Resource Management in Space-Air-Ground Integrated Vehicular Networks: SDN Control and AI Algorithm Design

Huaqing Wu, *Student Member, IEEE,* Jiayin Chen, Conghao Zhou, *Student Member, IEEE,* Weisen Shi, *Student Member, IEEE,* Nan Cheng, *Member, IEEE,* Wenchao Xu, *Member, IEEE,* Weihua Zhuang, *Fellow, IEEE,* and Xuemin (Sherman) Shen, *Fellow, IEEE*

***Abstract*—With its potential versatility and reliability, space-air-ground integrated vehicular network (SAGVN) is envisioned as a promising solution to deliver quality vehicular services anywhere at anytime. This article proposes a software defined framework for SAGVN to achieve flexible, reliable, and scalable network resource management. First, key applications and research challenges in resource management are identified. Then, we propose a hybrid and hierarchical SAGVN control architecture to balance the trade-off between system status acquisition and signaling overhead in different scenarios. Considering the dynamic networking environment with multi-dimensional resources and diverse services, it is challenging to make optimal resource management decisions in real time; thus, artificial intelligence (AI)-based engineering solutions are investigated to facilitate efficient network slicing, mobility management, cooperative content caching and delivery. A trace-driven case study is presented to demonstrate the effectiveness of the proposed SAGVN framework with AI-based methods in increasing the SAGVN throughput performance.**

I. Introduction

Connected and automated vehicle (CAV) techniques, which allow vehicles to communicate and autonomously make intelligent decisions without human intervention, are expected to revolutionize the transportation systems in the near future. By utilizing cutting-edge technologies (e.g., in advanced sensors, on-board processing, wireless communications and networking), CAVs will significantly improve driving safety, enhance transportation efficiency, reduce energy consumption, and facilitate environmental sustainability. To empower smart services for CAVs, worldwide network connectivity and reliable accessibility of various applications are expected. However, current terrestrial communication networks cannot guarantee ubiquitous and reliable service provisioning due to the unavoidable limitations including scarce radio spectrum resources, high operational expenditure, and fixed infrastructure deployment within constrained geographical areas.

Huaqing Wu, Jianyin Chen, Conghao Zhou, Weisen Shi, Weihua Zhuang, and Xuemin (Sherman) Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada. e-mail: {h272wu, j648chen, c89zhou, w46shi, wzhuang, sshen}@uwaterloo.ca.

Nan Cheng is with State Key Lab. of ISN, and with the School of Telecommunications Engineering, Xidian University, Xi'an 710071, China. e-mail: dr.nan.cheng@ieee.org.

Wenchao Xu is with the School of Computing and Information Sciences in Caritas Institute of Higher Education, Hong Kong. e-mail: wenchaoxu.ru@gmail.com.

Recently, communication networks via low earth orbit (LEO) satellites have attracted substantial attention from both academia and industry [1]. There are various initiatives to construct satellite constellations (e.g., Starlink, OneWeb, Hongyan) and launch thousands of LEO satellites by 2022 [2]. Benefiting from the global availability, the LEO networks have a potential to supplement worldwide seamless high-bandwidth Internet connectivity in a cost-effective way. In addition, the agility of aerial networks based on unmanned aerial vehicles (UAVs) capacitates the flexible deployment of UAV networks in response to dynamic service demands and unexpected or emergency situations, such as disaster relief or service congestion. On the other hand, the spatial and aerial networks have their own shortcomings, e.g., relatively long propagation delay for satellite connection and limited service endurance for UAV communication. Therefore, it is imperative to develop a space-air-ground integrated vehicular network (SAGVN) to exploit the complementary advantages of the three network segments [3].

With different network segments integrated, it becomes a significant challenge to efficiently manage the multi-dimensional resources (for communication, computing, and caching) from the space, air, and terrestrial networks to guarantee the quality-of-service (QoS) performance to vehicular users in the SAGVN. As a promising network architecture, software-defined networking (SDN) provides a unified platform for logically centralized control with a global view of the network via the separation of control and data planes [4]–[6]. Therefore, in this article, we focus on the efficient resource management in the SAGVN with an SDN-based control architecture to facilitate programmable, scalable, and adaptable network optimization. Considering that the SAGVN is extremely dynamic and complex because of the high mobility, large network scale, and heterogeneous devices/services, modeling such a system is painstaking, if not impossible. Therefore, the conventional model-based optimal resource management mechanisms are inapplicable for the SDN controllers. In addition, the urging growth in data traffic, unprecedented heterogeneity of devices and resources, and increasingly diversified and stringent QoS requirements in the SAGVN render traditional optimization approaches inadequate due to the high computational complexity and excessive decision-making delay. Therefore, innovative artificial intelligence (AI)-based engineering solutions are necessary to proactively make high-quality and real-time decisions to promptly respond to the dynamic SAGVN environment [7], [8].

In this article, we present an SDN-based SAGVN framework to facilitate flexible network management and expeditious harmonization of multi-dimensional resources. In particular, a hybrid and hierarchical SAGVN control architecture is proposed to balance the network controllability and signaling overhead in different communication scenarios. The characteristics and functionalities of different SDN controllers are further specified to demonstrate their cooperation in network slicing and associated resource allocation. AI-based techniques are then investigated to address intractable research problems in the SAGVN to enhance resource management efficiency. Specifically, the application of AI methods in network slicing, mobility management, and cooperative content caching and delivery are discussed in detail to provide insights on future intelligent SAGVN resource management.

II. SAGVN Applications and Research Challenges

A. SAGVN Scenarios and Applications

In the SAGVN, the space, air, and terrestrial networks have their pros and cons in providing services, in terms of coverage, transmission delay, throughput, and reliability. Through effective internetworking, the three network segments can complement each other to provide ubiquitous network access and support multifarious vehicular services. In the following, three typical SAGVN application cases are discussed to demonstrate the collaboration among different network segments in different scenarios.

Remote service provisioning and disaster relief (space-air-dominant): In areas without coverage of terrestrial networks (e.g., sparsely populated areas or remote mountainous areas), the SAGVN offers new opportunities for vehicular users to obtain high-speed Internet access through space-air communications. Further, satellites and UAVs, which can operate independently from terrestrial infrastructure, are robust to natural disasters and can effectively support post-disaster recovery. For instance, the satellites can capture wide-area disaster information via satellite images and UAVs can get a close visualization of real effects of disaster. Through effective SAGVN communication, the disaster-recovery center in terrestrial networks can implement efficient disaster loss estimation and reliable relief operations based on the information provided by satellites and UAVs. The SAGVN also has a potential to predict disasters and provides early warning, by utilizing surveillance UAVs/satellites to detect abnormalities.

Smart city development (space-air-assisted): In addition to supplementing the terrestrial networks in remote/disaster areas, the space-air networks in the SAGVN can assist the existing terrestrial network in supporting smart city applications. With main functionalities implemented and performed in terrestrial networks, the smart city applications can be enriched with the assistance from the space-air networks. Specifically, the SAGVN can provide ultra-reliable communications and facilitate reliable acquisition of city monitoring data. Benefiting from the mobility and/or high altitude, the satellites and UAVs are able to provide larger-scale monitoring data than stationary IoT devices. In smart transportation, for example, the satellites/UAVs can be utilized to collect data about road status, climate conditions, traffic congestions, and accidents. Such real-time data is the key enabler for efficient urban traffic management in smart city.

Data-craving service provisioning (space-air-terrestrial-integration): There are many data-intensive vehicular applications, such as self-driving software update, vehicular video streaming, online gaming, and mobile advertising. As the terrestrial networks are limited in coverage and capacity, the space-air networks can be integrated with the terrestrial networks to provide alternative radio access technologies (RATs) for data-craving services. With space-air-terrestrial integration, the three network segments are of equal importance, and each of them can play a leading role in service provisioning for different applications based on the network characteristics and vehicles' QoS requirements. For instance, satellite communications can be leveraged to support CAV software update simultaneously for millions of vehicles, benefiting from the inherent multicast/broadcast capability. UAVs can provide flexible and reliable connectivity for vehicles in traffic jam to alleviate the terrestrial network burden and to boost service capacity. In addition, the complementary properties of satellite links (wide coverage) and fiber-optical backbone (high data rate) can be considered as alternative backbone technologies to wireless backhaul, to mitigate the long-distance multi-hop backhaul.

B. Research Challenges in SAGVN Resource Management

To effectively manage network resources in the SAGVN, there exist many research challenges and opportunities that require further investigation, including:

Mobility management - In the SAGVN, there are multiple types of mobility introduced by LEO satellites, UAVs, and vehicular users. The diverse mobility complicates the management of the integrated network in terms of both access network association and backhaul selection. In access networks, the high mobility of the devices/APs can lead to 1) inaccurate channel estimation due to the fast time-varying characteristics; 2) frequent horizontal handoff within the homogeneous segment and vertical handoff among heterogeneous network segments; and 3) inefficient routing optimization due to difficulty in user localization. When considering an LEO satellite network as wireless backhaul, its capacity is constrained by the highly dynamic satellite constellation topology and the short contact duration among LEO satellites, which further obscures the decision process in selecting satellite or terrestrial backhaul. Therefore, innovative mobility management solutions are imperative yet challenging.

QoS-aware resource allocation - The SAGVN incorporates multi-dimensional resources with highly dynamic availability. Efficient resource management can be impeded by 1) significant signaling overhead to acquire the network state information; 2) high computational complexity due to the involvement of massive devices and limited resources; and 3) spatio-temporal dynamics in terms of QoS requirements, service request distribution, and resource availability. In addition, the relationship between QoS requirements and multi-

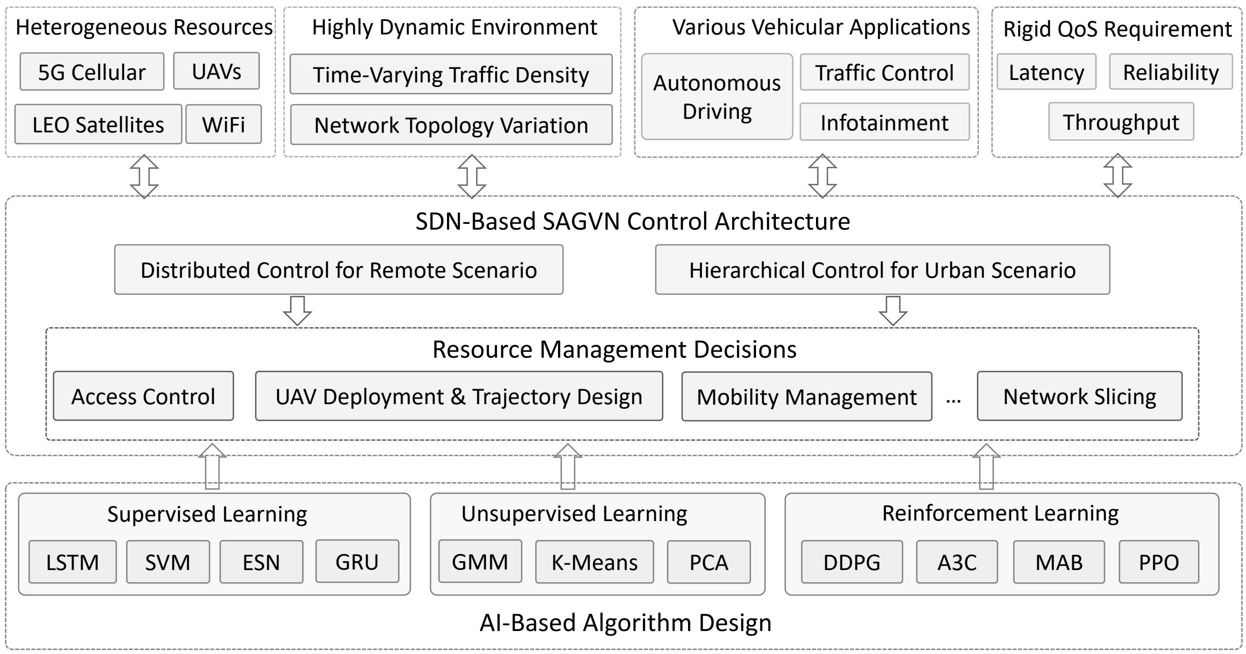


Fig. 1: SDN-based SAGVN framework

dimensional resources is yet to be established and modeled, which hinders the service-oriented resource allocation. Painstaking efforts are required to develop innovative resource orchestration solutions.

Energy efficiency - Different from the terrestrial infrastructure which has stable power supply, the satellites and UAVs are powered by batteries and/or solar energy. The limited energy capacities of satellites/UAVs constrain the service duration and further affect service functionalities including sensing, transmission, and processing. Therefore, vehicular users in the SAGVN can potentially suffer intermittent connection and service interruption due to satellite/UAV energy depletion. In addition, service-unrelated energy consumption for satellites (due to intense radiation and space-variant temperature) and UAVs (due to propulsion and direction adjustment) further deteriorates the service endurance. Therefore, improving energy efficiency to prolong the service duration of satellites/UAVs is critical to provide persistent vehicular services.

III. SDN-Based SAGVN Framework

This article focuses on the management of multi-dimensional SAGVN resources to support multifarious vehicular applications with QoS guarantee. To facilitate flexible and efficient resource allocation in the complex integrated networks, we propose an SDN-based SAGVN framework, as shown in Fig. 1, to address the aforementioned challenges. In particular, we propose an SDN-based hybrid and hierarchical control architecture to balance the trade-off between the network status acquisition and signaling overhead in different communication scenarios. By collecting information of the physical infrastructure (e.g., resource availability and communication link conditions) and user services (e.g., differentiated QoS requirements), the SDN controllers make centralized resource management decisions. Considering that the SAGVN is extremely dynamic and complex, AI technologies are envisioned as agile tools to autonomously optimize the resource utilization in real-time with enhanced network performance [9]. In this section, the SDN-based hybrid and hierarchical control architecture is introduced in detail, while AI-based algorithm design for the resource management is investigated in Section IV.

A. Hybrid and Hierarchical Control Architecture

The SAGVN control architecture, as shown in Fig. 2, includes space, air, and terrestrial network segments. Notice that the terrestrial and aerial communication systems (e.g., LTE, WiFi, and DSRC) generally use lower frequency bands, while the LEO communication networks operate in higher frequency bands (e.g., Ku-band and Ka-band) and require different radio frequency (RF) transceivers. Vehicles equipped with both kinds of transceivers can directly communicate with the satellites, while vehicles without a satellite transceiver can only access the satellite network via the terrestrial-satellite access points (TSAPs). The TSAPs, which support both the TSAP-satellite links over high-frequency bands and the TSAP-vehicle links over low-frequency bands, can relay the information between the satellites and vehicles.

Hybrid control architecture for different scenarios: The SDN controllers need to collect massive network information (from all the space, air, and ground segments) with high signaling overhead to perform network management. To reduce unnecessary signaling overhead, distributed control is adopted in remote scenarios. Due to low user density and limited available network resources (e.g., only space network is available in some areas), resource management in remote areas is relatively simple and distributed algorithms can generally work well without deploying SDN controllers.

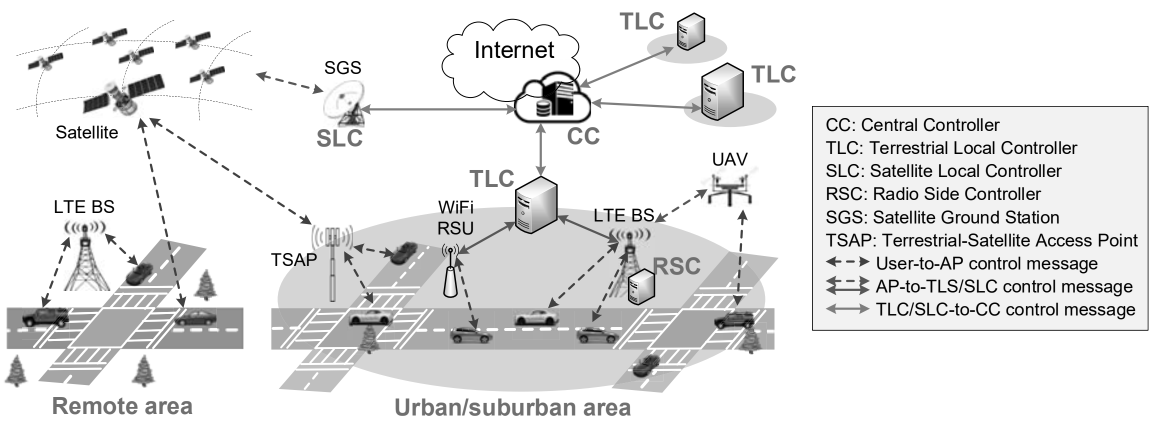


Fig. 2: An illustration of the SAGVN control architecture

In urban/suburban scenarios, on the other hand, the resource management is sophisticated and SDN controllers are required to enhance resource utilization efficiency and optimize the network performance.

Hierarchical control architecture for cross-domain resources: In urban/suburban areas, it is inefficient to rely solely on one SDN controller to perform network-wide resource management, due to the extensive signaling overhead and non-negligible control delay. For scalability, SDN controllers are organized as a hierarchy to manage the cross-domain resources, where SDN controllers at different levels target at network operations in different domains. More specifically, the detailed roles and characteristics of different tiers of SDN controllers are summarized as follows:

- Radio side controllers (RSCs): Deployed at the access points (APs) such as terrestrial cellular base stations (BSs), WiFi APs, and UAVs, RSCs are at the lowest level of the hierarchy and have the smallest coverage area. In close proximity, RSCs are able to manage and schedule the network resources for vehicular users in a timely manner.
- Terrestrial local controllers (TLCs): The TLCs cover a larger area and coordinate multiple RSCs to perform higher-level network operations. The hierarchical control architecture can have multiple layers of TLCs based on different coverage domains (e.g., local, regional, and national). The coverage of each TLC can be determined by considering factors including control delay requirements, implementation cost, control complexity, etc.
- Satellite local controllers (SLCs): Considering that each satellite can cover multiple TLCs, it is inefficient to leverage TLCs for satellite resource allocation. Therefore, the TLCs are only in charge of the terrestrial and aerial network resources within their scopes. The SLCs, which can be mounted on the satellite ground stations, are exploited to monitor satellite information (e.g., orbits, coverage, and resource availability) and allocate their resources accordingly.
- Central controller (CC): Based on the abstracted network status information from the LCs (including TLCs and SLCs), the CC rules the entire network to orchestrate the operations in all the network segments.

B. Network Slicing

With the hybrid and hierarchical SDN-based control architecture, the network resources should be orchestrated to support multifarious vehicle services. For flexible network management to ensure user satisfaction of different services, network slicing enables the coexistence of multiple independent virtual networks (i.e., network slices) atop a shared physical network infrastructure. The created network slices, each supporting a specific service, are isolated logically and do not interfere with each other [10]. In our SDN-based SAGVN control architecture, the SDN controllers can be utilized to construct and manage the virtual slices to support services with differentiated QoS requirements. Furthermore, network function virtualization (NFV), which decouples network functions from specialized hardware, can be leveraged to flexibly implement network functions as software instances in the network slices.

To perform network slicing, the SDN controllers (slice managers) have the following functionalities, as shown in Fig. 3: 1) Information collection: An upper-level SDN controller collects network information from its lower-level SDN controllers or the vehicular users. Different levels of abstracted network information are required in SDN controllers at different tiers, as shown in the example given in the left part of Fig. 3; 2) Slice creation: Stemmed from the network information from lower-level SDN controllers, an upper-layer SDN controller determines whether to create a network slice for a service; 3) Network function placement: For each created network slice, one or more virtual network functions (VNFs) can be implemented for service provisioning; 4) Resource allocation: Network resources are allocated to the created network slices to guarantee service level agreement (SLA) for different services. At the service planning stage, a network slice is allocated multiple-dimensional resources, each of which can be mapped to the physical SAGVN network infrastructure (e.g., LTE BSs, WiFi APs, UAVs, satellites); 5) Decision distribution: The network slicing decisions are then distributed from the upper-level SDN controller to its lower-level SDN controllers; 6) Resource scheduling: In the service operation stage, the

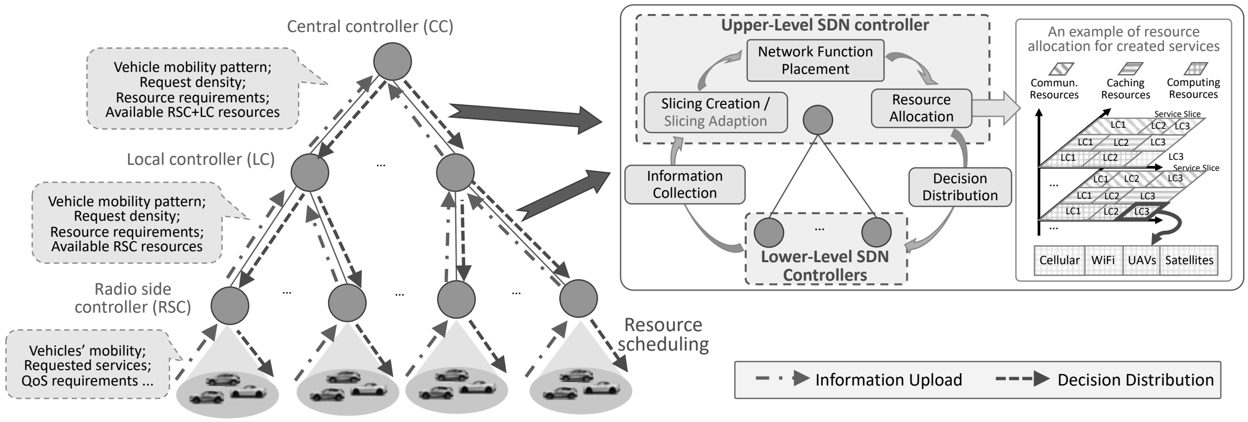


Fig. 3: Network slicing procedure in SAGVN.

RSCs perform physical resource scheduling for individual vehicular users based on the established network slice by its upper-level controller; 7) Slicing adaptation: The decisions on network slice creation, network function placement, and resource allocation should be adjusted or reconfigured in a relatively large time scale whenever necessary, to adapt to temporal-spatial data traffic load dynamics with high resource utilization.

In the network slicing and associated resource allocation, the upper-layer controllers generally have a coarser slicing granularity in the spatial domain and longer planning intervals in the temporal domain. Therefore, slices for different services should be managed at different levels of SDN controllers to achieve differentiated QoS requirements. For instance, eMBB services which require high data rates across a wide geographical area can be handled at the CC with centralized signal processing and resource allocation capabilities. In contrast, URLLC services in a local area with stringent requirements on latency and reliability can be managed by taking advantage of the edge caching and edge computing capabilities of the RSCs.

IV. AI Algorithm for Resource Management

Given the highly dynamic network environment and complicated operation and management in the SAGVN, conventional model-based optimization methods become inefficient for the SDN controllers. Recent advances in AI technologies provide an alternative approach to learn the complex networks and make intelligent decisions with low time complexity.

Basically, different AI methods are suitable for different SAGVN resource management problems because of the diversified approaches, input data requirements, and outputs. For instance, the supervised learning algorithms can be leveraged to learn network characteristics and predict vehicle mobility patterns by learning from historical labeled data. Unsupervised learning can be utilized for vehicle clustering by learning the hidden pattern from big data. Reinforcement learning (RL) can be adopted to make online decisions to adapt to environmental dynamics and address the stochastic network characteristics (e.g., uncertain channel conditions and data traffic) to maximize long-term service performance. In addition, the above-mentioned AI methods can be combined with the transfer learning techniques, which repurpose a model trained for one task to solve related ones, to accelerate the model training or improve performance.

In the following, we discuss how to leverage AI approaches for efficient resource management algorithms, as summarized in Table I.

A. Network Slicing

The intangible relationship between QoS requirements and the requisite multi-dimensional resources inhibits the effective QoS-oriented network slicing in the SAGVN. Based on the network slicing procedure in Subsection III-B, AI-based solutions can be exploited for efficient network slicing in the following aspects.

1) Network slice creation: The decision on slice creation is made based on various information of service request and network status, such as the QoS requirements, service request patterns, and resource availability. The high dimension of the input information makes the problem intricate and can hamper the convergence performance when adopting AI-based methods. Hence, based on the principal component analysis (PCA), the dimension reduction process can be applied by selecting the decisive components from the collected information. Further, clustering algorithms such as Gaussian mixture model (GMM) can be adopted to make the decision on whether to create a slice for service requests from a group of users;

2) Function placement and resource allocation: Generally, a vehicular service consists of several subtasks. For instance, a vehicle surveillance application can be completed by running four subtasks in sequence, i.e., video decoding, subtracting background, object detection, and object classification. These subtasks can be placed at different locations as virtual functions, and computing resources should be allocated to execute the functions. The sequential function placement and resource allocation for subtasks can be coordinated by the CC with global network information. Particularly, an RL algorithm can be implemented to solve the problem, in which the states include the QoS requirement for each function chain, service load prediction, locations of the source and destination nodes,

TABLE I: Various AI methods in solving SAGVN resource management problems

	Potential Algorithms	Network Slicing	Mobility Management	Cooperative caching & content delivery
Supervised Learning	GRU, ESN SVM, LSTM	• Service request prediction	• Vehicle mobility prediction • Vehicular traffic pattern prediction	• Content popularity prediction • Network state classification
Unsupervised Learning	K-means, GMM, PCA	• Input dimension reduction • Network slice creation	• Vehicle clustering	• Input dimension reduction
Reinforcement Learning	DDPG, A3C, MAB, PPO	• Function chain placement • Resource scheduling	• LC-RSC association	• Content placement optimization
Transfer Learning	Accelerating the learning process and improving prediction efficiency with enhanced performance			

and available resources; actions include the location of each function and the allocated resources; reward is the overall QoS satisfaction level and resource utilization. Considering the potentially enormous state space, we can adopt the deep deterministic policy gradient (DDPG) approach [11] to learn policies with high-dimensional and continuous action space;

3) Resource scheduling: Network resource scheduling can be implemented in the RSCs to dynamically allocate resources to vehicular users during service provisioning. Since different RSCs can have overlapping coverage areas, their resource scheduling decisions interact with each other. To minimize the signaling overhead of frequent interactions, the multi-agent RL or asynchronous actor-critic agents (A3C) can be leveraged to enable efficient resource scheduling. In particular, the RSCs can act as actors and independently make decisions based on local information, and LC functions as a global critic to provide performance feedback to RSCs within its coverage range.

B. Mobility Management

The diverse mobility, especially the vehicle mobility, significantly affects the service connectivity, routing path update, and resource allocation. To maintain stable network connection and satisfy the diverse QoS requirements, AI methods can be viable in the following aspects.

1) Individual vehicle mobility prediction: Predicting the individual vehicle mobility is of vital importance for high resource scheduling performance. Particularly, the time-series-based AI algorithms, such as gated recurrent unit (GRU) algorithm [12], long short-term memory (LSTM), and echo state network (ESN) [13], can be implemented at the RSCs for individual vehicle mobility prediction. Considering that vehicles in the same region have similar mobility patterns, the transfer learning [14] can be applied to improve the prediction efficiency by transferring the prediction result of one vehicle to its neighboring vehicles served by the same RSC;

2) Proactive on-demand UAV deployment: Based on the vehicle mobility predictions, RSCs can analyze each vehicle's connection performance with the available APs. For vehicles with potential connection loss or poor connection performance, clustering algorithms such K-means or hierarchical clustering can be used to cluster the vehicles into groups and on-demand UAVs can be dispatched to serve vehicles in need;

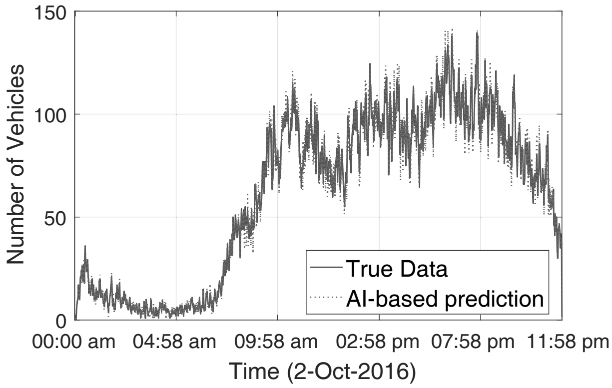
3) Traffic pattern prediction: Besides predicting the individual vehicle mobility, the prediction of vehicular traffic patterns (e.g., average speed and vehicle density) can be applied in the LCs or even CC to facilitate network planning in the complex and large-scale SAGVN system. Likewise, service request distributions in different regions can also be predicted. Based on the vehicular traffic flow and service request prediction, LCs can adaptively adjust the LC-RSC association to achieve load balancing. Particularly, multi-armed bandit (MAB) learning can be utilized to facilitate the LC-RSC association and proactive resource allocation, to maximize the long-term QoS performance for vehicular users.

C. Cooperative Caching and Content Delivery

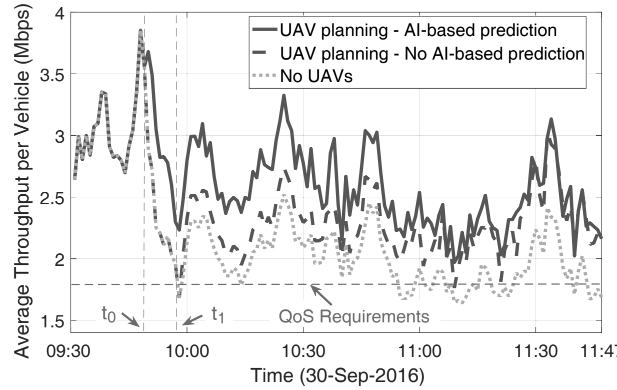
Content caching at the network edge is a promising paradigm to mitigate the backhaul transmission congestion and reduce response latency. One critical research problem in edge caching is content placement, i.e., the problem of what to cache and where to cache. Further, the parameters for content caching and content delivery (e.g., resolution and frame rate for video streaming services) should also be optimized to satisfy differentiated requirements in terms of service quality and resource demand. To guarantee diverse QoS requirements with minimal resource consumption, the content placement decision, content caching parameters, and content delivery parameters should be jointly optimized, which can base on AI algorithms.

We propose an AI-based two-layered optimization scheme to solve the joint optimization problem with different optimization granularities. Particularly, the content placement and caching parameters are determined by the content request distribution and available caching resources, and the decision update can be triggered by a QoS degradation level. On the other hand, the optimal delivery parameter selection is highly dependent on the time-varying communication environment and user QoS requirements. Thus the delivery parameters should be adjusted over time for different service requests, which has a finer granularity than the content placement and caching parameter optimization.

1) Content placement and caching parameter optimization: The caching parameters can be quantized to reduce decision dimension. Therefore, one content file with different caching parameters can be virtually viewed as a group of files, among which at most one can be cached. Then, the joint optimization of content placement and caching parameters can be reduced to the problem of content placement. To maximize long-term caching performance (e.g., cache hit rate) with uncertain

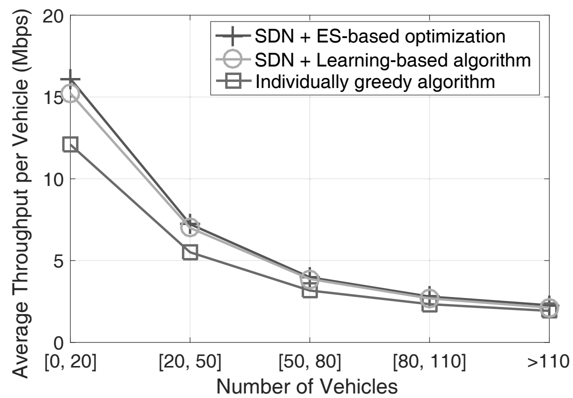


(a) Prediction of Number of Vehicles

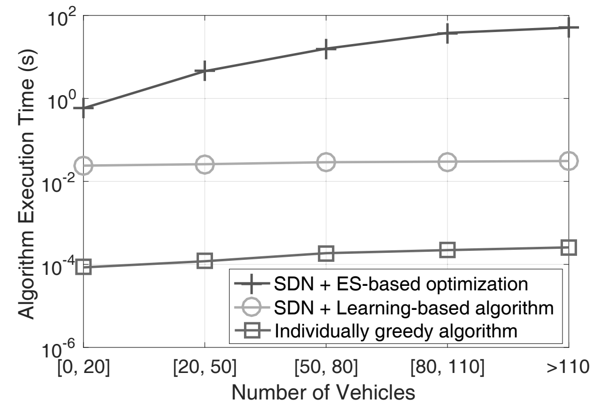


(b) UAV dispatching with/without traffic prediction

Fig. 4: Impact of AI-based vehicular traffic prediction on the effectiveness of UAV planning.



(a) Average throughput per vehicle



(b) Algorithm execution time

Fig. 5: Performance comparison for three RAT selection algorithms: SDN-based centralized optimization with ES algorithm, SDN-based centralized optimization with learning-based algorithm, and individually greedy algorithm without SDN controllers.

user request patterns, RL algorithms such as proximal policy optimization (PPO) [15] can be leveraged to optimize the content placement;

2) Content delivery parameter optimization: The content delivery parameters are determined by the content placement policy, the available communication resources, and the QoS requirements. With delivery parameter quantization, the content delivery parameter optimization can be solved by leveraging the classification algorithms such as support vector machine (SVM). The classification algorithms can classify the observed network state into different categories, each of which corresponds to a specific delivery parameter set. Furthermore, input dimension reduction methods can be applied to facilitate efficient learning.

V. A Case Study

In this section, we conduct trace-driven simulations to investigate the performance boost introduced by AI-based techniques and the effectiveness of the SDN-enabled central control architecture in the SAGVN. We adopt the Didi Chuxing GAIA Initiative dataset, which includes taxi GPS traces within the second ring road in the city of Xi'an, Shaanxi, China. The satellite orbiting movement information is obtained by using the STK simulator. In specific, we focus on service provisioning for vehicles within a 2000 m $\times$ 2000 m square area for the case study.

Fig. 4 shows the performance enhancement introduced by AI techniques. Focusing on the case of UAV planning in the SAGVN, AI techniques can be utilized to predict the vehicle traffic flow, and the model training can be conducted at the edge servers with powerful computing and processing capabilities. Then, the well-trained prediction model can be implemented on the UAVs, which can proactively perform UAV planning based on the predicted ground demand. As shown in Fig. 4a, the number of vehicles in the target scenario is predicted by exploiting the LSTM method. The LSTM method can achieve high prediction accuracy with less than 10% prediction error on average. The highly accurate prediction results are then utilized to make proactive decisions, e.g., UAV planning as shown in Fig. 4b, to enhance the network performance. Foreseeing a QoS degradation at time t_1, UAVs are dispatched at time t_0 to serve the demanding areas with the prediction-based UAV planning scheme, achieving significant throughput improvement when compared to the case without UAVs. However, with the UAV planning scheme without AI-based prediction, UAVs can only be reactively dispatched after detecting the QoS degradation at time t_1, which results in unsatisfactory throughput enhancement due to the response delay.

Fig. 5 shows the effectiveness of the AI-based methods and the superiority of the SDN-enabled central control architecture in the SAGVN. With deployed terrestrial BSs, UAVs, and satellites, vehicular users can be served by different RATs. We focus on RAT selection for the vehicles to optimize the overall network throughput. Particularly, with communication resources allocated by upper-level SDN controllers, RAT selection optimization is performed by RSCs deployed in LTE BSs. We compare the throughput and execution time of three RAT selection algorithms: 1) "SDN + ES-based optimization": SDN-enabled centralized decision-making with exhaustive search (ES) optimization algorithm; 2) "SDN + Learning-based algorithm": SDN-enabled centralized decision-making with DNN-based learning algorithm; and 3) "Individually greedy algorithm": Individually greedy decision-making without SDN controllers, where vehicles are connected to the AP with the best SNR. As shown in Fig. 5a, the average throughput per vehicle decreases with the number of vehicles for all the algorithms. However, the SDN-enabled central control architecture can effectively coordinate the RAT selection for vehicles and outperform the case without SDN controllers. In addition, the learning-based method can achieve a near-optimal throughput performance with a relatively low time complexity, as shown in Fig. 5b, which can facilitate intelligent resource management in a real-time manner.

VI. Conclusion and Future Research

In this article, we have proposed an SDN-based SAGVN framework, where a hybrid and hierarchical control architecture is illustrated to facilitate flexible and scalable resource management, and AI-based algorithms are applied to further enhance management efficiency. The proposed open network architecture can achieve network agility and flexibility, simplify network management and maintenance, and adapt to spatial-temporal network environment dynamics. Various aspects of network slicing, mobility management, and cooperative content caching and delivery are discussed. This article presents an initial step towards understanding how different types of AI algorithms can be applied for intelligent network management and resource orchestration in the future SAGVN.

For further research, the placement of SDN controllers should be investigated to guarantee reliable network control with low signaling overhead and short control latency. Considering the critical importance of the SDN controllers in the proposed SAGVN framework, the security of the SDN controllers should also be investigated. In addition, the proposed framework can be extended by involving edge intelligence techniques to empower federated learning and collaborative learning to reduce inference latency with better privacy protection.

References

[1] B. Di, L. Song, Y. Li, and H. V. Poor, "Ultra-dense LEO: Integration of satellite access networks into 5G and beyond," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 62–69, Apr. 2019.

[2] Federal Communications Commission, "SpaceX non-geostationary satellite system (attachment a)," 2016.

[3] N. Cheng, W. Quan, W. Shi, H. Wu, Q. Ye, H. Zhou, W. Zhuang, X. Shen, and B. Bai, "A comprehensive simulation platform for space-air-ground integrated network," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 178–185, Feb. 2020.

[4] N. Zhang, S. Zhang, P. Yang, O. Alhussein, W. Zhuang, and X. Shen, "Software defined space-air-ground integrated vehicular networks: Challenges and solutions," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 101–109, Jul. 2017.

[5] J. Wang, Y. Tan, J. Liu, and Y. Zhang, "Topology poisoning attack in SDN-enabled vehicular edge network," *IEEE Internet Things J.*, DOI: 10.1109/JIOT.2020.2984088, Mar. 2020.

[6] W. Zhuang, Q. Ye, F. Lyu, N. Cheng, and J. Ren, "SDN/NFV-empowered future IoV with enhanced communication, computing, and caching," *Proc. IEEE*, vol. 108, no. 2, pp. 274–291, Feb. 2020.

[7] N. Kato, Z. M. Fadlullah, F. Tang, B. Mao, S. Tani, A. Okamura, and J. Liu, "Optimizing space-air-ground integrated networks by artificial intelligence," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 140–147, Aug. 2019.

[8] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[9] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," *Proc. IEEE*, vol. 107, no. 8, pp. 1738–1762, Aug. 2019.

[10] X. Shen, J. Gao, W. Wu, K. Lyu, M. Li, W. Zhuang, X. Li, and J. Rao, "AI-assisted network-slicing based next-generation wireless networks," *IEEE Open J. Vehicular Technology*, vol. 1, no. 1, pp. 45–66, Jan. 2020.

[11] Y. Dai, D. Xu, S. Maharjan, G. Qiao, and Y. Zhang, "Artificial intelligence empowered edge computing and caching for internet of vehicles," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 12–18, Jun. 2019.

[12] A. B. Adege, H.-P. Lin, and L.-C. Wang, "Mobility predictions for IoT devices using gated recurrent unit network," *IEEE IoT J.*, vol. 7, no. 1, pp. 505–517, Jan. 2020.

[13] U. Challita, A. Ferdowsi, M. Chen, and W. Saad, "Machine learning for wireless connectivity and security of cellular-connected UAVs," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 28–35, Feb. 2019.

[14] W. Liu and Y. Shoji, "Edge-assisted vehicle mobility prediction to support V2X communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10 227–10 238, Oct. 2019.

[15] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, Aug. 2017.