A Security Traffic Measurement Approach in SDN-Based Internet of Things

Liuwei Huo[1], Dingde Jiang[2(✉)], and Hui Qi[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 Computer Science and Technology, CUST, Changchun 130022, China

Abstract. In the Internet of things (IoT), a large amount of data are exchanged through IoT networks between devices and cloud computing. However, the legacy architecture of IoT networks is not flexible and scalable for the increment of devices. Software defined networking (SDN) separates the control plane from the data plane in the legacy switches and centralizes the control plane as a logical control center, making network management more flexible and efficient. In SDN, the controller is very easy to be attacked, then, we use the Blockchain technology into the measurement framework to ensure the security and consistency of the statistics. To obtain the measurement results with low overhead and high accuracy, we collect the statistics of coarse-grained traffic of flows and fine-grained traffic of links and estimate the flow traffic with an ARIMA model. We propose an objective function to decrease the estimation errors. The objective function is an NP-hard problem, we present a heuristic algorithm to obtain the optimal solution of the fine-grained measurement. Finally, some simulations are performed to verify the validity of the proposed scheme.

Keywords: Internet of things · Software defined networking · Network measurement · Traffic matrix

1 Introduction

The Internet of things (IoT) connects many kinds of devices in urban areas, such as transportation, schools, and hospitals. A huge amount of data are integrated into cloud computing through the IoT network which is designed to provide the highest possible degree of flexibility, scalability, and security to all interconnected entities [1]. Cloud computing is a universal computing platform which has powerful computing, it supports the operation of different applications at the same time [2]. There are large amounts of data in networks should be processed and exchanged, so the network quality monitoring is very important for network maintenance.

In the legacy IoT network, its architecture is not scalable and not satisfy the requirement of the increment devices. Software Defined Networking (SDN) intends to simplify network management and improve the flexibility of the IoT. SDN separates the control plane from the underlying forwarding device and integrates the control

H. Song et al. (Eds.): SIMUtools 2019, LNICST 295, pp. 146–156, 2019.
https://doi.org/10.1007/978-3-030-32216-8_14

plane into the logically center to simplify network management and dynamically configure network rules [3]. In SDN, the controller has a global view of the networks, so the controller can program the global optimization rules for the traffic dispatching. SDN provides a novel flow-based statistical measurement method, it is very flexible and convenient to collect the traffic statistics information from switches.

SDN has one of the most important drawbacks is its increased attack surface compared to traditional networking deployments and the increased effect any successful attack will have. In distributed systems, Blockchain is a means of ensuring data security, reliability, and transparent exchange and storage, ensuring the security and consistency by having all participants share ledger. If the controller is hijacked or some nodes masquerade as controller nodes and send improper flow forwarding rules or malicious read switch information to the network, it may cause erroneous traffic or network storms in the network, causing the entire network to crash. In [4], Sharma et al. provide to use Blockchain to increase the security SDN. The information in the black chain cannot be modified unless all participants have reached a consensus to modify the ledger, this is because that all the participants share the same information in private ledger to maintain the consistency and completeness of the information. Traffic modeling [10, 11], traffic estimation [12], network selection [13], energy efficiency [14] and network behaviors [15] are studied in previous work.

According to the analysis above, we propose a pull-based and flow-based network traffic measurement in IoT network with lower measurement overhead. We measure some data of the network traffic directly and estimate the fine-grained network traffic. Then, we propose an objective optimization model to decrease the fine-grained measurement error inferred and present a heuristic algorithm to seek the optimal solution of the model. Our main contributions in this paper are as follows:

(1) We propose a framework for network traffic measurement in the IoT networks. To ensure the security and consistency of the statistics, we introduce the Blockchain technology into the measurement framework of SDN-based IoT networks. To obtain the measurement results with low overhead and high accuracy, we collect the statistics of coarse-grained traffic of flows and fine-grained traffic of links.
(2) We model the network traffic as an ARIMA model and forecast the flow of traffic with the coarse-grained flow measurement. Then, we propose an objective function to decrease the estimation errors.
(3) The objective function is an NP-hard problem, we present a heuristic algorithm to obtain the optimal solution of the fine-grained measurement. Finally, we conduct some simulations to verify the validity of the proposed measurement scheme.

The rest of this paper is organized as follows: Sect. 2 states the main security challenges of SDN networking, then we provide a fine-grained measurement scheme with Blockchain scheme, and describes the fine-grained traffic estimation and optimization in the IoT paradigm. Section 3 presents the simulation of the performance of the fine-grained measurement. Finally, Sect. 4 concludes our works in this paper.

2 Problem Statement

Cloud computing frequently requests the resource from cloud computing to decrease the network. Network measurements such as load balancing, path planning, and anomaly detection are required. In this section, we propose a cloud computing network measurement architecture based on SDN architecture.

2.1 System Model

In the IoT, a large number of devices are typically dynamically connected to the network, which requires the network to be very flexible and scalable. The flow-based measurement in SDN is much easier and more flexible than the legacy network. Network measurement is the key of the network management, then, we propose an SDN-based network measurement architecture for cloud computing networks that uses a sampling method to obtain coarse-grained measurements, then uses estimation and optimization methods to recover fine-grained measurements. In this architecture, we use Blockchain approach to ensure the security of information exchanged between the controller and switches. The measurement components are installed into the controller and are compatible with the other existing software defined measurement frameworks. The key technologies for the novel measurement architecture include coarse-grained measurements, traffic modeling and estimation, and traffic optimization. The flow-based coarse-grained measurements in SDN can be obtained by collecting the flow statistics in OpenFlow-based switches; the traffic matrix consists of link load, flow traffic, and the routing matrix, which reflects the traffic in the network, and has been widely researched in traffic engineering; estimation is a common method of data estimation; and the optimization methods are widely used to find the best solution for the complex issues.

2.2 Blockchain Application

In recent, most evidence preservation systems which based on a centralized repository structure have serious security issues. Centralized structures always require strong safety requirements. In SDN, the controller becomes a critical node, it will lead to serious consequences for the whole network once the controller is hijacked. For example, once a malicious application is installed in the controller or a malicious node inside the network masquerades as the controller, malicious traffic will constantly be generated in IoT networks, which takes up a large amount of bandwidth, causing link congestion and making the request of the legitimate node unresponsive, even bringing down the entire network. In order to ensure the authenticity and credibility of the network statistics collected by the controller, we propose to combine Blockchain technology with SDN in IoT networks. In SDN, all the device are accountants and update the data from the controller, and each accountant maintains a ledger. Each time data is written to the ledger, then a block is created. The header of each block contains the hash value of the previous block, blocks linked together forming a chain. Trusted timestamps can be immediately attached to newly created blocks. Most importantly, it

is possible to avoid trust issues by spreading the authority of the auditor. It demonstrates the integrity, accuracy, and timeliness required for preservation.

A special number needs to be attached to the end of each block and use a hash function to generate a hash value which is difficult to decipher, as Fig. 1 shows. So, the output data is completely unpredictable unless you know the full input data and the hash function. The node who first get proof of work will broadcast the block with its proof of work, and all other nodes in the network will append this block to their own Blockchain. So if a malicious node in the network tries to tamper the information and broadcast a fake block. So, there are two different blocks will be exchanged in the network, and the Blockchain will branch in a short time. The admissibility rule for Blockchain is always to trust the longest Blockchain. However, the Blockchain branch of the malicious node is only maintained by partial nodes, while the real branch chain is maintained by most of all nodes in the network together except malicious nodes. Based on this data structure, if a malicious node attempts to modify a previous block, the hash value of that block changes. In order to make subsequent blocks connect to it, all subsequent blocks must be modified in turn, otherwise, the altered block will not be accepted by other nodes. The Blockchain technology ensures that the network statistics information collected by the controller in SDN is truly unimpeded, which greatly improves the security of the SDN network.

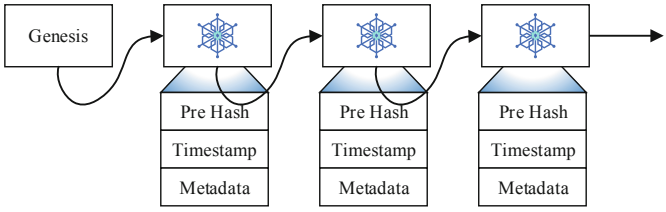

Fig. 1. The Blockchain model

2.3 Traffic Matrix and ARMA Model Construction

Origin-Destination (OD) traffic refers to traffic between any two nodes in the network and describes the distribution of network traffic between OD pairs. The flow traffic can be calculated based on the measured link traffic and the network routing matrix. The relationship can be expressed as a linear equation:

$$Y = AX \tag{1}$$

where Y is a column vector representing link traffic, X is also a column vector representing the traffic matrix, and A is the routing matrix. The problem of flow calculation is an inverse problem-solving of an underdetermined and ill-conditioned system.

In IoT networks, the volume of large amounts of data in IoT networks is very small and require low latency, and the flow traffic fluctuates sharply, but the flow traffic in IoT networks can be represented as time series. By studying the time series and discovering changes of flows, the modeling method belongs to the field of time series

analysis. The Autoregressive Integrated Moving Average Model (ARIMA) model is the most commonly used time series analysis model [5]. The ARIMA model is the extension of the ARMA model, and ARIMA model is mainly used for non-stationary time series modeling. The ARMA model can be written as

$$\begin{aligned} x(t) = \delta + \phi_1 x(t-1) + \phi_2 x(t-2) + + \cdots + \phi_p x(t-p) \\ + u(t) + \theta_1 u(t-1) + \theta_2 u(t-2) + \cdots \theta_q u(t-q) \end{aligned} \tag{2}$$

In time series analysis, the Lag operator (L) on a value of a time series to produce the previous value. So the traffic with the lag operator can be written as

$$x(t-k) = L^k x(t) \tag{3}$$

where L is the Lag operator, and k is the lag order. So, the ARIMA(p, d, q) model with the Lag operator can be written as

$$\begin{aligned} x(t) &= \delta + \phi_1 x(t-1) + \phi_2 x(t-2) + \cdots + \phi_p x(t-p) + u(t) \\ &\quad + u(t) + \theta_1 u(t-1) + \theta_2 u(t-2) + \cdots \theta_q u(t-q) \\ &= \delta + \sum_{i=1}^{p} \phi_i L^i x(t) + u(t) + \sum_{j=1}^{q} \theta_j L^j u(t) \\ &= \delta + \sum_{i=1}^{p} \phi_i L^i x(t) + (1 + \sum_{j=1}^{q} \theta_j L^j) u(t) \end{aligned} \tag{4}$$

where $x(t)$ is the predictive value; $x(t-i)(i = 1, 2, \ldots, p)$ are the actual values; $\phi_i(i = 1, 2, \ldots, p)$ and $\theta_j(j = 1, 2, \ldots, q)$ are the parameters of autoregressive coefficients and random errors coefficients, respectively; $u(t-j)(j = 0, 1, 2, \ldots, q)$ are the Gaussian white noise, their means and constant variance of 0 and σ^2, respectively; δ is a constant; p and q are the orders of the AR model and MA model, respectively. Then, the ARIMA model with d-order difference of data series can be expressed as

$$(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - L)^d x(t) \; = \; \delta + (1 + \sum_{j=1}^{q} \theta_j L^j) u(t) \tag{5}$$

where d is the difference order; $x(t)$, $u(t)$ are the estimation of the traffic and random errors following Gaussian distribution, respectively. ϕ, θ are the coefficients of autoregressive and random errors, respectively; δ is a constant.

In the ARIMA model, the d-order difference of data series is an ARMA model, then we use ARMA model to estimate the network traffic with the time series $(1 - L)^d x(t)$ instead of the original network traffic series $x(t)$. However, estimation results of flows have big errors with the actual flow traffic. In the edge network, the link load reflects the integrated traffic transmission in the network. So, we use the pull-based method to obtain the fine-grained link load Y in networks. The traffic of flows in the network can be written as

$$f = \alpha\left\|Y - A\hat{X}\right\|_2 + \beta\left\|\hat{X}\right\|_2 \tag{6}$$

where A is the routing matrix, α and β are the weight coefficient of the measurement results. In order to decrease the deviation between the estimations and the actual traffic results, we construct an objective function to optimize the estimation results. The objective function as the function (7) shows.

$$\begin{cases} \min\alpha\left\|Y - A\hat{X}\right\|_2 + \beta\left\|\hat{X}\right\|_2 \\ s.t. \\ \quad C1 : Y{=}A\hat{X} \\ \quad C2 : \hat{X} \geq 0 \\ \quad C3 : Y_m \geq \sum\limits_n a_{mn}\hat{X} \\ \quad C4 : \sum\limits_{i=1}^{N} x_{ij} = \sum\limits_{j=1}^{N} x_{ji} \end{cases} \tag{7}$$

where $\hat{X}$ is estimated by ARIMA. Constraint $C1$ represents the constraint between link load and flow traffic; constraint $C2$ shows the link load is non-negativity; constraint $C3$ is the limitation of flows on each link; constraint $C4$ represents that the traffic that input and output of the switch are constant, i is the source node and j is the destination node. In the IoT networks, the routing matrix A has M rows and N columns. However, the OD pairs are much larger than links, namely: $M \ll N$, then the routing matrix A is an underdetermined matrix, there are infinite traffic matrices X which satisfy the constraint $C1$. The objective function (7) is an NP-hard problem and is difficult to solve directly. Then, we use a heuristic method to solve it.

2.4 Ant Colony Algorithm

The ant colony algorithm (AC) is a probabilistic algorithm for solving the hard optimization problem [6], and it is inspired by the utilization of pheromone as a communication medium to find the optimal path in the food search process. Pheromone update is the core of AC. Ant colony algorithm process is as follows:

(a) Initialization: Initializing the control parameters, the maximum number of iterations, and the pheromone concentration value for each candidate.
(b) Construction of solutions: Each ant starts from its own start, and the go through each node in the network. Suppose that the pheromone concentration corresponding to the *ath* candidate traffic value of the *jth* variable is $\tau(i,j)$. When pheromone concentration is updated, the pheromone concentration value corresponding to each candidate traffic of each OD pair in the network will also be updated. The ant selects the candidate traffic according to the following equation:

$$S_i = \begin{cases} \arg(\max[\tau(i,j)]), & q \leq q_0 \\ \dfrac{\tau(i,j)}{\sum\limits_{k=1}^{N}\tau(i,j)}, & q > q_0 \end{cases} \tag{8}$$

(c) Pheromone concentration update: When ants pass through the solution space of traffic, change the pheromone concentration value corresponding to the candidate traffic value (s) selected by the ant in the passing traffic. The update rule of local pheromone concentration is

$$\tau(i,j) = (1-\rho)\tau(i,j) \tag{9}$$

(d) Global pheromone concentration update: After all ants complete the pathfinding process (G), updating the pheromone concentration in the network according to the global pheromone concentration updating rule which can be expressed as

$$\tau_k(j) = (1-\alpha)\tau_k(j) + \Delta\tau_k(j) \tag{10}$$

$$\Delta\tau_k(j) = \begin{cases} \alpha/f_{best}, & optimal \\ -\alpha'/f_{worst}, & worst \\ 0, & others \end{cases} \tag{11}$$

(e) Analysis and calculation: The solutions constructed by ants need to be analyzed, and after all ants have constructed solutions, it is necessary to judge whether these solutions are new ones. If it is a new path, the corresponding target function value is calculated. Otherwise, the function value will be set to given default values.

(f) Looping: Performing the loop from step (b) to (e) until the termination condition is reached.

3 Simulation Result and Analysis

3.1 Simulation Environment

In order to evaluate the performance of the measurement scheme proposed, we built an SDN test platform and wrote the measurement module in python. We use Ryu [7] as the controller and simulate the switches, hosts and links by Mininet [8]. We generate the flows in the network by Iperf and set the maximum link load as 10Gbps.

To verify the performance of the measurement scheme proposed, we introduce the Relative error (RE) and Relative Error (RE). The AE reflects the deviations between the actual traffic and the measurement results. The smaller the AE, the more accurate the measurement. The RE is the ratio of measured AE to the actual value, reflecting the credibility of the measurement. Root Mean Squared Error (RMSE) is usually used to evaluate the measurement accuracy in simulations.

3.2 Simulation Evaluation

We use Iperf to generate packets to fill each link from origin host to destination host in the network and measure all the flow traffic with different methods. We randomly choose two flows as an example for discussion. In this paper, we combine the ARIMA model and AC model to estimate and optimize the fine-grained network traffic and use

ARIMA-AC to it. Then, we compare ARIMA-AC to Fine-grained measurement (FGM), the Principal Component Analysis (PCA) and SRSVD [9].

Figure 2 shows the flow of traffic for different estimation method. The red line is the actual traffic generated by Iperf. The blue line, pink line, and green line represent the measurement methods ARIMA-AC, SRSVD and PCA, respectively. We notice that the ARIMA-AC, SRSVD and PCA measurement methods all reflect the change of flows f2 and f3. The PCA fluctuates more sharply than that of other methods. Figures 3 and 4 show the AE and RE cumulative distribution function (CDF) of flow f2 and f3, respectively. From Fig. 3, we can see that the about 80% AE of the ARIMA-AC, SRSVD are less than 1000 Mbps in the 40 min measurement process, while the performance of PCA method is the worst. However, the average AE of the ARIMA-AC is smaller than that of SRSVD and PCA. Figure 4 is the RE of different measurement methods, and we find that about 80% AE of the ARIMA-AC, SRSVD is less than 0.2 in the 40 min measurement process, which means that the network traffic measurement is stable and effective. The performance of PCA is the worst of four measurement methods.

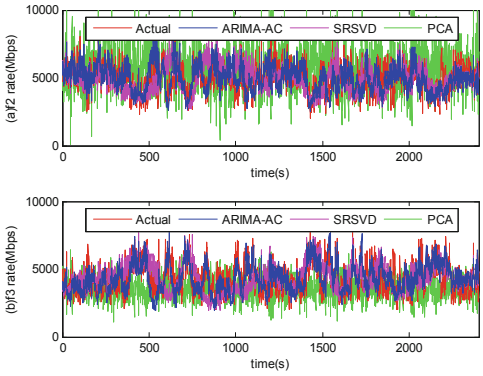

Fig. 2. The network traffic measurement of different methods. (Color figure online)

Figure 5 shows the curve of CCDF (Complementary Cumulative Distribution Function) of RMSE for the different measurement methods. CCDF reflects the RE RMSE of the measurement error, and it also shows that when RMSE is bigger than 0.2 of ARIMA-AC, the measurement error is very small. The performance of ARIMA-AC is better than that of ARSVD and PCA, and a little worse than that of FGM. So, when the measurement accuracy requirement is no more than 20%, the ARIMA-AC is feasible. In the measurement process, we just measure the coarse-grained flow traffic and the fine-grained link traffic, there is smaller overhead than that of FGM.

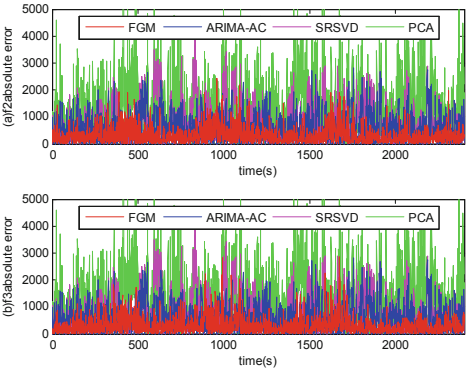

Fig. 3. The AE of different methods.

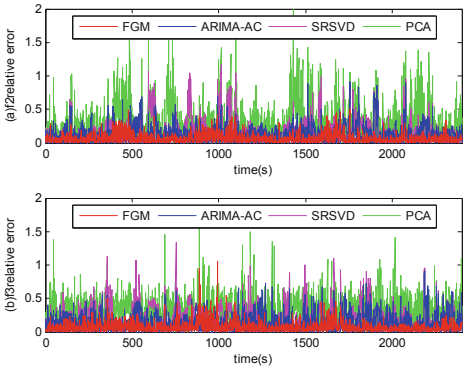

Fig. 4. The RE of different methods.

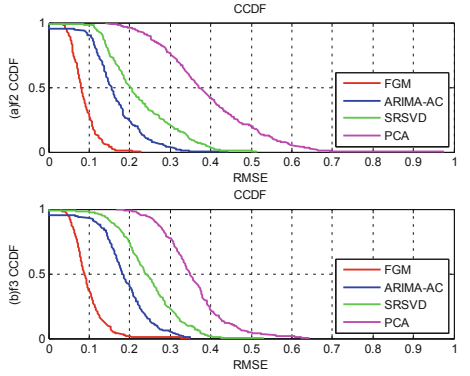

Fig. 5. The CCDF of RMSE of different methods.

4 Conclusions

In this paper, we propose a novel low-overhead measurement scheme and construct the measurement architecture of IoT network. Security is one of the most important issues in SDN, then we introduce the Blockchain to ensure the security and consistency of the statistics exchanged between the controller and switches. Then, we measure the coarse-grained traffic of flows and fine-grained traffic of links. In order to obtain the fine-grained flow traffic, we model the network traffic as an ARIMA model and forecast the network traffic, and propose an objective function to decrease the estimation errors. The objective function is a NP-hard problem, we use a heuristic algorithm to obtain the optimal solution. At last, we perform some simulations to verify the measurement architecture and scheme proposed in this paper.

Acknowledgment. This work was supported by 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. 2018510007 000134). The authors wish to thank the reviewers for their helpful comments.

References

1. Ray, P.P.: A survey on Internet of Things architectures. J. King Saud Univ. Comput. Inf. Sci. **30**(3), 291–319 (2018)
2. Li, S., Xu, L., Zhao, S.: 5G Internet of Things: a survey. J. Ind. Inf. Integr. **10**, 1–9 (2018)
3. Xu, H., Yu, Z., Qian, C., et al.: Minimizing flow statistics collection cost of SDN using wildcard requests. In: Proceedings OF INFOCOM 2017, pp. 1–9 (2017)
4. Sharma, P.K., Singh, S., et al.: Distblocknet: a distributed Blockchains-based secure SDN architecture for IoT networks. IEEE Commun. Mag. **55**(9), 78–85 (2017)
5. Unnikrishnan, J., Suresh, K.K.: Modelling the impact of government policies on import on domestic price of Indian gold using ARIMA intervention method. Int. J. Math. Math. Sci. **2016**, 1–6 (2016)
6. Liu, J., Yang, J., Liu, H., et al.: An improved ant colony algorithm for robot path planning. Soft. Comput. **21**(19), 5829–5839 (2017)
7. The Mininet Platform. http://mininet.org/. Accessed Dec 2018
8. The Ryu Platform. https://github.com/osrg/ryu/. Accessed Dec 2018
9. Roughan, M., Zhang, Y., et al.: Spatio-temporal compressive sensing and internet traffic matrices. IEEE/ACM Trans. Netw. **20**(3), 662–676 (2012)
10. Huo, L., Jiang, D., Zhu, X. et al.: An SDN-based fine-grained measurement and modeling approach to vehicular communication network traffic. Int. J. Commun. Syst. 1–12 (2019)
11. 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)
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)

14. Jiang, D., Zhang, Y., Song, H., et al.: Intelligent optimization-based energy-efficient networking in cloud services for multimedia big data. In: Proceedings of IPCCC 2018, pp. 1–6 (2018)
15. Jiang, D., Huo, L., Song, H.: Understanding base stations' behaviors and activities with big data analysis. In: Proceedings Globecom 2018, pp. 1–7 (2018)