Traffic Measurement Optimization Based on Reinforcement Learning in Large-Scale ITS-Oriented Backbone Networks

LAISEN NIE\textsuperscript{1,2}, HUIZHI WANG\textsuperscript{1}, XIN JIANG\textsuperscript{3}, YI GUO\textsuperscript{4}, AND SHENGTAO LI\textsuperscript{5}

\textsuperscript{1}School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China
\textsuperscript{2}Qingdao Research Institute, Northwestern Polytechnical University, Qingdao 266000, China
\textsuperscript{3}Department of Geriatrics, The Second Clinical Medical College (Shenzhen People's Hospital), Jinan University, Shenzhen 510632, China
\textsuperscript{4}Department of Neurology, The Second Clinical Medical College (Shenzhen People's Hospital), Jinan University, Shenzhen 518020, China
\textsuperscript{5}College of Information Science and Engineering, Shandong Normal University, Jinan 250014, China

Corresponding authors: Xin Jiang (jiangxinsz@163.com) and Yi Guo (xuanyi_guo@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61701406 and Grant 61803238, in part by the Applied Basic Research Programs of Qingdao City under Grant 18-2-2-36-jch, in part by the Shaanxi Provincial Innovation and Entrepreneurship Training Program for Undergraduates under Grant S201910609404, in part by the Shenzhen Science and Technology Innovation Commission Basic Research Project under Grant JCYJ20170818111012390, and in part by the Sanming Project of Medicine in Shenzhen under Grant SYJY201905 and Grant SYJY201906.

ABSTRACT The end-to-end network traffic information is the basis of network management for a large-scale intelligent transportation systems-oriented backbone network. To obtain exact network traffic data, a prevalent idea is to deploy NetFlow or sFlow on all routers of the network. However, this method not only increases operational expenditures, but also affects the network load. Motivated by this issue, we propose an optimized traffic measurement method based on reinforcement learning in this paper, which can collect most of the network traffic data by activating NetFlow on a subset of interfaces of routers in a network. We use the $Q$-learning-based approach to deal with the problem of the interface-selection. We propose an approach to compute the reward, furthermore a modified $Q$-learning approach is proposed to handle the problem of interface-selection. The method is evaluated by the real data from the Abilene and GEANT backbone networks. Simulation results show that the proposed method can improve the efficiency of traffic measurement distinctly.

INDEX TERMS Network measurement, reinforcement learning, intelligent transportation systems, IP backbone network.

I. INTRODUCTION

Intelligent Transportation Systems (ITS) use various communication technologies to build a comprehensive integrated system, which can provide an available solution for the road safety and management problems. In recent years, kinds of emerging techniques have been leveraged in networking for constructing an ITS, such as the Vehicular Ad-hoc Network (VANET) and Internet of Things (IoT) [1]. These wireless and wired networks provide an open access for each user. In this case, the network management and security functions are much more crucial for the complex and heterogeneous networks [2]–[9]. In addition, effective network management can provide users with a high quality of service and keep our networks away from network congestion, Distributed Denial of Service (DDoS) and other network attacks [4], [6], [10].

As a crucial input parameter, a traffic matrix (TM) describes the dynamic traces of network traffic between origin-destination (OD) nodes in the network. There are many taxonomies for achieving the TM. Generally, it can be divided into two classifications, which are the direct measurement methods and the network traffic estimation methods [11]–[13]. The network traffic estimation methods infer the TM in terms of the relationship between the TM and link loads. However, it suffers from a highly ill-posed feature so that it is significantly difficult to acquire an accurate traffic matrix estimator directly. On the other hand, although the direct measurement can obtain a precise traffic matrix, it arises a burden in the communication bandwidth. Besides,
it also consumes many resources of routers (e.g., CPU and memory). The main reasons can be summarized as follows:

- Switches and routers applied to implement direct measurement in networks need to be updated and maintained in time, which increases the cost of network operations.
- The construction of TM requires frequently collecting of network traffic measurements. Storing such information occupies a large amount of storage capacity of network nodes.
- The collected network traffic data needs to be transmitted to the network manager for further processing and analysis, which will add the redundant communication overhead for the network.

To improve the high consumption of direct measurement, we focus on the problem of traffic measurement optimization in large-scale ITS-oriented backbone networks, and propose a method based on reinforcement learning (RL). The details of the traffic measurement optimization are presented in Fig. 1. A novel method based on Q-learning to realize the network measurement optimization is proposed in this paper. Since the routing matrix of an IP backbone network is available via the routing table, we compute the immediate reward (denoted by the matrix R) and the cumulative reward (denoted by the matrix Q) according to the routing matrix. Then, we can achieve a subset of interfaces, in which the NetFlow has to be activated. By enabling the NetFlow on these interfaces, we can measure a large proportion of end-to-end network traffic via a small number of routers. The main contributions of this paper can be summarized as follows:

- The network traffic measurement optimization is an NP-hard problem. Hence, we use the Q-learning algorithm to search the optimal solution of the traffic measurement optimization model. In detail, an immediate reward is first defined to describe the objective of the model. Moreover, the optimal solution with the highest cumulative reward is calculated by using the Q-learning algorithm.
- To decrease the number of routers activating NetFlow, the correlation of each interface is considered. Meanwhile, to acquire the optimal solution, we involve a weight in the immediate reward so that the number of routers can be decreased obviously.
- An algorithm to select the routers via the cumulative reward is devised. According to the cumulative reward, we propose an algorithm to gain a specified proportion of network traffic data.

The reminder of this paper is organized as follows. Section II presents the related work in the traffic measurement optimization problem. Section III introduces the system model of the traffic measurement optimization problem. In section IV, the traffic measurement optimization method based on RL is proposed. In section V, we evaluate the proposed method by real network traffic data sets. Section VI summarizes the main work of this paper.

II. RELATED WORK

To reduce the consumption of direct measurement for our network, many methods have been proposed to optimize the deployment mechanism of direct measurement [11], [14]–[21]. Ghode et al. focused on the problem of traffic measurement optimization in mobile ad-hoc networks, and proposed an improved ZRP protocol. Meanwhile, they added the energy constraints of the optimization model to make it work efficiently. The authors in [17] proposed a dynamic cooperative monitoring interface-selection mechanism based on social network analysis, which can effectively reduce the network cost of direct measurement. Besides, the authors in [18] studied the problem of selecting a set of monitoring interfaces, and proposed a Cross-Layer Security Monitoring selection algorithm based on Traffic Prediction (CLSM-TP). The authors in [19] proposed a mechanism to support the Software Defined Network (SDN) to distribute the monitoring load on multiple switches in the network by monitoring traffic in a distributed manner. They distributed the monitoring entries over switches assigning the monitoring tasks and eliminating the duplicated monitoring entries. Thus, it non-trivially reduces the amount of monitoring overhead in switches, the controller, and the control channels. Besides, Shin et al. in [20] proposed a distributed on-line algorithm for optimal sniffer channel allocation for passive monitoring in multi-channel wireless networks.

By using the advantages of SDN, some novel methods have been proposed to monitor the TM directly. The authors in [22] proposed a flow monitoring algorithm called lonely flow first (LFF). The LFF method polled switches with traffic flows that pass through only a single switch at first. Then, they also proposed a weight function to decide the polling order. Furthermore, the polling mechanism can select the lower cost. Besides, Tian Yang in [23] presented a new framework for solving the problem of network traffic measurement in SDN-based IP networks. Based on the low rank and time-varying characteristics of traffic flows, the authors in [24] proposed a novel mechanism for the on-line network traffic measurement by utilizing the flow measurement ability of the SDN technique. In [25], an SDN-based fine-grained measurement method in vehicular communication network was proposed. The proposed method contained two phases. At first, they proposed a monitoring method to measure the statistics of links and flows, and then obtained the coarse-grained measurements depending...
on the measured links and flows. Then, they performed the interpolation algorithm on the coarse-grained measurements to gain the fine-grained network traffic measurements. The work in [26] presented an on-line measurement mechanism for SDN-based data center networks. A comprehensive monitoring system, which was used to poll the switches to collect the real-time statistics for higher accuracy and lower overhead, was designed.

Although a lot of methods have been proposed to optimize the consumption of direct measurement, it is significantly difficult to acquire an outstanding optimal solution due to the lack of scalability. Thus, we propose a traffic measurement optimization method based on RL, which realizes the requirement of obtaining a large proportion of traffic data while activating NetFlow on fewer routers.

III. PROBLEM DEFINITION

When network operators perform network management decisions, they need to know the current states of our network, such as delay, packet loss rate, bandwidth, and network traffic, etc. Network measurement techniques provide feasible solutions and technical support for network operators to obtain reliable network states. Since the network tomography technology was applied to the field of TM estimation, many TM estimation methods based on the network tomography technology have emerged. The network tomography technology uses easily measured network data to estimate the TM. At the same time, it is widely used in network delay estimation and other fields. In the field of TM estimation, the network tomography model describes the linear relationship among the link load $Y$, the routing matrix $\Theta$, and the TM $X$, which is shown as follows:

$$Y = \Theta X.$$  \hspace{1cm} (1)

The routing matrix (shown by Fig. 2) describes the transmission path of each OD flow in the network, thus the linear relationship between the link load and the TM can be described by the routing matrix. If each element of the routing matrix is represented as $\Theta(l, o, d)$, and then $\Theta(l, o, d) = 1$ indicates that the flow from node $o$ to node $d$ passes the link $l$, otherwise it is 0. In a large-scale IP backbone network, the link load is an available network data, which can be measured by way of the Simple Network Management Protocol (SNMP) [11], [27]. The routing matrix is built via the routing table in each router. Therefore, for large-scale IP backbone networks, the routing matrix is also available data. Although the network tomography method has the above convenience, there is also a key problem that restricts its applications. In detail, for a large-scale IP backbone network with $N$ routers, the number of OD flows is $N^2$ ($N^2 >> L$), where $L$ is the total number of links in the network. Therefore, the network traffic estimation method based on the network tomography model is an inverse problem with highly ill-posed characteristics (i.e., underdetermined characteristics). So the TM cannot be estimated directly by the network tomography estimation model without other prior constrains.

Assume that, an OD flow in the network can be numbered as $x(o, d)$, and $o, d = 1, 2, 3, \ldots, N$. In addition, the link can be denoted as link $(l)$, and $l = 1, 2, 3, \ldots, L$ (see Fig. 2). NetFlow is a solution that can measure all of OD flows in the network. The problem of traffic measurement optimization is selecting a subset of routers activating NetFlow so that we can gain the network traffic information as much as possible. We assume that each router in the network has $K(n)$ interfaces on the router $n$, and the variable $r_{n,k}$ takes 1 if the NetFlow is enabled on the $k$-th interface of the $n$-th router, otherwise it is 0. Furthermore, the volume of OD flows that can be measured by the $k$-th interface of the $n$-th router is $S_{n,k}$. Then the problem of traffic measurement optimization can be defined as:

$$\min \sum_{n \in N, k \in K(n)} r_{n,k}$$
$$s.t. \bigcup_{n \in N, k \in K(n)} S_{n,k} \geq \alpha S,$$
$$r_{n,k} \in \{0, 1\}$$, \hspace{1cm} (2)

where $S$ is all the traffic data in the network, and $\alpha$ is a ratio that needs to be measured. Generally, some network management functions can be carried out using incomplete network traffic data. Namely, the network management functions have a tolerability for the error of network traffic data, which arises the ratio $\alpha$. From the above model, we see that it limits the number of interfaces activating NetFlow. To make the number of routers lower, the optimization model can be further deduced by:

$$\min \sum_{n \in N} d_n$$
$$s.t. \bigcup_{n \in N, k \in K(n)} S_{n,k} \geq \alpha S,$$
$$r_{n,k} \in \{0, 1\},$$
$$d_n = \bigcup_{k \in K(n)} r_{n,k},$$ \hspace{1cm} (3)

where $d_n = 1$ if the NetFlow is enabled on the $n$-th router, otherwise it is 0.
IV. REINFORCEMENT LEARNING AND OUR METHODOLOGY

A. REINFORCEMENT LEARNING

RL [28], [29] is one of the most prevalent machine learning methods. The RL is viewed as the Markov decision process that can be denoted by $(S, A, P, R)$, where $S$ and $A$ are the sets of states and actions of an agent, respectively. $P$ is the matrix made up of the probability of transition from the current state to the next state after executing an action. In RL, the agent interacts with the environment to obtain the suitable evaluation value (reward) of the action, and the policy is updated to obtain the maximum reward (see Fig. 3). At a certain state $s$, the agent selects an action $a$ to act on the environment, and causes the environment to change and return an immediate reward $R$. $R(s,a)$ is the reward function, namely, the evaluation given by the environment after the agent selects the action $a$ according to the policy. After the agent receives the reward $R$, it makes the next step. Since each step of action $a$ not only affects the single enhancement value, but also the cumulative reward value after multiple learning, the selection policy of each step can increase the probability of obtaining the maximum reward. After each attempt to reach the destination state, the value of the function $Q$ indicating the cumulative reward will be updated. If $Q$ is enhanced, the agent will find the fastest route to reach the target state.

The main algorithm of RL consists of the TD algorithm, the $Q$-learning algorithm, the Sarsa algorithm, and the $R$ learning algorithm, etc. The destination of the $Q$-learning algorithm known as a model-free algorithm is to learn a policy to make the value function optimal [28]. The optimal policy for $Q$-learning is to make the agent repeatedly explore the cumulative value function to maximize $Q$, which can be expressed as:

$$
\pi^*(s,a) = \arg \max Q(s,a). \tag{4}
$$

$\pi^*(s,a)$ is the optimal policy given the state $s$ and the action $a$. In other words, after the value function $Q$ is generated, given a state $s$, the agent only needs to select a series of actions in the direction, in which the value function is largest, and the global optimal solution can be obtained. The $Q$-learning iteration is shown as follows:

$$
Q(s_t, a_t) = Q(s_t, a_t) + \eta \left[ R(s_t, a_t) + \gamma \cdot \max_a Q(s_{t+1}, a_t+1) \right], \tag{5}
$$

where $\eta$ is the learning factor, and $Q(s_t, a_t)$ is the cumulative reward for $t$ steps. $R(s_t, a_t)$ is the one-step reward value, and $\gamma$ is the discount factor.

B. TRAFFIC MEASUREMENT OPTIMIZATION BASED ON REINFORCEMENT LEARNING

The problem of traffic measurement optimization is defined by selecting a subset from all routers activating NetFlow to collect most (or all) of the network traffic data. Hence, we assume that, there are $L$ links, $N$ routers, and $P$ interfaces in the network. We refer to each interface as the state $i$, and the process of selecting the next interface is viewed as the action $j$.

Each row of the routing matrix shown in Fig. 2 describes all the OD flows passing the corresponding link. We express each row of routing matrix by vector $\theta(l)$, where $l = 1, 2, 3, \ldots, L$. When an interface is selected, we hope that this interface is much more different from the previous selected interface. In other words, we prefer to select an interface consisting of much more unmeasured OD flows. Therefore, we define the reward function $R(l_1, l_2)$ as follows:

$$
R(l_1, l_2) = \sum \left( \theta(l_1) \otimes \theta(l_2) \right), \tag{6}
$$

where $l_1, l_2 \in \{1, 2, 3, \ldots, L\}$. Then, the reward is an $L \times L$ matrix.

An IP backbone network contains two types of links, that is the internal links and the external links. When we select an interface in terms of the corresponding link, a crucial problem arises since an internal link corresponds to two interfaces. Furthermore, the reward is extended as a $P \times P$ matrix, and

$$
P = 2L_{\text{internal}} + L_{\text{external}}, \tag{7}
$$

where $L_{\text{internal}}$ is the number of internal links and $L_{\text{external}}$ is the number of external links. In this case, the element of the extended reward $R(i,j)$ is calculated by:

$$
\begin{cases}
R(i,j) = \sum \left( \theta(l_1) \otimes \theta(l_2) \right), \\
i \in \text{link}(l_1), \\
j \in \text{link}(l_2).
\end{cases} \tag{8}
$$

Note that, $i \in \text{link}(l_1)$ means that $i$ is an interface of the link $l_1$. Besides, when an interface is determined, we also consider the number of routers and limit it small enough. Thereby, a weight is involved in the reward, which is computed as:

$$
\begin{cases}
\forall (i, n_1) \in PN, (j, n_2) \in PN \\
PN = \{(p, n) | p \in P_n, n \in N\} \\
w_{ij} = W, \text{ if } n_1 = n_2 \\
w_{ij} = 0, \text{ if } n_1 \neq n_2
\end{cases} \tag{9}
$$

We define $P_n$ as the interface $P$ of the router $n$. Then, the reward $R(i,j)$ is:

$$
R(i,j) = R(i,j) + \lambda w_{ij}, \tag{11}
$$

where $\lambda$ is the factor that expresses the impact of the weight.

The cumulative reward can be calculated by carrying out multiple iterations via the immediate reward. In this paper, we use the $Q$-learning algorithm to compute the cumulative reward. The specific steps of calculating the cumulative reward $Q$ are as follows:
Algorithm 1 Calculating the Cumulative Reward $Q$

Step 0: Set parameter $\gamma$ and let $Q = 0$.
Step 1: Calculate the matrix $R$ according to Eq. 11.
Step 2: Select a starting state $i$.
Step 3: Check the target status: if it is reached, and stop.
Step 4: Select one possible action $j$ from all possible actions of the current state $i$. Use this possible action to reach the next state. For the next state, choose the largest $Q$ value based on all possible actions. Calculate $Q(s_t, a_t) = R(s_t, a_t) + \gamma \times \max Q(s_{t+1}, a_{t+1})$.
Step 5: Set the next state as the current state, and go to step 3.

After obtaining the cumulative reward matrix $Q$, we propose an algorithm for traffic measurement optimization. Our goal is to select several interfaces to characterize a proportion ($\alpha \in [0, 1]$) of the total network traffic. We denote the total traffic amount by $t \geq 0$, and the constraint can be written as:

$$t \geq \alpha \times N^2.$$ (12)

The details of the proposed traffic measurement optimization algorithm are described as follows:

Algorithm 2 Traffic Measurement Optimization

Step 0: Set parameter $\alpha$, and $t = 0$.
Step 1: Calculate the correlation coefficient of between any two links, and select the link with the largest correlation coefficient with other links. Select the initial state $i$ according to the relationship between the link and the interface.
Step 2: Check the requirement shown in Eq. 12. If it is satisfied, and the STOP.
Step 3: Select the action $j$ with the highest $Q$ value corresponding to $i$.
Step 4: Record selected $i$ and $j$, and delete the $i$ and $j$ column in the matrix $Q$ to prevent duplicate selection.
Step 5: Record the router $n$, where $i$ and $j$ are located.
Step 6: Compute the traffic amount $t'$.
Step 7: Update the total traffic amount $t$: $t = t + t'$. Set the next state $i$ as the current state $j$, and go to step 2.

V. EVALUATION AND RESULTS

A. DATA SET

In this section, we evaluate the performance of the proposed algorithm based on two real network traffic data sets with different configurations. The Abilene backbone network is made up of 12 routers and 54 links consisting of 30 internal links and 24 external links, see Fig. 4. The data records the network traffic with 2016 time slots. The $54 \times 144$ routing matrix is a matrix with elements of 1 or 0. The GÉANT network contains 45 routers and 142 links (shown by Fig. 5). Meanwhile, all the links of the GÉANT network are internal links.

B. PARAMETER SELECTION AND PERFORMANCE INDICATORS

We first evaluate the sensitivity of the proposed method with various parameters. Figure 6 shows the number of routers activating NetFlow with various weights $W$ and the proportion of traffic $\alpha$ in the Abilene backbone network. We find that the number of routers is much lower, if the weight value increases. In Fig. 7, we set $\alpha = 95\%$, and show the number of routers with various weights and factors. From Fig. 7, we can gain the same conclusion as Fig. 6.
FIGURE 7. Sensitivity for the weight $W$ and the factor $\lambda$ in Abilene.

FIGURE 8. Sensitivity for the weight $W$ and the proportion of traffic $\alpha$ in GÉANT.

FIGURE 9. Sensitivity for the weight $W$ and the factor $\lambda$ in GÉANT.

Figures 8 and 9 show the simulation results in the GÉANT backbone network. Comparing with the Abilene network, the GÉANT is a larger scale backbone network. Figures 8 and 9 show that the proposed method has a better optimization solution and much more insensitive.

To evaluate the performance of the proposed method, we mainly consider the following performance metrics:

- To minimize the extra overhead, we hope that the number of routers activating NetFlow is small enough. Hence, the number of routers activating NetFlow is an important metric, when we want to collect different proportions of traffic. As a result, when measuring the same proportion of traffic, we choose the number of routers as a metric.
- When we measure the same number of OD flows, we want to obtain traffic information as much as possible. Therefore, the proportion of measurable traffic information of the total traffic is another performance metric chosen in this paper.
- We also measure the performance using the Normalized Mean Absolute Error (NMAE) defined as:

$$NMAE = \frac{\sum_{i,j \in \Omega} |X(i,j) - \hat{X}(i,j)|}{\sum_{i,j \in \Omega} |X(i,j)|},$$

where $X(i,j)$ is the element of TM, and $\hat{X}(i,j)$ is its estimator. $\Omega$ represents the unestimated set of OD flows.

C. COMPARISON ANALYSIS

To evaluate the performance of the proposed method, we compare our method (RL for short) with the method in [10]. In [10], the authors propose a method to solve the traffic measurement optimization problem by using linear programming (LP for short). When implementing the Algorithm 1, we set $W = 100$.

Figures 10 and 11 plot the number of routers activating NetFlow, when we collect different proportions of traffic for the Abilene and the GÉANT networks, respectively. The x-axis and y-axis are the proportion of traffic and the number of routers, respectively. As shown in Fig. 10, to get 95% of the total traffic information, we only need 50% of the routers for the proposed method. In Figure 11, we need to activate less than 5 routers to obtain more than 95% of the total traffic information using the proposed method in GÉANT. By contrast, the LP method requires more than 11 and 40 routers in Abilene and GÉANT, respectively. Moreover, we can find...
that RL is good at optimizing a large-scale network. As shown by above simulations, the RL method needs less routers than the LP method. At the same time, the solution of LP takes more time in practical, and this process is more complex.

Moreover, we compare our method with four data-driven network traffic measurement approaches, i.e., the iSTAMP method [30], the Bernoulli method (BM) [31], the compressive sensing over graph method (CSOG) [32], and the random walk method (RW) [33]. In [30], they proposed an efficient MAB-based flow sampling algorithm to select the most rewarding flows. It’s worth noting that the RL method we proposed can select the most important flows in the network as well. Figures 12 and 13 show the proportion of traffic that can be sampled for the Abilene and the GÉANT networks, respectively. The x-axis and y-axis are the parameter $\alpha$ that has been proposed before and the proportion of traffic, respectively. As shown in Fig. 12, when we measure the same number of OD flows, the method we proposed can obtain more traffic information than iSTAMP, BM, CSOG and RW. For instance, if we select 80% of OD flows in the Abilene network, we can only obtain 35% of traffic information using iSTAMP method and nearly 70% of traffic can be measured based on the RL method we proposed. Meanwhile, they are 14%, 13%, and 8% in turn for BM, RW, and CSOG. When we apply these methods to the GÉANT network, which is larger than the Abilene network, the advantage is much more obvious (see Fig. 13).

We also compare the NMAE of the proposed algorithm with the other four methods, as shown in Figures 14 and 15. Obviously, with the number of OD flows that will be
measured increases, the NMAE decreases rapidly. However, comparing with the other four methods, the NMAE of the RL method is almost an invariant near zero with the increase of $\alpha$, which means that the RL method may be an optimal solution when selecting the most rewarding flow of a network.

VI. CONCLUSION

This paper studies the problem of traffic measurement optimization, and propose a RL-based method to decrease the cost of direct measurement which is the main contribution of this paper. In our method, we first model the problem of traffic measurement optimization as a Markov decision process, and propose a RL-based method to select the subset of routers to implement end-to-end network traffic measurement. First, we propose an approach to construct the immediate reward via the routing matrix. Then, an iteration algorithm based on $Q$-learning is proposed to compute the cumulative reward. By means of the cumulative reward, the optimal subset of routers activating NetFlow to measure most of network traffic.

We evaluate the performance of the proposed method by real traffic data sets from the Abilene and GÉANT backbone networks. The simulation results indicate that the proposed method can observably reduce the number of routers activating traffic measurement. Specially, aiming at the large-scale backbone network (e.g., the GÉANT backbone network), the RL method reveals to be an excellent solution.

REFERENCES

[1] Z. Ning, Y. Feng, M. Collotta, X. Kong, X. Wang, L. Guo, X. Hu, and B. Hu, "Deep learning in edge of vehicles: Exploring trirelationship for data transmission," IEEE Trans Ind. Informat., vol. 15, no. 10, pp. 5737–5746, Oct. 2019.

[2] H. Yang, X. Xie, and M. Kadoch, "Intelligent resource management based on reinforcement learning for ultra-reliable and low-latency IoV communication networks," IEEE Trans. Veh. Technol., vol. 68, no. 5, pp. 4157–4169, May 2019.

[3] A. M. R. Tolba, "Trust-based distributed authentication method for collision attack avoidance in VANETs," IEEE Access, vol. 6, pp. 62747–62755, 2018.

[4] A. M. Khakhlof, N. Boudriga, and M. S. Obaidat, "Radio-based cooperation scheme for DDoS detection," in Proc. 14th IEEE Int. Conf. Electron., Circuits Syst., Dec. 2007, pp. 1051–1054.

[5] Z. Ning, P. Dong, and X. Wang, "Mobile edge computing enabled 5G health monitoring for Internet of Medical Things: A decentralized game theoretic approach," IEEE J. Sel. Areas Commun., to be published.

[6] N. B. M. Obaidat, Security of E-Systems and Computer Networks. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[7] A. Tolba, "Content accessibility preference approach for improving service optimality in Internet of vehicles," Comput. Netw., vol. 152, pp. 78–86, Apr. 2019.

[8] Z. Ning, K. Zhang, X. Wang, M. S. Obaidat, L. Guo, X. Hu, B. Hu, Y. Guo, B. Sadow, and R. Y. K. Kwok, "Joint computing and caching in 5G-envisioned Internet of vehicles: A deep reinforcement learning based traffic control system," IEEE Trans. Intell. Transp. Syst., to be published.

[9] H. Liu, B. Xu, D. Lu, and G. Zhang, "A path planning approach for crowd evacuation in buildings based on improved artificial bee colony algorithm," Appl. Soft Comput., vol. 68, pp. 360–376, Jul. 2018.

[10] M. Boudhtou and O. Klopfenstein, "Robust optimization for selecting NetFlow points of measurement in an IP network," in Proc. IEEE Global Telecommun. Conf. (GLOBECOM), Nov. 2007, pp. 2581–2585.

[11] H. Yang, A. Alphones, W.-D. Zhong, C. Chen, and X. Xie, "Learning-based energy-efficient resource management by heterogeneous RF/VLC for ultra-reliable low-latency industrial IoT networks," IEEE Trans Ind. Informat., to be published.

[12] Z. Ning, Y. Li, P. Dong, X. Wang, M. S. Obaidat, X. Hu, L. Guo, Y. Guo, H. Huang, and B. Hu, "When deep reinforcement learning meets 5G-enabled vehicular networks: A distributed offloading framework for traffic big data," IEEE Trans Ind. Informat., vol. 16, no. 2, pp. 1352–1361, Feb. 2020.

[13] H. Liu, B. Liu, H. Zhang, L. Li, X. Qin, and G. Zhang, "Crowd evacuation simulation approach based on navigation knowledge and two-layer control mechanism," Inf. Sci., vol. 436–437, pp. 247–267, Apr. 2018.

[14] Z. Ha, Y. Qiao, and J. Luo, ATME: Accurate traffic matrix estimation in both public and private datacenter networks," IEEE Trans. Cloud Comput., vol. 6, no. 1, pp. 60–73, Jan. 2018.

[15] Z. Ning, J. Huang, X. Wang, J. J. P. C. Rodrigues, and L. Guo, "Mobile edge computing-enabled Internet of vehicles: Toward energy-efficient scheduling," IEEE Netw., vol. 33, no. 5, pp. 198–205, Sep. 2019.

[16] S. Salsano, F. Patria, F. E. Preti, F. L. Ventre, and V. M. Gentile, "Accurate and efficient measurements of IP level performance to drive interface selection in heterogeneous wireless networks," IEEE Trans. Mobile Comput., vol. 17, no. 10, pp. 2223–2235, Oct. 2018.

[17] Y. Yu, L. Guo, J. Huang, F. Zhang, and Y. Zong, "A cross-layer security monitoring selection algorithm based on traffic prediction," IEEE Access, vol. 6, pp. 35382–35391, 2018.

[18] Y. Yu, Z. Ning, Q. Song, L. Guo, and H. Liu, "A dynamic cooperative monitor node selection algorithm in wireless mesh networks," in Proc. IEEE 17th Int. Conf. High Perform. Comput. Commun. 7th Int. Symp. Cyberspace Saf. Secur., IEEE 12th Int. Conf. Embedded Softw. Syst., Aug. 2015, pp. 1800–1805.

[19] X. T. Phan, I. D. Martinez-Casaneva, and K. Fukuda, "Adaptive and distributed monitoring mechanism in software-defined networks," in Proc. 13th Int. Conf. Netw. Service Manage. (CNSM), Nov. 2017, pp. 1–5.

[20] D.-H. Shin, S. Bagchi, and C.-C. Wang, "Toward optimal distributed monitoring of multi-channel wireless networks," IEEE Trans. Mobile Comput., vol. 15, no. 7, pp. 1826–1838, Jul. 2016.

[21] Z. Ning, P. Dong, X. Wang, J. J. P. C. Rodrigues, and F. Xia, "Deep reinforcement learning for vehicular edge computing: An intelligent offloading system," ACM Trans. Intel. Syst. Technol., vol. 10, no. 6, pp. 66, Oct. 2019, doi: 10.1145/3317572.

[22] Z. Yang and K. L. Yeung, "An efficient flow monitoring scheme for SDN networks," in Proc. IEEE 30th Can. Conf. Electr. Comput. Eng. (CCECE), Apr. 2017, pp. 1–4.

[23] T. Tian, W. Chen, and C.-T. Lea, "An SDN-based traffic matrix estimation framework," IEEE Trans. Netw. Service Manage., vol. 15, no. 4, pp. 1435–1445, Dec. 2018.

[24] D. Li, N. Dai, F. Li, C. Xing, and F. Dai, "Estimating SDN traffic matrix based on online informative flow measurement method," in Proc. 5th Int. Conf. Adv. Cloud Big Data (CBD), Aug. 2017, pp. 75–80.

[25] L. Huo, D. Jiang, X. Zhu, Y. Wang, and S. Singh, "A SDN-based fine-grained measurement and modeling approach to vehicular communication network traffic," Int. J. Commun. Syst., no. 9, p. e4092, 2019.

[26] H. Tahari, R. Saleh, S. Khan, R. Izard, K.-K.-R. Choo, and N. B. Anuar, "A multi-objective software defined network traffic measurement," Measurement, vol. 95, pp. 317–327, Jan. 2017.

[27] X. Wang, Z. Ning, X. Hu, L. Wang, L. Guo, B. Hu, and X. Wu, "Future communications and energy management in the Internet of vehicles: Toward intelligent energy-harvesting," IEEE Wireless Commun., vol. 26, no. 6, pp. 87–93, Dec. 2019.

[28] N. Dilokthanakul, C. Kaplanis, N. Pawlowski, and M. Shanahan, "Feature selection in hierarchical reinforcement learning," IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 11, pp. 3409–3418, Nov. 2019.

[29] A. Tolba, O. Said, and Z. Al-Makhadmeh, "MDS: Multi-level decision system for patient behavior analysis based on wearable device information," Comput. Commun., vol. 147, pp. 180–187, Nov. 2019.

[30] M. Malboubi, L. Wang, C.-N. Chuah, and P. Sharma, "Intelligent SDN based traffic (de)Aggregation and measurement paradigm (STAMP)," in Proc. IEEE Conf. Comput. Commun. (INFOCOM), Apr. 2014, pp. 934–942.

[31] L. Nie, D. Jiang, and L. Guo, "A compressive sensing-based approach to end-to-end network traffic reconstruction utilising partial measured origin-destination flows," Trans. Emerg. Telecommun. technol., vol. 26, no. 8, pp. 1108–1117, Apr. 2014.
[32] W. Xu, E. Mallada, and A. Tang, “Compressive sensing over graphs,” in Proc. IEEE INFOCOM, Apr. 2011, pp. 2087–2095.
[33] L. Nie, D. Jiang, and L. Guo, “A power laws-based reconstruction approach to end-to-end network traffic,” J. Netw. Comput. Appl., vol. 36, no. 2, pp. 898–907, Mar. 2013.

LAISEN NIE received the Ph.D. degree in communication and information system from Northeastern University, Shenyang, China, in 2016. He is currently an Associate Professor with the School of Electronics and Information, Northwestern Polytechnical University, Xi’an, China. His research interests include network measurement, network security, and cognitive networks.

HUIZHI WANG is currently pursuing the bachelor’s degree in communication engineering with Northwestern Polytechnical University. Her research interests include machine learning and network measurement.

XIN JIANG received the Ph.D. degree from the Tongji Medical College, Huazhong University of Science and Technology, in 2008. From 2009 to 2010, she was a Visiting Scholar with the Prince of Wales Hospital, The Chinese University of Hong Kong. From 2015 to 2016, she studied at the Global Clinical Scholars Research Training Program. She is currently the Chief Physician, an Associate Professor, and the Deputy Director of Geriatrics with Jinan University.

YI GUO received the Medical and M.Med. degrees from the Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, in 1985 and 1991, respectively, and the Ph.D. degree from the University of Greifswald, Greifswald, Germany, in 1997. He is currently the Chief of Neurology with The Second Clinical Medical College, Jinan University. His major research areas are cerebrovascular diseases, dementia, movement disorder diseases, sleep disorder, depression, and anxiety. He is a member of the Cerebrovascular Disease Group, Chinese Medical Association Neurology Branch, the Chairman of the Shenzhen Medical Association of Neurology, and the Shenzhen Medical Association of Psychosomatic Medicine.

SHENGTAO LI received the M.S. degree in operational research and cybernetics from Ludong University, China, in 2010, and the Ph.D. degree in control theory and control engineering from Northeastern University, China, in 2013. He is currently an Associate Professor with the School of Information Science and Engineering, Shandong Normal University. His research interests include nonlinear systems theory, optimal switch-time control theory of switched stochastic systems, and robust control of systems with time delay.