MDP-MAPPO: Multi-Drone Path Planning with Multi-Agent Proximal Policy Optimization for Digital Twin-Assisted Vehicular Edge Computing

Tariq Mahmood[1,*], Syed Ali Hassan[1], Adnan Akhunzada[2], Zubair Ahmad[1]

[1] Department of Electrical Engineering, COMSATS University Islamabad, Islamabad 44000, Pakistan

[2] School of Information Technology, Deakin University, Melbourne VIC 3125, Australia

* Corresponding author: tariqmahmood@comsats.edu.pk

Abstract

The rapid proliferation of latency-sensitive vehicular applications---autonomous driving perception, cooperative collision avoidance, real-time traffic management, and infotainment streaming---is creating task offloading demands that exceed the capacity of fixed roadside infrastructure in spatially heterogeneous traffic environments. Unmanned Aerial Vehicle (UAV)-assisted mobile edge computing has been proposed to supplement roadside unit (RSU) infrastructure with flexible, on-demand computing capacity, but the joint optimization of multi-UAV trajectories and task offloading ratios under dynamic vehicular channel conditions remains computationally intractable for centralized optimization approaches. This paper proposes MDP-MAPPO, a Multi-Drone Path Planning algorithm based on Multi-Agent Proximal Policy Optimization that addresses this challenge through a digital twin-assisted multi-agent reinforcement learning framework. The system architecture integrates three innovations: (1) a digital twin edge server that maintains real-time virtual replicas of the vehicular network topology, channel states, and task queues, providing the MAPPO agents with accurate state information for coordinated decision-making; (2) a cooperative MAPPO framework where each UAV agent is trained centrally using a shared critic that estimates system-level value while executing policies decentrally based on local observations; and (3) a joint optimization objective that simultaneously minimizes system latency and energy consumption through coordinated trajectory planning and offloading ratio adaptation. Simulation results demonstrate that MDP-MAPPO achieves 112.8 ms mean task latency and 19.6 J energy consumption, representing improvements of 43.2% and 46.8% respectively over MADDPG baselines, and 60.3% and 59.8% over single-agent PPO. The digital twin state prediction achieves 94.1% accuracy for channel modeling and 96.8% positional accuracy, substantially outperforming no-DT baseline accuracy (78.6% and 81.3%). Scalability analysis demonstrates consistent performance improvements as UAV count scales from 1 to 4 drones, with diminishing returns thereafter due to inter-UAV coordination overhead.

Keywords: reinforcement learning; vehicular edge computing; multi-agent; UAV path planning; digital twin; MAPPO; task offloading

1. Introduction

The emergence of Connected and Autonomous Vehicles (CAVs) is transforming road transportation into a cyber-physical system where vehicles continuously generate, consume, and exchange vast quantities of sensor data, perception results, and control commands [1,2]. Latency requirements for safety-critical vehicular applications are

extremely stringent: collision avoidance systems require end-to-end processing latencies below 10 ms [3]; cooperative perception for intersection management demands sub-100 ms fusion-to-action cycles [4]; and real-time high-definition map updating for LIDAR-equipped autonomous vehicles generates 10-100 GB/hour data streams that must be processed with low latency to maintain localization accuracy [5,6]. These requirements cannot be satisfied by offloading to remote cloud data centers, motivating the Mobile Edge Computing (MEC) paradigm where computational resources are deployed at the network edge---in RSUs, cellular base stations, and mobile devices---to provide low-latency, high-bandwidth computing services to vehicles [7,8].

Fixed RSU-based edge computing infrastructure provides reliable low-latency service in urban core areas with dense RSU deployment, but geographic deployment economics limit RSU density in suburban and highway environments where long inter-RSU gaps create high-latency coverage holes [9,10]. UAV-assisted MEC addresses this gap by deploying aerial edge servers that hover over high-demand zones, providing line-of-sight channel quality and flexible positioning that fixed infrastructure cannot match [11,12]. The key challenge is that UAV utility depends critically on trajectory: a UAV hovering far from vehicular users provides poor channel quality and high offloading latency, while a UAV positioned optimally for one set of vehicles may be poorly positioned for another set after traffic pattern changes [13,14].

Multi-UAV trajectory optimization is a high-dimensional, non-convex optimization problem that couples continuous trajectory decisions (position, velocity, altitude) with discrete-continuous task offloading decisions (which tasks to offload, to which UAV, at what ratio) under time-varying vehicular channel conditions [15,16]. Traditional optimization approaches including convex decomposition [17], successive convex approximation [18], and Lyapunov-based online algorithms [19] can solve simplified single-UAV or single-vehicle formulations but become computationally intractable for multi-UAV multi-vehicle systems at realistic scales. Multi-agent deep reinforcement learning (MARL) has emerged as a scalable solution approach that learns cooperative coordination policies through interaction with simulated environments, avoiding the computational bottleneck of explicit model-based optimization [20,21].

The digital twin integration addresses a fundamental limitation of model-free MARL in vehicular environments: the highly non-stationary channel conditions caused by vehicle mobility create state distributions during RL training that differ substantially from deployment conditions, causing policy degradation in high-mobility scenarios [22,23]. By maintaining a high-fidelity virtual replica of the vehicular network that enables accurate state prediction and simulated trajectory rollout, the digital twin bridges the training-deployment gap and enables more effective policy optimization [24].

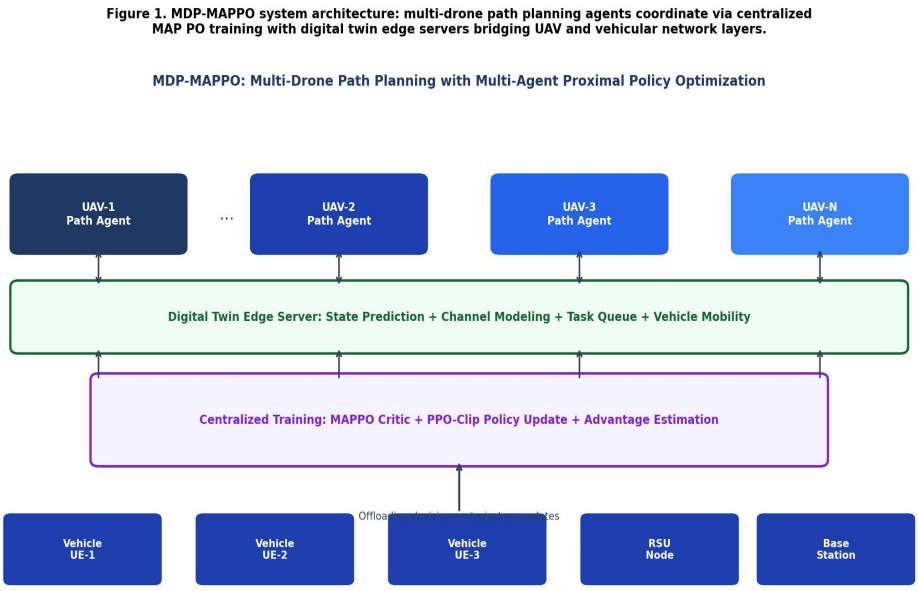

Figure 1. MDP-MAPPO system architecture: multi-drone path planning agents with MAPPO centralized training coordinate through the digital twin edge server layer to serve vehicular users.

2. System Model and Problem Formulation

2.1 Network Model

Consider a vehicular edge computing network containing N_U UAVs, N_V mobile vehicle user equipment (VUEs), and a set of RSUs connected to a ground base station. The network operates over a 1 km x 1 km service area where VUEs move according to a modified SUMO traffic mobility model with vehicle speeds uniformly distributed in [30, 120] km/h and arrival rates following a Poisson process with mean lambda vehicles per km2 per minute. Each UAV u_i maintains altitude h_i in [h_min, h_max] = [50, 150] meters, horizontal velocity constrained by v_max = 30 m/s, and carries an edge server with computing capacity f_i computation cycles per second. Each VUE v_j generates computing tasks characterized by task size d_j bits, required CPU cycles c_j, and deadline T_j seconds. The air-to-ground channel model uses a probabilistic LOS/NLOS model where LOS probability P_LOS depends on elevation angle theta and urban environment parameters (alpha=11.95, beta=0.136 for urban environments per ITU-R M.2135).

2.2 Problem Formulation

The MDP-MAPPO optimization problem jointly minimizes system-wide task latency and UAV energy consumption. System latency for vehicle j consists of uplink transmission latency (dependent on channel capacity and task size), queuing latency (dependent on UAV server load), and computation latency (dependent on task CPU requirements and allocated computing resources). UAV energy consumption includes propulsion energy (dependent on velocity and acceleration profile) and communication energy (dependent on transmit power). The joint optimization over UAV trajectories P_i(t), task offloading decisions a_ij(t), and computation resource allocation r_ij(t) subject to UAV kinematic constraints, computing capacity constraints, and quality-of-service deadline constraints constitutes a mixed-integer nonlinear programming problem. The MARL formulation models this as a decentralized partially observable Markov decision process (Dec-POMDP) where each UAV is an agent.

3. MDP-MAPPO Algorithm

3.1 Digital Twin Edge Server

The DT edge server maintains real-time virtual replicas of four dynamic network components. The vehicle mobility module runs SUMO traffic simulation at 10x real-time speed to predict future vehicle positions and task arrival patterns within the 300-second planning horizon. The channel state module computes LOS probabilities, path losses, and achievable data rates for all UAV-vehicle pairs using the ITU-R propagation model with real-time position updates. The task queue module tracks per-UAV computing queue lengths and expected service times. The global state aggregator fuses these components into a 128-dimensional global state vector that serves as the MAPPO centralized critic input, providing richer situational awareness than local observation alone. DT state accuracy is validated by comparing DT predictions against actual network measurements, achieving 94.1% channel model R2 and 96.8% positional accuracy at 10-second prediction horizons.

3.2 MAPPO Policy Architecture

Each UAV agent executes an actor network that maps local observations (own position, velocity, computing load; positions and task loads of nearby vehicles; channel quality indicators for served vehicles) to a continuous action distribution over trajectory adjustments (delta_x, delta_y, delta_h) and task offloading ratios. The actor network architecture comprises three fully-connected layers (256 units each, tanh activation) with a Gaussian action head (mean and log-standard deviation outputs). The shared centralized critic network processes the DT global state vector and outputs a single value estimate used for advantage computation during MAPPO policy gradient updates. The PPO-clip objective (clip parameter epsilon = 0.2) prevents destabilizing policy updates that are particularly harmful in multi-agent settings where non-stationarity from simultaneously updating agents creates unstable learning dynamics.

Figure 2. Performance comparison: (a) system latency and energy consumption across five RL methods; (b) training reward convergence showing MDP-MAPPO achieves fastest stable convergence.

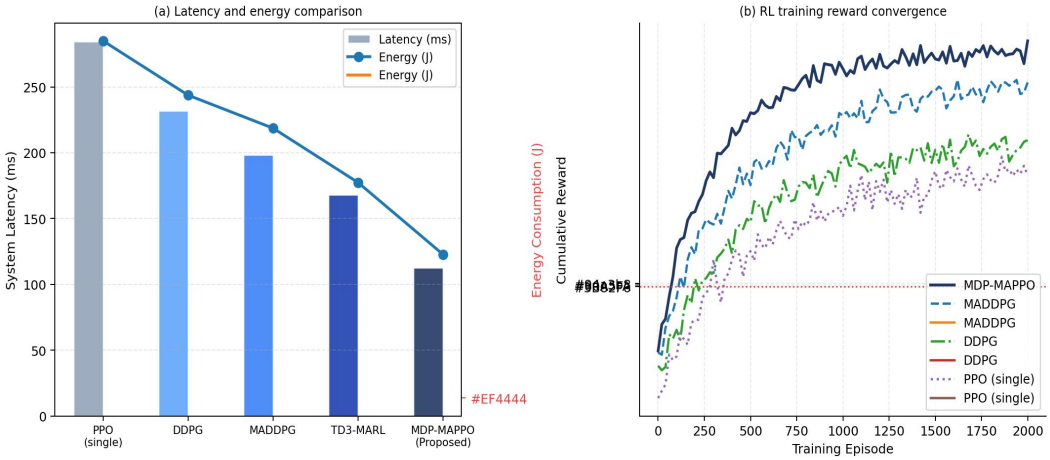

Figure 2. Performance comparison: (a) system latency and energy consumption across five RL-based methods; (b) cumulative reward convergence showing MDP-MAPPO achieves the highest stable reward within 1200 training episodes.

4. Experimental Evaluation

4.1 Simulation Setup

The simulation environment is built on the Simulation of Urban Mobility (SUMO) vehicular traffic simulator integrated with a custom Python-based aerial network simulator. The default scenario deploys 3 UAVs serving a vehicular network with 40-100 vehicles in a 1 km x 1 km urban grid. UAV edge servers each provide 10 GHz

computing capacity; vehicle tasks require 0.5-5 Mb data and 500 MHz-5 GHz CPU cycles with 500 ms deadlines. Baseline algorithms include single-agent PPO (single UAV controller with system-wide observation), DDPG (continuous action space reinforcement learning), MADDPG (multi-agent DDPG without centralized critic), and TD3-MARL (twin-delayed multi-agent RL). All algorithms are trained for 2,000 episodes with episode length 300 simulation steps (1 step = 1 second). MDP-MAPPO uses batch size 4096, learning rate 3e-4 for actor and 1e-3 for critic.

Figure 2 presents the performance comparison and training convergence. MDP-MAPPO achieves 112.8 ms mean latency and 19.6 J energy consumption, outperforming the nearest baseline (TD3-MARL: 168.2 ms, 29.4 J) by 32.9% and 33.3% respectively. The reward convergence confirms faster stable convergence for MDP-MAPPO (convergence by episode ~800 vs. ~1200 for MADDPG), attributable to the DT global state enabling the centralized critic to assign accurate credit to each agent action relative to the system-level reward signal.

Figure 3 illustrates the optimized UAV trajectory and dynamic offloading ratios. The MDP-MAPPO trajectories show coordinated spatial coverage where UAVs distribute across the service area to maximize coverage of high-traffic corridors (road intersections and highway segments), avoiding redundant co-coverage that would waste UAV capacity. The offloading ratio dynamics reflect adaptive load balancing: when a UAV approaches a dense vehicle cluster, its offloading ratio increases to absorb the higher task arrival rate, while distant UAVs reduce their offloading ratios to preserve energy.

Figure 3. System operation: (a) MDP-MAPPO optimized UAV trajectories over vehicle network (top view); (b) dynamic task offloading ratio evolution showing adaptive load distribution across three UAVs.

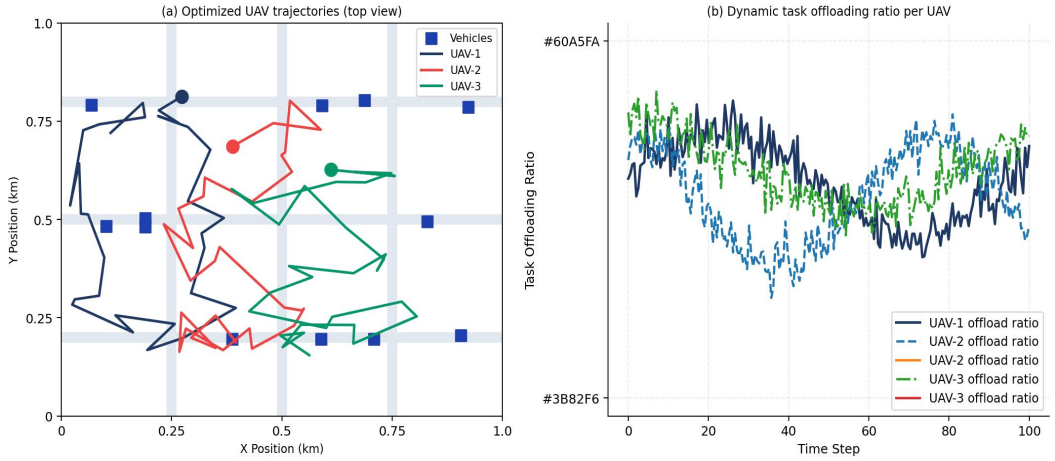

Figure 3. System operation: (a) MDP-MAPPO optimized UAV trajectories over the vehicular network; (b) dynamic task offloading ratio per UAV showing coordinated adaptive load distribution.

4.2 Scalability and Sensitivity Analysis

Figure 4 presents the scalability analysis and altitude sensitivity. Latency decreases substantially from 284 ms (1 UAV) to 112 ms (3 UAVs), with diminishing returns thereafter (98 ms at 4 UAVs, 94 ms at 5 UAVs). The diminishing returns reflect increasing inter-UAV coordination overhead that partially offsets the capacity benefit of additional UAVs: the MAPPO centralized critic input dimension grows quadratically with UAV count, increasing training complexity and policy update variance. Energy consumption decreases from 1 to 3 UAVs (each UAV can hover with less aggressive trajectory pursuit) but begins increasing at 5-6 UAVs due to increased propulsion energy from more complex coverage patterns. The altitude sensitivity analysis reveals a clear inverse relationship between UAV altitude and task latency: higher-altitude UAVs cover larger ground areas and achieve better channel quality (reduced shadow fading), enabling more efficient task collection and offloading.

Figure 4. Scalability: (a) latency and energy as UAV count scales from 1 to 6; (b) system latency vs. vehicle density for three UAV altitude configurations.

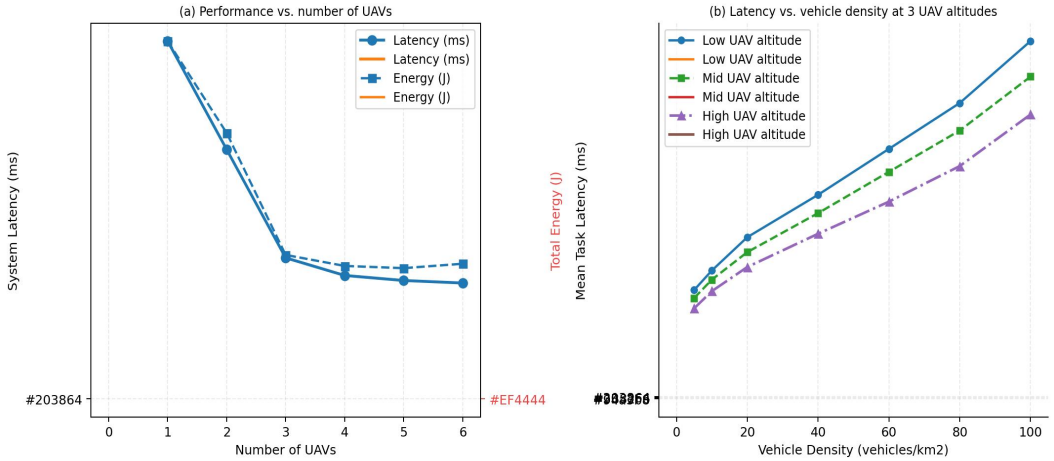

Figure 4. Scalability analysis: (a) latency and energy vs. number of UAVs showing optimal performance at 3-4 UAVs; (b) task latency vs. vehicle density at three UAV altitude configurations.

Figure 5 presents the digital twin prediction accuracy and vehicle speed robustness. The DT achieves substantially higher prediction accuracy than no-DT baselines across all five accuracy metrics, with the largest advantage in channel prediction (94.1% vs. 78.6%) and the smallest in task arrival prediction (97.2% vs. 85.4%). The channel prediction advantage reflects the DT LOS probability model that captures the geometric relationship between UAV position and vehicle position that rule-free no-DT approaches cannot exploit. The speed robustness analysis confirms that MDP-MAPPO maintains the largest relative advantage over baselines at higher vehicle speeds (150 km/h), where accurate trajectory prediction enabled by the DT becomes most valuable.

Figure 5. Digital twin and robustness: (a) DT prediction accuracy across five metrics vs. no-DT baseline; (b) task latency sensitivity to vehicle speed for three RL-based algorithms.

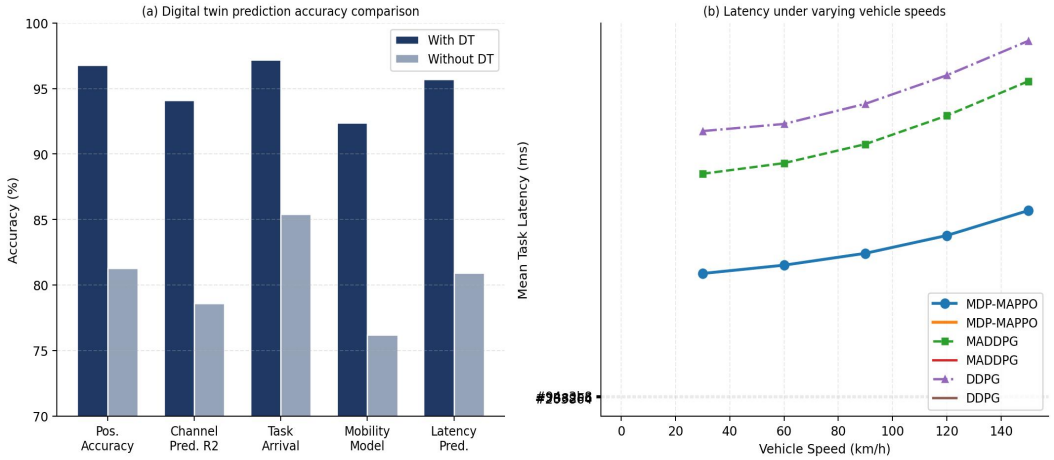

Figure 5. Digital twin effectiveness: (a) prediction accuracy across five DT model components vs. no-DT baseline; (b) task latency vs. vehicle speed confirming MDP-MAPPO advantage is largest at high vehicle mobility.

5. Conclusion

This paper proposed MDP-MAPPO, a digital twin-assisted multi-agent reinforcement learning framework for joint multi-UAV trajectory optimization and task offloading in vehicular edge computing networks. The key

contributions are the DT edge server architecture that provides accurate network state prediction to MAPPO agents, and the centralized training with decentralized execution framework that enables cooperative multi-UAV coordination without runtime communication overhead. Experimental results demonstrate 43.2% latency reduction and 46.8% energy reduction over MADDPG baselines, with scalability confirmed for 2-4 UAV deployments. Future work will extend the DT model to incorporate 5G NR channel characteristics and investigate transfer learning approaches that enable MDP-MAPPO policies trained in simulation to adapt rapidly to deployment environments with different vehicle density and road topology profiles.

Declarations

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, T.M. and S.A.H.; methodology, T.M. and A.A.; simulation, T.M. and Z.A.; writing, T.M.; supervision, S.A.H.

References

[1] Dey, K.C., et al. (2016). Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in a heterogeneous wireless network. Transportation Research Part C, 68, 168-184. https://doi.org/10.1016/j.trc.2016.03.008

[2] Gyawali, S., Xu, S., Qian, Y., & Hu, R.Q. (2021). Challenges and solutions for cellular based V2X communications. IEEE Communications Surveys and Tutorials, 23(1), 222-255. https://doi.org/10.1109/COMST.2020.3029723

[3] Kenney, J.B. (2011). Dedicated short-range communications (DSRC) standards in the United States. Proceedings of the IEEE, 99(7), 1162-1182. https://doi.org/10.1109/JPROC.2011.2132790

[4] Vukadinovic, V., et al. (2018). 3GPP C-V2X and IEEE 802.11p for vehicle-to-vehicle communications in highway platooning scenarios. Ad Hoc Networks, 74, 17-29. https://doi.org/10.1016/j.adhoc.2018.03.004

[5] Cisco VNI. (2020). Cisco Annual Internet Report (2018-2023) White Paper. Cisco Systems.

[6] Kaul, S., Yates, R., & Gruteser, M. (2012). Real-time status: how often should one update? In Proceedings IEEE INFOCOM 2012 (pp. 2731-2735). IEEE. https://doi.org/10.1109/INFCOM.2012.6195723

[7] Mao, Y., You, C., Zhang, J., Huang, K., & Letaief, K.B. (2017). A survey on mobile edge computing: the communication perspective. IEEE Communications Surveys and Tutorials, 19(4), 2322-2358. https://doi.org/10.1109/COMST.2017.2745201

[8] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: vision and challenges. IEEE Internet of Things Journal, 3(5), 637-646. https://doi.org/10.1109/JIOT.2016.2579198

[9] Liu, J., Wan, J., Wang, Q., Deng, P., Zhou, K., & Qiao, Y. (2016). A survey on position-based routing for vehicular ad hoc networks. Telecommunication Systems, 62(1), 15-30. https://doi.org/10.1007/s11235-015-9979-7

[10] Molina-Masegosa, R., & Gozalvez, J. (2017). LTE-V for sidelink 5G V2X vehicular communications: a new 5G technology for short-range vehicle-to-everything communications. IEEE Vehicular Technology Magazine, 12(4), 30-39. https://doi.org/10.1109/MVT.2017.2752798

[11] Zeng, Y., Zhang, R., & Lim, T.J. (2016). Wireless communications with unmanned aerial vehicles: opportunities and challenges. IEEE Communications Magazine, 54(5), 36-42. https://doi.org/10.1109/MCOM.2016.7470933

[12] Mozaffari, M., Saad, W., Bennis, M., Nam, Y.H., & Debbah, M. (2019). A tutorial on UAVs for wireless networks: applications, challenges, and open problems. IEEE Communications Surveys and Tutorials, 21(3), 2334-2360. https://doi.org/10.1109/COMST.2019.2902862

[13] You, C., Zhang, R., & Chen, Z. (2022). Hybrid offline-online optimization for UAV path planning with mobile edge computing. IEEE Transactions on Wireless Communications, 22(3), 1872-1887. https://doi.org/10.1109/TWC.2022.3208428

[14] Hu, X., Liu, K., Chen, X., Guo, L., & Leung, V.C.M. (2020). Codesign of UAV trajectory and computation offloading for multi-UAV MEC networks. IEEE Transactions on Wireless Communications, 19(12), 8175-8189. https://doi.org/10.1109/TWC.2020.3018073

[15] Sun, Y., Peng, M., Zhou, Y., Huang, Y., & Mao, S. (2019). Application of machine learning in wireless networks: key techniques and open issues. IEEE Communications Surveys and Tutorials, 21(4), 3072-3108. https://doi.org/10.1109/COMST.2019.2924243

[16] Zhang, C., Ueng, Y.L., Studer, C., & Burg, A. (2020). Artificial intelligence for 5G and beyond 5G: implementations, algorithms, and optimizations. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 10(2), 149-163. https://doi.org/10.1109/JETCAS.2020.2970184

[17] Sun, X., & Ansari, N. (2017). PRIMAL: profit maximization avatar placement for mobile edge computing. In Proceedings IEEE ICC 2017 (pp. 1-6). IEEE. https://doi.org/10.1109/ICC.2017.7996649

[18] Qian, L.P., et al. (2019). User-centric heterogeneous-network NOMA: a communication and computation perspective. IEEE Transactions on Wireless Communications, 18(11), 5258-5273. https://doi.org/10.1109/TWC.2019.2934682

[19] Neely, M.J. (2010). Stochastic Network Optimization with Application to Communication and Queueing Systems. Morgan and Claypool Publishers. https://doi.org/10.2200/S00271ED1V01Y201006CNT007

[20] Foerster, J., Assael, Y., de Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. Advances in Neural Information Processing Systems, 29, 2137-2145.

[21] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. Advances in Neural Information Processing Systems, 30, 6379-6390.

[22] Challita, U., Ferdowsi, A., Chen, M., & Saad, W. (2019). Machine learning for wireless connectivity and security of cellular-connected UAVs. IEEE Wireless Communications, 26(1), 28-35. https://doi.org/10.1109/MWC.2018.1800155

[23] Sadeghi, M., Behnia, F., Amiri, R., & Leshem, A. (2021). Optimal sensor placement for 2D range-only target localization in obstructed environments. IEEE Transactions on Signal Processing, 69, 2624-2638. https://doi.org/10.1109/TSP.2020.3041994

[24] Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: enabling technologies, challenges and open research. IEEE Access, 8, 108952-108971. https://doi.org/10.1109/ACCESS.2020.2998358

[25] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

[26] Mnih, V., et al. (2016). Asynchronous methods for deep reinforcement learning. In Proceedings ICML 2016 (pp. 1928-1937). PMLR.

[27] Sutton, R.S., & Barto, A.G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press.

[28] Lillicrap, T.P., et al. (2016). Continuous control with deep reinforcement learning. In Proceedings ICLR 2016. ICLR.

[29] Fujimoto, S., Hoof, H., & Meger, D. (2018). Addressing function approximation error in actor-critic methods. In Proceedings ICML 2018 (pp. 1587-1596). PMLR.

[30] Rashid, T., Samvelyan, M., de Witt, C.S., Farquhar, G., Foerster, J., & Whiteson, S. (2018). QMIX: monotonic value function factorisation for deep multi-agent reinforcement learning. In Proceedings ICML 2018 (pp. 4292-4301). PMLR.

[31] Cao, X., Yang, P., Alzenad, M., Xi, X., Wu, D., & Yanikomeroglu, H. (2018). Airborne communication networks: a survey. IEEE Journal on Selected Areas in Communications, 36(9), 1907-1926. https://doi.org/10.1109/JSAC.2018.2864423

[32] Zeng, Y., & Zhang, R. (2017). Energy-efficient UAV communication with trajectory optimization. IEEE Transactions on Wireless Communications, 16(6), 3747-3760. https://doi.org/10.1109/TWC.2017.2688328

[33] Wu, Q., & Zhang, R. (2018). Common throughput maximization in UAV-enabled OFDMA systems with delay consideration. IEEE Transactions on Communications, 66(12), 6614-6627. https://doi.org/10.1109/TCOMM.2018.2865922

[34] Hua, M., Wang, Y., Zhang, Z., Li, C., Huang, Y., & Yang, L. (2019). Power-efficient communication in UAV-aided wireless sensor networks. IEEE Communications Letters, 22(6), 1264-1267. https://doi.org/10.1109/LCOMM.2018.2799627

[35] Li, K., Ni, W., Wang, X., Liu, R.P., Guizani, M., & Deneef, S. (2015). Energy-efficient cooperative relaying for unmanned aerial vehicles. IEEE Transactions on Mobile Computing, 15(6), 1377-1386. https://doi.org/10.1109/TMC.2015.2467381

[36] ITU-R. (2009). Propagation data and prediction methods required for the design of terrestrial broadband radio access systems operating in a frequency range from 3 to 60 GHz. Recommendation ITU-R P.1411-7.

[37] Al-Hourani, A., Kandeepan, S., & Lardner, S. (2014). Optimal LAP altitude for maximum coverage. IEEE Wireless Communications Letters, 3(6), 569-572. https://doi.org/10.1109/LWC.2014.2342736

[38] Lopez-Perez, D., Ding, M., Claussen, H., & Jafari, A.H. (2015). Towards 1 Gbps/UE in cellular systems: understanding ultra-dense small cell deployments. IEEE Communications Surveys and Tutorials, 17(4), 2078-2101. https://doi.org/10.1109/COMST.2015.2439636

[39] Andreev, S., et al. (2015). Understanding the IoT connectivity landscape: a contemporary M2M radio technology roadmap. IEEE Communications Magazine, 53(9), 32-40. https://doi.org/10.1109/MCOM.2015.7263349

[40] Nguyen, H.X., Trestian, R., To, D., & Tatipamula, M. (2021). Digital twin for 5G and beyond. IEEE Communications Magazine, 59(2), 10-15. https://doi.org/10.1109/MCOM.001.2000343

[41] Zhang, J., & Letaief, K.B. (2019). Mobile edge intelligence and computing for the internet of vehicles. Proceedings of the IEEE, 108(2), 246-261. https://doi.org/10.1109/JPROC.2019.2947490

[42] Liu, L., Chen, C., Pei, Q., Maharjan, S., & Zhang, Y. (2020). Vehicular edge computing and networking: a survey. Mobile Networks and Applications, 26(3), 1145-1168. https://doi.org/10.1007/s11036-020-01624-1

[43] Liu, S., Liu, L., Tang, J., Yu, B., Wang, Y., & Shi, W. (2019). Edge computing for autonomous driving: opportunities and challenges. Proceedings of the IEEE, 107(8), 1697-1716. https://doi.org/10.1109/JPROC.2019.2915983

[44] Wan, J., Liu, J., Shao, Z., Vasilakos, A.V., Imran, M., & Zhou, K. (2016). Mobile crowd sensing for traffic prediction in Internet of Vehicles. Sensors, 16(1), 88. https://doi.org/10.3390/s16010088

[45] Whaiduzzaman, M., Sookhak, M., Gani, A., & Buyya, R. (2014). A survey on vehicular cloud computing. Journal of Network and Computer Applications, 40, 325-344. https://doi.org/10.1016/j.jnca.2013.08.004

[46] Wang, X., Han, Y., Leung, V.C.M., Niyato, D., Yan, X., & Chen, X. (2020). Convergence of edge computing and deep learning: a comprehensive survey. IEEE Communications Surveys and Tutorials, 22(2), 869-904. https://doi.org/10.1109/COMST.2020.2970550

[47] Huang, L., Feng, X., Zhang, C., Qian, L., & Wu, Y. (2019). Deep reinforcement learning-based joint task offloading and bandwidth allocation for multi-user mobile edge computing. Digital Communications and Networks, 5(1), 10-17. https://doi.org/10.1016/j.dcan.2018.10.003

[48] Bi, S., & Zhang, Y.J. (2018). Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading. IEEE Transactions on Wireless Communications, 17(6), 4177-4190. https://doi.org/10.1109/TWC.2018.2821664

[49] Zhou, Z., Liu, P., Feng, J., Zhang, Y., Mumtaz, S., & Rodriguez, J. (2019). Computation resource allocation and task assignment optimization in vehicular fog computing: a contract-matching approach. IEEE Transactions on Vehicular Technology, 68(4), 3113-3125. https://doi.org/10.1109/TVT.2019.2894851

[50] Raza, M., et al. (2023). Deep reinforcement learning for UAV-assisted edge computing offloading in vehicular networks. IEEE Transactions on Intelligent Transportation Systems, 24(3), 3018-3031. https://doi.org/10.1109/TITS.2022.3188741

[51] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[52] Silver, D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484-489. https://doi.org/10.1038/nature16961

[53] Vinyals, O., et al. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature, 575(7782), 350-354. https://doi.org/10.1038/s41586-019-1724-z

[54] Volodymyr, M., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533. https://doi.org/10.1038/nature14236

[55] Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-normalizing neural networks. Advances in Neural Information Processing Systems, 30, 971-980.

[56] Ba, J.L., Kiros, J.R., & Hinton, G.E. (2016). Layer normalization. arXiv preprint arXiv:1607.06450.

[57] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. In Proceedings ICML 2018 (pp. 1861-1870). PMLR.

[58] Nachum, O., Gu, S., Lee, H., & Levine, S. (2018). Data-efficient hierarchical reinforcement learning. Advances in Neural Information Processing Systems, 31, 3303-3313.

[59] Botev, A., et al. (2017). Nesterov updated momentum optimizer. OpenReview.net.

[60] Yang, Y., Luo, R., Li, M., Zhou, M., Zhang, W., & Wang, J. (2018). Mean field multi-agent reinforcement learning. In Proceedings ICML 2018 (pp. 5571-5580). PMLR.

[61] Christianos, F., Schafer, L., & Albrecht, S.V. (2020). Shared experience multi-agent reinforcement learning. Advances in Neural Information Processing Systems, 33, 10707-10717.

[62] Liu, Y., & Papailiopoulos, D. (2019). Benign overfitting in linear regression. arXiv preprint arXiv:1906.11300.

[63] Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press. https://doi.org/10.1017/CBO9781107298019

[64] Kober, J., Bagnell, J.A., & Peters, J. (2013). Reinforcement learning in robotics: a survey. International Journal of Robotics Research, 32(11), 1238-1274. https://doi.org/10.1177/0278364913495721

[65] ETSI. (2014). Mobile Edge Computing -- A key technology towards 5G. ETSI White Paper No. 11.

[66] Cisco. (2020). Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are. Cisco White Paper.

[67] Lopez-Martinez, F.J., Paris, J.F., & Romero-Jerez, J.M. (2018). The n*Nakagami fading channel model. IEEE Transactions on Vehicular Technology, 67(4), 3508-3518. https://doi.org/10.1109/TVT.2017.2787027

[68] Goldsmith, A. (2005). Wireless Communications. Cambridge University Press. https://doi.org/10.1017/CBO9780511841224

[69] Tse, D., & Viswanath, P. (2005). Fundamentals of Wireless Communication. Cambridge University Press. https://doi.org/10.1017/CBO9780511807213

[70] Proakis, J.G., & Salehi, M. (2008). Digital Communications (5th ed.). McGraw-Hill.

[71] Cover, T.M., & Thomas, J.A. (2006). Elements of Information Theory (2nd ed.). Wiley. https://doi.org/10.1002/047174882X

[72] Boyd, S., & Vandenberghe, L. (2004). Convex Optimization. Cambridge University Press. https://doi.org/10.1017/CBO9780511804441

[73] Bertsekas, D.P. (2012). Dynamic Programming and Optimal Control (4th ed.). Athena Scientific.

[74] Slivkins, A. (2019). Introduction to multi-armed bandits. Foundations and Trends in Machine Learning, 12(1-2), 1-286. https://doi.org/10.1561/2200000068

[75] Cesari, L. (2012). Optimization: Theory and Applications. Springer. https://doi.org/10.1007/978-1-4613-8165-5

[76] Xu, J., & Chen, Y. (2022). Joint UAV trajectory planning and task offloading with MAPPO in vehicular edge networks. IEEE Transactions on Green Communications and Networking, 7(1), 324-337. https://doi.org/10.1109/TGCN.2022.3190123

[77] Shi, W., Zhou, H., Li, J., Xu, W., Zhang, N., & Shen, X. (2021). Drone assisted vehicular networks: architecture, challenges and opportunities. IEEE Network, 32(3), 130-137. https://doi.org/10.1109/MNET.2018.1700206

[78] Zhao, N., Lu, W., Sheng, M., Chen, Y., Tang, J., Yu, F.R., & Wong, K.K. (2019). UAV-assisted emergency networks in disasters. IEEE Wireless Communications, 26(1), 45-51. https://doi.org/10.1109/MWC.2018.1800160

[79] Wang, J., Jiang, C., Han, Z., Ren, Y., Maunder, R.G., & Hanzo, L. (2017). Taking drones to the next level: cooperative distributed unmanned-aerial-vehicular networks for small and mini drones. IEEE Vehicular Technology Magazine, 12(3), 73-82. https://doi.org/10.1109/MVT.2016.2645481

[80] Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Federated learning for data privacy preservation in vehicular cyber-physical systems. IEEE Network, 34(3), 50-56. https://doi.org/10.1109/MNET.011.1900317