FlowCGAN: Exploratory Study of Class Imbalance for Encrypted Traffic Classification Using CGAN

Pan Wang*, Shuhang Li*, Feng Ye†, Zixuan Wang* and Moxuan Zhang‡
*School of Modern Posts, Nanjing University of Posts & Telecommunications, Nanjing, China
†Department of Electrical & Computer Engineering, University of Dayton, Dayton, OH, USA
‡Schools of International Education, Jinling Institute of Technology, Nanjing, China
Email: *wangpan@njupt.edu.cn, *lish@runtrend.com.cn, †fye001@udayton.edu, *wangzx@runtrend.com.cn
, ‡zhangmoxuan_7@126.com

***Abstract*—With more and more adoption of Deep Learning (DL) in the field of image proessing, computer vision and NLP, researchers have begun to apply DL to tackling with encrypted traffic classification problems. Although these methods can automatically extract traffic features to overcome the difficulty of traditional classification methods like DPI in terms of feature engineering, a large amount of data is needed to learn the characteristics of various types of traffic. Therefore, the performance of classification model always significantly depends on the quality of dataset. Nonetheless, the building of dataset is a time-consuming and costly task, especially encrypted traffic data. Apparently, it is often more difficult to collect a large amount of traffic samples of those unpopular encrypted applications than well-known ones, which leads to the problem of class imbalance between major and minor encrypted application in dataset. In this paper, we proposed traffic data augmentation method called FlowCGAN using Conditional GAN, one of a genre of Generative Adversarial Network (GAN) . As a generative model, FlowCGAN takes the advantage of GAN to generate new traffic samples by learning the characteristics of the traffic data and thereby balancing between major and minor classes of the data set. To verify the feasibility and evaluate the performance of this method, Convolutional Neural Networks(CNN) was designed to classify three datasets: the original unbalanced dataset, the dataset based on Random Over Sampling (ROS) method and the dataset based on FlowCGAN respectively. The experimental evaluation results show that the FlowCGAN method can achieve better performance than the other two in terms of encrypted traffic classification.**

***Index Terms*—encrypted traffic classification, deep augmentation, Conditional Generative Adversarial Network, traffic identification, class imbalance**

I. Introduction

With the rapid development of network technology, the types and quantity of traffic data in cyberspace are increasing. Network traffic identification and classification is a crucial research task in the area of network management and security. It is the footstone of dynamic access control, network resources scheduling, content based billing, intrusion and malware detection etc. High efficient and accurate traffic classification is of great practical significance to provide service quality assurance, dynamic access control and abnormal network behaviors detection. With the widespread adoption of encryption techniques for internet, especially 5G and IoT applications, the growth of portion of encrypted traffic has dramatically posed a huge challenge for QoS, network management and security monitoring. Therefore, studies on encrypted traffic classification not only help to improve the fine-grained network resource allocation based on application, but also enhance security level of network and application.

Traditionally, the evolution of encrypted traffic classification technology has gone through three stages: port matching based, payload matching based and flow statistical characteristics based. Port matching based classification method infers applications' types by assuming that most applications consistently use 'well known' TCP or UDP port numbers, however, the emergence of port camouflage, dynamic port, proprietary protocols with user-defined ports and tunneling technology makes these methods lose efficacy quickly. Payload matching based methods, namely, DPI (Deep Packet Inspection) technology cannot deal with encrypted traffic because of invisible packet content of encrypted traffic, in addition, it incurs high computational overhead and requires manual signatures maintenance [1–3]. As a result, in order to attempt to solve the aboved problems of encrypted traffic identification , flow-based methods emerged, which usually combine statistical or time series traffic features and Machine Learning (ML) algorithms, such as naive bayes(NB), support vector machine(SVM), decision tree, Random Forest(RF), k-nearest neighbor(KNN) [4–7]. Although classical machine learning approach can solve many issues that port and payload based methods cannot solve, it still has some limitations, such as handcrafted traffic features driven by domain-expert, time-consuming, lack of ability of automation, rapidly outdated when compared to the evolution. Unlike most traditional ML algorithms, Deep Learning performs automatic feature extraction without human intervention, which undoubtedly makes it a highly desirable approach for traffic classification, especially encrypted traffic. Recent research work has demonstrated the superiority of DL methods in traffic classification [8], such as MLP [9], CNN [10–14], SAE [15], LSTM [16, 17].

In this paper, we proposed an unbalanced dataset solution based on FlowCGAN, which utilizes the advantage of GAN in data augmentation, and generates a certain amount of minor class through feature learning to achieve the purpose of balancing the traffic data set. In this paper, the classical deep learning model CNN was used to classify the unbalanced data set, the random oversampled data set and the CGAN balanced

data set to verify the feasibility of FlowCGAN data generation.

The subsequent chapters of this paper are organized as follows: Section **??** describes the related works about solving the unbalanced data set; Section **??** briefly describes the principles and network architecture of GAN and CGAN; Section **??** elaborates on the methodology of FlowGAN, including data preprocessing, model architecture, and related algorithms; Section **??** describes the experimental environment and experimental results; Section **??** provides conclusions about our work and an introduction to the future work.

Problem: However, due to the different popularity of various applications, the class imbalance problem of traffic samples often occurs when building traffic datasets. That is, the number of popular application samples is much larger than others, which always leads to the misclassifying problems of minor applications and thereby decrease of classifier performance. Imbalanced class distribution of a dataset has posed a serious challenge to most ML based classifiers which assume a relatively balanced distribution [18]. Network traffic classification is no exception due to the imbalanced property of network traffic data [19, 20], especially encrypted traffic. Therefore, it plays a very crucial role to deal with the problem of imbalanced class distribution of traffic dataset for network traffic classification. However, there are very few studies focusing on traffic data augmentation used for traffic classification to overcome the limitation of class imbalace.

In this paper, we present the design and development of Generative Adversarial Network (GAN) based traffic data augmenting method called FlowGAN to generate synthesized samples for encrypted traffic classification. The synthesized data is then combined with the original (viz. real) data to construct the new traffic training dataset. As a proof of concept, we adopted three state-of-the-art deep learning based encrypted traffic classification methods, Multi-perceptron (MLP), Convolutional Neural Network (CNN) and Stack AutoEncoders (SAE) respectively on our new augmented training dataset synthesized from our FlowGAN. The experimental results demonstrate that classical deep learning based encrypted traffic classification algorithms over our new dataset can achieve better performance compared with other traditional data augmenting methods like over sampling [21] or generating artificial data [22].

The rest of this paper is organized as follows. Section II introduces the preliminaries and related works of traffic classification, some current methods for tackling with the problem of imbalanced class data and GAN. Section III illustrates the algorithm of FlowGAN. Section IV describes the design of the encrypted traffic classification method based on FlowGAN. The experimental results are provided and discussed in Section V. Section VI concludes our work and presents some future works.

Many related works have been carried out recently. Mohammad Aazaml et al. [**?**] proposed a Fog Computing and Smart Gateway Based Communication for Cloud of Things. Mohammad Abdullah Al Faruque et al. [**?**] brought forward an energy management platform based on Fog Computing architecture. Mohamed Saleem Haja Nazmudeen et al. [**?**] introduced a distributed processing framework for data aggregation based on fog computing architecture. Feyza Yildirim Okay et al. [**?**] presented an Smart Grid model based on fog computing. The model is divided into Smart Grid layer, fog layer and the cloud layer from bottom to top. Smart Grid layer mainly consists of smart meters, smart appliances and other intelligent equipments. The fog layer consists of multiple fog computing nodes. The cloud layer is mainly responsible for data storage, analysis and mining. Based on the concept of fog computing someone proposed a portable data storage and processing solution applying to Advanced Metering Infrastructure (AMI) [**?**]. In addition, many research works about programming model about IoT application have been carried out in recent years. I.Satoh [**?**] proposed a framework for data processing at the edges based on Mobile Agent and mapreduce. S.Cherrier et al. [**?**] introduced a distributed logic for IoT services based on OSGi to improve the modularizaion programming. K.Hong et al. [**?**] brought forward a programming model called Mobile Fog for large scale applications on IoT to try to develop IoT application by fog computing. However, existing schemes proposed before can not meet the new requirements of IoT application in smart grid, especially distributed coordination between fog computing nodes.

There are two main contributions in this paper. Firstly, we propose a new distributed Fog Computing architecture for IoT application in smart grid. To improve the application latency, we integrate distributed coordination capacity called FCC(Fog Computing Coordinator) in our architecture, which gather information of FC nodes in the same area periodically. In addition, FCC also devotes itself to assigning jobs to FC nodes so that all nodes can fulfill some complex tasks collaboratively. Secondly, a programming model is proposed to realize the architecture.

The remaining of the paper is organized as follows. The new Fog Computing architecture is presented in Section 2. In Section 3, the programming model corresponding to the architecture is disscussed. Section 4 presents the evaluation of our architecture and programming model. Finally, Section 5 draws conclusion of this paper.

II. Fog Computing Based Architecture for IoT application in Smart Grid

A. The new requirements for IoT application in Smart Grid

In this section we will identify several requirements that need to be taken into consideration to effectively deploy IoT application in Smart Grid.

1) Latency Sensitivity: Many IoT applications in Smart Grid depend on instant decisions and even second level latencies are not tolerable [**?**]. For example, electric substations in smart grid systems are equipped with various sensors to monitor status of power transmission. In this scenario, any latency may lead to serious accident like power failure.

2) Distributed Coordination: There are always a lot of sensors distributed in geographic area, e.g., charging piles for electric vehicles. It is very important that coordinating multiple sensors or nodes distributed in several areas to provide electric vehicles charging services of IoT application in Smart Grid.

3) Locations awareness and Mobility Support: In many IoT applications end devices of Smart Grid are mobile and geographic distribution such as electric vehicles. As the main goal of the Fog Computing is to move computing power close to where the data is generated, it is necessary to be able to aggregate data at the closest network element while the end devices are moving.

B. Fog Computing Based Architecture for IoT application in Smart Grid

As shown in Fig. 1, our fog computing based architecture for IoT application in smart grid is still divided into terminal layer, fog layer and cloud layer from bottom to top.

1) Terminal nodes layer: This is the bottom layer which is made up of smart devices, which are responsible for sending raw sensed data and event logs to upper layer.

2) Fog layer: The middle layer consists of fog nodes deployed at the edge of network to extend the processing ability of cloud center. Compared with the traditional Fog Computing model, our fog layer is divided into fog nodes(FN) sub-layer and fog nodes coordination(FNC) sub-layer. With the ability of computing and storage, these fog nodes of FN sub-layer provide a mechanism for migrating processing logic to the edge of the network. The FN sub-layer also has the aggregation capability for the sensed data from terminal nodes layer. After gathering and analyzing raw sensed data, part of them directly feed back to the active nodes in terminal nodes layer to complete the real-time response and process to the emergency event, the other part is transmitted to FNC sub-layer. FNC sub-layer consists of multiple coordinators located in the geographical areas. Fog nodes are divided into several clusters, where there are a few equipments with computing and storage selected by some principle. We call this equipments FCN(Fog Computing Coordinators), which focus on coordinating the fog nodes to deal with some complex tasks due to the problem of distributed collaboration during the service, for example, query a suitable charging station for moving electric vehicles. In addition to undertaking data analysis and processing of the sub-region, the coordinator is responsible for coordinating all fog computing nodes in the region to further improve application performance by using the parallel computing capabilities.

3) The cloud layer: This layer is the upper layer in this architecture. It is composed of servers, such as Data Centers which are responsible for analyzing massive historical data.

C. System Model

As shown in Fig. 2, the system model consists of sensors, action devices, communication nodes, fog computing nodes, cloud computing servers, FNC, service orchestration and scheduling servers called OSS servers. It is different with

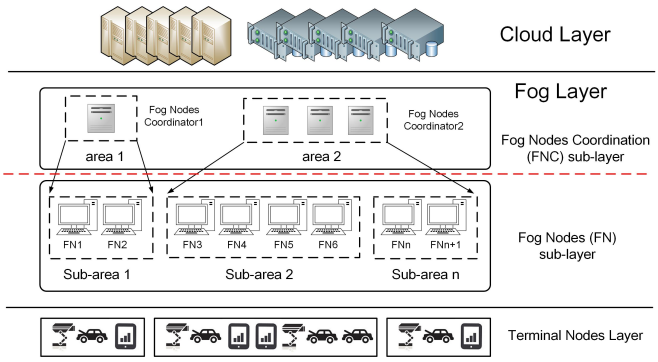

Fig. 1: Fog Computing Based Architecture for IoT application in Smart Grid

the system model proposed in papers, we introduce FNC sub-layer composed by FNC from each area. FNC accept requests from TN and decompose the services data flow according to the resource usage and service flow capacity of each FN and cloud computing servers, and then dispatch the jobs execution to related Fog nodes. Finally, FNC collect all the execution results to make the final decisions and instructions to the action devices. FN within the same layer have to achieve a complete user request under the coordination of FNC because there are no direct interaction between them. The interaction between FN and FNC can be realized by the way like relay communication, or Flow Table in SDN controller [**?**].

Besides, we introduce OSS servers in Cloud Computing Center, which are able to decompose services data flow based on resource usage and capacity collecting from computing nodes. OSS servers dispatch job execution images to computing nodes by the way like traditional virtual machine with better security isolation or Docker container [**?**] with less start latency. OSS servers mainly aim at initialization deployment of new applications provided by service providers and setup service related execution logic on computing nodes. In contrast, FNC provide application services that meet the functionality and QoS requirements for end users by resource allocation and service logic coordination deployed in various computing nodes after receiving service requests from end users.

III. Programming Model

A. Programming Model Introduction

The IoT application in Smart Grid can be considered as a distributed application system essentially. However, the traditional programming model of distributed system was based on the model of ”request-response” message interaction, which cannot meet the requirements of realtime processing large amount of data generated by devices. Therefore, we need to design a programming model based on Data Flow programming [**?**]. The typical programming framework is WoTKit processor based on WoTKit platform [**?**] and NR of IBM [**?**]. WoTKit is developed on JAVA Spring framework. Developer can run data flow program by creating wire between modules.

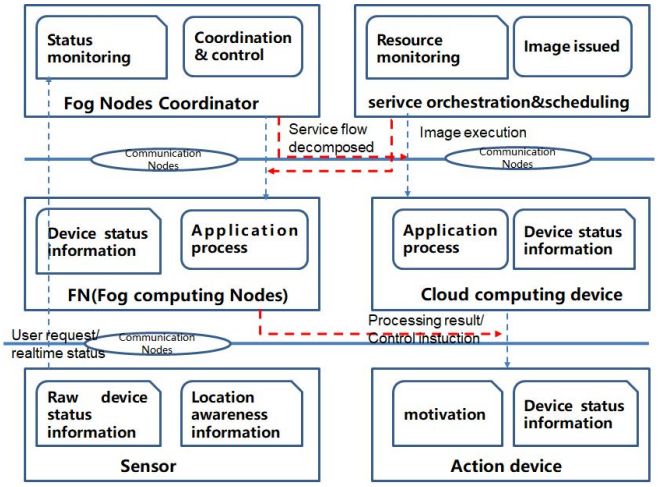

Fig. 2: system model

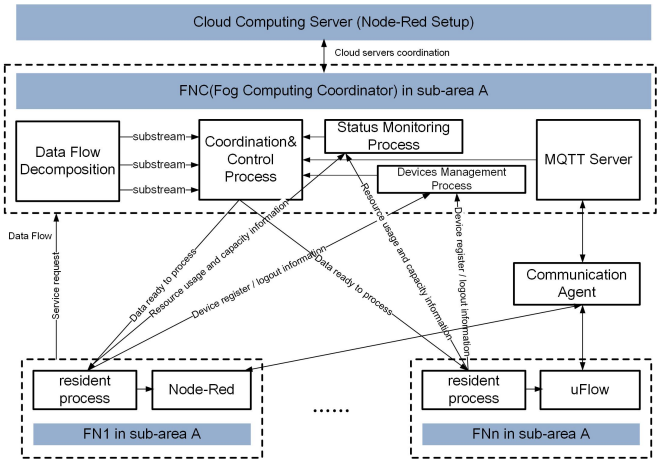

Fig. 3: programming model

However, WoTKit is designed for deployment of server level, it is better that using NR framework in IoT application, which was designed for application development in single computing unit, including nodes of input, output, processing and visual developing environment based on Web.In this paper [**?**], author added some feature about distribution to the traditional NR framework so that a data flow can be deployed in several computing nodes in the manner of multiple data flow slices.

B. Distributed Coordination Dataflow Programming model

Distributed Coordination Dataflow Programming model is shown in Fig. 3. Under the control of FNC, cloud servers and FN distributed in different geographic area together accomplish data analysis and processing of application services. There are two types of computing nodes, one is FN with rich computing resource,which choose Node-Red as distributed data flow computing framework. The other is FN with limited resource, which choose uFlow [**?**] as flow processing framework.

There are resident processes in every distributed computing nodes, which are responsible for collecting information like resource and capacity and then reporting to upper layer so that FNC will make better decision and instruction. After receiving substream and data ready to process, FN translate these data flow into instructions that can be identified and executed in terminal nodes. For example of parking service of electric vehicles, when a vehicle under the control of a sub-area FN apply a vehicle parking service, FNC of this sub-area will dispatch the request to all FN of this sub-area. After distributed coordinative processing of all FN, FNC can provide a best parking information for vehicles. As for online monitoring service all the time, we should consider the collaboration of Cloud Layer.After processing of FN in local area, FNC should report all the statistical data to cloud servers for future analysis.

Because of mobility of Terminal Nodes, the final data flow made by FNC may not suitable to current network condition, FN need coordinate themselves to adapt this situation at that time. That is called Fog Computing Nodes Migration. According to the initiator of migration, we can devide the migration into two types: one is initiated by FN; the other is Terminal Nodes. Two API used during the process of migration are following:

1)State on_migration_start(nodeID): Invoked by migration device before migration process start. It returns a state object for flow information waiting for migration.

2)Void on_migration_end(state_s): Invoked by migration target device after receiving request message from migration initiator.

The following is pseudocode describing the migration process for example of FN initiator.

1 procedure migration_source(nodeID)
2: obtain the actual latency T to node whose ID is nodeID
3: compare T with threshold Tupper
4: if T less than Tupper
5: return;
6: else
7: obtain the best candidate Vb from candidate group V
8: V = V - Vb
9: send a Start Migration message to node Vb
10: wait for response Resp
11: if Resp is ACCEPT then
12: S = on_migration_start(nodeID)
13: send Object State message to node Vb
14: release local resource of dataflow relate to nodeID
15: return
16: else
17: if V is empty
18: warn the message(can not migrate)
19: return
20: endif
21: endif
22: goto 7
23: endif
24: end procedure

Compared with QoS of Terminal Nodes and FN's default threshold during process of substream computing, FN will

decide whether migrate or not, see line 2-5.The detail are following: 1) choose the best nodes from node group with proper capacity, resource and QoS requirements, see line 7-8; 2) Send a migration start request message to alternative nodes and wait for response, see line 9-10; 3) if alternative nodes accept, then call on_migration_start to get all status information of current nodes and send to alternative nodes. Meanwhile, free all related resources of data flow ready to migrate, see line 11-15; 4) if alternative nodes do not accept, send the warning message,like can not migrate, and quit. Otherwise, choose the next alternative node from node group and repeat above-mentioned. target nodes will call on_migration_end to take over following work after receiving migration status message from migration initiator.

IV. Evaluation

A. Simulation Setup

The electric vehicle intelligent service system is composed of electric vehicle, charging pile, regional coordinator, regional application server, cloud service center, communication proxy server and basic communication network. As sensing devices of network at the edge of network, electric vehicles report relevant realtime information and request for services, as well as to provide services to users. The charging pile device is located at the edge of the network as a fog computing device. It processes the data according to the established application logic, and transfers the processed data through communication proxy server to application server of the region or romote cloud service center according to control of area coordinator. The regional application server joined with cloud service center provides related services to users. The difference is that the regional application server is more emphasis on providing some geographically related or strict requirements of delay services (such as navigation, intelligent parking, etc.), and the cloud center server provides some long-term analysis and forecasting services. According to density of charging pile and the size of traffic flow, we set a certain service area, and set up a regional coordinator in each area. The coordinator completes the data flow chart and transmits datas to relevant devices according to the service requests from users and the resources, capabilities, location information reported by the equipments.It coordinates these devices to complete the related service logic. For services that require cross-domain provisioning, coordination is done by regional application servers in different regions.

This section provides an electric vehicle intelligent service experiment system which is used to simulate and analyze the performance of the fog computing architecture presented in this paper. This paper chooses IBM's NR framework as an implementation tool for application development, and deployes a stream-based micro-runtime environment uFlow over resource-limited IoT devices. In the uFlow environment, we use Lua as the programming language, MQTT as a communication protocol. In this experiment, in order to supply charging service to the electric vehicle, the system selects the most suitable charging pile to the electric vehicle. There are two cases: one is based on the traditional fog computing architecture, in this case electric vehicles directly transmit the requests to all charging piles in a certain range, after the calculation, the charging pile met returns the confirmation to the electric vehicle. The other case is based on the fog computing coordinator architecture, in this case electric vehicles will directly send requests to the fog computing coordinator in the area. According to the information of the charging piles, the coordinator will forward the request to some charging piles with the possibility of providing the service and then the response is returned to the electric vehicle by the final eligible charging pile.

B. Simulation Results

The experimental system is mainly used to evaluate the contrast of application latency between the traditional fog computing architecture and our architecture based on fog computing coordinator. We used 20 software terminal nodes running on embeded system as electric vehicles, 10 software Fog nodes as charging piles and 2 FNC nodes. The range of all entity is located about 2000 meters. Fig. 4 shows the relationship between the application latency and the query distance of the two architectures. Clearly, it can be seen that when the query distance is short, the application latency is similar. However, when the query distance is gradually increased, our architecture has a lower application latency. In Fig. 5, we discuss the relationship between the application delay and the number of service requests from the electric vehicles. Obviously, as the numbers of service request increase, the proposed architecture delay is significantly smaller than traditional one. Fig. 6 describes the relationship between the number of fog computing coordinator and application latency. When the numbers of coordinators increase, the application latency decrease significantly. As shown above, the proposed architecture effectively reduces the application latency of IoT application in smart grid.

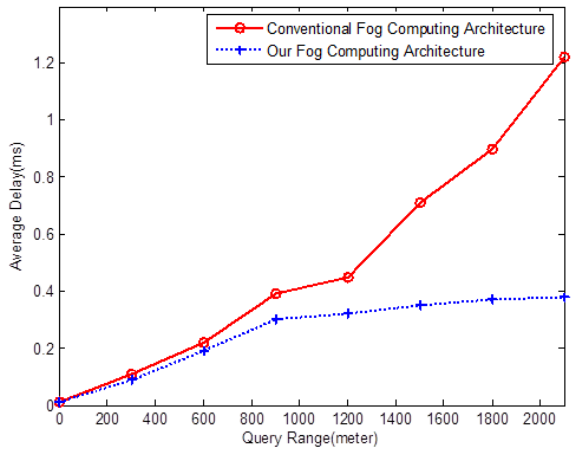

Fig. 4: application delay with query range

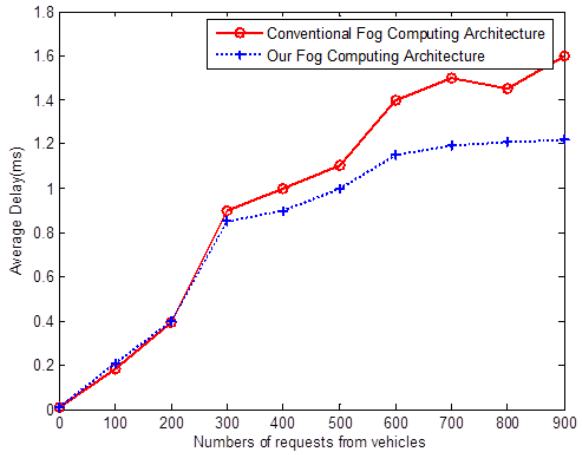

Fig. 5: application delay with service requests

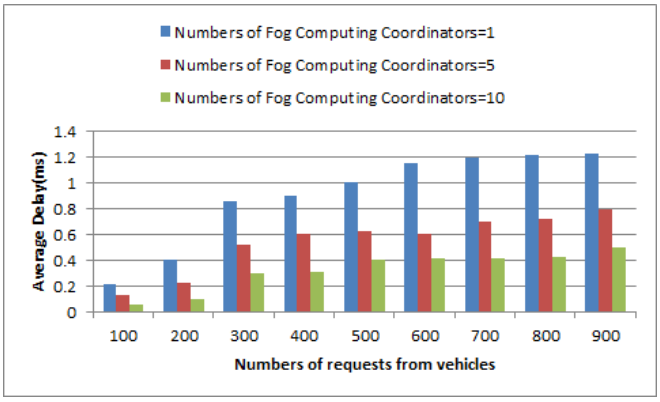

Fig. 6: application delay with FNC

V. Conclusion

The IoT applications in Smart Grid need higher requirements in terms of response time and location-awareness. It can be met well by distributed computing architecture based on fog computing (edge computing) as a supplement to the traditional "cloud - link - end" architecture. In this paper, the proposed distributed fog computing architecture and programming model for IoT applications in smart grid can effectively reduce service latency. The application development in this paper is mainly based on the Data flow mechanism to realize communication between devices. Next we will find a more appropriate communication protocol; and study the optimization of resource allocation algorithm in this architecture. At the same time we will carry out the studies about the handover between the mobile nodes in high-speed mobility.

Reference

[1] M. Finsterbusch, C. Richter, E. Rocha, J. Muller, and K. Hanssgen, "A survey of payload-based traffic classification approaches," *IEEE Communications Surveys Tutorials*, vol. 16, no. 2, pp. 1135–1156, Second 2014.

[2] P. Wang, F. Ye, and X. Chen, "A smart home gateway platform for data collection and awareness," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 87–93, Sep. 2018.

[3] P. Wang, X. Chen, F. Ye, and Z. Sun, "A smart automated signature extraction scheme for mobile phone number in human-centered smart home systems," *IEEE Access*, vol. 6, pp. 30 483–30 490, 2018.

[4] A. Dainotti, A. Pescape, and K. C. Claffy, "Issues and future directions in traffic classification," *IEEE Network*, vol. 26, no. 1, pp. 35–40, January 2012.

[5] G. Sun, Y. Xue, Y. Dong, D. Wang, and C. Li, "An novel hybrid method for effectively classifying encrypted traffic," in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, Dec 2010, pp. 1–5.

[6] P. Velan, M. Čermák, P. Čeleda, and M. Drašar, "A survey of methods for encrypted traffic classification and analysis," *Netw.*, vol. 25, no. 5, pp. 355–374, Sep. 2015. [Online]. Available: http://dx.doi.org/10.1002/nem.1901

[7] D. J. Arndt and A. N. Zincir-Heywood, "A comparison of three machine learning techniques for encrypted network traffic analysis," in *2011 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, April 2011, pp. 107–114.

[8] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescap, "Mobile encrypted traffic classification using deep learning," in *2018 Network Traffic Measurement and Analysis Conference (TMA)*, June 2018, pp. 1–8.

[9] P. Wang, F. Ye, X. Chen, and Y. Qian, "Datanet: Deep learning based encrypted network traffic classification in sdn home gateway," *IEEE Access*, vol. 6, pp. 55 380–55 391, 2018.

[10] M. J. S. M. S. Mohammad Lotfollahi, Ramin Shirali Hossein Zade, "Deep packet: A novel approach for encrypted traffic classification using deep learning," *Available from http://www.arxiv.org*, 2017.

[11] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, "End-to-end encrypted traffic classification with one-dimensional convolution neural networks," *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 43–48, 2017.

[12] and, , and and, "Malware traffic classification using convolutional neural network for representation learning," in *2017 International Conference on Information Networking (ICOIN)*, Jan 2017, pp. 712–717.

[13] Z. Chen, K. He, J. Li, and Y. Geng, "Seq2img: A sequence-to-image based approach towards ip traffic classification using convolutional neural networks," in *2017 IEEE International Conference on Big Data (Big Data)*, Dec 2017, pp. 1271–1276.

[14] X. Chen, J. Yu, F. Ye, and P. Wang, "A hierarchical approach to encrypted data packet classification in smart home gateways," in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, Aug 2018, pp. 41–45.

[15] Z.Wang, "The application of deep learning on traffic identification," *Available from http://www.blackhat.com*, 2015.

[16] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Network traffic classifier with convolutional and recurrent neural networks for internet of things," *IEEE Access*, vol. 5, pp. 18 042–18 050, 2017.

[17] W. Wang, Y. Sheng, J. Wang, X. Zeng, X. Ye, Y. Huang, and M. Zhu, "Hast-ids: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection," *IEEE Access*, vol. 6, pp. 1792–1806, 2018.

[18] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intell. Data Anal.*, vol. 6, no. 5, pp. 429–449, Oct. 2002. [Online]. Available: http://dl.acm.org/citation.cfm?id=1293951.1293954

[19] L. Vu, C. T. Bui, and Q. U. Nguyen, "A deep learning based method for handling imbalanced problem in network traffic classification," in *Proceedings of the Eighth International Symposium on Information and Communication Technology*, ser. SoICT 2017. New York, NY, USA: ACM, 2017, pp. 333–339. [Online]. Available: http://doi.acm.org/10.1145/3155133.3155175

[20] L. Vu, D. Van Tra, and Q. U. Nguyen, "Learning from imbalanced data for encrypted traffic identification problem," in *Proceedings of the Seventh Symposium on Information and Communication Technology*, ser. SoICT '16. New York, NY, USA: ACM, 2016, pp. 147–152. [Online]. Available: http://doi.acm.org/10.1145/3011077.3011132

[21] H. Guo and H. L. Viktor, "Learning from imbalanced data sets with boosting and data generation: The databoost-im approach," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 30–39, Jun. 2004. [Online]. Available: http://doi.acm.org/10.1145/1007730.1007736

[22] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Int.*

Res., vol. 16, no. 1, pp. 321–357, Jun. 2002. [Online]. Available: http://dl.acm.org/citation.cfm?id=1622407.1622416