# An Adaptive Measurement Method for Flow Traffic in Software Defined Networking

Liuwei Huo[1], Dingde Jiang[2](✉), Xiangnan Zhu[2], and Huibin Jia[3]

[1] School of Computer Science and Engineering, NEU, Shenyang 110819, China
[2] School of Astronautics and Aeronautic, UESTC, Chengdu 611731, China
jiangdd@uestc.edu.cn
[3] School of Electrical and Electronic Engineering, NCEPU, Baoding, China

**Abstract.** In Software Defined Networking (SDN), the fine-grained measurements are crucial for network management and design. However, the measurement overhead and accuracy are contradiction, how to accurately measure the network traffic with low overhead has become a hot topic. Artificial Intelligence (AI) has been used to predict the traffic in networks. Then, we propose an AI-based Lightweight Adaptive Measurement Method (ALAMM) for traffic measurement in SDN with low overhead and high measurement accuracy. Firstly, we use measurements in the front to train the AI-based traffic prediction model and utilize the model to predict traffic in SDN. Then, we obtain the sequence of sampling points by judging the change of traffic prediction and send the measurement primitive to switches to obtain coarse-grained measurements. At last, we utilize the interpolation theory to fill the coarse-grained measurement and propose an optimization function to optimize the fine-grained measurement. Simulation results show that the ALAMM is feasible, and the measurement overhead of ALAMM is low.

**Keywords:** Software Defined Networking · Adaptive network measurement · Traffic matrix · Artificial Intelligence

## 1 Introduction

Accurate traffic measurement is the foundation for network planning and management. It not only displays the current status of networks but also helps operators manage networks to detect network failure and abnormal traffic. The network traffic measurement is the basis of network monitoring and management. With the expansion of network scale and the emergence of new network applications such as cloud computing, edge computing and big data, this poses a huge challenge to the management of the network. SDN decouples the data plane and control plane of the traditional switch, and centralizes the control plane into a controller for unified management, improving network scalability and management flexibility. The network traffic measurement of SDN is different from the traditional network.

In networks, there are some direct measurement methods, such as sFlow, NetFlow, they need the support of network devices and additional software, and consume a lot of storages and computing resource in network devices. In contrast to the measurement


H. Song et al. (Eds.): SIMUtools 2019, LNICST 295, pp. 115–124, 2019.
https://doi.org/10.1007/978-3-030-32216-8_11

scheme of traditional networks, SDN provides flow-based measurement methods by collecting the statistics from switches, this scheme is more convenient, efficient, and flexible. The pull-based scheme is an active measurement mode and the push-based scheme is a passive measurement mode. However, when the network scale and the number of active flows are very huge in SDN, the flow-based traffic measurement will face an enormous challenge due to a large number of flow statistics from switches and increase huge overhead in the network components. So we pay attention to the pull-based mechanism with low overhead in SDN.

Artificial Intelligence (AI) has been widely used in smartphones, voice recognition and authentication, which has been changed human's behavior patterns and lifestyle. AI is a junior intelligent system that requires some knowledge and reasoning to be added to the existing applications, database, and environment to make it friendlier, smarter and more sensitive to the environment. There is a large amount of data in the communication system, which provides rich history data for training the AI model. The application of AI in communication system has attracted the interest of many researchers [1]. Javier et al. have a comprehensive survey of AI-based optical networking, from low-level devices to high-level management [2]. AI in the optimal network not only improves the utilization of the wavelength but also improve management efficiency. Proietti et al. utilize machine learning-aided Quality of Transmission (QoT) estimation for lightpath configuration of intra-inter-domain traffic and obtain high accurate Optical Signal to Noise (OSNR) prediction [3]. Hagos et al. present a robust, scalable and generic machine learning-based method which may be of interest for network operators that experimentally infers congestion window and the underlying variant of loss-based TCP algorithms within flows from passive traffic measurements collected at an intermediate node [4]. Latah et al. investigated the application of AI to SDN paradigms, such as load balancing, network security, and intelligent network applications [5]. We also have some researches about the traffic matrix prediction and estimation with the deep learning in the data center network [6]. Our previous work can be found in [11–13].

Inspired by the AI-based traffic prediction and the adaptive flow traffic measurement, we propose an adaptive lightweight measurement scheme by predicting the traffic characteristics to measure the traffic effectively and accurately. ALAMM is pull-based active flow measurement. The main contribution of this paper as follows:

(1) We propose that using the measurement data in the front in the network to train the AI-based model.
(2) We use the trained AI-based model to predict the traffic in SDN and obtain the sampling points. Then, we use the sequence of sampling points to obtain the coarse-grained measurement.
(3) We use interpolation method to fill the coarse-grained measurement and construct an optimization function which has multiple constraints to optimal the obtained fine-grained measurement.
(4) We do some simulations to verify the performance of the proposed method.

The rest of this paper is organized as follows. Section 2 describes the measurement model ALAMM and introduces the adaptive sampling frequency and fine-grained interpolation and optimization of measurements. Section 3 makes simulations to verify the performance of ALAMM and the conclusion are stated in Sect. 4.

## 2 Problem Statement

In SDN, the control plane runs in the controller which is independent of switches. So, the flow-based measurement in SDN is much easier and more flexible than traditional networks, but the overhead of the measurement is a key issue that should be considered in the measurement process. We consider a simple mesh network with a controller $\{C_0\}$ and four switches $\{S_1, S_2, S_3, S_4\}$, as shown in Fig. 1. Each switch has at least one physical link which connects with the other switches. There are two flows $\{f_1, f_2\}$ in the network, $f_1$ through switches $(S_1, S_2, S_4)$, $f_1$ through switches $(S_1, S_2, S_3, S_4)$, and there are five physical links in the network $\{L_1, L_2, L_3, L_4, L_5\}$.

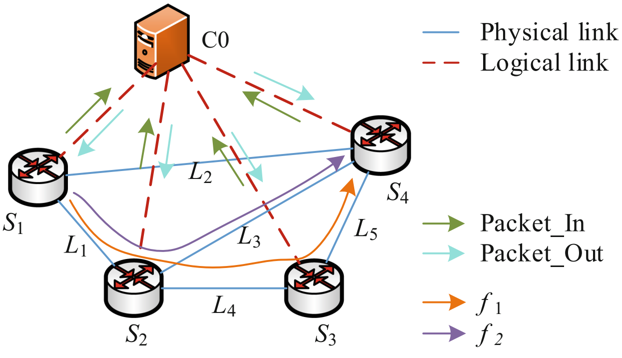


**Fig. 1.** The network topology and flows of SDN

Each switch connects into the controller directly in logical. There are two kinds of methods to deploy the controllers in SDN, in-band and out-of-band. In the In-band deployment scenario, the controllers are deployed inside the network, some switches directly connect into the controller. Control messages and data messages are exchanged between controllers and switches over the same network. In out-of-band deployment scenario, the controllers are external to the network and each switch is connected to the controller through a dedicated link. Data messages and control messages exchanged between the controllers and switches over different links. In out-of-band deployment scenario, controllers can exchange messages directly with switches. For simplicity, we consider an out-of-band scenario here.

### 2.1 Network Traffic

In this deployment scenario, each switch is directly connected to the controller, and exchange control messages through the control channel. The controller periodically sends LLDP packets to discover the links in the network, so the controller has a global view of the network topology. Flow is the traffic between each pair of the source node and the destination node, and the flow forwarding action in switches are programmed by the controller, so the controller knows all the routing information of the networks, we use $A$ to represent the routing matrix. $x$ and $y$ are the traffic of flow and link, respectively. So we can represent the traffic in the network as

$$y_j = \sum_i a_{ij} x_i \tag{1}$$

where $x_i$ is the traffic of flow $i$ and $y_i$ is the traffic on the link $j$. $a_{ij}$ is the route of flow $i$. If $a_{ij} = 1$, it means that the flow $i$ through the link $j$.

The traffic of flows and links in the network has a relationship that

$$\begin{cases} y_1 = x_1 + x_2 \\ y_3 = x_1 \\ y_4 = x_2 \\ \cdots \end{cases} \tag{2}$$

From Eq. (2), we have the relationship of traffic in switch $S_2$ as

$$y_{1+} - y_{3-} - y_{4-} + y_{2+} - y_{2-} \le \theta \tag{3}$$

where $\theta$ is the error threshold of flow traffic. $y_{2+}$ and $y_{2-}$ are the traffic of $S_2$ which transmitted from $S_2$ to $S_2$. $y_{2+}$ and $y_{1+}$ are the input traffic, $y_{2-}$, $y_{3-}$ and $y_{4-}$ are the output traffic. In each switch, the traffic meets the principle of conservation. Then we represent (3) the traffic of switch $k$ as

$$\left| \sum y_{k+} - \sum y_{k-} \right| \le \theta \tag{4}$$

The traffic in the network changes over time, so we can represent the traffic of links and flows in the network as

$$\begin{cases} x_i = \{x_i(t)\} = \{x_i(1), x_i(2), x_i(3), \ldots.\} \\ y_j = \{y_j(t)\} = \{y_j(1), y_j(2), y_j(3), \ldots.\} \end{cases} \tag{5}$$

where $x_i(t)$ is the traffic of flow $i$ at time $t$, $y_j(t)$ is the traffic of flow $i$ at time $t$.

The Eq. (1) can be rewritten as

$$y_j(t) = \sum_i a_{ij} x_i(t) \tag{6}$$

In the network, measuring all the traffic of links will consume much computing and transmit resource. Su et al. proposed the CeMon method which selects the subset of switches which coverage the most active flows [7], then we can sample the flow traffic in some switches to reflect all the traffic in the network. However, the traffic measurement in switches should last a long time, so a lot of measurement overhead will be generated in this process. Then, how to find the optimal sampling sequence with high measurement accuracy and low overhead become a key issue which should be studied. There are many methods, such as unified sampling, random sampling, but both of them has high measurement overhead and not flexible. In this paper, we proposed a lightweight measurement scheme, we train an AI model to predict the network traffic feature, and sample network traffic based on the prediction. Then, the sampling method to measure the network can be written as

$$\hat{x} = \sum_{i} x(t)\delta(t) \tag{7}$$

where $\delta(t)$ is a sampling sequence.

## 2.2 Adaptive Sampling

Flows with the features of high density and high dynamic bring about a huge challenge for the accurate, fast, and fine-grained traffic measurement. Through short time slot sampling, we can obtain the instantaneous rate of flows and links. However, the traditional flow-based fine-grained network measurements in SDN require the controller to frequently send Read-state messages to OpenFlow-based switches and also generate a large number of report messages to the controller, which would consume much computing resource of the controller. So, we use a coarse-grained measurement method to reduce measurement overhead.

ANN is one of the most widely used methods of AI, it is a popular model for solving the multi-dimension traffic prediction issues, such as network traffic, vehicle traffic. The structure of ANN is flexible, users can change the ANN structure based on their requirement. ANN consists of one input layer and N hidden layers and one output layer, it is a stack of many neurons. In addition to the input layer, each neuron is a weighted sum of the previous layer of neurons, the neurons in the hidden and output layer are statistics variables. So the traffic prediction model of ANN can be written as

$$\begin{cases} h_m(t) = F(\sum\limits_{n=1}^{N} w_{m-1n}h_{m-1n}(t)), m = 2, 3, \ldots, M \\ R_p(t) = F(\sum\limits_{n=1}^{N} w_{Mn}h_{Mn}(t)) \end{cases} \tag{8}$$

where $w_{mn}$ are weighted factors between neurons in a different layer, and $h_m(t)$ are the middle results. There are $M$ hidden layers in the ANN model. $F(\cdot)$ is the activation function of neurons, $R_p(t)$ are prediction results of network traffic at time slot $t$.

The traffic prediction $R_p(t)$ has the features of the traffics, so we can use the features to adaptively adjust the sampling points, and help us to improve the measurement accuracy and decrease the measurement overhead. Then, we design an adaptive sequence of the sampling points in the following, it can be written as

$$\delta(t) = \begin{cases} 1, R_p(t) - R_p(t-1)\Delta \ or \ R_p(t+T) - R_p(t) > \Delta \\ 0, otherwise \end{cases} \tag{9}$$

where $\delta(t)$ is a sampling point or not at time slot $t$, $R_p(t)$ is the traffic prediction at time slot $t$. If the change of flow prediction values is bigger than the threshold $\Delta$, we send a sampling primitive; otherwise, we think the flow is stable, and not send the sampling primitive; $T$ is a fixed period which is used to ensure the maximum sampling interval not exceed $T$.

### 2.3 Fine-Grained Matrix Filling and Optimization

The fine-grained measurement result of flow $j$ is $x_i$, it is obtained by filling the coarse-grained measurement with the cubic interpolation method, and the actual flow traffic of flow $j$ is $\hat{x}_i$. Due to the cubic interpolation is the smoothest possible approximations of actual flow traffic, so there is a gap between the measurement results and the actual flow traffic. In order to obtain accurate measurements, we optimize the filling data to decrease the gap between the measurement results and the actual results of flows. Then, we propose an optimal function as follows:

$$\begin{cases} \min \quad \left\|Y - A\hat{X}\right\|_2 + \lambda\|X\|_2 \\ \quad s.t. \\ \qquad C1: \quad Y_i \geq A_i\hat{X}, \quad A = (A_1, A_2, \ldots, A_M)^T \\ \qquad C2: \quad x_j \geq 0, Y_i \geq 0 \\ \qquad C3: \quad \left|\sum x_{i+} - \sum x_{i-}\right| \leq \theta \\ \qquad C4: \quad \sum x_{i+} - \sum x_{i-} > 0 \end{cases} \tag{10}$$

where $\lambda$ is a Lagrange multiplier. Contrast to Eq. (10), the equation above can easily be solved by the following heuristic algorithm proposed in this paper.

Constraint C1 represents the constraint between link load and flow traffic; Constraint C2 denotes that the traffic in the network is non-negative. Constraint C3 and C4 means that the output traffic of node $i$ is no more than input traffic of node $i$, this is the traffic conservation principles. Under constraint $C1$, we know that link load and flow traffic mapping relationship matrix $A$ has $M$ rows and $N$ columns, and $M \geq N$ when multiple flows transmission through a link. Then, the routing matrix $A$ is an underdetermined matrix, there are infinite solutions for the linear constraint $C1$. Then, we use some heuristic method to solve it.

## 3 Simulation Result and Analysis

We evaluate the performance of the proposed measurement scheme by building a SDN test platform. In the simulation scenario, we use Ryu [8] as the controller and utilize Mininet [9] to construct the network topology. For simplicity, the network topology as

Fig. 1 shows. Iperf is used to generate TCP packets to fill each link from origin host to destination host, and all the links in the network are set as the duplex transmission mode. We analyze the traffic of flow f1 and f2, and compare the ALAMM to uniform sampling method under different intervals (Uniform60, Uniform240) and Principal Component Analysis (PCA) method, where Uniform60 and Uniform240 are the uniform sampling method with the sampling interval 60 and 240 slots, respectively. It is well known that the measured granularity is usually inversely proportional to the measurement interval. For the uniform sampling method under different sampling interval, when the sampling interval is small, we think it as the fine-grained measurement, and the sampling interval is big as the coarse-grained.

Relative Errors (RE) and Root Mean Square Error (RMSE) are mainly used parameters depict the performance of the methods [10]. For the ALAMM proposed, the sampling sequence is very important traffic measurement. Figure 2 shows the average RE of measurement under different threshold and interval of the measurement step. We find that when the interval and measurement interval are both small, the average RE of measurement is very small. When the sampling interval is larger than 150 slots, the average measurement RE trends to stable, but when the interval is larger than 200 slots, the average measurement RE becomes fluctuating. In addition, as the measurement threshold increases, the average RE also increases. When the measurement threshold is larger than 200, the average RE becomes fluctuating. Then, in the following, we use the measurement threshold and interval as 50 and 200, respectively.

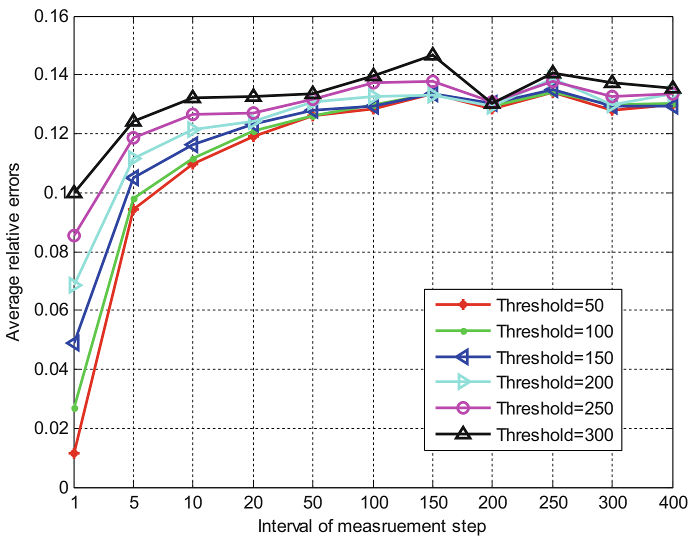


**Fig. 2.** The average RE under different threshold and interval of the measurement step

Figure 3 shows the RE cumulative distribution function (CDF) of flow f1 and f2, respectively. From Fig. 3(a), we can see that the about 80% RE of the ALAMM and Uniform sampling method are less than 0.3, while the RE is about 60% for the PCA method. Figure 3(b) has a similar trend with Fig. 3(a). However, the average RE of ALAMM is smaller than Uniform60 and larger than Uniform240 and PCA, the curves show that the ALAMM is slightly inferior to Uniform60 but better than Uniform240 and PCA. Figure 3(b) shows that the ALAMM is better than Uniform60, Uniform240, and PCA. Figure 4 curves the RMSE of different measurement methods, we note that RMSE of the ALAMM is much steeper than the other methods, this means the average measurement RE of ALAMM is much more stable than other methods. As well as, we find that the RMSE of ALAMM is close to the performance of Uniform60.

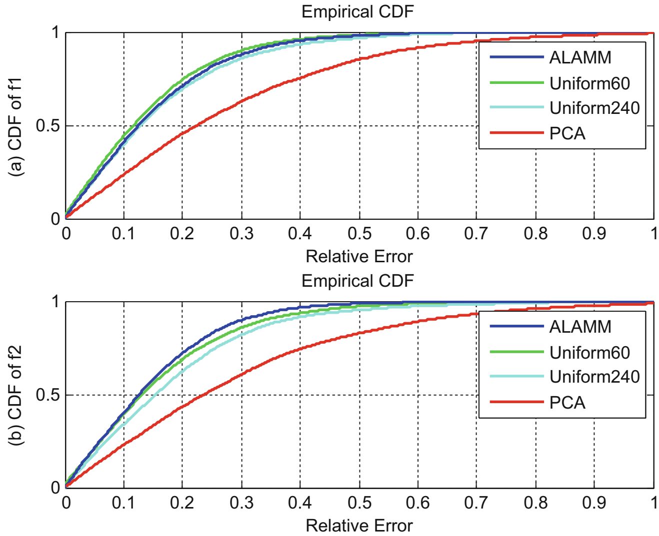


**Fig. 3.** The RE CDF of the flows f1 and f2

We compare the measurement overhead and the measurement error in Fig. 5. The bars are the measurement overhead, the left y-axis is their scale. The star on the blue line is the measurement errors of the corresponding measurement method, and the right y-axis is their scale. From Fig. 5, we note that the measurement errors of ALAMM are a little larger than the Uniform60 and smaller than the Uniform240, however, the measurement overhead of ALAMM is similar as the Uniform240 and far less than Uniform60. We know that ALAMM has lower overhead than Unifrom60, however, the average measurement errors are similar to it. For PCA, its average measurement error is about 0.35, which are larger than the ALAMM, Uniform60, and Uniform240. Although the measurement overhead of PCA is almost zero, its measurement performance is poor. Through the above analysis, we know that our ALAMM is feasible and it accurately measures the network traffic with low overhead.

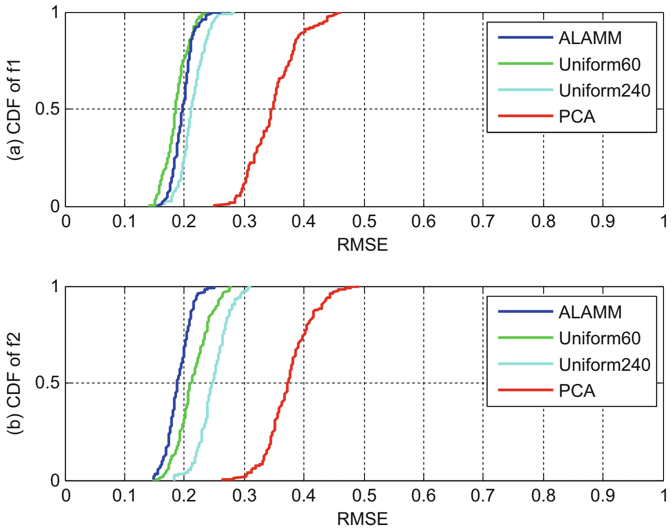


**Fig. 4.** The RMSE CDF of the flows f1 and f2

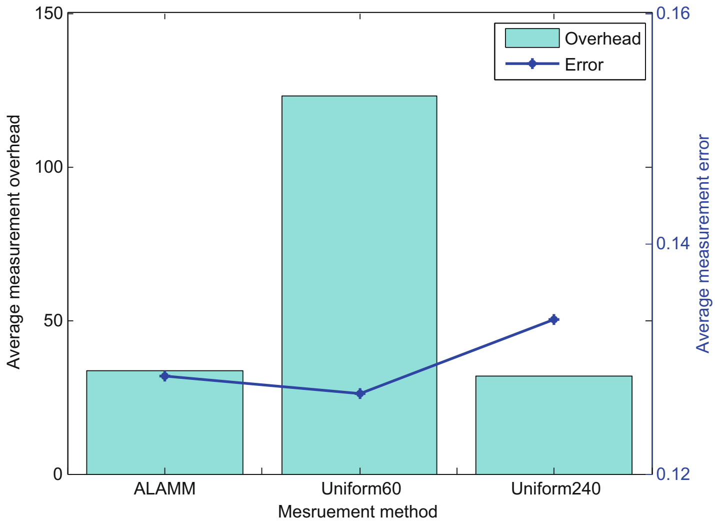


**Fig. 5.** The overhead and average error for different measurement methods

## 4 Conclusions

Accurate flow-based network measurement has a great impact on network traffic management in SDN. We propose the ALAMM for traffic measurement in SDN. In ALAMM, we use measurement results in the front to train the AI model, then use the model to predict the traffic in the network. Then, we obtain the sequence of sampling points based on prediction results and send sampling primitives to switches to obtain the coarse-grained measurement. Then, we perform the interpolation method on the coarse-grained measurement and utilize the optimization method to decrease the fine-grained measurement errors. At last, we make some simulations to verify the measurement method proposed in this paper. The simulation results show that the proposed ALAMM can accurately measure the traffic with low overhead.

**Acknowledgment.** This work was supported by the National Natural Science Foundation of China (No. 61571104), Sichuan Science and Technology Program (No. 2018JY0539), Key projects of the Sichuan Provincial Education Department (No. 18ZA0219), Fundamental Research Funds for the Central Universities (No. ZYGX2017KYQD170), and Innovation Funding (No. 2018510 007000134). The authors wish to thank the reviewers for their helpful comments.

## References

1. Li, Y., Zhang, Y., et al.: Ultra-dense HetNets meet big data: green frameworks, techniques, and approaches. IEEE Commun. Mag. **56**(6), 56–63 (2018)
2. Mata, J., de Miguel, I., Duran, R.J., et al.: Artificial intelligence (AI) methods in optical networks: a comprehensive survey. Opt. Switch. Netw. **28**, 43–57 (2018)
3. Proietti, R., Chen, X., Zhang, K., et al.: Experimental demonstration of machine-learning-aided QoT estimation in multi-domain elastic optical networks with alien wavelengths. J. Opt. Commun. Netw. **11**(1), A1–A10 (2019)
4. Hagos, D., Engelstad, P., Yazidi, A., et al.: General TCP state inference model from passive measurements using machine learning techniques. IEEE Access **6**, 28372–28387 (2018)
5. Latah, M., Toker, L.: Application of artificial intelligence to software defined networking: a survey. Indian J. Sci. Technol. **9**(44), 1–7 (2016)
6. Jiang, D., Wang, W., Shi, L., et al.: A compressive sensing-based approach to end-to-end network traffic reconstruction. IEEE Trans. Netw. Sci. Eng. **5**(3), 1–12 (2018)
7. Su, Z., Wang, T., Xia, Y., Hamdi, M.: CeMon: a cost-effective flow monitoring system in software defined networks. Comput. Netw. **92**, 101–115 (2015)
8. The Ryu Platform. https://github.com/osrg/ryu/. Accessed Dec 2017
9. The Mininet Platform. http://mininet.org/. Accessed Dec 2017
10. He, Q., Wang, X., Huang, M.: OpenFlow-based low-overhead and high-accuracy SDN measurement framework. Trans. Emerg. Telecommun. Technol. **29**(2), e3263 (2018)
11. 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. **1**(1), 1–12 (2018)
12. Jiang, D., Huo, L., Li, Y.: Fine-granularity inference and estimations to network traffic for SDN. PLoS ONE **13**(5), 1–23 (2018)
13. Jiang, D., Huo, L., Lv, Z., et al.: A joint multi-criteria utility-based network selection approach for vehicle-to-infrastructure networking. IEEE Trans. Intell. Transp. Syst. **PP**(99), 1–15 (2018)