A 5G Network Slice Based Edge Access Approach with Communication Quality Assurance

Fanbo Meng[1](✉), Huan Li[2], Bin Lu[1], Shuai Ren[2], and Dongdong Wang[1]

[1] State Grid Liaoning Electric Power Supply Co., Ltd., Shenyang 110004, China
mengfanbo221@sina.com

[2] Electric Power Research Institute of State Grid Liaoning Electric Power Supply Co., Ltd., Shenyang 110006, China

Abstract. With the development of various vertical industry services such as autonomous driving, energy Internet, and smart cities, mobile communication networks need to provide users with ubiquitous high-speed access while using limited network resources to provide differentiated and customized services, the 5G network satisfies the requirement of the future . However, for the access network, there are many types services accessing the network. In order to provide users with diverse personalized services, network slicing scheme is introduced into 5G network. Network slicing is based on the technology of network function virtualization, which can establish multiple virtual private networks in the device according to the needs of users. Each slice is a private network, and different virtual networks are kept isolated from each other. This article studies the access network of the 5G network, in order to ensure the quality of user access, we study the mapping scheme of network slices and NFV to ensure the communication quality of access networks of different user types. Finally, we perform some simulations to verify the proposed method, and the result shows that our proposed can ensure the communication quality for the users which connect into the 5G network.

Keywords: Network slicing · 5G · NFV

1 Introduction

With the widespread use of various new mobile terminals such as smart terminals and the rapid development of Internet technologies, the number of smart terminal devices connected to the network will explode, this poses a huge challenge to the existing network architecture and network capacity. Obviously, due to the inherent defects of the traditional network, such as the network architecture cannot be flexibly configured and adjusted, so it cannot cope with the future needs of users for mobile networks [1–3]. At the same time, new applications such as augmented reality, smart homes, and intelligent transportation enable people to experience new lifestyles. However, these applications place higher demands on system throughput, end-to-end delay and network reliability.

H. Song and D. Jiang (Eds.): SIMUtools 2020, LNICST 369, pp. 12–23, 2021.
https://doi.org/10.1007/978-3-030-72792-5_2

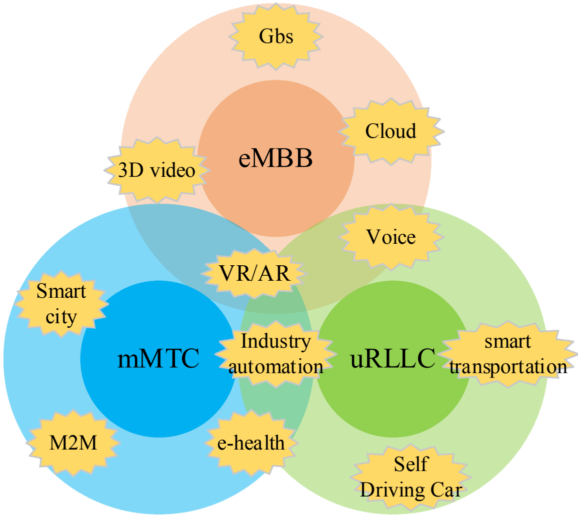

Fig. 1. The application scenarios of 5G.

At the same time, the emergence of new technologies such as autonomous driving and virtual reality has made the types of communication services more complicated [4–6]. Increasing data traffic, network device connections, and increasingly complex service scenarios make the 4G network face huge challenges [7–9]. Due to the limitations of network architecture, spectrum resources, and energy consumption, existing 4G networks cannot meet users' needs in terms of system capacity, data transmission rate, transmission delay, and resource utilization. In order to meet the data growth demand for users to access the network, the wireless communication group of International Telecommunication Union (ITU) propose the fifth generation mobile communication technology (5G) [10–12]. The 5G network meet the goal of connection between humans and things in the industry, agriculture, medical, education, transportation, etc. Compared with 4G network, 5G network has a big improvement in performance and function. It can provide greater throughput, ultra-high link density, lower latency and more secure and reliable communication [13–15]. At the same time, network energy consumption and network deployment costs are more low. The ITU has determined that the three major application scenarios of 5G, namely: enhanced mobile broadband (eMBB), mass machine communication (mMTC), low latency and highly reliable communication (uRLLC). The application scenarios as Fig. 1 shows.

(1) eMBB: Compared with 4G networks, the transmission rate and network coverage of users are greatly improved, and the user's communication is further improved in the existing broadband scenario Experience
(2) mMTC: The mMTC is the basic of large-scale IoT, it mainly oriented to the communication between objects, suitable for the scenarios of Internet of Things communication. The main advantage lies in supporting higher peak rate of uplink and downlink on the basis of ensuring coverage and power consumption; supporting

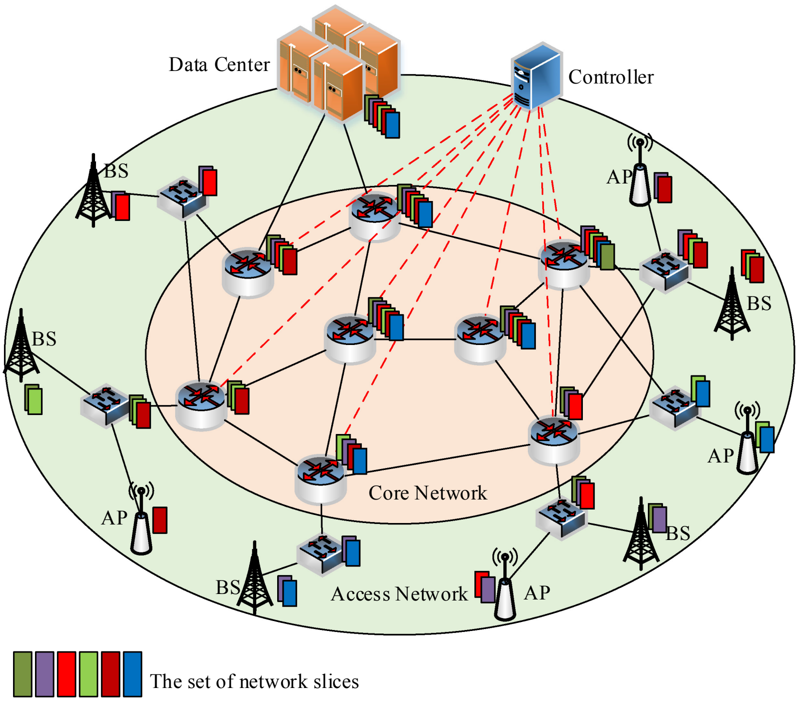

Fig. 2. The Network slicing service in 5G.

the mobility of the connected state to improve the performance of handover and improve the user experience;

(3) uRLLC: The uRLLC mainly used for Internet of Vehicles, precise positioning and virtual reality services. Such services have high reliability and latency requirements and need to provide millisecond end-to-end latency and extremely high availability for users.

Due to the diverse application scenarios and business requirements of 5G, the traditional unified network deployment method cannot flexibly provide users with multiple business services [16, 17]. The network slicing technology shown in Fig. 2 is introduced into 5G networks, it has the following advantages:

(1) Sharing resources, reducing costs, and improving efficiency: multiple network slices operating simultaneously on a unified infrastructure can significantly reduce network construction and operating costs, and improve the resource utilization of general infrastructure;

(2) Logical isolation, safe and reliable: each slice has its own independent virtual resources and life cycle, the creation and destruction of network slices will not affect other slices;

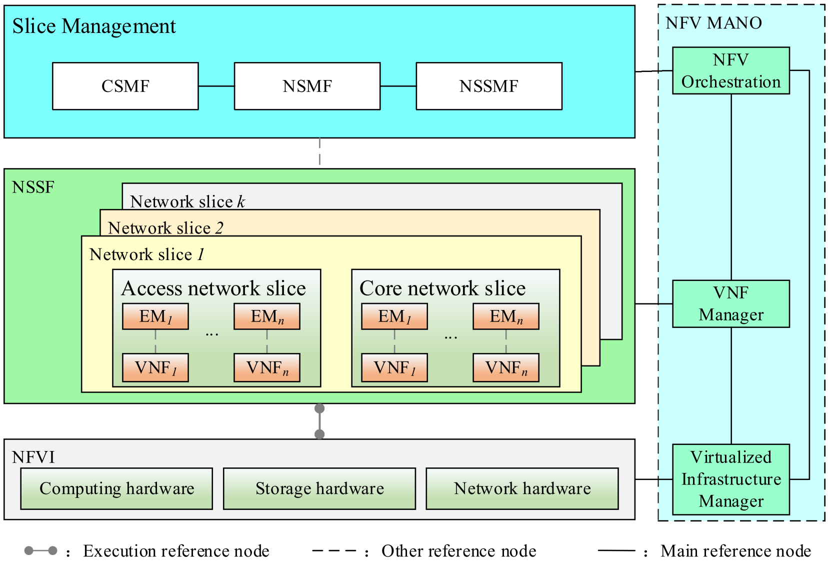

Fig. 3. The architecture of network slicing.

(3) Customized on demand, flexible expansion: the cloud-based network slicing native architecture provides services for different businesses, and infrastructure-as-a-service (IaaS) resources can be customized according to different business scenarios; cloud monitoring can achieve network Real-time monitoring of resource utilization provides high reliability and flexible expansion;

(4) End-to-end to meet differentiated needs: Different business needs determine the performance of network slicing. Through end-to-end deployment, the core network, transmission network and access network are divided to fully meet the needs of 5G diversified services.

The diversification of business and application requirements in the 5G network requires the network to allocate resources more flexibly according to different business characteristics [19–21]. Therefore, 5G networks require a completely innovative network architecture to achieve flexible and alternative networking methods, then the network slicing become as one of the best alternative technologies. Network slicing is an on-demand networking method that allows operators to separate multiple virtual end-to-end networks on a unified infrastructure. Each network slice is logical from the wireless access network bearer network to the core network [22–24]. Isolation to suit various types of applications. The network slicing as Fig. 2 shows.

The architecture of network slicing is shown in Fig. 3, which mainly includes four parts: Network Functions Virtualization Infrastructure (NFVI), NFV Management and Orchestration (MANO), Network Slice Selection Function (NSSF), and Slice management function. There are three functions of network slice management, namely communication service management function (CSMF), network slicing management function (NSMF), and network slicing subnet management function (NSSMF).

2 Problem Statement

At present, the mobile network industry is rapidly evolving to 5G, and three new application scenarios of eMBB, mMTC, and uRRLC play an important role. The 5G network has high flexibility to respond to mobile operators' business changes [25–27]. In particular, the concept of SDN and NFV makes the infrastructure flexible to meet the diversification of vertical application requirements[28, 29]. 5G network slicing promotes the innovation of 5G network architecture with the help of SDN and NFV technology. Compared with traditional networks, 5G networks pay more attention to user experience and have higher requirements for service quality (such as reliability, latency, security, etc.).

5G mobile communication systems as shown in Fig. 4, especially in low-latency scenarios, have high requirements for the latency of communication services, and need to support end-to-end network services in milliseconds for users. In order to reduce the end-to-end delay, it can be achieved by reducing the degree of VNF redundancy, at the expense of reliability. In order to meet the low-latency and high-reliability requirements of service function chain deployment, we take the processing delay and transmission delay as end-to-end delays, and considers the impact of node load when analyzing the processing delay, which is more realistic. The link reliability is used as a constraint to establish a model, and a service function chain deployment algorithm based on QoS guarantee is proposed. The algorithm includes two stages: a virtual network function deployment algorithm based on node invulnerability and a link mapping algorithm based on reliability [30, 31]. We consider the link's resource capacity and reliability conditions during the VNF deployment phase, and realizes two-phase collaborative deployment, which reduces the complexity of the algorithm and improves the utilization of resources while achieving the overall goal of optimizing delay and reliability. In order to more effectively achieve load balancing and protect the underlying nodes and link failures. The continuous iterations of the system obtain a smooth distribution of scores, and the final result can more reasonably reflect the importance of the node in the entire network. Then select the node with the highest score to deploy VNF, effectively balance the node load, reduce processing delay, and at the same time improve the reliability of service function request deployment without reserving resources. Finally, the reliability-based link mapping algorithm is used to select the shortest delay path that meets the reliability requirements to further optimize the performance of the deployment. The proposed algorithm effectively achieves load balancing, maximizes the resource utilization of the underlying network, and guarantees QoS.

Considering the low latency and high reliability requirements of 5G network slicing, we study the deployment of service function chains based on QoS guarantees in the access network of 5G [32, 33]. The bottom layer of the 5G slicing network is a network composed of OpenFlow-based switches, and we use the undirected graph $G^s = (N^s, L^s)$ to represent it, where N^s is the node set of switches, each switch can deploy one or more NFV functions; L^s represents the link set of the network.

$$LS(j) = \sum_{i=1}^{n} r_i LR(i) \tag{1}$$

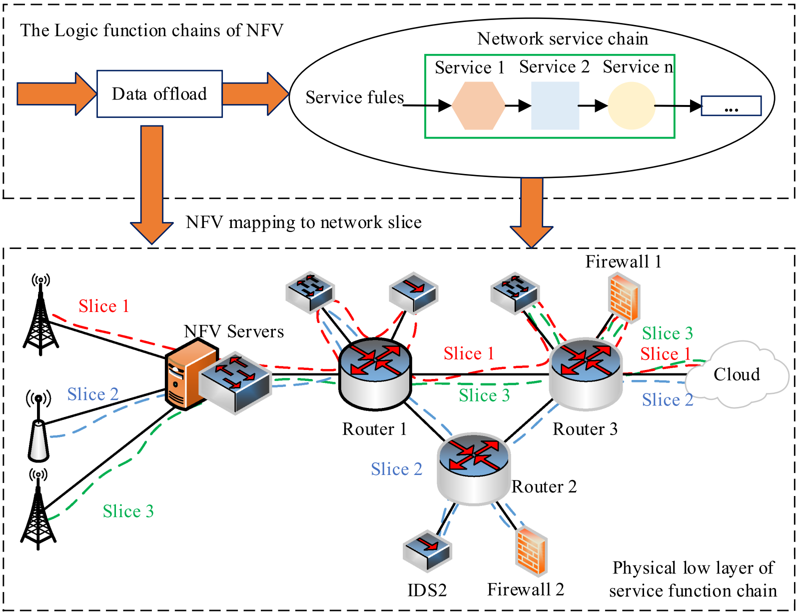

Fig. 4. Mapping of access network and service chain based on network slice.

where $LS(j)$ is the total overload of the SWITCH j, n represent the number of overload factor, r_i represent the coefficient of the weight, $LR(i)$ represent the usage state of the controller overload factor i. Assuming that the load of the controller close value of the threshold is $MAX(j)$, there are two situations about the controller migration:

In the process of controller migration, we set the signal strength from BS to the Switch as one of the triggering conditions, the signal strength from BS to the satellite can be written as

$$P_{mean}(t) = \frac{\sum_{k=1}^{K} \alpha_k P(t - k\Delta T)}{K} \tag{2}$$

where K is the window duration of the signal sampling; ΔT is the sampling period; α_k is the weight when sampling the signal. When the signal strength $P_{mean}(t)$ is smaller than the threshold P_T, therefore, signal strength is a necessary condition for controller migration in the 5G network.

The controller migration cost based on dynamic feedback adopted in this paper not only considers dynamic load information but also static load information when considering migration cost. When calculating the comprehensive cost value and selecting the target controller and switch, the corresponding main influencing factors are mainly considered.

The ***controller migration cost*** (η) is the switch that the message number controller needs to collect statistics from each second at the rate they related. The controller migration overhead in time slot t can be quantified as:

$$\eta_1(t) = \sum_{i=1}^{N} a_i(t)\lambda(t)_i \tag{3}$$

where $\lambda(t)_i$ is the cost of the switches which connect into the controller; $a_{ij}(t)$ is factor of the signal strength between satellite and the BS on the ground, and it can be determined as that

$$a_{ij}(t) = \begin{cases} 1, P_{mean}(t) > P_T \\ 0, P_{mean}(t) \leq P_T \end{cases} \tag{4}$$

The ***switches access cost*** (**ξ**) is that the switches connect into controller needs to handwork with the controller to reconstruct the link between controller and switches. The switch reconstruct overhead can be rewritten as that

$$\xi(t) = \sum_{i=1}^{N} \lambda(t)_i x(t)_i \tag{5}$$

where $\lambda(t)_i$ is the cost of the switches which connect into the controller; $x(t)_i$ is the view duration that the switch access cost when connect into the controller in the SWITCH network.

The migration cost of the controller is the issue that we should pay attention to at present. When the controller is migrated from one Switch into another satellite, switches on the transmission path needs to be re-connected to the new controller periodically to adjust the control plane of the network. The goal of the controller migration in 5G network is to minimize the response time and switching time of the controller, while keeping control flow overhead low [34, 35]. Therefore, we apply the weighting factor to the response time in the objective function.

$$\begin{aligned} \min \eta_1(t) + \xi(t) \\ s.t. \\ C1: a_i(t) \in \{0, 1\} \\ C2: \lambda(t) \in \{0, 1\} \\ C3: P_{mean}(t) < P_T \\ C4: \alpha_k \in [0, 1] \end{aligned} \tag{6}$$

where the condition $C1$, $C2$ and $C3$ ensure that the BS on the ground can be connected into the satellite or not. Constraint $C3$ is the necessary condition of the BS which can be connected into the satellite in the 5G network. Then, the controller migration step can be written as

Step 1 The measurement configuration information delivered by the source satellite through the BS includes the neighbor measurement trigger threshold P_T.

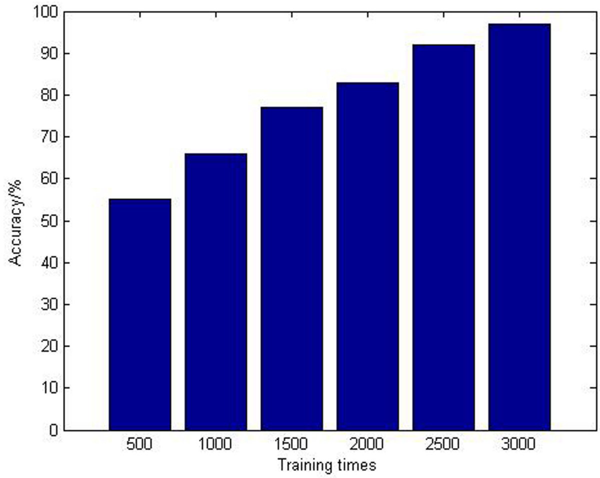

Fig. 5. Relationship between business perception accuracy and training times

Step 2 The BS measures the signal strength of the serving satellite.
Step 3 When the signal strength of the serving satellite is lower than the neighboring satellite measurement trigger threshold, the BS calculates the service satellite sub-satellite point according to the ephemeris information, predicts the handover trigger time handover, and further calculates the neighbor satellites list at the handover trigger time.
Step 4 The BS simultaneously measures the signal of the neighboring satellite and the serving satellite.
Step 5 When the signal strength of the serving satellite is lower than the handover trigger threshold, the handover trigger is determined. The BS calculates the link duration of all neighbor satellites in the neighbor satellite list, filters out the neighbor satellites whose link duration is too long, and then calculates the switching weight of each neighbor satellite.
Step 6 If the link duration with the largest switching weight in the neighbor list is still the current service satellite, and it is determined that the result of switching to the current service satellite, the switching is aborted, and the switching trigger process is interrupted; otherwise, migrating the controller to the target satellite.
Step 7 Return Step 2.

3 Simulation Result and Analysis

In order to verify the proposed mechanism, this paper builds a simulation system on the edge of the network based on the NS2 simulation software. The simulation system consists of 32 network elements mounted on a terminal, designing an SDN network, adding edge nodes, and verifying the proposed service perception technology at the edge of the network can achieve satisfactory results. Through simulation, we compare the network edge with service awareness mechanism and the network edge without service awareness [21, 36]. In the simulation, the service types are divided into two types: packet loss rate sensitive service and delay sensitive service.

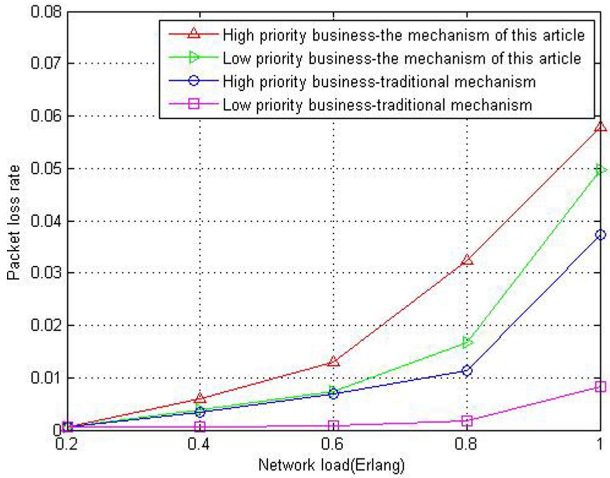

Fig. 6. Comparison of packet loss rate

Figure 5 is a diagram of the relationship between the accuracy of service perception and the number of trainings. It can be seen that as the number of trainings increases, the accuracy of business perception also increases. The fully trained echo state network algorithm can ensure the accuracy of business classification.

At the edge network, this method (service-aware) and the traditional non-service-aware mechanism are used for simulation. Figure 6 and Fig. 7 are the results of comparison. Combining the simulation results of Fig. 6 and Fig. 7, with the increase of the network service load, the packet loss rate and real-time performance of our constructed system show a tendency to deteriorate. Under high load, the network edge system adopting the mechanism of this paper is superior to the traditional EPON system in two important indicators of high-priority service, such as packet loss rate and transmission delay. On the other hand, low-priority services have low requirements on transmission delay and packet loss rate, the mechanism of this paper will reduce the performance of services to a certain extent. In exchange for overall business service quality, especially high priority QoS requirements for high-level services.

The comparison of packet loss rate and delay shows that the business-aware method proposed in this paper can ensure that different types and different priorities of services at the edge of the network can be matched with the lower computational complexity, and ensure the overall service quality of the business.

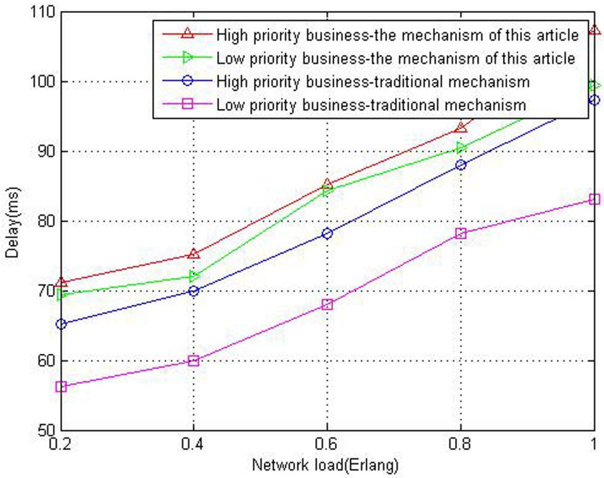

Fig. 7. Delay comparison

4 Conclusions

In this paper, we study the 5G network slice based edge access approach with the communication quality insurance. In response to the requirements of 5G network low-latency and high-reliability scenarios and the deficiencies of traditional service function chain deployment methods, we take into account the invulnerability of nodes and the reliability of links while studying latency issues, ensuring the reliability of service function chain deployment. The feasibility and effectiveness of the method are verified. This method evaluates the node through comprehensive factors such as node invulnerability and link reliability, and then selects the node for service function chain deployment to achieve load balancing and reduce processing delay. At the same time, the shortest path that meets the reliability constraints is found to reduce the shortest path. Transmission delay reflects the overall performance of optimizing delay and reliability. The simulation results show that the algorithm in this chapter significantly improves the service function chain request acceptance rate, end-to-end delay, node and link utilization, and node and link reliability performance.

References

1. Guo, Y., Wang, Z., Yin, X., et al.: Traffic engineering in hybrid SDN networks with multiple traffic matrices. Comput. Netw. **126**, 187–199 (2017)
2. Liu, G., Guo, S., Zhao, Q., et al.: Tomogravity space based traffic matrix estimation in data center networks. Transp. Res. Part C: Emerg. Technol. **86**, 39–50 (2018)
3. Jiang, D., Wang, Y., Lv, Z., Wang, W., Wang, H.: An energy-efficient networking approach in cloud services for IIoT networks. IEEE J. Sel. Areas Commun. **38**(5), 928–941 (2020)
4. Hashemi, H., Abdelghany, K.F., et al.: Real-time traffic network state estimation and prediction with decision support capabilities: Application to integrated corridor management. Transp. Res. Part C: Emerg. Technol. **73**, 128–146 (2016)

5. Kawasaki, Y., Hara, Y., Kuwahara, M.: Traffic state estimation on a two-dimensional network by a state-space model. Transp. Res. Part C: Emerg. Technol. **5**, 1–17 (2019)
6. Jiang, D., Wang, W., Shi, L., Song, H.: A compressive sensing-based approach to end-to-end network traffic reconstruction. IEEE Trans. Netw. Sci. Eng. **7**(1), 507–519 (2020)
7. Dias, K.L., Pongelupe, M.A., Caminhas, W.M., et al.: An innovative approach for real-time network traffic classification. Comput. Netw. **158**, 143–157 (2019)
8. Ermagun, A., Levinson, D.: Spatiotemporal short-term traffic forecasting using the network weight matrix and systematic detrending. Transp. Res. Part C: Emerg. Technol. **104**(5), 38–52 (2019)
9. Jiang, D., Huo, L., Li, Y.: Fine-granularity inference and estimations to network traffic for SDN. PLoS ONE **13**(5), 1–23 (2018)
10. Keshavamurthy, P., Pateromichelakis, E., Dahlhaus, D., et al.: Cloud-enabled radio resource management for co-operative driving vehicular networks. In: Proceedings of the WCNC'19, pp. 1–6 (2019)
11. Wang, Y., Jiang, D., Huo, L., Zhao, Y.: A new traffic prediction algorithm to software defined networking. Mob. Netw. Appl. (2019). https://doi.org/10.1007/s11036-019-01423-3
12. Qi, S., Jiang, D., Huo, L.: A prediction approach to end-to-end traffic in space information networks. Mob. Netw. Appl. (2019). https://doi.org/10.1007/s11036-019-01424-2
13. Jiang, D., Zhang, P., Lv, Z., et al.: Energy-efficient multi-constraint routing algorithm with load balancing for smart city applications. IEEE Internet of Things J. **3**(6), 1437–1447 (2016)
14. Li, J., Shen, X., Chen, L., et al.: Service migration in fog computing enabled cellular networks to support real-time vehicular communications. IEEE Access **7**(2019), 13704–13714 (2019)
15. Wang, F., Jiang, D., Qi, S., et al.: A dynamic resource scheduling scheme in edge computing satellite networks. Mob. Netw. Appl. (2019). https://doi.org/10.1007/s11036-019-01421-5
16. El-sayed, H., Sankar, S., Prasad, M., et al.: Edge of things: the big picture on the integration of edge, IoT and the cloud in a distributed computing environment. IEEE Access **6**, 1–12 (2018)
17. Jiang, D., Huo, L., Song, H.: Rethinking behaviors and activities of base stations in mobile cellular networks based on big data analysis. IEEE Trans. Netw. Sci. Eng. **7**(1), 80–90 (2020)
18. Zhang, K., Mao, Y., Leng, S., et al.: Mobile-edge computing for vehicular networks. IEEE Veh. Technol. Mag. **12**, 36–44 (2017)
19. Pu, L., Chen, X., Mao, G., et al.: Chimera: an energy-efficient and deadline-aware hybrid edge computing framework for vehicular crowdsensing applications. IEEE Internet of Things J. **6**(1), 84–99 (2019)
20. Jiang, D., Wang, Y., Lv, Z., Qi, S., Singh, S.: Big data analysis based network behavior insight of cellular networks for industry 4.0 applications. IEEE Trans. Ind. Inf. **16**(2), 1310–1320 (2020)
21. Eldjali, C., Lyes, K.: Optimal priority-queuing for EV charging-discharging service based on cloud computing. In: Proceedings of the ICC'17, pp. 1–6 (2017)
22. Jiang, D., Huo, L., Lv, Z., Song, H., Qin, W.: A joint multi-criteria utility-based network selection approach for vehicle-to-infrastructure networking. IEEE Trans. Intell. Transp. Syst. **19**(10), 3305–3319 (2018)
23. Xie, R., Tang, Q., Wang, Q., et al.: Collaborative vehicular edge computing networks: architecture design and research challenges. IEEE Access **7**(2019), 178942–178952 (2019)
24. Yang, Y., Niu, X., Li, L., et al.: A secure and efficient transmission method in connected vehicular cloud computing. IEEE Netw. **32**, 14–19 (2018)
25. Jiang, D., Li, W., Lv, H.: An energy-efficient cooperative multicast routing in multi-hop wireless networks for smart medical applications. Neurocomputing **2017**(220), 160–169 (2017)

26. Kaur, K., Garg, S., Kaddoum, G., et al.: Demand-response management using a fleet of electric vehicles: an opportunistic-SDN-based edge-cloud framework for smart grids. IEEE Netw. **33**, 46–53 (2019)
27. Guo, H., Zhang, J., Liu, J.: FiWi-enhanced vehicular edge computing networks. IEEE Veh. Technol. Mag. **14**, 45–53 (2019)
28. Liu, H., Zhang, Y., Yang, T.: Blockchain-enabled security in electric vehicles cloud and edge computing. IEEE Netw. **32**(3), 78–83 (2018)
29. Wang, F., Jiang, D., Qi, S.: An adaptive routing algorithm for integrated information networks. China Commun. **7**(1), 196–207 (2019)
30. Wang, J., He, B., Wang, J., et al.: Intelligent VNFs selection based on traffic identification in vehicular cloud networks. IEEE Trans. Veh. Technol. **68**(5), 4140–4147 (2019)
31. Huo, L., Jiang, D., Qi, S., et al.: An AI-based adaptive cognitive modeling and measurement method of network traffic for EIS. Mob. Netw. Appl. (2019). https://doi.org/10.1007/s11036-019-01419-z
32. Li, M., Si, P., Zhang, Y.: Delay-tolerant data traffic to software-defined vehicular networks with mobile edge computing in smart city. IEEE Trans. Veh. Technol. **67**(10), 9073–9086 (2018)
33. Garg, S., Kaur, K., Ahmed, S., et al.: MobQoS: mobility-aware and QoS-driven SDN framework for autonomous vehicles. IEEE Wirel. Commun. **26**, 12–20 (2019)
34. Huo, L., Jiang, D., Lv, Z., et al.: An intelligent optimization-based traffic information acquirement approach to software-defined networking. Comput. Intell. **36**, 1–21 (2019)
35. Lin, C., Deng, D., Yao, C.: Resource allocation in vehicular cloud computing systems with heterogeneous vehicles and roadside units. IEEE Internet of Things J. **5**(5), 3692–3700 (2018)
36. Garg, S., Singh, A., Batra, S., et al.: UAV-empowered edge computing environment for cyber-threat detection in smart vehicles. IEEE Netw. **32**, 42–51 (2018)