Cross-Domain Resource Orchestration for the Edge-Computing-Enabled Smart Road

Quan Yuan, Jinglin Li, Haibo Zhou, Guiyang Luo, Tao Lin, Fangchun Yang, and Xuemin (Sherman) Shen

Abstract

Intelligent driving plays a role in significantly improving the safety and efficiency of transportation systems. As the onboard capabilities of perception, comprehension, and decision making are limited, vehicles can employ the edge computing infrastructure of the smart road to enhance their intelligence. Therefore, the smart road is considered an intelligent Internet of Things system. It provides vehicles with not only the road space in the transportation domain, but also the communication, sensing, and computing resources in the information domain to improve the composite quality of intelligent driving. However, the resources in the information and transportation domains are complicatedly coupled, and the orchestration of these cross-domain resources is confronted with the huge state-action space, which cannot be solved in a real-time manner. In this article, we investigate the fundamental research challenges in cross-domain resource orchestration for the smart road, and design a multi-agent-based framework. Within the framework, each vehicle is associated with an exclusive agent on the edge cloud, and the agents utilize swarm intelligence to jointly optimize the traffic flow and information flow for their respective vehicles. Specifically, a value iteration network is used by agents to learn the routing behavior of vehicles, and a multi-agent deep reinforcement learning method is proposed, enabling agents to cooperatively learn decentralized resource optimization policies. To verify the effectiveness of the proposed framework, a cross-domain resource orchestration prototype is implemented and evaluated.

Introduction

Intelligent and connected vehicles (ICVs) are the key to safer and more efficient transportation. Existing intelligent vehicles on the road use a combination of sensors, such as cameras, millimeter-wave (mmWave) radars, and lidars, to understand the ambient environments and achieve automated driving. However, these onboard sensors are limited to line of sight, which cannot provide panoramic information for vehicles to maneuver correctly. Furthermore, intelligent vehicles usually rely on their respective "brains" (i.e., computing platforms including powerful CPUs and GPUs) to independently make driving and routing decisions. Under the circumstances, each vehicle tries to maximize its own profit, which can hardly achieve the system optimum for road traffic.

With the deployment of advanced roadside infrastructure, which contains plenty of communication, sensing, and computing resources, the road is becoming smarter to boost intelligent driving [1]. Thus, the smart road is an intelligent Internet of Things (IoT) system. By leveraging the multi-dimensional resources, the smart road can enhance vehicles' environment perception and decision making capabilities:

- 5G and beyond: With greater speed, higher reliability, and lower latency, the fifth generation (5G) mobile network can interconnect the intelligence/knowledge of vehicles and roadside infrastructure [2], which empowers distributed artificial intelligence (AI) to efficiently cooperate with them.
- Cognitive sensing: The roadside sensors usually have better views of the traffic environment, which can provide additional information for vehicles to maneuver correctly. As the cooperative comprehension of the environment is more efficient than the exchange of raw sensing data, the smart road provides on-demand features extracted from raw sensing data to enhance vehicles' environment perception.
- Ubiquitous computing: By leveraging edge computing, the smart road can take the compute-intensive tasks (e.g., semantic image segmentation, motion planning, and route planning) over from intelligent vehicles. This paradigm exploits advanced wireless technology to extend the computational power of vehicles. Furthermore, the smart road can jointly solve offloaded tasks from multiple vehicles, thus coordinating their actions to improve traffic safety and efficiency.

Therefore, the smart road is an essential factor to ensure the quality of intelligent driving, while reducing the hardware requirements of intelligent vehicles. In this article, the complex interaction of cross-domain resources in the smart road is studied. We aim to identify and address the key challenges in the orchestration of these cross-domain resources. To this end, we propose a multi-agent learning framework and implement a prototype to verify its feasibility. It is expected that our research will promote more activities in this important area.

Cross-Domain Resources in the Smart Road

In the transportation systems dominated by intelligent vehicles, the smart road consists of cross-domain resources to facilitate intelligent driving:

Digital Object Identifier: 10.1109/MNET.011.2000007

Quan Yuan, Jinglin Li (corresponding author), Guiyang Luo, Tao Lin, and Fangchun Yang are with the Beijing University of Posts and Telecommunications; Haibo Zhou is with Nanjing University; Xuemin (Sherman) is with the University of Waterloo.

 0890-8044/20/$25.00 © 2020 IEEE

- In the transportation domain, it provides road space for vehicle traveling.
- In the information domain, it leverages communication, sensing, and computing resources to enhance vehicles' perception and planning capabilities.

Therefore, the resource scheduling in both the transportation domain and the information domain are critical to the quality of service (QoS) of intelligent driving. Specifically, road traffic scheduling concerns the resource of road capacity to achieve traffic equilibria and improve traffic efficiency, while the resource allocation in roadside infrastructure for ICVs determines the reliability and effectiveness of intelligent driving. The cross-domain resources in the smart road and their interactions are shown in Fig. 1. It should be noted that the computation tasks unrelated to intelligent driving are not considered in this article.

There are many studies on the resource scheduling problem in either the transportation domain or the information domain. In this section, we first review related works, and then analyze the mutual influence of cross-domain resource scheduling.

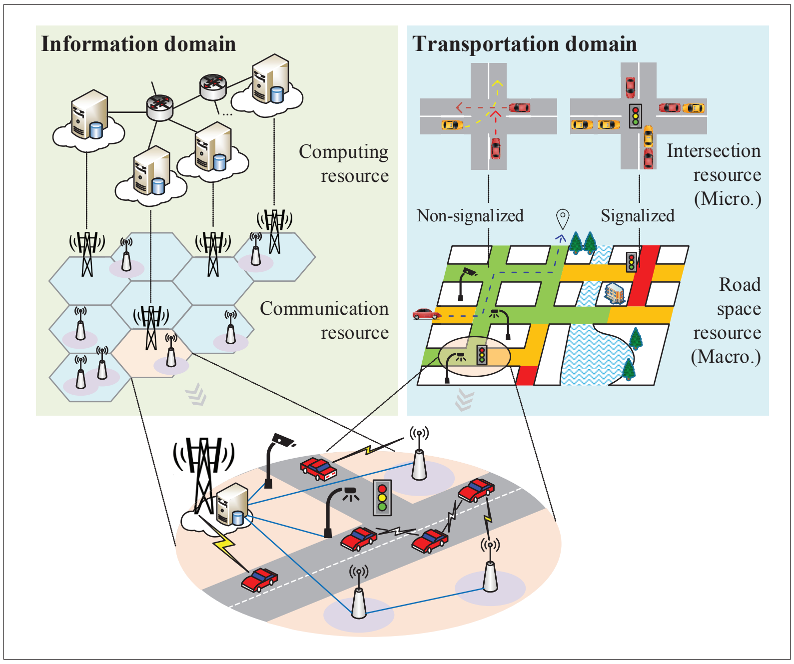

FIGURE 1. The cross-domain resources in the smart road.

Resource Scheduling in the Transportation Domain

Road space is the resource provided for vehicles to travel safely and efficiently. From a macroscopic perspective, the resource scheduling of road space aims to optimize routes for vehicles and optimize traffic flow distribution over the road network. From a microscopic perspective, signalized and non-signalized intersections are scheduled to reduce traverse delay while ensuring safety.

Route planning is a real-time large-scale optimization problem. To reduce computational complexity, some studies have modeled vehicles as distributed agents and performed cooperative route planning. In [3], Lin *et al.* exploited evolutionary game among vehicles to calculate optimal routes, and proved that its strategy selection converges to Nash equilibrium. However, the convergence time of multiplayer stochastic games will be unacceptable for citywide planning. To achieve scalability, Cao *et al.* [4] trained vehicle routing via deep reinforcement learning, which maximizes the probability of arriving on time. Considering the trade-off between system optimum and user optimum, Groot *et al.* [5] adopted a reverse Stackelberg game between the road authority and vehicles to optimize traffic flow distribution.

As for intersection scheduling, Chu *et al.* [6] applied multi-agent deep reinforcement learning to adaptive traffic signal control. Qian *et al.* [7] utilized vehicular networks to implement safe and efficient driving at non-signalized intersections for autonomous vehicles.

Resource Scheduling in the Information Domain

Multi-dimensional resource scheduling has been well investigated in edge-computing-assisted vehicular networks. Software-defined networking (SDN) [8] takes advantage of centralized intelligence to manage and allocate network resources. Based on SDN architecture, Li *et al.* [9] modeled multi-dimensional resource scheduling as a partially observable Markov decision process (POMDP) and used value iteration to jointly optimize networking, caching, and computing. Due to the complicated coupling of multi-dimensional resources, the central controller can hardly know a priori the effects of its actions on system performance. To this end, He *et al.* [10] proposed a deep reinforcement learning (DRL)-based resource orchestration method, with which the central controller learns an effective policy via trial-and-error search.

The central controller's action space grows exponentially with the number of users, which is not scalable to large networks. Through cooperative networks [11], the central controller can be decomposed into multiple simpler controllers to reduce the complexity of a large action space. In this way, Nasir *et al.* [12] leveraged multi-agent deep Q-learning to distributedly schedule power allocation in wireless networks. Considering the spectrum competition among selfish vehicles, Tian *et al.* [13] utilized the evolutionary game to optimize channel access for cognitive vehicular networks. In addition, the high mobility of vehicles poses great challenges to effective resource scheduling in vehicular networks. Aissioui *et al.* [14] investigated resource allocation for service migration across multiple edge clouds, which supports service continuity for autonomous vehicles.

Resource Interaction between the Transportation Domain and the Information Domain

The previous subsections have shown that existing studies usually optimize the resources in the transportation domain and information domain separately. However, the resources in these two domains intensely interact with each other, so the separate optimization cannot guarantee the composite QoS of intelligent driving. On one hand, the already deployed base stations and edge servers cannot always accommodate the offloaded tasks of nearby vehicles. If the resource scheduling (e.g., route planning) of road space does not take into account the resource status in the information domain, the edge computing of

Multi-scale spatiotemporal optimization creates new opportunities for improving the utility of edge computing, while increasing the complexity to solve the problem. Specifically, the coordination between different spatiotemporal scales is challenging.

vehicles may deteriorate due to the shortage of roadside communication and computing resources. This is because the route planning may make the distribution of vehicles as well as their edge computing demands not conform to the capacity distribution of base stations and edge servers. The imbalance between resource demand and supply in the information domain will prevent vehicles from comprehending correctly and reacting rapidly to the traffic environment, which decreases the availability and reliability of intelligent driving. On the other hand, if the allocation of communication, computing, and caching resources does not consider the mobility and routing of vehicles, the self-directed movements of vehicles will cause disorder of resource allocation in the information domain. Although existing studies have used statistic mobility patterns of vehicles to optimize edge computing, the dynamic route planning will break the predicted mobility patterns and make the resource scheduling in edge clouds deviate from the system optimum. This phenomenon will diminish the utility of edge computing in vehicular networks, and eventually worsen the timeliness and correctness of vehicles' offloaded tasks. Therefore, it is necessary to coordinate the information flow and the traffic flow. Only by orchestrating the cross-domain resources for the smart road can the QoS of intelligent driving be guaranteed.

Key Research Challenges

In this section, we discuss some key research challenges of cross-domain resource orchestration for the smart road, resulting from the intense interaction of the information flow and the traffic flow.

Multi-Scale Spatiotemporal Optimization

Existing resource allocation methods for edge computing in vehicular networks are usually performed according to real-time resource status. As the influence of edge resource allocation is spatiotemporally limited, the respective optimization problem is not hard to solve. However, in a smart road system, the vehicles offload intelligent driving related tasks constantly, so the distribution of vehicles has a great impact on the distribution of edge resource demands. Therefore, the edge resource demands can be redistributed in both space and time by guiding the traffic flow. Cross-domain resource orchestration becomes a multi-scale spatiotemporal optimization problem, considering not only the short-term regional allocation of edge resources, but also the long-term global redistribution of edge resource demands.

Multi-scale spatiotemporal optimization creates new opportunities for improving the utility of edge computing, while increasing the complexity of solving the problem. Specifically, the coordination between different spatiotemporal scales is challenging. First, it is not trivial to determine whether to reactively "borrow" resources from nearby edge clouds or proactively reshape the distribution of resource demands. Second, the rewards for optimizations on different scales are asynchronous. The allocation of edge resources can receive immediate reward, while the redistribution of edge resources can only receive delayed reward. These asynchronous rewards can hardly be handled by conventional optimization methods, which remains a research issue.

Bounded Rationality of Individuals

In vehicular networks, SDN controllers can rationally perform resource allocation to optimize networking, computing, and caching. By leveraging this centralized scheme, the SDN controller can accurately anticipate the performance of edge resource allocation. However, the routing decisions of intelligent vehicles exhibit bounded rationality as a result of an individual vehicle's limited knowledge and formed preference. The vehicles do not always follow the suggested routing from the central controller, causing the distribution of edge resource demands to deviate from system optimum. To improve routing compliance, the cross-domain resource orchestration has to consider the bounded rationality of individuals. Only in this way can the resource orchestration perform as expected and approximate the system optimum.

There are mainly two issues when dealing with the bounded rationality of individuals. To begin with, it is essential to model the routing behavior of each vehicle. The models, which are learned from vehicles' historical trajectories, should be able to predict routing actions in different traffic environments (e.g., traffic status, traffic events, and traffic control schemes). Based on these routing behavior models, the central controller can induce vehicles to select desired routes by using traffic control schemes or game theoretic methods. Additionally, the user optimum and system optimum should be balanced for cross-domain resource orchestration. For example, it is unacceptable to pursue system-level resource utility at the expense of numerous big detours. Therefore, Nash equilibria for the central controller and intelligent vehicles should be found, where everyone can converge to a satisfactory policy.

Coupling of Cross-Domain Resources

The coupling of cross-domain resources requires the joint optimization of information flow and traffic flow. Developing from single-domain resource allocation to cross-domain resource orchestration, the system will be confronted with the curse of dimensionality. The super controller orchestrating cross-domain resources will see a huge state space of the environment and determine the best reaction from a huge action space. Specifically, the state space covers every factor that affects orchestration policies. It is a compound state space including road traffic related state (e.g., multi-grained traffic status and traffic demands) and vehicular network related state (e.g., wireless link status, the loads of edge servers, and the demands for computation offloading). Accordingly, the compound action space contains the resource scheduling for a large number of vehicles in both the information domain (e.g., how to transmit, where to compute, and what to cache) and the transportation domain (e.g., traffic signal timing and routing). Furthermore, the action space grows exponentially with the number of vehicles.

The super controller can hardly solve such a complicated optimization problem in real time. Even though the super controller can use a deep neural network to comprehend the high-dimensional state of the environment, the huge action space is still an obstacle. To this end, it is necessary to decompose the joint optimization problem into independent subproblems. However, finding a natural decomposition is not trivial due to the tight coupling of cross-domain resources. With swarm intelligence, distributed controllers instead of a super central controller will make decisions based on their local observations, which reduces the dimensionality of the action space. Nevertheless, it is an important issue to nominate capable entities of the smart road as the distributed controllers. Furthermore, how to coordinate these distributed controllers to achieve the overall optimization objective remains a great challenge.

General Framework and Potential Solutions

In this section, we propose the general framework and potential solutions, which take advantage of swarm intelligence and deep learning, to address the above-mentioned research challenges.

A Cross-Domain Resource Orchestration Framework

As a single controller cannot cope with the huge state space and action space, we propose a multi-agent-based cross-domain resource orchestration framework, which is shown in Fig. 2. In this framework, each vehicle is associated with an always-on exclusive agent on the edge cloud. The agent is a bundle of individualized data (e.g., preference and knowledge) of the vehicle/driver it serves and the processing logic (e.g., deep neural networks) based on the data. These agents not only help their respective vehicles make driving and routing decisions, but also schedule edge resources to make themselves work normally and efficiently. In other words, the agents orchestrate cross-domain resources for their respective vehicles. It should be noted that ICVs and non-ICVs will coexist on the smart road. The ICVs can access cross-domain resources, and the non-ICVs only access the resources in the transportation domain. Therefore, in this framework, only ICVs are associated with agents, while the road traffic formed by non-ICVs is considered as part of the environment.

Understanding the environment is a necessary condition of resource orchestration. By analyzing the sensing data of a vehicle and the resource status of the edge cloud, each agent has a local view of its environment. Furthermore, multi-agent interactions enable each agent to obtain a broader and more accurate view of the environment. Then the agents cooperatively perform real-time resource scheduling to achieve system-wide resource optimization. Furthermore, with the strict constraint on service delay, the agents should be migrated across edge servers to follow the mobility of vehicles, which involves cross-edge resource scheduling. Finally, edge clouds execute the optimized resource allocation, and vehicles perform the planned driving and routing actions. It can be seen that the proposed multi-agent-based framework is a typical cyber-physical system. To prevent multi-agent interactions from growing exponentially with the number of agents, the interactions can be approximated by those between a single agent and the total effect of the population.

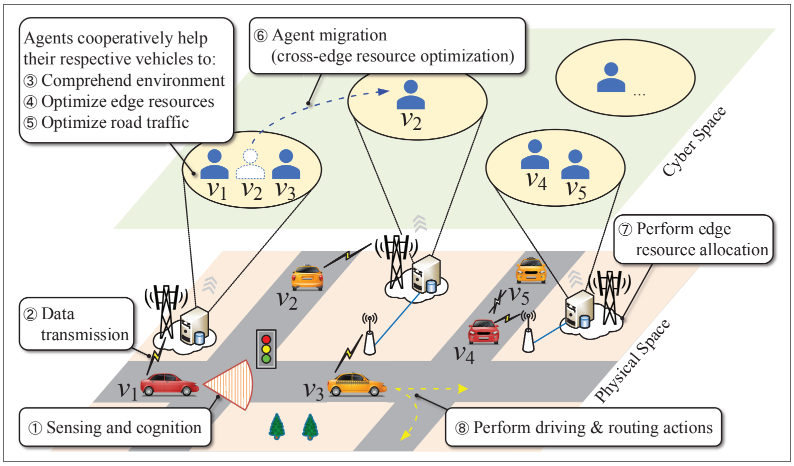

FIGURE 2. Multi-agent-based cross-domain resource orchestration.

Routing Behavior Learning

Each agent communicates and cooperates with other agents to make routing decisions on behalf of the vehicle it serves. To make the vehicle comply with the decision, the agent should learn the vehicle's routing preference with respect to various traffic statuses. In this article, we design a value iteration network (VIN) to train the agent's routing policy, which is supervised by the vehicle's trajectories. The VIN can learn the nature of the routing behavior rather than cloning it. Therefore, the VIN can generalize to plan accurate routing actions under unseen scenarios. The sketch of the method is given as follows, and the details can be found in [15].

The architecture of VIN is shown in Fig. 3, and it maps the destination and real-time traffic status to the values of possible routing actions. Specifically, the traffic status is represented in a macroscopic way: the urban area is partitioned into disjoint grids, and the traffic status is denoted as the average traverse time between adjacent grids. The VIN uses a neural network to compute the following value iteration,

$$V_{i+1}(s) = \max_{a} Q_i(s,a), \tag{1}$$

$$Q_i(s,a) = R(s,a) + \gamma \sum_{s'} P(s' \mid s,a) V_i(s'), \tag{2}$$

where state $s \in \mathcal{S}$ denotes the vehicle position; action $a \in \mathcal{A}$ denotes a possible routing action; $Q(\cdot)$ and $V(\cdot)$ are the state-action value function and the state value function, respectively; $R(\cdot)$ is the reward function; $P(\cdot)$ is the state transmission probability function; and γ is a discount factor. Let V^* denote the state value function for optimal routing policy. By value iteration, V_i can converge to V^* when $i \to \infty$, and the vehicle's routing behavior can be learned. In Fig. 3, $R(\cdot)$ is converted from the routing destination and current traffic status, and $P(\cdot)$ is implemented through the convolution operator. Next, Eq. 2 is calculated via sum-pooling, which obtains the value of each state-action

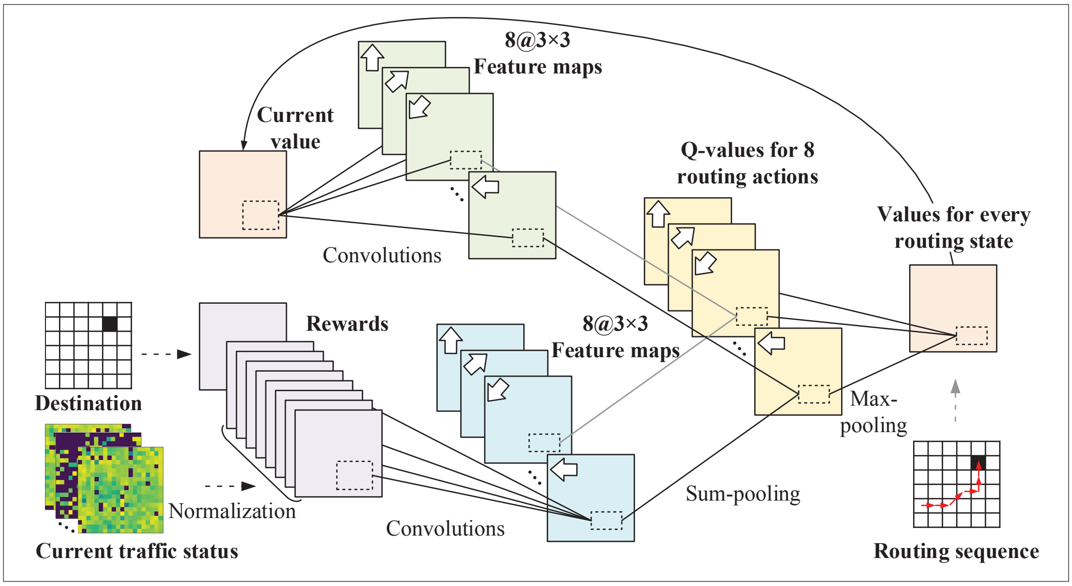

FIGURE 3. VIN-based routing behavior learning.

pair; and Eq. 1 is calculated via max-pooling, which obtains the value of each state. Then the state values are fed back to perform the next round of value iterations. After N iterations, the final state value V^* is achieved, and the cross-entropy loss between the routing action derived from V^* and the actual action of the vehicle is back-propagated. By training with the historical trajectory data, VIN can provide the state value that reflects the vehicle's routing preference. Let $\mathcal{N}_s \subset \mathcal{S}$ be the set of neighboring grids of s; then $\hat{s}$ = arg $\max_{s' \in \mathcal{N}_s} V^*(s')$ will be the next grid selected by the vehicle according to the learned routing preference from s to the destination.

Multi-Agent Deep-Reinforcement-Learning-Based Resource Orchestration

Based on the framework in Fig. 2, we propose a multi-agent deep reinforcement learning (MADRL) method, which enables a set of agents to learn resource orchestration policies by interacting with the complex environment. Compared to traditional optimization methods, MADRL has better scalability because its complexity grows linearly instead of exponentially with the number of vehicles. In this subsection, the basic ideas of MADRL-based resource orchestration are introduced. The smart road network is segmented into hexagonal grids based on the wireless coverage of macrocells. We focus on the grid-level orchestration, where the resources in the information domain and those in the transportation domain can be abstracted at the same granularity. It is considered that the intra-grid resources are allocated effectively by adopting state-of-the-art methods (e.g., game-theory-based route planning, and DRL-based communication, computing, and caching). In the remainder of this subsection, the key points to construct state-action space for the MADRL are described, and then a convolution-based deep Q-network is designed.

With the observed environment states, all agents learn to select actions that jointly maximize a single reward signal accumulated over time. As the environment involves cross-domain resources, the state space should contain the vehicle's origin-destination (OD), the agent's host server, macroscopic traffic conditions, the loads of base stations and edge servers, and so on. It can be seen that the state space reflects the total effect of population and the status of the ego agent. As for the action space, it consists of vehicle routing and agent migration, which coordinates traffic flow and information flow.

A deep Q-network (DQN) is used to map the current observed environment state to the state-action values. As the environment state has spatial structures, we use multi-channel matrix signals to represent the environment state and then take advantage of convolution to capture the spatial features. The architecture of DQN for each agent is shown in Fig. 4. The DQN consists of two branches: an agent migration network branch and a route planning network branch. The training of these two branches is governed by a joint reward,

$$r_v = \tilde{r}_v - p_v - \left(c_v^{mgt} + \omega c_v^{tfc}\right), \quad (3)$$

where $\tilde{r}_v$ is a reward for vehicle v reaching its destination, p_v is the penalty carried when the service delay (i.e., wireless transmission delay, backhaul delay, and computing delay) is not satisfied, c_v^{mgt} is the agent migration cost, c_v^{tfc} is the travel time for the OD pair, and ω is the weight parameter. The joint reward coordinates two branches to achieve the joint optimization of the information and transportation domains.

MADRL takes advantage of efficient cooperation among agents to solve complex optimization problems. A straightforward method enables each agent to independently learn an individual Q function, but this method may not converge because each agent's learning is confused by the exploration and exploitation of others. To overcome this problem, we train decentralized policies in a centralized way, that is, centralized training with decentralized execution. Considering the homogeneity of agents, we train a single DQN using the experiences gathered from all agents, which effectively increases the amount of training data generated per step of the environment. In addition, to decrease convergence time, we just train

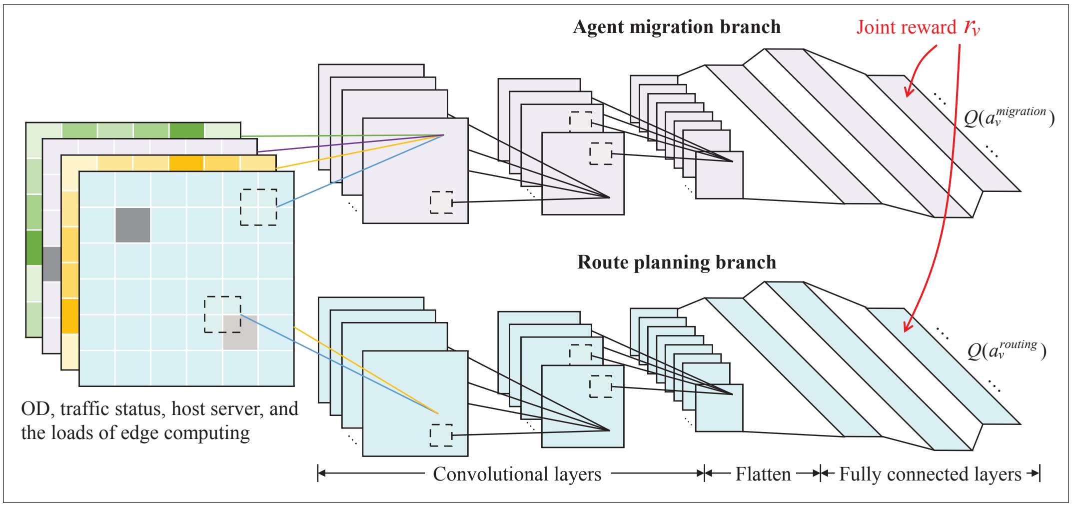

FIGURE 4. The deep Q-network architecture for MADRL-based resource orchestration.

one branch during a period of time and keep the other branch fixed during this period. In this way, the optimal migration is learned under a stable route planning policy, and then the optimal routing is learned under a stable agent migration policy. This procedure should be repeated until the training converges.

As the number of intelligent vehicles increases, more and more agents are created in the system. The newly created agents can obtain model parameters directly from the central trainer. Since each agent's policy depends on the total effect of the population, the system can admit a number of new agents while keeping the overall performance nearly unchanged. Furthermore, the continuous centralized training based on the newly gathered experience should periodically transfer the up-to-date model parameters to all agents. In this way, the system can smoothly adapt to the increase in the number of agents.

Performance Evaluation

In this section, we implement a prototype of the proposed multi-agent-based cross-domain resource orchestration framework, and the performance of the discussed techniques is evaluated using trace-driven simulations.

Evaluation of Routing Behavior Learning

This section evaluates the proposed VIN-based routing behavior learning method. To represent the macroscopic traffic status, the area within the 4th Ring Road of Beijing is partitioned into 20 × 20 grids. The trajectory dataset was recorded by over 12,000 taxicabs in November 2012. Focusing on the routing patterns with respect to general traffic conditions, we only use the trajectories of occupied taxicabs in peak and mid-peak hours (i.e., 8:00–20:00) on weekdays in the experiments. The average traverse time of all vehicles passing through a grid in each direction is used to represent the reward of the corresponding routing action. The VIN can use the learned state value function to predict a route from origin to destination by selecting the neighboring grid with maximum state value in a stepwise manner. Thus, the VIN can imitate the routing behavior of taxicab drivers.

The performance of VIN-based routing behavior learning is shown in Fig. 5. From Figs. 5a–c, it can be seen that VIN is able to imitate vehicles' routing actions with possible deviation. Figure 5d shows the accuracy of predicted routes; top-k accuracy measures the accuracy of the correct prediction being in the top-k predictions. Top-1 accuracy is not high due to the intrinsic uncertainty of route planning, while top-2 accuracy indicates that the VIN has successfully learned most routing patterns considering alternative routes.

Evaluation for Cross-Domain Resource Orchestration

As shown in Fig. 6a, we consider a road network covered by hexagonal grids, which are determined by the coverage areas of macrocells. The

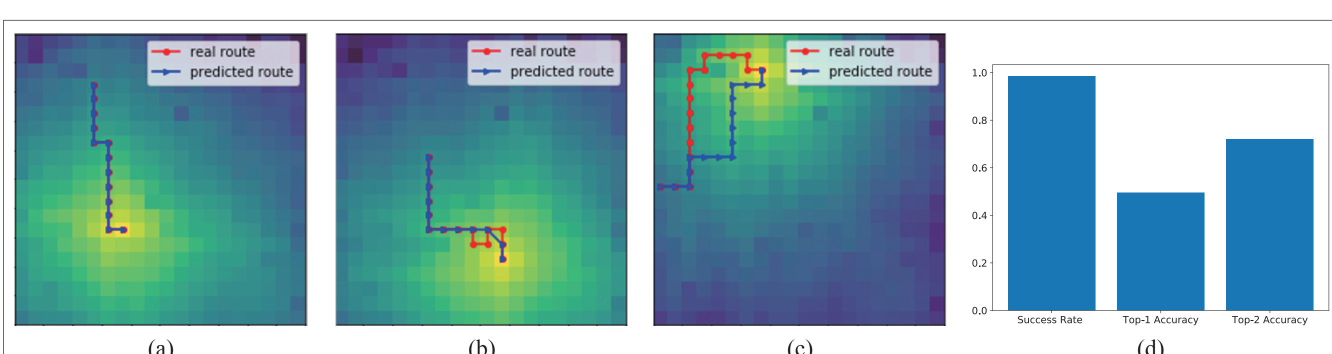

FIGURE 5. Performance of VIN-based routing behavior learning: a–c) examples of VIN's predicted routes; d) accuracy of VIN's predicted routes.

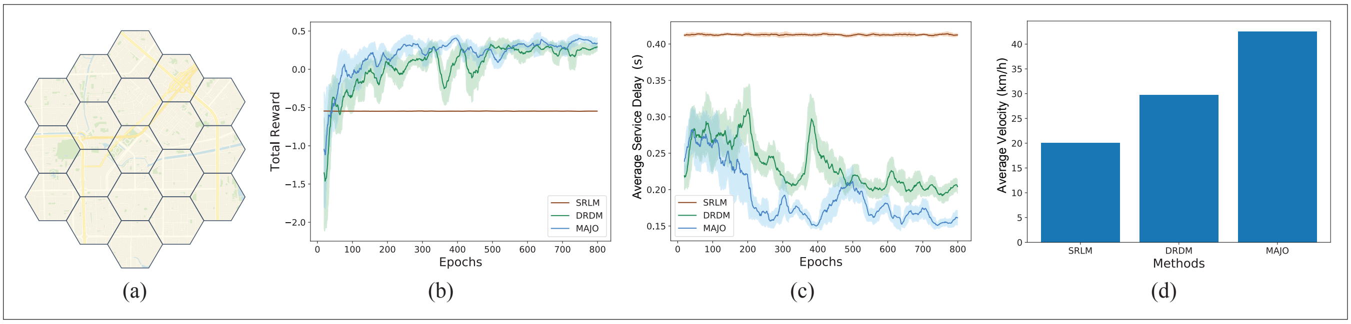

FIGURE 6. Performance of MADRL-based cross-domain resource orchestration: a) simulation scenario; b) joint reward; c) average service delay; d) average travel velocity. In b) and c), the solid lines are smoothed by window size of 20 epochs, while shaded areas show the standard deviations.

traffic simulation is performed based on the Macroscopic Bureau of Public Roads (MBPR) function, which models the relationship between travel time and traffic flow over a region-wide road network. In the following simulations, the computing capacity of each edge server is 100 GHz, and the communication system has 50 resource blocks, each with bandwidth of 180 kHz. There are 30 ICVs taking computation offloading. The size of each offloaded task is 500 MB, while the required CPU cycles follow a uniform distribution $\mathcal{U}(2,4) \times 10^8$. The proposed MADRL-based joint optimization method (MAJO) is compared to the following methods:

- Shortest-distance routing and lazy migration (SRLM): Dijkstra's algorithm is used to find the shortest-distance route. In the meantime, once the service delay exceeds the upper bound, the agent will be migrated to the edge server nearest to its serving vehicle.
- DQN-based routing and DQN-based migration (DRDM): One DQN is used to plan routes based on resource status in the transportation domain, and another DQN is used to migrate agents based on the resource status in the information domain. These two DQNs are trained separately.

The performance of MADRL-based cross-domain resource orchestration is shown in Figs. 6b–d. The joint reward defined in Eq. 3 reflects how well the cross-domain resources are orchestrated, so the joint reward accumulated by all vehicles is used as a metric in Fig. 6b. The results show that MAJO converges quickly and achieves the highest reward. It is worth noting that MAJO outperforms DRDM, although their network architecture is nearly the same. This demonstrates that cross-domain resource orchestration can largely improve the quality of edge-assisted intelligent driving. In Figs. 6c and 6d, the average service delay and average travel velocity for different methods are shown. It can be seen that MAJO has the ability to balance the performance across the information and transportation domains.

Conclusion and Future Directions

In this article, we have studied the key issues of cross-domain resource orchestration for smart road. To meet the challenges of jointly optimizing information flow and traffic flow, we have proposed a multi-agent-based cross-domain resource orchestration framework. In this framework, a value iteration network has been used to learn the routing behavior of each vehicle, which makes the vehicle comply with the agent's routing decision. Furthermore, a resource orchestration method based on multi-agent deep reinforcement learning has been proposed to improve the quality of intelligent driving. To fully leverage the smart road to assist intelligent driving, further research issues are discussed as follows.

Cooperative sensing and communication: In existing intelligent transportation systems, sensing and communication work separately; only their application data may interact. This poses the challenge of matching the sensed physical entity (e.g., vehicle, pedestrian) with its network identity (i.e., IP address). Therefore, the smart road can hardly monitor and control the vehicle behavior in a specific scenario (e.g., non-signalized intersection). In 5G/6G systems, mmWave and terahertz have unique features such as directional communications, high-quality imaging, and high-accuracy localization. Therefore, sensing, communication, and localization can be deeply integrated at the signal level. In this way, the smart road can simultaneously monitor vehicle/pedestrian behavior and send guiding instructions to the exact vehicle/pedestrian (without knowing its IP address), which improves system reliability and reduces end-to-end delay.

Air-ground integrated networking and transportation: With the vision of seamless and multi-modal mobility, urban transportation will move into the third dimension, exploring vast skies to relieve the congested roads. To provide required network coverage for the multi-modal transportation, air-ground integrated vehicular networks are needed for future smart roads. An aerial network formed by quasi-stationary high-altitude platforms is employed to assist the terrestrial network provide broadband connectivity. Therefore, the resources in both the information and transportation domains are expanded to a three-dimensional space, which increases the complexity of cross-domain resource orchestration.

Energy flow scheduling: Electric vehicles (EVs) are going to replace gas vehicles to contribute to greener transportation. Besides the resources in the information and transportation domains, energy resources will become a new domain to be scheduled for the smart road. The distribution of EV charging stations and wireless charging roads will have quite an impact on the route planning of EVs. In addition, the powerful sensors, CPUs, and GPUs on EVs consume a lot of energy, so the scheduling of energy flow affects the computation offloading policies. Orchestrating cross-domain resources for the smart road to coordinate information flow, traffic flow, and energy flow remains a challenge.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China under Grant 61902035, Grant 61876023, and Grant 61871211, in part by the Natural Science Foundation of Beijing under Grant 4181002, and in part by China Postdoctoral Science Foundation under Grant 2019M660565.

References

[1] Q. Yuan *et al.*, "Toward Efficient Content Delivery for Automated Driving Services: An Edge Computing Solution," *IEEE Network*, vol. 32, no. 1, Jan./Feb. 2018, pp. 80–86.

[2] H. Zhou *et al.*, "Evolutionary V2X Technologies Toward the Internet of Vehicles: Challenges and Opportunities," *Proc. IEEE*, vol. 108, no. 2, Feb. 2020, pp. 308–23.

[3] K. Lin *et al.*, "Vehicle Route Selection Based on Game Evolution in Social Internet of Vehicles," *IEEE Internet of Things J.*, vol. 5, no. 4, Aug. 2018, pp. 2423–30.

[4] Z. Cao *et al.*, "Maximizing the Probability of Arriving on Time: A Practical Q-Learning Method," *Proc. AAAI*, San Francisco, CA, 2017, pp. 4481–87.

[5] N. Groot, G. Zaccour, and B. De Schutter, "Hierarchical Game Theory for System-Optimal Control: Applications of Reverse Stackelberg Games in Regulating Marketing Channels and Traffic Routing," *IEEE Control Sys. Mag.*, vol. 37, no. 2, Apr. 2017, pp. 129–52.

[6] T. Chu *et al.*, "Multi-Agent Deep Reinforcement Learning for Large-Scale Traffic Signal Control," *IEEE Trans. Intell. Transp. Sys.*, vol. 21, no. 3, Mar. 2020, pp. 1086–95.

[7] B. Qian *et al.*, "Toward Collision-Free and Efficient Coordination for Automated Vehicles at Unsignalized Intersection," *IEEE Internet of Things J.*, vol. 6, no. 6, Dec. 2019, pp. 10,408–20.

[8] W. Zhuang *et al.*, "SDN/NFV-Empowered Future IoV with Enhanced Communication, Computing, and Caching," *Proc. IEEE*, vol. 108, no. 2, Feb. 2020, pp. 274–91.

[9] M. Li, P. Si, and Y. Zhang, "Delay-Tolerant Data Traffic to Software-Defined Vehicular Networks with Mobile Edge Computing in Smart City," *IEEE Trans. Vehic. Tech.*, vol. 67, no. 10, Oct. 2018, pp. 9073–86.

[10] Y. He, N. Zhao, and H. Yin, "Integrated Networking, Caching, and Computing for Connected Vehicles: A Deep Reinforcement Learning Approach," *IEEE Trans. Vehic. Tech.*, vol. 67, no. 1, Jan. 2018, pp. 44–55.

[11] S. Han *et al.*, "Dense-Device-Enabled Cooperative Networks for Efficient and Secure Transmission," *IEEE Network*, vol. 32, no. 2, Mar. 2018, pp. 100–06.

[12] Y. S. Nasir and D. Guo, "Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks," *IEEE JSAC*, vol. 37, no. 10, Oct. 2019, pp. 2239–50.

[13] D. Tian *et al.*, "Channel Access Optimization with Adaptive Congestion Pricing for Cognitive Vehicular Networks: An Evolutionary Game Approach," *IEEE Trans. Mobile Comp.*, vol. 19, no. 4, Apr. 2020, pp. 803–20.

[14] A. Aissioui *et al.*, "On Enabling 5G Automotive Systems Using Follow Me Edge-Cloud Concept," *IEEE Trans. Vehic. Tech.*, vol. 67, no. 6, June 2018, pp. 5302–16.

[15] J. Li *et al.*, "A Traffic Prediction Enabled Double Rewarded Value Iteration Network for Route Planning," *IEEE Trans. Vehic. Tech.*, vol. 68, no. 5, May 2019, pp. 4170–81.

Biographies

Quan Yuan [S'17, M'20] (yuanquan@bupt.edu.cn) received his Ph.D. degree in computer science and technology from Beijing University of Posts and Telecommunications (BUPT), China, in 2018. He is currently a lecturer at the State Key Laboratory of Networking and Switching Technology, BUPT. His current research interests include mobile computing, crowdsensing, and vehicular networks.

Jinglin Li [M'16] (jlli@bupt.edu.cn) received his Ph.D. degree in computer science and technology from BUPT. He is currently a professor at the State Key Laboratory of Networking and Switching Technology, BUPT. His research interests are mainly in the areas of network intelligence, mobile Internet, the Internet of Things, and the Internet of Vehicles.

Haibo Zhou [M'14, SM'18] (haibozhou@nju.edu.cn) received his Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University, China, in 2014. From 2014 to 2017, he was a postdoctoral fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada. He is currently an associate professor with the School of Electronic Science and Engineering, Nanjing University, China. His research interests include resource management and protocol design in cognitive radio networks and vehicular networks.

Guiyang Luo (luoguiyang@bupt.edu.cn) received his Ph.D. degree in computer science and technology from BUPT in 2020.. He is currently a postdoctoral fellow at the School of Computer Science, BUPT. His current research interests include software-defined networks, medium access control, and intelligent transportation systems.

Tao Lin (lintaomffl@gmail.com) is currently an M.S. candidate at the State Key Laboratory of Networking and Switching Technology, BUPT. His current research interests include deep reinforcement learning and the Internet of Vehicles.

Fangchun Yang (fcyang@bupt.edu.cn) received his Ph.D. degree in communications and electronic systems from BUPT. He is currently a professor at the State Key Laboratory of Networking and Switching Technology, BUPT. He is a Fellow of the IET. His current research interests include network intelligence, service computing, and the Internet of Vehicles.

Xuemin (Sherman) Shen [M'97, SM'02, F'09] (sshen@uwaterloo.ca) received his Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, New Jersey, in 1990. He is currently a university professor with the Department of Electrical and Computer Engineering, University of Waterloo. His research focuses on network resource management, wireless network security, social networks, 5G and beyond, and vehicular ad hoc and sensor networks. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, a Chinese Academy of Engineering Foreign Fellow, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.