 depicts the throughput performance of TMT and all benchmark approaches as the number of training episodes grows. The proposed TMT framework achieves the highest throughput, thanks to the synergy between the UAV and ground O-RUs and the traffic-aware optimization. These properties indeed lead to improved load balancing, task offloading, and resource allocation decisions, thereby increasing throughput. Similar to TMT, the traditional MATD3 demonstrates improved throughput performance compared to MADDPG and GA, but it has lower throughput than TMT and converges more slowly due to its limited adaptability to complex and dynamic vehicular O-RAN environments.

From the compared learning-based algorithms, the MADDPG algorithm achieves moderate throughput. However, it requires a longer training duration to achieve stable results because of its less effective policy exploration and coordination among agents in this hierarchical and traffic-aware framework. Conversely, the GA algorithm shows the lowest throughput with the slowest improvement compared to the learning-based algorithms. This is because the GA method lacks a learning mechanism to optimally adapt to such complex and dynamic resource allocation strategies. In summary, the proposed TMT substantially outperforms its alternatives in throughput, achieving gains of 12.6%, 28.6%, and 96.5% over MATD3, MADDPG, and GA, respectively, demonstrating its efficiency and adaptability to dynamic and heterogeneous vehicular O-RAN environments.

Energy consumption. Given that mMTC traffic demands highly energy-efficient operation due to its massive low-power device connections, we evaluate the energy consumption performance of the proposed TMT framework in comparison to

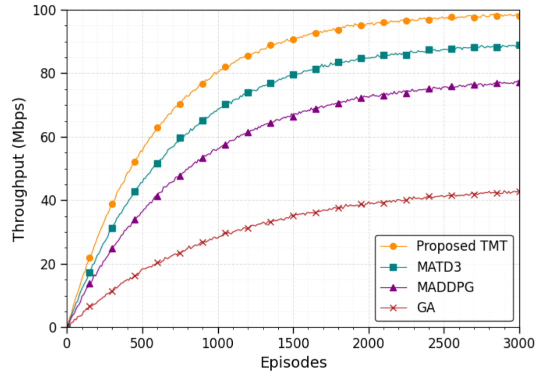

Fig. 7: Analysis of the data transmission performance of the compared algorithms in terms of throughput over the number of learning episodes.

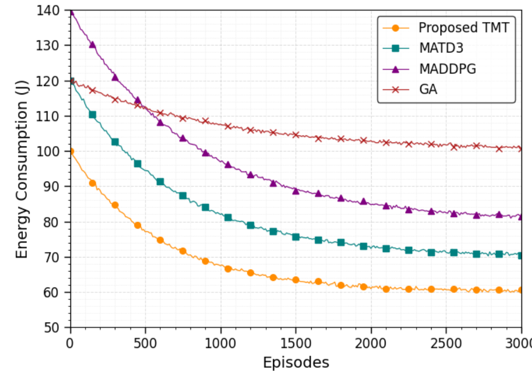

Fig. 8: Energy consumption analysis of all algorithms over increasing training episodes up to 3000.

benchmark methods. With an increase in training episodes, TMT achieves the lowest energy consumption throughout the learning process due to its adaptive task offloading and intelligent and traffic-aware resource coordination between the UAV and ground O-RUs. The hierarchical and traffic-aware design of TMT enables adaptive coordination between UAV and ground O-RUs, allowing dynamic adjustment of task offloading, transmission power, and resource allocation based on traffic conditions. This helps TMT minimize redundant computation and communication overhead, leading to reduced overall energy consumption while maintaining balanced performance across URLLC, eMBB, and mMTC services. In comparison, MATD3 demonstrates a significant reduction in energy consumption compared to MADDPG and GA methods. However, due to its flat multi-agent structure and lack of hierarchical coordination, MATD3 limits its ability to balance computation and communication loads between ground and UAV O-RUs, leading to high energy consumption under dynamic network conditions. While MADDPG moderately minimizes energy consumption compared to TMT and MATD3, it is significantly better than that of the GA approach. This higher energy consumption of MADDPG than TMT and MATD3 comes from its less effective policy coordination and limited ability to adapt to dynamic traffic settings. Since MADDPG lacks hierarchical coordination and traffic prioritization structure, it struggles to balance computational and communication loads across the ground and UAV O-RUs, resulting in redundant processing and high energy consumption. In contrast, GA achieves the highest energy consumption due to its static optimization nature and lack of learning and real-time feedback mechanisms to dynamically adjust computation, communication, and power resource allocation strategies of the agents under heterogeneous and dynamic vehicular network conditions. The proposed TMT algorithm achieves notable energy-efficiency gains, reducing energy consumption by 14.3%, 26.8%, and 40% compared to MATD3, MADDPG, and GA, respectively.

Reliability. We analyze the success rate of the proposed TMT compared to MATD3, MADDPG, and GA. As seen in Fig. 9, all the considered algorithms show increasing performance over time as the number of training episodes increases. However, the proposed TMT achieves a superior success rate than the its alternatives by reaching nearly 99% reliability by the end of the training episodes. This success rate achievement is significant, particularly for URLLC services that demand ultra-low latency and nearly perfect packet delivery. TMT exhibits this remarkable reliability through its hierarchical and traffic-aware architecture, facilitating dynamic resource allocation, coordination between ground and UAV O-RUs, and prioritization of essential URLLC traffic, while adequately supporting eMBB and energy-constrained mMTC traffic, thereby effectively reducing packet losses and service interruptions under dynamic vehicular network conditions. The traditional MATD3 achieves the second-highest success rate, reaching around 95%. The MATD3 has lower reliability than the TMT due to its flat multi-agent structure, which limits coordination between agents and traffic prioritization. The MADDPG algorithm achieves approximately 92%. In addition to its flat multi-agent structure and lack of explicit traffic prioritization, its limited policy exploration makes MADDPG achieve suboptimal resource allocation for latency-sensitive URLLC and throughput-intensive eMBB traffic. Conversely, the GA algorithm demonstrates the lowest success rate across all traffic classes, attaining approximately 80% reliability. Since the GA algorithm uses a static optimization approach, it cannot adapt to dynamic traffic and network conditions, inducing frequent packet drops and service interruptions, particularly for URLLC services.

Ablation Study of TMT Components. To quantify the contribution of each component in the proposed hierarchical TMT algorithm, we conduct a comprehensive ablation study by systematically removing key components, namely, the hierarchical policy structure, inter-agent communication, traffic-aware weighting, and the twin-critic mechanism, and then we evaluate all variants under identical simulation settings using throughput, E2E latency, energy consumption, and reliability. As indicated in Table II, the full TMT model consistently achieves higher performance, with a throughput of 92 Mbps, an E2E latency of 82 ms, an energy usage of 65 J, and a reliability of 99.6%. Notably, removing the hierarchical policy structure reduces scalability and delays convergence, particularly under

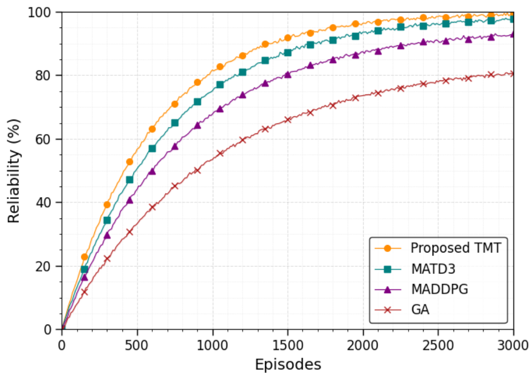

Fig. 9: Analysis of system reliability for all compared algorithms in terms of task completion success rate across the number of training episodes.

TABLE II: Ablation study of TMT components

Model Variant	Throughput (Mbps)	Latency (ms)	Energy (J)	Reliability (%)
Full TMT	92	82	65	99.6
W/o Twin-Critic	88	90	69	98.6
W/o Traffic-Aware Weighting	84	102	74.6	97.8
W/o Hierarchy	81	105	78	97.2
W/o Agents' Communication	77	115	84	96.5

high-traffic conditions. In addition, avoiding inter-agent communication leads to a considerable throughput reduction and a marked rise in latency due to lower coordination among agents, underlining the significance of cooperative behavior in distributed systems. Moreover, removing the traffic-aware weighting mechanism significantly impacts adaptability to dynamic traffic loads, resulting in inefficient resource allocation, lower throughput, and higher latency and energy consumption. Similarly, omitting the twin-critic method produces learning instability and value overestimation, ultimately reducing both throughput performance and overall system efficiency. Collectively, these findings offer convincing evidence that each component performs a distinct yet synergistic role, and most importantly, their seamless integration is key for the TMT algorithm to attain resilient, scalable, and high-efficiency performance in dynamically changing vehicular network settings.

Sensitivity Analysis. To evaluate the robustness and trade-off behavior of the proposed TMT framework, we conduct a sensitivity analysis on weighting parameters associated with the formulated multi-objective, namely latency, energy consumption, throughput, and reliability. The simulation results, summarized in Table III, demonstrate that the proposed TMT algorithm maintains stable performance across all considered weighting configurations, showing its strong robustness to parameter variations. As the latency weight increases, the framework increasingly prioritizes delay minimization, achieving a significantly lower latency of 78 ms. This improvement comes at the expense of higher energy consumption and a significant reduction in throughput. In contrast, increasing energy efficiency weight reduces power consumption to 58 J, while slightly maximizing latency to 96 ms and reducing throughput to 90 Mbps. Similarly, increasing throughput weight allows the system to achieve the highest data rate of 99 Mbps, which leads to moderate increases in energy consumption to 74 J, latency to 101 ms, and a slight reduction in reliability to 94.6%. Moreover, increasing the reliability weight significantly improves successful transmission performance, achieving a 99.9% success rate, while maintaining balanced behavior between latency, energy, and throughput. Furthermore, the balanced weighting configuration provides almost equal importance across URLLC, eMBB, and mMTC objectives to ensure fairness and stable performance across heterogeneous traffic classes. The full TMT prioritizes overall system-level performance through optimized weighting and effectively manages the inherent trade-offs among competing objectives by maintaining low latency for URLLC traffic, high throughput for eMBB services, and reliable large-scale connectivity for mMTC devices, thereby ensuring robust, adaptive, and consistently strong performance under diverse weighting configurations.

TABLE III: Sensitivity analysis of multi-objective weighting parameters in TMT

Case	Latency	Energy	Throughput	Reliability
Full TMT	82 ms	65 J	92 Mbps	99.6%
High Latency Weight	78 ms	71 J	88 Mbps	96.5%
High Energy Weight	96 ms	58 J	90 Mbps	95.8%
High Throughput Weight	101 ms	74 J	99 Mbps	94.6%
High Reliability Weight	90 ms	69 J	90 Mbps	99.9%
Balanced (Equal Weights)	93 ms	63 J	90 Mbps	97.1%

Furthermore, we conduct a sensitivity analysis on key causal intervention parameters, including latency violation thresholds, hysteresis margins, TTT values, and SINR thresholds under heterogeneous traffic classes (URLLC, eMBB, and mMTC). This sensitivity analysis comprehensively evaluates the robustness, adaptability, and stability of the proposed TMT under diverse network dynamics and varying causal intervention parameter configurations. The simulation results, presented in Table IV, show that tighter latency-violation thresholds under URLLC configurations (7–12 ms) improve reliability and throughput by enforcing more aggressive prioritization of delay-sensitive traffic, while slightly increasing energy consumption due to more frequent control interventions. In contrast, relaxed latency-violation thresholds for eMBB and mMTC (40–60 ms and 80–120 ms, respectively) achieve a more balanced energy consumption and higher system efficiency at the cost of latency increase. Increasing hysteresis margins, particularly in mMTC scenarios, improves handover stability by mitigating ping-pong effects and reducing unnecessary switching operations, thereby lowering signaling overhead and energy consumption. However, this introduces slight degradation in throughput and increased latency due to reduced responsiveness under rapidly varying channel conditions. Similarly, higher TTT values enhance connection stability by reducing frequent handovers and improving delay consistency but may slightly reduce throughput due to delayed adaptation to channel dynamics. Finally, SINR threshold variations significantly influence overall system performance. Higher SINR requirements in URLLC improve reliability and throughput by enforcing stricter link quality conditions, although they introduce moderate increases in latency and energy consumption. In contrast, lower thresholds in mMTC improve connectivity density at the expense of reduced link quality. Overall, the proposed TMT framework demonstrates stable and robust performance, effectively balancing latency, reliability, throughput, and energy efficiency across varying parameter settings and highlighting low-to-moderate sensitivity to causal intervention thresholds in highly dynamic vehicular O-RAN environments.

The proposed TMT framework is inherently designed to operate under constrained and highly dynamic vehicular O-RAN environments, including harsh interference conditions and limited computational resources. In scenarios with degraded SINR caused by dense traffic, channel congestion, or increased inter-cell interference, the traffic-aware hierarchical policy adaptively reallocates communication and computation resources to maintain reliable connectivity and QoS performance. In particular, the integration of SINR-aware decision-making and adaptive traffic prioritization enables the framework to respond to lower link conditions while minimizing performance degradation dynamically. Furthermore, the distributed multi-agent architecture and centralized training with decentralized execution (CTDE) paradigm provides scalability and robustness under computationally constrained environments by enabling decentralized execution and localized decision-making with reduced processing overhead. Thus, the framework can be extended to city-scale vehicular O-RAN scenarios with increased UAV density and vehicle population.

Generalization Capability. The proposed TMT framework is modular and generalizable, enabling extension to various transportation scenarios beyond the UAV-assisted vehicular O-RAN setting. Built on the CTDE paradigm, it does not rely on scenario-specific heuristics, which enhances its transferability. For applications such as urban intersections or mixed traffic environments, only the state representation needs to be extended (e.g., lane-level positions, signal phases, or vehicle-type indicators), while the underlying policy architecture remains unchanged. Overall, the framework requires modifications only in environment-specific components, such as state/action spaces and reward design, demonstrating its high level of flexibility.

VII. Conclusion

This paper presented a Traffic-Aware Dynamic Resource Allocation (TADRA) architecture for UAV-assisted vehicular O-RAN to address the challenges of dynamic traffic conditions, infrastructure failures, and stringent QoS requirements. The UAVs are deployed as intelligent aerial O-RUs to enhance service continuity during ground infrastructure failures. We modeled the resource allocation and traffic prioritization problem as a MOMDP to optimize latency, energy consumption, throughput, and reliability under heterogeneous traffic conditions. To solve the formulated optimization problem, the proposed framework integrates RAN Intelligence with a Hierarchical Multi-Agent DRL-based solution (TMT), ensuring adaptive, resilient, and efficient O-RAN resource orchestration. The hierarchical design enables the system to dynamically adapt to vehicular mobility, varying load, and service requirements. Our results demonstrate that the proposed TMT significantly outperforms conventional DRL-based approaches, i.e., MATD3 and MADDPG, as well as the GA heuristic scheme, achieving 17% lower latency, 10% higher throughput, 14% lower energy consumption, and 6.5% higher reliability. Future research directions include extending the framework to federated multi-agent learning and cross-layer optimization and further improving robustness, explainability, and real-time adaptability in 6G O-RAN deployments.

References

[1] R. Firouzi and R. Rahmani, "Delay-sensitive resource allocation for iot systems in 5g o-ran networks," *Internet of Things*, vol. 26, p. 101131, 2024.

[2] H. Ko, J. Lee, and S. Pack, "Pdras: Priority-based dynamic resource allocation scheme in 5g network slicing," *Journal of Network and Systems Management*, vol. 30, no. 4, p. 68, 2022.

[3] M. Kalntis, G. Iosifidis, and F. A. Kuipers, "Adaptive resource allocation for virtualized base stations in o-ran with online learning," *IEEE Transactions on Communications*, vol. 73, no. 3, pp. 1787–1800, 2025.

[4] H. Li, X. Tang, D. Zhai, R. Zhang, B. Li, H. Cao, N. Kumar, and A. Almogren, "Energy-efficient deployment and resource allocation for o-ran-enabled uav-assisted communication," *IEEE Transactions on Green Communications and Networking*, vol. 8, no. 3, pp. 1128–1140, 2024.

[5] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding o-ran: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023.

[6] C. Pham, F. Fami, K. K. Nguyen, and M. Cheriet, "When ran intelligent controller in o-ran meets multi-uav enable wireless network," *IEEE Transactions on Cloud Computing*, vol. 11, no. 3, pp. 2245–2259, 2023.

[7] M. Mehdaoui and A. Abouaomar, "Dynamics of resource allocation in o-rans: An in-depth exploration of on-policy and off-policy deep reinforcement learning for real-time applications," *arXiv preprint arXiv:2412.01839*, 2024.

[8] A. Tuerxun and A. Nakao, "Sora: Energy-efficient resource allocation in open radio access network with online learning," *IEEE Access*, vol. 13, pp. 113 686–113 701, 2025.

[9] T. Demir, M. Masoudi, E. Bjornson, and C. Cavdar, "Cell-free massive mimo in o-ran: Energy-aware joint orchestration of cloud, fronthaul, and radio resources," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 356–372, 2024.

[10] Q. Wang, Y. Liu, Y. Wang, X. Xiong, J. Zong, J. Wang, and P. Chen, "Resource allocation based on radio intelligence controller for open ran toward 6g," *IEEE Access*, vol. 11, pp. 97 909–97 919, 2023.

[11] R. M. Sohaib, S. Tariq Shah, and P. Yadav, "Towards resilient 6g o-ran: An energy-efficient urllc resource allocation framework," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 7701–7714, 2024.

[12] Y. Chen, R. Yao, H. Hassanieh, and R. Mittal, "Channel-aware 5g ran slicing with customizable schedulers," in *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*, 2023, pp. 1767–1782.

[13] V.-D. Nguyen, T. X. Vu, N. T. Nguyen, D. C. Nguyen, M. Juntti, N. C. Luong, D. T. Hoang, D. N. Nguyen, and S. Chatzinotas, "Network-aided intelligent traffic steering in 6g o-ran: A multi-layer optimization framework," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 389–405, 2024.

[14] Q. Ren, O. Abbasi, G. K. Kurt, H. Yanikomeroglu, and J. Chen, "Caching and computation offloading in high altitude platform station (haps) assisted intelligent transportation systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9010–9024, 2022.

[15] A. Alsharoa and M.-S. Alouini, "Improvement of the global connectivity using integrated satellite-airborne-terrestrial networks with resource optimization," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5088–5100, 2020.

[16] Y. A. Ergu, V.-L. Nguyen, R.-H. Hwang, Y.-D. Lin, C.-Y. Cho, H.-K. Yang, H. Shin, and T. Q. Duong, "Efficient adversarial attacks against drl-based resource allocation in intelligent o-ran for v2x," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 1, pp. 1674–1686, 2025.

TABLE IV: Sensitivity analysis of key causal intervention parameters in the TMT framework

Intervention Parameter	Tested values for each traffic class (URLLC, eMBB, mMTC)	Latency	Reliability	Throughput	Energy consumption
Latency violation thresholds	[7–12, 40–60, 80–120]ms	∼4.3%↓	∼2.1%↑	∼3.5%↑	∼2.8%↑
Hysteresis margins	[1–3, 2–4, 3–5] dB	∼3.8%↑	±1.5%	∼2.6%↓	∼2.1%↓
TTT thresholds	[15–20, 25–75, 50–100] ms	∼4.9%↑	∼2.4%↓	∼3.1%↓	∼1.9%↓
SINR thresholds	[10–15, 5–10, 3–8] dB	∼6.1%↑	∼3.4%↑	∼2.9%↑	∼2.5%↑

[17] Y. Cui, X. Yang, P. He, R. Wang, and D. Wu, "Urllc-embb hierarchical network slicing for internet of vehicles: An aoi-sensitive approach," *Vehicular Communications*, vol. 43, p. 100648, 2023.

[18] B. Zheng, K. Zhuo, H. Wu, and T. Xie, "Multi-priority queueing mechanism for channel threshold based multiple access in fanets," in *2021 IEEE 9th International Conference on Information, Communication and Networks (ICICN)*, 2021, pp. 15–20.

[19] Y. Cui, X. Yang, P. He, D. Wu, and R. Wang, "O-ran slicing for multi-service resource allocation in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 7, pp. 9272–9283, 2024.

[20] A. Staffolani, V.-A. Darvariu, L. Foschini, M. Girolami, P. Bellavista, and M. M. Foschini, "Prorl: Proactive resource orchestrator for open rans using deep reinforcement learning," *IEEE Transactions on Network and Service Management*, vol. 21, no. 4, pp. 3933–3944, 2024.

[21] X. He and C. Lv, "Towards energy-efficient autonomous driving: A multi-objective reinforcement learning approach," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 5, pp. 1329–1331, 2023.

[22] L. Wu, C. Zhang, B. Zhang, J. Du, and J. Qu, "Toward energy-efficiency: Integrating matd3 reinforcement learning method for computational offloading in ris-aided uav-mec environments," *IEEE Internet of Things Journal*, vol. 12, no. 14, pp. 26 582–26 595, 2025.

[23] A. M. Seid, G. O. Boateng, B. Mareri, G. Sun, and W. Jiang, "Multi-agent drl for task offloading and resource allocation in multi-uav enabled iot edge network," *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4531–4547, 2021.

[24] J. Ackermann, V. Gabler, T. Osa, and M. Sugiyama, "Reducing overestimation bias in multi-agent domains using double centralized critics," *arXiv preprint arXiv:1910.01465*, 2019.

[25] J. Chen, J. Kang, M. Xu, Z. Xiong, D. Niyato, C. Chen, A. Jamalipour, and S. Xie, "Multiagent deep reinforcement learning for dynamic avatar migration in aiot-enabled vehicular metaverses with trajectory prediction," *IEEE Internet of Things Journal*, vol. 11, no. 1, pp. 70–83, 2024.

PLACE PHOTO HERE

Hayla Nahom Abishu Biography text here.

PLACE PHOTO HERE

Ahmed Badawy Biography text here.

PLACE PHOTO HERE

Amr Mohamed Biography text here.

PLACE PHOTO HERE

Carla Fabiana Chiasserini Biography text here.