Deploying Edge Computing Nodes for Large-scale IoT: A Diversity Aware Approach

C .Suguna Devi, R.Rana Prathap, Shaik.Salmabegum

[1,2,3] *Asst..Prof, CSE, AITS-R (India)*

ABSTRACT

The current advances in microelectronics and correspondences have prompted the improvement of huge scale IoT networks, where gigantic tactile data is produced and should be handled. To help ongoing handling for vast scale IoT, conveying edge servers with capacity and computational ability is a promising methodology. In this paper, we precisely break down the affecting elements and key difficulties for edge node organization. We at that point propose a novel three-stage organization approach which thinks about both activity assorted variety and the remote decent variety of IoT. The proposed work goes for giving ongoing handling administration to the IoT system and diminishing the required number of edge nodes. We led broad reenactment tries, the outcomes demonstrate that contrasted with the current works that neglected the two sorts of decent varieties, the proposed work incredibly lessens the quantity of edge nodes and enhances the throughput amongst IoT and edge nodes.

Keywords—IoT, Large-scale, Edge Computing, Deployment

I. INTRODUCTION

The recent advances in low-power wireless communications and computing technologies have enabled the largescale implementation of Internet of Things (IoT) systems [1], where massive sensors, micro-controllers and transceivers are embedded to the facilities of buildings, vehicles, wearable items and wild areas [2], [3], [4]. The IoT aims at making the Internet even more immersive and pervasive, providing interactive cyber-physical access and control services [5], [6]. Based on the IoT infrastructure, various large-scale real-time applications emerge, which makes the real-time processing a fundamental and critical service for IoT [7], [8]. For example, the smart building system [4] consists various types of IoT sensory nodes including HD cameras, wearable sensors, localization anchor sensors, gym equipment sensors, etc. Those sensors need to keep collecting the sensory data continuously and provide real-time response to the upper level applications. For example, the health monitoring system needs to collect various health data from the wearable sensors from users and alarm when abnormal phenomenon is detected.Figure 1 shows the edge architecture for IoT systems. Multiple edge computing servers are deployed to cover part of the IoT nodes. The computational tasks and sensory data from the IoT nodes are sent to the edge nodes (ENs) for processing. The results are then returned to the IoT nodes or transmitted to the cloud for big data analytics [10].

Deployment of the edge nodes is a fundamental problem for the above architecture. Different from the existing works on sink deployment in multi-sink sensor networks, the deployment of edge nodes have several distinct challenges. First, compared to traditional sensor network nodes, IoT nodes are more diverse and have largely different traffic demands. For example, the video cameras produce much more data than the equipment maintenance sensor nodes. Second, unlike the mobile edge computing where WiFi/Cellular communications are utilized, IoT nodes often employ the low-power radios [11], [12] and are more prone to the wireless interference. Considering the edge servers are responsible for collecting data as well as disseminating data (e.g., for software update or computational feedback), the wireless interference can significantly affect the EN deployment.

The main contribution of this paper is summarized as follows.

1) We propose a discretization scheme to generate candidate positions, where traffic diversity is considered and the demands/resources of IoT nodes are normalized.
2) We propose a novel utility metric to evaluate the candidate positions, where the wireless diversity is considered.
3) Based on the above schemes, we propose a deployment algorithm which improves the IoT-Edge throughput and reduces the number of edge nodes.

II.RELATEDWORK

A. Preliminaries and System model

Our aim is to deploy a number of edge nodes to a large-scale IoT network, where diverse IoT sensor nodes are in an area possibly with pedestrians and wireless interference. Figure 1 shows a typical IoT network in an airport, which consists a number of HD camera sensors with high traffic demands and a number of ordinary sensors with low traffic demands. The ordinary sensors are used for building monitoring, indoor navigation, equipment monitoring, etc. Although those IoT sensor nodes generates large amount of data, they usually have very limited computational resources for real-time data processing [7]. To support real-time data processing for the large-scale IoT network, a promising alternative is to deploy a number of edge nodes hierarchically with the IoT network, which are connected to the IoT nodes and processes the IoT data in real time. All IoT nodes then send the sensory data to the connected edge servers for data processing.

Compared to the traditional large-scale sensor network, the large-scale IoT network has two main differences.

1) First, the IoT networks are heterogeneous rather than homogeneous, which consists much more diverse IoT nodes. For example shown in the figure, the camera sensors and equipment sensors have largely different demands on data traffic and data processing.

2) Second, the IoT networks are often deployed in indoor environments rather than unmanned areas. Considering WiFi has been pervasively deployed for wireless access, the co-existence problem of edge nodes and the environmental wireless networks also needs to be considered. Specifically, as low power radios are often employed in the IoT nodes, they can be easily affected by WiFi communications, Bluetooth communiations, etc .

Our goal now is to deploy a number of edge nodes to the IoT network to cover all the IoT nodes. The problem of minimizing the number of edge nodes is equivalent to the problem of Knapsack problem , which is NP Complete. Therefore in this paper, we design a heuristic to reduce the number of edge nodes and provide high-throughput data collection/dissemination service for real-time data processing in the IoT networks.

B. Diversity Aware Deployment of the Edge Servers

i) Challenges:

There are two challenges to deploy edge nodes for realtime data processing in large-scale IoT – Traffic diversity and wireless diversity.

1) Traffic diversity. The IoT nodes are diverse in data types and traffic demands. Different types of data have different processing requirements and will require different amount of computational resources. Considering the edge nodes are often powerful, in this paper, we mainly consider the diversity of traffic demands. Different amount of traffic demands will directly affect the deployment of edge nodes in two ways. First, the edge nodes are targeted to receive data from the IoT nodes. Intuitively, the edge nodes should be deployed nearer to the IoT nodes with more traffic demands. Second, the traffic demands may not be consistent with the node density because different nodes have different demands. The two factors need to be jointly considered in the deployment process.
2) Wireless diversity. For wireless diversity we mainly consider link quality and link correlation. For many large-scale IoT systems, especially for those deployed in indoor environments, the WiFi networks, Bluetooth communications, even microwave ovens can have large impact on the transmission quality between the edge nodes and the IoT nodes as they typically use lowpower radios (e.g., ZigBee). Considering that edge nodes are used to collect sensory data and disseminate remote commands and maintenance instructions, both inbound and outbound performance can be largely impacted by the interference. Therefore, the link quality/correlation distribution and the impact on the performance of both collection/dissemination should also be considered in the deployment process.

ii) Overview:

To address the above challenges and deploy the edge nodes effectively, we incorporate the two kinds of diversities into the deployment process and propose a three-phase deployment approach. Figure 2 shows the overview of the proposed approach.

1) Discretization. Before determining the positions for deploying edge nodes, we first discretize the whole IoT network area into many small sections and the centroid of each section is a candidate position. In the discretization, we combine both wireless transmissions and the data traffic demands to define “effective” transmission levels, with which all nodes’ levels are normalized and the traffic diversity is incorporated. The details are described in Section III-C.
2) A utility metric. Next, we propose a comprehensive metric to evaluate the impact of each candidate position. The utility metric calculates the expected performance gain of the candidate position regarding

the expected number of transmissions required for both data collection and message dissemination. Wireless diversity including link quality and link correlation among multiple links are considered in the metric. The detailed design of the metric is described in Section III-D.

3) The deployment algorithm. Based on the proposed utility metric, we further devise a heuristic to select the best candidate positions for deploying the edge nodes. The input is the traffic demands and wireless

iii)Deployment algorithm:

Based on the metric ρ proposed , we can sort all the candidate positions. The problem is NP complete. We then propose a heuristic solution by selecting the candidate positions from the position with the best ρ. When a candidate position is selected, the IoT nodes within the m-th utility level are included as its subscribing receivers. Then we exclude all the covered subscribing receivers, update the ρ values for all candidate positions (expected the chosen positions) and select the position with the highest ρ for deploying the next edge node. The above process continues until all IoT nodes are covered by the edge nodes. The detailed deployment algorithm is described in Alg. 1.

Algorithm 1 The deployment algorithm Input:

1) The set of all candidate positions, P_c;

2) The set of all IoT nodes, N;

3) The link quality/correlation for all positions which is used forcalculating the ρ metric for the candidate positions;

Output: The selected positions for edge node deployment P_s while *There exists* $n \in N$ *that is not covered by any* $p_i \in P_c$. positions in each iteration. With a larger τ, the algorithm runs fast but the selected positions may have worse utilities; With a smaller τ, the algorithm runs slow but the selected positions have better utilities.

III. INDENTATIONS AND EQUATIONS

A. Evaluation

i) Experimental Settings:

We tune the following parameters to see in which cases the proposed work performs better or worse.

1) The fraction of high-demand IoT nodes. The highdemand IoT nodes generate five times traffic of the ordinary IoT nodes. We tune the fraction of the highdemand nodes and see the performance gain achieved by the proposed work.
2) The fraction of dissemination tasks. Different IoT networks may have different designing goals, leading to different fractions of dissemination tasks . This fraction has impact on the calculation of ρ.

The wireless interference. We change the number of interfering wireless APs and compare the performance gains. The interference impacts the wireless diversities, which further impacts the selection of candidate positions.

ii) Simulation Results:

We can see that 1) compared to the existing work based solely on wireless communications, the proposed work always reduces the number of edge nodes. The reason is that although link quality is important, the throughput may not be good as high-demand nodes may be assigned to poorquality links. 2) The reduction increases and then decreases, which means our work better suits for the case with more diverse traffic demands from IoT nodes. The reason is that we explicitly consider the traffic diversity in the candidate positions. As a result, when the IoT nodes are more diverse, we have more room for optimization.

Figure 5 depicts the performance gain for both collection tasks and dissemination tasks with varying fraction of the dissemination tasks. We set that the IoT network contains 60% high-demand nodes. Recall that the dissemination tasks are used for network maintenance or periodic network update. From the results we can see that 1) the performance for data collection is consistently improved. The reason is two-fold. First, wireless link diversity is considered in the proposed work, which reduces the expected packet losses. Second, although the fraction of collection tasks deceases, for each specific collection task, the link diversity is still considered and thus the throughput is improved; 2) the performance for data dissemination significantly increases. This is because the existing works overlook the dissemination task demands. When the fraction of the dissemination tasks increase, the throughput gain increases accordingly. It is also worth noting that in most IoT networks, dissemination is not the dominating

Figure 6 shows the throughput gains for collection and dissemination under different number of wireless APs (WiFi). Similar to [25], the impact of WiFi interference is introduced in the simulation by deliberately failing some packet transmissions. The packet losses generated at the sender side will be the correlated packet losses and the packet losses generated at the receiver side will be the independent packet losses. According to the studies in [25], WiFi interference is a dominating reason for correlated packet losses, as a result the packet loss link correlation becomes stronger when WiFi interference becomes stronger. From the results it can be inferred that as the interference becomes stronger, the throughput gain of dissemination becomes larger and the throughput gain of collection remains similar. From the calculation process in Section III-D we can see that link correlation mainly impacts the performance of data dissemination. When link correlation becomes stronger with the interference, there are more optimization space for dissemination.

Figure 7 depicts the reduction of edge nodes with varying number of interfering wireless APs. Different from the experiment in Figure 4, the fraction of high-demand IoT nodes is fixed and the number of interfering nodes is varying. We set 40% nodes with high traffic demand and 20% dissemination tasks. It can be inferred that 1) the reduction increases as the interference becomes stronger. From the above analysis on dissemination, we can infer that the increments come from the portion of nodes that have 20% dissemination tasks. In order to meet the dissemination throughput threshold, more edge nodes will be required for the work without considering dissemination performance. 2) Compared to the results in Figure 4, the reduction changes are much smaller.

Therefore, we can conclude that the number of edge nodes is mainly determined by the traffic diversity. The impact of interference on reducing the edge nodes is limited.

B. Related Work

The combination of mobile edge servers and IoT as well as the corresponding benefits are discussed . Our work differs from the scenario described in the following ways. First, we consider a large-scale and more practical IoT network, where different IoT nodes are with more diverse demands. Second, we focus on the deployment problem of edge nodes for IoT network while the authors in consider a general idea of combining mobile edge computing and IoT applications.

Although there are few existing works on deploying edge servers for large-scale IoT network, the problem is closely related to the powerful node deployment problem in largescale sensor networks, where the powerful nodes can be either relay nodes or sink nodes . Next, we mainly introduce and discuss the literature for deploying relay nodes or multiple sink nodes in largescale sensor networks. Bredin *et al.* studied the relay node deployment problem which should meet a survivability requirement. Cheng *et al.* considers the connectivity constraint in the relay node deployment. Similar to our work, each IoT node is required to be connected to a relay node. Misra *et al.* additionally consider limiting the candidate positions and propose to select candidate positions before deployment. Our work differs from in that we incorporate the traffic diversity (traffic demand distribution) in the candidate position generation process, therefore providing more reasonable and efficient candidate positions. Nikolov *et al.* aim at deploying a given number of relays to the network to maximize the communication gains. Bagaa *et al.* is a recent work that achieves optimal placement of the relays over limited candidate positions. Different from these works, IoT networks contains more diverse nodes and experience more wireless interference. Therefore in our work, we jointly consider the traffic diversity and wireless diversity (especially the link correlation characteristic). As a result, the proposed work is more suitable for large-scale heterogeneous IoT networks and can achieve better throughput gains.Some works have specific requirements according to the target scenarios. Wu *et al.*consider the relay node deployment with pipeline inspection. Ma *et al.* additionally consider the delay constraint for the deployment. Our work is orthogonal to these works, i.e., the above constraints can be easily added into our scheme. Besides, the traffic diversity and link correlation are overlooked in these works, which may lead to performance degradation under strong interference scenarios.

IV.FIGURES AND TABLES

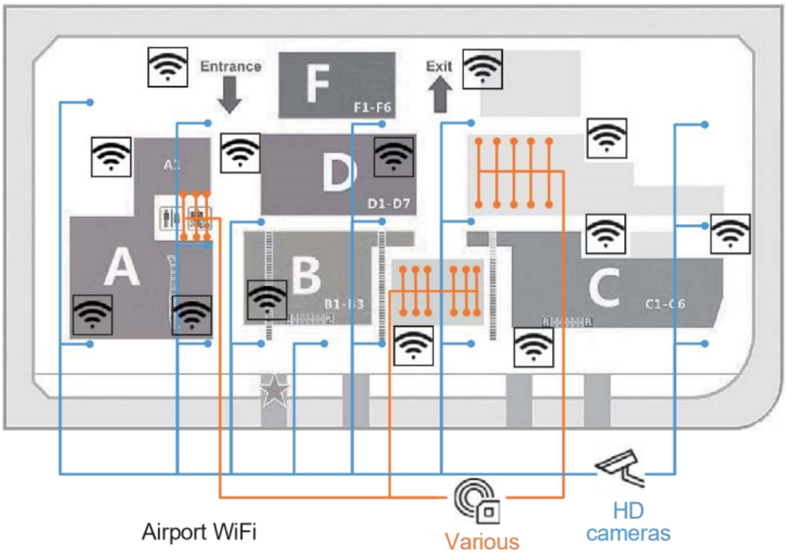

Fig. 1. An illustrative example for the system model. There are various types of IoT sensor nodes and wireless access points (AP) deployed in the target area. The edge nodes needs to cover the IoT nodes and try avoiding the interference from the wireless APs.

Traffic demands aware discretization	A utility metric for candidate positions	Deployment based on the metric
Discretizing the network area into several candidate positions	Evaluating the candidate postions considering both kinds of diversities	Deploying edge nodes to the best candidate postions based on the proposed metric

Fig. 2. Overview of the proposed work

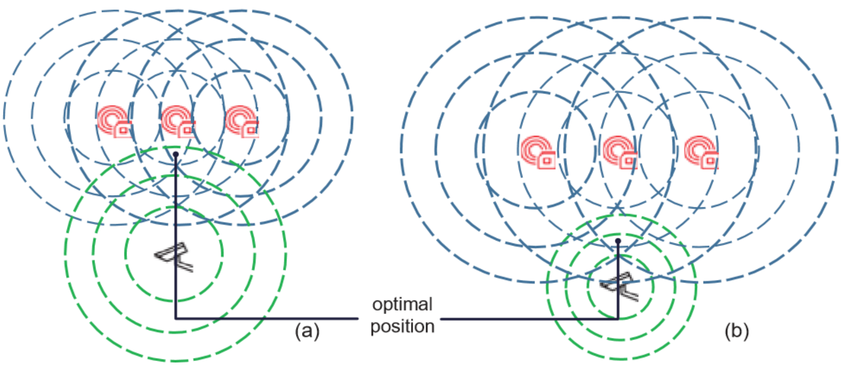

Fig. 3. The discretization approach with different leveling schemes. (a) shows the case that uses the transmission-rate levels for discretization; (b) shows the normalized levels which consider the data traffic demands from different IoT nodes.

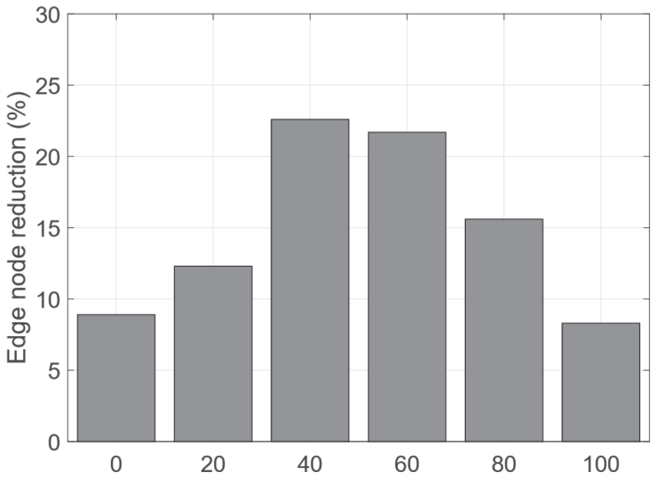

Fig. 4. The reduction on the number of edge nodes with different high-demand IoT nodes.

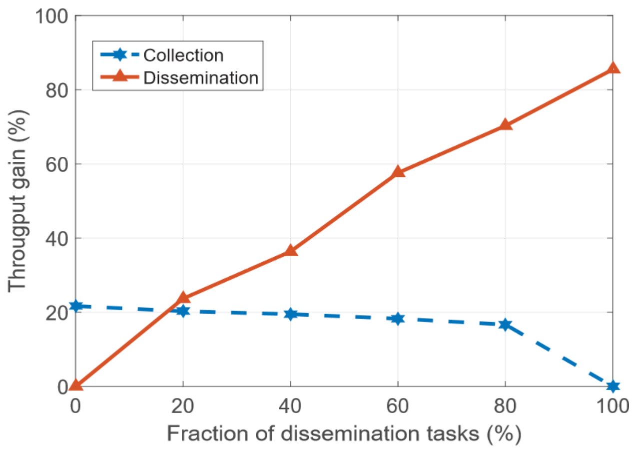

Fig. 5. The throughput gains with varying fractions of dissemination tasks.

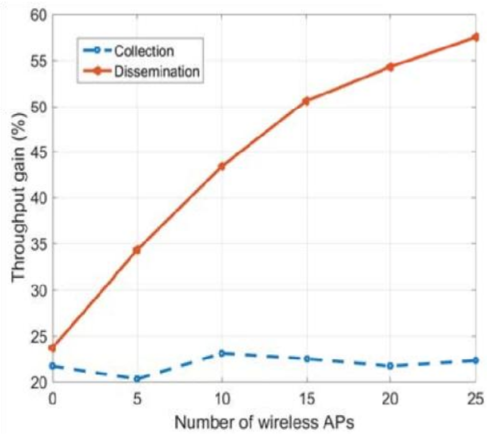

Fig. 6. The throughput gains with varying interfering wireless APs.

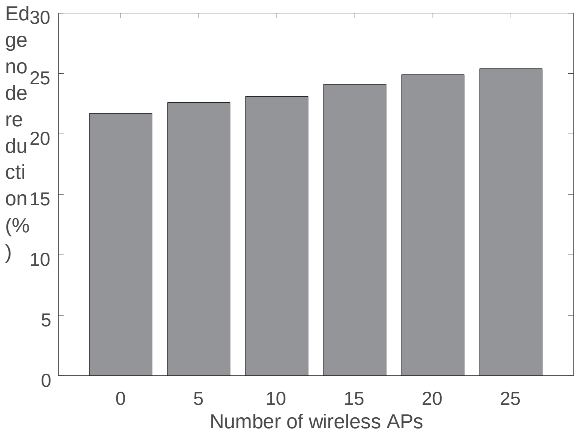

Fig. 7. The reduction on the number of edge nodes with varying interfering wireless APs

V. CONCULSION

In this paper, we propose to deploy edge nodes for realtime data processing in large-scale IoT networks. We identified the key challenges for edge node deployment – the traffic diversity and the wireless diversity. We then propose a novel three-phase deployment approach considering both kinds of diversities. The proposed work aims at minimizing the number of edge nodes and providing real-time processing service for the IoT network. We have conducted simulation experiments and the results show that compared to the existing works that overlooked the two kinds of diversities, the proposed work greatly reduces the number of edge nodes and improves the throughput for both data collection and dissemination.

VI. ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China the National Key Research and Development Program of China.

REFERENCES

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future generation computer systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[2] W. Du, Z. Xing, M. Li, B. He, L. H. C. Chua, and H. Miao, "Sensor placement and measurement of wind for water quality studies in urban reservoirs," *ACM Transactions on Sensor Networks (TOSN)*, vol. 11, no. 3, p. 41, 2015.

[3] Z. Li, W. Du, Y. Zheng, M. Li, and D. Wu, "From rateless to hopless," *IEEE/ACM Transactions on Networking (TON)*, vol. 25, no. 1, pp. 69– 82, 2017.

[4] M. Razmara, G. R. Bharati, M. Shahbakhti, S. Paudyal, and R. D. Robinett III, "Bilevel optimization framework for smart building-to-grid systems," *IEEE Transactions on Smart Grid*, 2016.

[5] A. Bader, H. ElSawy, M. Gharbieh, M.-S. Alouini, A. Adinoyi, and F. Alshaalan, “First mile challenges for large-scale iot,” *IEEE Communications Magazine*, vol. 55, no. 3, pp. 138–144, 2017.

[6] M. Vogler,¨ J. M. Schleicher, C. Inzinger, and S. Dustdar, “A scalable framework for provisioning large-scale iot deployments,” *ACM Transactions on Internet Technology (TOIT)*, vol. 16, no. 2, p. 11, 2016.

[7] P. Barnaghi, R. Tonjes,¨ J. Holler,¨ M. Hauswirth, A. Sheth, and P. Anantharam, “Citypulse: Real-time iot stream processing and large-scale data analytics for smart city applications,” in *Europen Semantic Web Conference (ESWC)*, vol. 2014, 2014.

[8] F. Chen, Z. Zhao, G. Min, W. Gao, J. Chen, H. Duan, and P. Yang, “Speed control of mobile chargers serving wireless rechargeable networks,” *Future Generation Computer Systems*, vol. 80, pp. 242–249, 2018.

[9] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, “Mobile edge computinga key technology towards 5g,” *ETSI White Paper*, vol. 11, no. 11, pp. 1–16, 2015.

[10] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, “Collaborative mobile edge computing in 5g networks: New paradigms, scenarios, and challenges,” *IEEE Communications Magazine*, vol. 55, no. 4, pp. 54–61, 2017.

[11] U. Raza, P. Kulkarni, and M. Sooriyabandara, “Low power wide area networks: An overview,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 855–873, 2017.

[12] K. Nair, J. Kulkarni, M. Warde, Z. Dave, V. Rawalgaonkar, G. Gore, and J. Joshi, “Optimizing power consumption in iot based wireless sensor networks using bluetooth low energy,” in *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on*. IEEE, 2015 , pp. 589–593.