Computing Resource Allocation in Three-Tier IoT Fog Networks: a Joint Optimization Approach Combining Stackelberg Game and Matching

Huaqing Zhang*, Yong Xiao‖, Shengrong Bu†, Dusit Niyato§, Richard Yu‡, and Zhu Han*
*Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA.
‖Department of Electrical and Computer Engineering at the University of Arizona, Tucson, Arizona, USA.
†School of Engineering, University of Glasgow, UK.
§School of Computer Science and Engineering, Nanyang Technological University, Singapore.
‡Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada.

Abstract—Fog computing is a promising architecture to provide economical and low latency data services for future Internet of Things (IoT)-based network systems. Fog computing relies on a set of low-power fog nodes that are located close to the end users to offload the services originally targeting at cloud data centers. In this paper, we consider a specific fog computing network consisting of a set of data service operators (DSOs) each of which controls a set of fog nodes (FNs) to provide the required data service to a set of data service subscribers (DSSs). How to allocate the limited computing resources of fog nodes (FNs) to all the DSSs to achieve an optimal and stable performance is an important problem. Therefore, we propose a joint optimization framework for all FNs, DSOs and DSSs to achieve the optimal resource allocation schemes in a distributed fashion. In the framework, we first formulate a Stackelberg game to analyze the pricing problem for the DSOs as well as the resource allocation problem for the DSSs. Under the scenarios that the DSOs can know the expected amount of resource purchased by the DSSs, a many-to-many matching game is applied to investigate the pairing problem between DSOs and FNs. Finally, within the same DSO, we apply another layer of many-to-many matching between each of the paired FNs and serving DSSs to solve the FN-DSS pairing problem. Simulation results show that our proposed framework can significantly improve the performance of the IoT-based network systems.

Index Terms — Fog computing, Stackelberg game, matching theory, Internet of things.

I. INTRODUCTION

With the rapid development of Internet of things (IoT), the number of connected devices has increased at an unprecedented speed [1]. It is known that analyzing the big data generated by all kinds of IoT devices requires a large amount of computing resources. In order to meet the demand of the data computing services, a large number of large-scale data centers has been deployed. In addition, cloud computing has been proposed recently to provide flexible and efficient services to the data service subscribers (DSSs). In cloud computing, the data service operator (DSO) is able to organize a shared pool of configurable computing resources (such as servers, storage, networks, applications and services), which can be easily accessed by DSSs on demands.

Generally speaking, large-scale data centers or clouds are typically built in remote areas far from the DSSs. This results in high transmission cost, transmission congestions and service latency, which can be intolerable for the applications that require real-time interaction or mobility. Accordingly, it is beneficial and necessary to pull the cloud closer to the users. In wireless radio access network, the concept of mobile edge computing is developed by ETSI (European Telecommunications Standards Institute), aiming to bring computation power into mobile radio access network. In mobile edge computing, the network edge is able to run in an isolated environment from the rest of the network and creates access to resources in the local neighborhood [5]. Moreover, in IoT, fog computing, put forward by Cisco, is proposed as a promising solution. In fog computing, multiple low-power computing devices, commonly referred to as the fog nodes (FNs), at the edge of the networks are deployed to offload the data computing services from the cloud. With the property of small-scale, low construction cost and mobility support, the FNs are generally deployed much closer to the DSSs and therefore can provide low-latency, fast-response and location-awareness service [6].

This work was partially supported by the National Science Foundation under Grants CNS-1646607, ECCS-1547201, CCF-1456921, CNS-1443917, and ECCS-1405121.

This work was supported in part by Singapore MOE Tier 1 (RG18/13).

Fig. 1: System architecture
able to receive data services from the FNs in the neighborhood or from remote data centers in remote areas.

In the fog computing network, the concept of network virtualization is also applied. As the large number of FNs and their computing resources are invisible for the DSSs, the DSSs can only contact and purchase the data services from the DSOs. Therefore, there is a virtualized network between DSOs and DSSs. When receiving the service requests from all DSSs, each DSO is able to collect the computing resources from the FNs and provide virtual data services for the DSSs. Based on the network requirements, each DSO is able to allocate different amount of computing resources from different FNs to different DSSs. Thus, the computing resource can be efficiently and effectively utilized by the nearby DSSs.

Fog computing networks can consist of a large number of FNs deployed by different DSOs at different locations to provide various data services and applications to the DSSs. When DSSs can choose their DSOs as well as the corresponding FNs to further enhance their quality-of-experience, how to allocate the limited computing resources of all the fog nodes to the DSSs is still an open problem. In this paper, we further extend our previous work [7] and focus on the resource selection and allocation problem between the FNs, DSOs and DSSs. We propose a joint optimization framework for all FNs, DSOs and DSSs in a distributed fashion. In the framework, we first formulate a Stackelberg game to model the interaction between DSOs and DSSs, where the DSOs set their service price first, and the DSSs purchase the optimal number of computing resource blocks (CRBs). Once the prices of DSOs and the purchased resources of DSSs have been obtained, each DSS can know how many CRBs are required and can then try to compete for the CRBs owned by the nearby FNs. Thus, we propose a many-to-many matching game to investigate the interaction between DSOs and FNs where each DSO has a set of CRBs to offload, and each FN has many vacant CRBs to sell. After all the DSOs decide their DSO-FN pairs, FNs will compete with each other to allocate their CRBs to the DSSs of the DSOs. We also adopt another layer of many-to-many matching framework to solve the FN-DSS pairing problem within the same DSO. Simulation results show that our proposed framework can significantly improve the performance of the fog computing networks.

The rest of this paper is organized as follows. We describe the system model in Section II and formulate the problems in Section III. Based on the formulated problem, we analyze the system with the proposed framework in Section IV, where the interaction between DSOs and DSSs is considered in Section IV-A, the interaction between FNs and DSOs is analyzed in Section IV-B, and the interaction between FNs and DSSs is discussed in Section IV-C. Finally, we evaluate the performance of our work in Section V, show related works in Section VI and summarize the paper in Section VII.

II. SYSTEM MODEL

Consider a fog computing network where each DSS can submit its data computing service to a set of neighboring FNs deployed by a set of DSOs as illustrated in Figure 1.

Accordingly, we consider a three-tier fog network, where the DSOs locate in the middle layer, managing the FNs in the upper layer and serving DSSs in the bottom layer. Without loss of generality, we assume there are $M$ DSOs, labeled as $\Psi = \{d_1, d_2, \ldots, d_M\}$ and $N$ DSSs, denoted as $\Upsilon = \{s_1, s_2, \ldots, s_N\}$. Let $\lambda_j$ be the workload arrival rate of DSS $s_j$, $\forall s_j \in \Upsilon$. We assume each DSS has a normalized preference list, denoted as $L_j$ over all DSOs. Moreover, $K$ FNs, labeled as $\Omega = \{f_1, f_2, \ldots, f_K\}$, locate in the area of consideration. We define the unit amount of computing resources that can be distributed by each FN as the “computing resource block (CRB)” [7], each of which can provide computing service at the rate of $\mu$. The physical data transmission network between FNs and DSSs satisfies the SecondNet topology [8], where the network facilities can provide the guaranteed quality-of-service (QoS) for the DSSs. Accordingly, in order to reduce the risk of potential network congestion and achieve real-time fast-response interaction, each DSO tries to offload the data services submitted by the DSSs to the large-scale data centers to the local FNs. However, as the DSSs cannot have the authorization to access the CRBs directly, the DSSs are required to receive the virtualized services from the DSOs, and with the management of DSOs,
the CRBs of the FNs can finally be allocated to the DSSs. We assume that different DSOs offer data services with different requirements. Based on the preference list $L_j$, the DSS $s_j$, $\forall s_j \in \mathcal{Y}$ is required to subscribe to at most one DSO. The network architecture is illustrated in Fig. 1. Assume the DSSs apply real-time interactive applications, where QoS is measured by the service delay. In this paper, for DSS $s_j$, $\forall s_j \in \mathcal{Y}$, we measure the cost of the service delay as $t_j = h_j + o_j$, which consists of the cost incurred by the queuing delay $o_j$ at the servers as well as the cost incurred by the network delay $h_j$ from the DSS to the physical service provider plus the distance from the physical service provider to DSS $s_j$. We suppose the workload of each DSS $s_j$ follows Poisson arrival process. According to the M/G/1 queuing delay model, the mean response time for unit transmitted data is $\frac{1}{\mu - \frac{\lambda}{q_j}}$ [9], [11]. Thus, the cost of queuing delay when serving DSS $s_j$ is $o_j = \frac{\lambda_j}{\mu - \frac{\lambda_j}{q_j}}$, where $q_j$ is the number of CRBs purchased by DSS $s_j$. Moreover, as the network delay is related to the transmission distance, traffic condition in the network and many other unpredicted factors, in practice, we suppose the network delay can be evaluated from training data periodically sent from sensors to the physical service provider and from the physical service provider to the DSS. In this paper, we set the distance between the farthest sensor to the physical service provider plus the distance between the physical service provider (e.g., FN $f_k$) to DSS $s_j$ to be $l_{kj}$. For simplicity, we assume the network resource from the physical service provider to DSS $s_j$ is always sufficient, and the cost incurred by the network delay $h_j$ generally follows a linear function of the distance from the sensor to the physical service provider plus the distance from the physical service provider to the DSS $s_j$, i.e., $h_j = \theta l_{kj}$, where $\theta$ is a scalar.

As the DSSs in the network pay DSOs for the service, following the structure of [7], the utility of DSS $s_j$, $\forall s_j \in \mathcal{Y}$, can be denoted as the revenue received from the workload data minus both the cost of service delay and payment to the DSOs, which can be expressed as follows.

$$W_j = \sum_{i=1}^{M} \tau_{ij} (\alpha_j \lambda_j - \beta_j q_j r_i - \gamma_j t_j),$$

where $\alpha_j$ denotes the revenue that DSS $s_j$ can obtain for unit received data rate. $\alpha_j \lambda_j$ indicates the total revenue for receiving the data with workload rate $\lambda_j$. $r_i$ is the price set by DSS $d_i$ for each unit of the virtualized CRB. Thus, when the DSS $s_j$ purchases $q_j$ CRBs from the DSO $d_i$, the total payment from the DSS $s_j$ to the DSO $d_i$ can be expressed as $q_j r_i$. $\beta_j$ is the weight factor indicating the importance of payment in the utility function of DSS $s_j$. $\gamma_j$ is weight factor indicating the importance of data service delay in the utility function of DSS $s_j$. We set $\tau_{ij}$ to be the boolean variable determining whether DSO $d_i$ serves DSS $s_j$ or not. If $\tau_{ij} = 1$, DSS $s_j$ is served by the DSO $d_i$. Otherwise, DSS $s_j$ prefers to be served by the other DSO. The value of $\tau_{ij}$ follows the preference list $L_j$ of DSS $s_j$, and each DSS can at most choose one DSS, i.e., $\sum_{j=1}^{M} \tau_{ij} = 1$, $\forall j \in \mathcal{Y}$. For each DSS, we assume there is an upper bound $t_{th}$ for the service delay. When the service delay is larger than the threshold, the DSS will regard it as an unsuccessful connection. Corresponding, we set $\tau_{ij}$ as the lower bound of CRBs required for DSS $s_j$ to guarantee the service delay within the acceptable thresholds.

Based on the number of virtualized CRB purchased by serving DSSs, the utility of each DSO is the revenue received from the DSSs minus the payment to the facilities that are able to provide the physical CRBs. Here, each DSO prefers to offload its services to the FNs nearby. However, if there are not sufficient available CRBs from the FNs which can meet the requirements of all DSSs. Some DSSs will be served by the remote data centers, which are located far away from the DSSs. We suppose the increment of the energy cost in the remote data center is $e_i$. Therefore, for the DSO $d_i$, if $q_{ik}$ CRBs are offloaded to the FN $f_k$, and $q_{ik}^o$ CRBs are still served by the remote data centers, the utility function of DSO $d_i$, $\forall i \in \Psi$, can be denoted as

$$W_i^d = \sum_{j=1}^{\mathcal{N}} \tau_{ij} (r_i q_{ij} - \sum_{k=1}^{K} p_k q_{ik}^f - e_i q_{ik}^o).$$

where $\sum_{j=1}^{\mathcal{N}} \tau_{ij} (r_i q_{ij})$ refers to the total revenue that DSO $d_i$ receives from DSSs for its data services. As $p_k$ is the price set by the FN $f_k$, which is determined by the cost and current traffic of FN $f_k$, $\sum_{k=1}^{K} p_k q_{ik}^f$ denotes the total payment from the DSO $d_i$ to all FNs. $e_i q_{ik}^o$ is the total cost for DSO $d_i$ to serve DSSs by itself.

For FN $f_k$ in the network, the utility is the payment received from the DSOs minus the transmission cost. We set $c_{kj}$ as the transmission cost for each unit CRB, which has a linear relationship with the distance $l_{kj}$. Moreover, we set $\eta_{ki}$ as the normalized preference to the DSO $d_i$. Accordingly, considering the preference to different DSOs, the discounted utility function for each DSO is

$$W_k^f = \sum_{k=1}^{K} \eta_{ki}^f (p_k - c_{kj}) q_{kj}^f,$$

where $\sum_{k=1}^{K} \eta_{ki}^f (p_k - c_{kj}) q_{kj}^f$ is the expected rewards received from DSOs, and $\sum_{i=1}^{\Psi} \eta_{ki}^f c_{kj} q_{kj}^f$ is the expected costs for serving DSSs in fog computing.

### III. Problem Formulation

According to the modeled architecture of fog network, with tradings between FNs and DSOs and between DSOs and DSSs, it is impossible to reach the maximum utilities for all FNs, MDCOs and DSSs simultaneously. Accordingly, we
Moreover, each FN in the network would like to choose its preferred DSOs and serve its DSSs with low distance. Thus, competing with other FNs, it is required to determine the number of CRBs allocated to DSOs and its serving DSSs, respectively. The optimization problem is denoted as

\[
\max_{q_{ik}, q_{kj}} W_k^f(q_{ik}, q_{kj} | q^s, q^{fs}_{i-k}, q^{fs}_{j-k}, p, \mathbf{r}) , \quad \forall k \in \Omega,
\]

s.t.,

\[
\sum_{i=1}^{M} q_{ik}^d \leq q_k^{d_{th}},
\]

\[
\sum_{j=1}^{N} q_{kj}^s \leq q_k^{s_{th}},
\]

\[
\sum_{k=1}^{K} q_{ik}^d \leq q_i^{d_{th}}, \quad \forall i \in \Psi,
\]

\[
\sum_{k=1}^{K} q_{kj}^s \leq q_j^{s_{th}}, \quad \forall j \in \Omega,
\]

where \(q_{ik}^d\) is the optimal number of CRBs rented to DSO \(d_i\) for all other FNs, \(q_{kj}^{fs}\) is the optimal number of CRBs allocated to DSS \(s_j\) for all other FNs. During the service, the total CRBs distributed to all DSOs or all DSSs cannot exceed its total available CRBs \(q_k^{d_{th}}\) for FN \(f_k\). Furthermore, the total CRBs purchased from DSO \(d_i\) or DSS \(s_j\) should not exceed its demand \(q_i^{d_{th}}\) or \(q_j^{s_{th}}\), respectively.

In summary, following the relationships among all FNs, DSOs and DSSs, we focus on the following problems:

1) **Resource purchasing problem for the DSSs:** In the network, as the DSSs can only access to the DSOs in a virtualized fashion, they are required to purchase the optimal number of CRBs from the DSOs. Following the system model, as different DSSs have different tolerances of service delay, when the upper bound of service delay is high, the DSSs are able to purchase a small number of CRBs to achieve satisfying services. However, when the upper bound of service delay is low, the DSSs should not purchase a large number of CRBs to guarantee the service delay is within the tolerated region. Moreover, the service price set by the DSOs affects the utility of DSSs. When the price is high, the DSSs may not be satisfying. Therefore, considering both the delay tolerance and setting prices of DSOs, the optimal number of CRBs should be determined for high utilities.

2) **Pricing problem for the DSOs:** In the data service with fog, the DSOs are required to provide virtualized CRBs to the DSSs and try to rent the CRBs from the FNs to serve DSSs in the physical network. Therefore, how to do the pricing for the DSOs is a problem. Considering the announced rent from all FNs, the DSOs need to set a price which can bring profits for themselves. However, if the price is set too high, the serving DSSs will not purchase a large number of CRBs. Therefore, predicting the reactions of CRBs and observing the rent of FNs, the DSOs are required to determine its service price so as to receive the maximum revenues.
3) **DSO-FN pairing problem**: As FNs may be private computing devices, which are small-scale and unable to communicate with DSSs directly, the FNs are accessible to the DSOs only. In the multi-DSO scenario, as the FNs are accessible to all DSOs, it is a problem for all DSOs in the network to pair FNs distributed so as to serve their DSSs with low latency. With different relations and trading history, each FN has different preference on all DSOs. Observing the rent of all FNs, each DSO also has preference on each FN. Based on the preferences of all DSOs’ and FNs’, it is required to reach a stable DSO-FN pairing result, where any DSO or FN is able to switch its current pairing result for a higher utility.

4) **FN-DSS pairing problem**: After the pairing between DSOs and FNs, each FN has allocated its CRBs to all DSOs. However, within one DSO, it is still a problem for the FNs to allocate their CRBs to all DSSs. As the distance between each FN and each DSS is various, with a longer transmission distance, the FNs have to pay more on the transmission cost. Thus, based on the transmission distance, each FN has a preference over all DSSs. Moreover, each DSS also have preference over FNs based on the rent. Therefore, following the preference of all FNs’ and DSSs’, a stable FN-DSS pairing result should be achieved.

According to the formulated problems, all FNs, DSOs and DSSs are rational and autonomous individuals, who observe the behaviors of others and make decisions to improve their own utilities. Therefore, in order to reach the optimal and stable solutions for all FNs, DSOs and DSSs, we model a three-stage joint optimization framework, as shown in Fig. 2. In the framework, we first model the Stackelberg games to solve the pricing problem for the DSOs and resource purchasing problem for the DSSs. When the DSOs know the expected amount of resource purchased by the DSSs, a many-to-many matching is proposed between the DSOs and the FNs to deal with the DSO-FN pairing problem. Finally, within the same DSO, we apply another many-to-many matching between its paired FNs and serving the DSSs to solve the FN-DSS pairing problem.

**IV. System Analysis**

In this section, we analyze the optimal strategies for FNs, DSOs and DSSs. Based on the analysis of the formulated framework, in the following sub-sections, we first investigate the interactions between the DSOs and DSSs to determine how many CRBs are required during the service. Given the optimized behaviors of the DSOs and DSSs, we analyze the interactions between the FNs and DSOs based on different preferences. Finally, with the obtained results, we discuss the interactions between the FNs and DSSs within the same DSO for better services.

**A. The Interaction between DSOs and DSSs**

In the virtualized network, the DSOs provides CRBs for the DSSs. Following the formulated problems for both DSOs and DSSs, there is a Stackelberg game, where the DSOs act as leaders and DSSs act as followers. In the game, when all DSSs choose their serving DSOs with their preferences, the DSO first sets the service price. Then, based on the price all DSSs determine optimal number of CRBs to purchase. Accordingly, considering the optimization problem of DSSs, we have the following lemma.

**Lemma 1.** In the modeled Stackelberg game between DSO $d_i$ and DSS $s_j$, when the DSO announces its service price $r_i$, the optimal number of CRBs $q_j$ purchased by the DSS is

$$q_j = \frac{\lambda_j}{\mu \sqrt{r_i \beta_j^2 / \gamma_j}} + \frac{\lambda_j}{\mu}$$  \hspace{1cm} (9)

**Proof:** According to the utility function of DSS $s_j$ 3, the second derivative of $W_j^s$ with respect to $q_j$ is

$$\frac{\partial^2 W_j^s}{\partial q_j^2} = -\frac{2 \lambda_j^2 \mu}{(\mu q_j - \lambda_j)^3}.$$  \hspace{1cm} (10)

As $\frac{\partial^2 W_j^s}{\partial q_j^2} < 0$, $W_j^s$ is a quasi-concave function with respect to $q_j$. Furthermore, the first derivative of $W_j^s$ with respect to $q_j$ is

$$\frac{\partial W_j^s}{\partial q_j} = \left( \frac{\lambda_j}{\mu b_j - \lambda_j} \right)^2 - r_i \frac{\beta_j^2}{\gamma_j}.$$  \hspace{1cm} (11)

We set the first derivative equal to zero and obtain the optimal number of CRBs to purchase so as to achieve the maximum utility, i.e.,

$$q_j^* = \frac{\lambda_j}{\mu \sqrt{r_i \beta_j^2 / \gamma_j}} + \frac{\lambda_j}{\mu}$$  \hspace{1cm} (12)

Therefore, considering the reactions of the DSSs, we adjust the optimization problem for the DSO $d_i$, $\forall i \in \Psi$, as follows.

$$\max_{r_i} \widehat{W}_i^d(r_i|q^*, p, \textbf{r}_s^*), \quad \forall i \in \Psi,$$

$$\text{s.t.} \quad \sum_{j=1}^{N_i} \tau_{ij}(\frac{\lambda_j}{\mu} \sqrt{\frac{\beta_j^2}{\gamma_j} r_i} + \frac{\lambda_j}{\mu} \sqrt{\frac{\beta_j^2}{\gamma_j} r_i}) \geq \sum_{k=1}^{K_i} p_k q_{ik}^d + e_i q_{io}, \quad r_i \geq 0,$$

where

$$\widehat{W}_i^d = \sum_{j=1}^{N_i} \tau_{ij}(\frac{\lambda_j}{\mu} \sqrt{\frac{\beta_j^2}{\gamma_j} r_i} + \frac{\lambda_j}{\mu} \sqrt{\frac{\beta_j^2}{\gamma_j} r_i}) - \sum_{k=1}^{K_i} p_k q_{ik}^d - e_i q_{io}.$$  \hspace{1cm} (13)

In the formulated problem (13), we take the first derivative of $\widehat{W}_i^d$ with respective to $r_i$ and discover it is a monotonous increasing function with respective to $r_i$. Furthermore, as the service delay cannot surpass $t_{th}$ for the DSSs. The CRB purchased by DSSs has the following low bound

$$q_j \geq \frac{\lambda_j t_{th}}{\mu t_{th} - \lambda_j}.$$  \hspace{1cm} (15)

Thus, following the relation in (9), the maximum and optimal price for the DSO $d_i$ to DSS $s_j$ is, $\forall i \in \Psi, \forall j \in \Upsilon,$
\[ r_i = \frac{\gamma_j}{\beta_j} \left( \frac{\mu_{th} - \lambda_j}{\lambda_j} \right)^2. \]  

(16)

**B. The Interaction between FNs and DSOs**

According to the predicted number of CRBs ordered by the DSSs, the DSOs try to offload the services from the massive data centers to the DSSs nearby. Observing the service prices set by all FNs, DSO \( d_i \), \( i \in \Psi \), has a preference list \( L^d_{ik} = [L^d_{i1k}, \ldots, L^d_{iMk}] \) on all FNs in the neighborhood. As the DSOs prefer to choose the FNs with a low price, we set

\[ L^d_{ik} = -p_k, \quad \forall i \in \Psi, \forall k \in \Omega. \]  

(17)

Furthermore, each FN also has different preferences over DSOs. Thus, we set the preference list of FN \( f_k, \forall k \in \Omega \), on all DSOs as \( L^f_{ik} = [L^f_{i1k}, \ldots, L^f_{iMk}] \), satisfying,

\[ L^f_{ik} = \eta^f_{ik}, \quad \forall i \in \Psi, \forall k \in \Omega. \]  

(18)

Considering the preference lists of FNs and DSOs, i.e., \( L^d_{ik} \) and \( L^f_{ik} \), respectively, we design a many-to-many matching algorithm for the DSO-FN pairing problem. As shown in Algorithm 1, after the preference lists are constructed, we set a pointer for each FN in its preference list. Initially, we set the pointer at the first DSO in the list. During the each round of matching, each FN first propose to the DSO in the pointer of the preference list. Based on behaviors of the FNs, each DSO chooses its most preferred FNs in its preference list until the required CRBs are all supplied by the FNs. If the FNs supply more CRBs than the DSO requires, the DSO will reject the CRBs from less favourite FNs. If the FNs supplies less CRBs than the DSO requires, the DSO will choose massive data centers for the rest of the services. At the end of each round, if all of the CRBs from the FN have been allocated to the DSOs, the pointers of the FN will keep unchanged. Otherwise, the pointers of the rejected FNs will move to the next DSO in the preference list. In the next round, the FNs which still have available CRBs will propose to the new DSOs according to their pointers. Specifically, if the CRBs of FN are chosen by the DSO in the last round, but discarded in the current round, we suppose the pointer of the FN doesn’t change its position, considering the pointed DSO in the current round may need more CRBs from the FN. The matching repeats in circulations until all the pointers of the FNs have moved out their preference list. According to the algorithm, we have the following lemmas.

**Lemma 2.** For each FN in the matching algorithm, the pointer of the FN in its preference list moves in one direction. In other words, when its pointer has moved to the next DSOs in the preference list, whatever the matching results of other FNs, the FN cannot achieve a higher utility by moving the pointer back.

**Proof:** As shown in Algorithm 2, when the DSOs determines which FNs to choose, they choose the CRBs from the most preferred amount. If some CRBs from FN \( f_k \) is discarded by DSO \( d_i \), the current accepted CRBs belong to the FNs which is more preferred than FN \( f_k \). In the future rounds, when there are new FNs proposing to DSO \( d_i \), if the DSO would like to change its current accepted FNs, the DSO can only choose the FNs that is even better than the FNs in the current accepted list. Therefore, for FN \( f_k \) which has been rejected once from DSO \( d_i \), it is impossible for it to be accepted by the same DSO in the future rounds.

**Lemma 3.** Following the Algorithm 2, the DSO-FN pairing problem will ultimately converge and achieve a stable matching result.

**Proof:** As proved in the Lemma 2, the pointer of the FNs can only move in one direction. Therefore, in the perspective of the FNs, when the pointer of each FN has moved to the end of the preference list, the FN has distributed its available

---

**Algorithm 1 Many-to-Many Matching Algorithm for DSO-FN pairing problem.**

1: for FN \( f_k \) do
2: Construct the preference list \( L^d_{ik} \) on all DSOs according to (18);
3: One pointer is set as the indicator pointing at the first DSO in the preference list;
4: end for
5: for DSO \( d_i \) do
6: Construct the preference list \( L^f_{ik} \) on all FNs according to (17);
7: end for
8: We set flag\(_k\), \( \forall k \in \Omega \), as the indicator to show if the CRBs of FN \( f_k \) were chosen by the DSO in the last round, but discarded in the current round. Initially, flag\(_k\) = 1;
9: while the pointers of all FNs have not scanned all the DSOs in their preference list do
10: FNs propose to DSOs with their service price;
11: for FN \( f_k \) who still have available FNs to purchase do
12: if flag\(_k\) = 1 then
13: The pointer keep current position in the list;
14: else
15: The pointer moves to the next position in the list;
16: end if
17: The FN proposes to pointed DSO in its preference list with its available FNs;
18: We set flag\(_k\) = 0;
19: end for
20: DSOs determines which FNs to choose;
21: for DSO \( d_i \) do
22: if The total available number of CRBs proposed by the FNs exceed the requirements then
23: The DSO \( d_i \) chooses the most preferred number of CRBs from FNs, and rejects the rest;
24: For CRBs of the FN \( f_k \) which is chosen by the DSO in the last round, but rejected in the current round, we set flag\(_k\) = 1;
25: end if
26: end for
27: end while
CRBs to the DSOs in an optimized way. In other words, the FN is unable to change its pairing results unilaterally for higher utilities. Furthermore, in the perspective of the DSOs, when the pointers of all FNs have moved to the end of the preference lists, each DSO has evaluated all proposals from the available FNs. Therefore, it also cannot unilaterally change its pairing results, get accepted by the FNs and achieve a higher utility for itself. Furthermore, according to [10], when every agents preference list is substitutable, a pairwise stable matching always exists. Based on the above, the DSO-FN pairing will ultimately converge and achieve a stable matching result in Algorithm 2.

C. The Interaction between FNs and DSSs

When the CRBs from FNs have been rent to all DSOs, within each DSO, how to allocate the CRBs to all DSSs still remains a problem. According to the rent of all FNs, DSS $s_j$, $\forall j \in T$, has a preference list $L_j^{sf} = [L_j^{s_1}, \ldots, L_j^{s_K}]$ on all FNs in the neighborhood, satisfying,

$$L_{kj}^{sf} = -p_k, \quad \forall j \in T, \forall k \in \Omega. \quad (19)$$

Furthermore, according to the utility function of the FN, the distance between the FN and its serving DSS affect the revenues the FN received. With a longer distance, the FN has to pay more for the data transmission in the network. Therefore, we set a preference list $L_i^{sf} = [L_i^{s_1}, \ldots, L_i^{s_K}]$ for FN $f_k, \forall k \in \Omega$, over all DSSs, i.e.,

$$L_{kj}^{sf} = -l_{kj}, \quad \forall k \in \Omega, \forall j \in T. \quad (20)$$

Considering the preference lists of DSSs and FNs, i.e., $L_{kj}^{sf}$ and $L_{kj}^{sf}$, respectively, we design a many-to-many matching algorithm for the FN-DSS pairing problem within DSO $d_i$, $\forall i \in \Psi$. As shown in Algorithm 2, after the preference lists are constructed, we set a pointer for each DSS in its preference list. Initially, we set the pointer at the first FN in the list. During the each round of matching, each DSS first proposes to the FN in the pointer of the preference list. Based on behaviors of the DSSs, each FN chooses its most preferred DSSs in its preference list until the maximum CRBs available in the DSO $d_i$ are reached. If the DSSs request more CRBs than the FN can supply, the FN will reject the less favourite DSSs. At the end of each round, if the DSSs have been allocated all of its requested CRBs from the FNs, the pointers of the DSS will keep unchanged. Otherwise, the pointers of the rejected DSS will move to the next FN in the preference list. In the next round, the DSSs which require CRBs will propose to the new FNs according to their pointers. Specifically, if some CRBs of FN are allocated to the DSS in the last round, but changed to other DSSs in the current round, we suppose the pointer of the DSS in the next round doesn’t change its position, considering the pointed FN in the current round may be able to supply more CRBs to the DSS. The matching repeats in circulations until all the pointers of the DSSs have moved out their preference list. Following Lemma 2 and Lemma 3 in a similar way, the FN-DSS pairing problem can ultimately achieve a stable matching result.

**Algorithm 2** Many-to-Many Matching Algorithm for FN-DSS pairing problem.

1. for DSS $s_j$ do
2. Construct the preference list $L_j^{sf}$ on all DSOs according to (19);
3. One pointer is set as the indicator pointing at the first FN in the preference list.
4. end for
5. for FN $f_k$ do
6. Construct the preference list $L_k^{sf}$ on all FNs according to (20);
7. end for
8. We set $flag_j, \forall j \in T$, as the indicator to show if the FNs allocate CRBs to the DSS $s_j$ in the last round, but adjusted in the current round. Initially, $flag_j = 1$;
9. while the pointers of all DSSs have not scanned all the FNs in their preference list do
10. DSSs propose to FNs for their services;
11. for DSS $s_j$ which has not been allocated required CRBs do
12. if $flag_j = 1$ then
13. The pointer keep current position in the list;
14. else
15. The pointer moves to the next position in the list;
16. end if
17. The DSS proposes to pointed FN in its preference list for its data services;
18. We set $flag_j = 0$;
19. end for
20. Each FN determines which DSSs to choose;
21. for FN $f_k$ do
22. if The total available CRBs requested by the DSSs exceed the available volume then
23. The FN $f_k$ allocate the CRBs to the most preferred DSSs, and rejects the rest;
24. For CRBs allocated to the DSS $s_j$ in the last round, but adjusted in the current round, we set $flag_j = 1$;
25. end if
26. end for
27. end while

Based on the above, the optimal strategy of each DSO, FN or DSS is a function with respect to the network information, e.g., channel state information and queue state information. When the network information changes dynamically, based on the function of optimal strategy, each DSO, FN or DSS is able to adjust its strategy immediately to achieve optimal utility. However, in practice, the sensing process for network information may be performed periodically, while the time within each period is very short and we can assume the network information as constant. Accordingly, each DSO, FN or DSS is able to iteratively update its optimal strategy to achieve high utility based on the observed network information within each period.
V. Simulation Results and Discussions

In this section, we present simulation results to evaluate our proposed framework with MATLAB. In the simulated IoT scenario, without specific explanation, there are 120 DSSs, 4 DSOs and 20 FNs allocated randomly in a circle district with a diameter of 10 kilometers. As we focus on the IoT scenarios where DSSs are closely located with its sensors, without loss of generality for the methods in this paper, we suppose the sensors of each DSS are located at the same position with the DSS. We follow the settings in [11] and set the service rate of each computing resource block is set to 50 km/ms. The delay tolerance of all DSSs is set to 60 ms. Furthermore, for each FN, we set its preference to each DSO as a random number satisfying the uniform distribution between [0, 1]. Based on the usage of its computing resources and its service cost, we set the announced rent as a random number averaged 0.5 (ms)$^{-1}$. The data transmission speed is 50 km/ms. The delay tolerance of all DSSs is set to 60 ms. Furthermore, for each FN, we set its preference to each DSO as a random number satisfying the uniform distribution between [0, 1]. Based on the usage of its computing resources and its service cost, we set the announced rent as a random number averaged 0.5 (ms)$^{-1}$. The data transmission speed is 50 km/ms. The delay tolerance of all DSSs is set to 60 ms. Furthermore, for each FN, we set its preference to each DSO as a random number satisfying the uniform distribution between [0, 1]. Based on the usage of its computing resources and its service cost, we set the announced rent as a random number averaged 0.5 (ms)$^{-1}$. The data transmission speed is 50 km/ms. The delay tolerance of all DSSs is set to 60 ms. Furthermore, for each FN, we set its preference to each DSO as a random number satisfying the uniform distribution between [0, 1].

As shown in Fig. 3, we evaluate the utility of all FNs when the number of DSSs increases. When the number of FNs is fixed, we discover with the number of DSSs increasing, the utility of all FNs generally increases, and the increasing speed first increases then gradually decreases to zero. The reason is that when the number of DSSs increases, but the number of FNs is fixed, the FNs can serve more favourable DSSs with a low transmission distance. However, when all of the available CRBs of FNs are allocated to the DSSs nearby and the number of DSS keep increasing, the DSS will be allocated with CRBs from the massive data centers. Thus, the total utility of the FNs stop increasing. Nevertheless, when we increase the number of FNs, we discover that with more FNs, the utility can converge to a higher bound ultimately. Furthermore, because of the competition between FNs, when the number of DSS is small, with more FNs, the increasing speed is smaller.

In Fig. 4, we consider the utility of all DSSs with the number of DSSs increasing. In the simulation, we compare the performance of the DSSs in our proposed framework with the performance of the DSSs which is served by the massive data centers only. With the same amount of workload, we discover that when the number of DSSs increases, the utility of DSSs generally increases, and the utility with FNs achieve a higher value than the one without FNs. Furthermore, when the number of DSSs is fixed and the average workload for each DSS increases, the DSSs are able to receive more revenues from the services. Thus, the utility of DSSs increases.

In Fig. 5, we analyze the relation between the utility of FNs and average workload arrive rate for DSSs. As shown in the figure, when the number of FNs is fixed and the average value of workload $\lambda$ increases, the utility of FNs first increases then gradually converge to a fixed value. The reason is that when the workload of all DSSs increases, the FNs are able to allocate more of its CRBs to the DSSs nearby. However, when all the available CRBs of FNs are allocated to the DSSs, the utility of the FNs stops improving and converges to one specific value. When the number of FNs increases, with the same value of $\lambda$, as the FNs are able to provide more CRBs to the DSSs, the converged value is higher.

In Fig. 6, we observe the utility of all DSOs when the value of $\mu$ increases. As shown in the figure, when $\mu$ increases,
Utility of DSOs

In the literature, fog computing has been advocated to be beneficial for the IoT. In [23], the authors overview the opportunities and challenges of fog, especially the applications of fog computing in IoT. In [24], the authors devise the method of MEdia FOg Resource Estimation (MeFoRE), to provide resource estimation based on the service give-up ratio and to enhance QoS based on the previous QoE. [25] addresses the utility based pairing problem between the fog nodes and IoT devices with the Irving’s matching algorithm. In [26], the authors propose a distributed dataflow programming model for IoT devices to optimize resource allocation on computing infrastructures across the fog and the cloud. In [27], the authors consider issues of resource prediction, customer type based resource estimation and reservation, advance reservation, and pricing in the fog computing for IoTs. [28] considers the requirements of mobility, scalability, reliable control and actuation in some challenging scenarios of IoT to show the benefits and significance of fog computing. Considering the advantages of fog computing, the authors in [29] discuss and propose a procedure to be implemented in smart phones for UV measurement.

In order to solve the resource management problems in a network system with a distributed fashion, game theory has been shown as a powerful tool [30]. In the literature, most of the cases, the network system is normally modeled as a bipartite or a multi-tier graph. Based on this model, in [31], a Stackelberg game theoretic model was shown for dynamic bandwidth allocation between virtual networks. In [32], the authors consider a Stackelberg game between data center and buses in the smart city, where each buses collect data along its route and compete with other buses for the reward forwarding to the data center. In the game, following the proposed heuristic algorithm, the Stackelberg equilibrium is shown to be achieved where the data center and each bus are

Each DSS is able to be served with a less number of CRBs. Therefore, the DSO is able to set a higher price to the DSSs and receive high revenues. Moreover, when the number of DSSs increases, as the DSO is able to serve more DSSs at the same time, the total utility of DSOs also increases.

In Fig. 7, we evaluate the relationship of utilities and the tolerance service delay of DSSs. As shown in the Fig. 7a, when the value of \( t_{th} \) increases, each DSS is able to be served with less CRBs. Thus, the FNs will supply less CRBs in the network, and the utility of the FNs generally decreases. In Fig. 7a, as the DSO is able to set a high price for its services, with the value of \( t_{th} \) increasing, the utility of DSO generally increases. Moreover, in Fig. 7c, even though the DSS is able purchase less CRBs with high \( t_{th} \), the price of CRBs set by the DSOs also increases. Furthermore, as the DSS suffers a lot with high delay, the utility of DSS generally decreases with the value of \( t_{th} \) increasing.

VI. RELATED WORKS

In the literature, fog computing has been advocated to be the promising future of the cloud. The concept of pulling the cloud closer to the users has been widely considered in previous work. In [12], the authors put forward the concept of mist computing, aiming to distribute the cloud and its benefits deep into the network. In [13], the deployment of edge cloud was proposed. From the DSSs’ side, the edge cloud was able to surrogate the requirements and simplify the management of the network. From the servers’ side, the edge cloud can exploit content and support service delivery in an efficient way. Without deploying massive data centers with high cost and latency, [14] on the other hand strengthened the importance of small distributed data center designs. The authors took email delivery as an example and showed the advantages of geo-diversity characteristics of micro data centers. In [15], a novel and distributed computing platform, called nano data centers was proposed. The authors evaluated the energy consumption of nano data centers and showed a significant improvement on energy efficiency, compared with the traditional data centers. In [16], the authors propose an intermediary framework where there exists an intermediary between multiple cloud providers and users. The intermediary first rents the cloud service from cloud providers and then provides streaming processing service to users with low cost and delay. In [17], the authors propose a deduplication-based energy efficiency storage system for VM storage (EEVS) and implement it with existing cloud platform. In the EEVS, an online-deduplication mechanism is designed to decrease the redundant data without service interruptions, and a deduplication selection algorithm is introduced to minimize the storage energy consumption with limited computing resources for deduplication. In [18], considering the existing telecommunication and Internet service providers, the authors showed that it was required and beneficial to leverage the existing infrastructure and provide value-added services with FNs. In [19], the authors outlined the vision of fog computing and overviewed the important features of fog computing. In [20], the network optimization with fog computing was considered. As the data centers were aware of the locations of DSSs with fog computing, dynamic adaptation of computing resources to the DSSs’ conditions was proposed. In [21], the authors compared the cloud computing with the fog computing and showed some significant characteristics of fog, which was required for current data services. In [22], the authors elaborated the role of fog computing in six important scenarios and surveyed the security issues with fog computing. Moreover, fog computing has been widely considered to be beneficial for the IoT. In [23], the authors overview the opportunities and challenges of fog, especially the applications of fog computing in IoT. In [24], the authors devise the method of MEdia FOg Resource Estimation (MeFoRE), to provide resource estimation based on the service give-up ratio and to enhance QoS based on the previous QoE. [25] addresses the utility based pairing problem between the fog nodes and IoT devices with the Irving’s matching algorithm. In [26], the authors propose a distributed dataflow programming model for IoT devices to optimize resource allocation on computing infrastructures across the fog and the cloud. In [27], the authors consider issues of resource prediction, customer type based resource estimation and reservation, advance reservation, and pricing in the fog computing for IoTs. [28] considers the requirements of mobility, scalability, reliable control and actuation in some challenging scenarios of IoT to show the benefits and significance of fog computing. Considering the advantages of fog computing, the authors in [29] discuss and propose a procedure to be implemented in smart phones for UV measurement.

In order to solve the resource management problems in a network system with a distributed fashion, game theory has been shown as a powerful tool [30]. In the literature, most of the cases, the network system is normally modeled as a bipartite or a multi-tier graph. Based on this model, in [31], a Stackelberg game theoretic model was shown for dynamic bandwidth allocation between virtual networks. In [32], the authors consider a Stackelberg game between data center and buses in the smart city, where each buses collect data along its route and compete with other buses for the reward forwarding to the data center. In the game, following the proposed heuristic algorithm, the Stackelberg equilibrium is shown to be achieved where the data center and each bus are
able to reach maximum utility. In [33], the authors formulate a Stackelberg game for power allocation of data centers in the cloud. In the game, the power grid controller acts as the leader and sets prices of the provided energy based on the current amount of renewable energy and costs. Observing the prices, the cloud controller, i.e., the follower, determines the optimal amount of energy to purchase and do resource allocation for its data centers. With backward induction, the near-optimal strategies of both players in the game can be achieved. In [34], the authors model the interaction between the monopolistic data center operator and the customers as a Stackelberg game. In the game, the pricing strategies of the monopolistic data center operator the corresponding behavior of data service customers is detailedly analyzed in both homogeneous and heterogeneous customer scenarios. In [35], the authors adopted the Stackelberg game to solve the problem of minimizing energy consumption in the data center networks. In [36], the authors proposed a multi-leader multi-follower Stackelberg game to address and cooperation problems among Wi-Fi, small cell and macrocell networks. In [37], the authors combined the Stackelberg game and the bargaining to design a resource allocation problem in a multi-tier LTE unlicensed network. Furthermore, the auction mechanisms are also powerful tools to solve the problem. In [38]–[40], the resource management problem could be perfectly optimized, but it requires high communication and computation overhead. In [41], the authors adopted the generic game theoretic framework to identify important edges in the context of k-edge connectivity between certain pairs of nodes in a general given network. In [42], the graphical game was put forward to analyze the optimized behaviors of each node in a general graph.

VII. CONCLUSIONS

In this paper, we proposed a joint optimization framework in the multi-FN, multi-DSO and multi-DSS scenario for IoT fog computing. In the framework, we first modeled the Stackelberg games to solve the pricing problem of the DSOs and resource purchasing problem of the DSSs. Then a many-to-many matching was proposed between the DSOs and the FNs to deal with the DSO-FN pairing problem. Finally, we applied another many-to-many matching between its paired FNs and serving DSSs to solve the FN-DSS pairing problem within the same DSO. For each stage of the problem, all participants were able to achieve the equilibrium or stable results where no one was able to change its behavior unilaterally for a higher utility. Simulation results showed that all FNs, DSOs and DCOs were able to reach optimal utilities for themselves, and high performance of the proposed framework could be achieved compared with the data services without fog nodes. For the future work, firstly the dynamic computing resource allocation problem can be considered in the three-tier IoT fog network with dynamic Stackelberg game, where each DSO is able to predict its future demands and to rent the computing resources of FNs in advance. Secondly, the analysis of cooperative and competitive behaviors among FNs may provide groupings strategies for FNs to achieve higher revenues. Correspondingly, the effective strategies are required for each DSO to prevent the severe competition for some FNs with other DSOs.

REFERENCES


Shengrong Bu received the PhD degree in electrical and computer engineering from Carleton University in 2012, and held a research position at Huawei Technologies Canada Inc, Ottawa as a NSERC IRDF until 2014. She is now a Lecturer (Assistant Professor equivalent) with the School of Engineering at the University of Glasgow, Scotland. Her research interests include energy-efficient networks, smart grids, big data analytics, wireless networks, wireless network security, cloud computing, game theory and stochastic optimization. She received three best paper awards at IEEE International conferences. Highlights of her professional activities include duties as an editor for IEEE TCGCC NewsLetters, an associate editor for the journal Springer Wireless Networks, TPC co-chair for six international workshops or conference symposiums, and duties as N2Women mentoring Co-Chair. She has sat on the TPC for more than twenty leading international conferences and workshops and served as a reviewer for more than ten leading journals.

Dusit Niyato (M’09-SM’15-F’17) is currently an Associate Professor in the School of Computer Science and Engineering, at Nanyang Technological University, Singapore. He received B.Eng. from King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand in 1999 and Ph.D. in Electrical and Computer Engineering from the University of Manitoba, Canada in 2008. His research interests are in the area of energy harvesting for wireless communication, Internet of Things (IoT) and sensor networks.

Zhu Han (S’01-M’04-SM’09-F’14) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively. From 2000 to 2002, he was an R&D Engineer of JDSU, Germantown, Maryland. From 2003 to 2006, he was a Research Associate at the University of Houston, Texas. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. Dr. Han received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the Journal on Advances in Signal Processing in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (best paper award in IEEE JSAC) in 2016, and several best paper awards in IEEE conferences. Currently, Dr. Han is an IEEE Communications Society Distinguished Lecturer.

F. Richard Yu (S00-M04-SM08) received the PhD degree in electrical engineering from the University of British Columbia (UBC) in 2003. From 2002 to 2006, he was with Ericsson (in Lund, Sweden) and a start-up in California, USA. He joined Carleton University in 2007, where he is currently a Professor. He received the IEEE Outstanding Service Award in 2016, IEEE Outstanding Leadership Award in 2013, Carleton Research Achievement Award in 2012, the Ontario Early Researcher Award (formerly Premiers Research Excellence Award) in 2011, the Excellent Contribution Award at IEEE/IFIP TrustCom 2010, the Leadership Opportunity Fund Award from Canada Foundation of Innovation in 2009 and the Best Paper Awards at IEEE ICC 2014, Globecom 2012, IEEE/IFIP TrustCom 2009 and Int’l Conference on Networking 2005. His research interests include cross-layer/cross-system design, security, green ICT and QoS provisioning in wireless-based systems.

He serves on the editorial boards of several journals, including Co-Editor-in-Chief for Ad Hoc & Sensor Wireless Networks, Lead Series Editor for IEEE Transactions on Vehicular Technology, and IEEE Transactions on Green Communications and Networking, IEEE Communications Surveys & Tutorials. He has served as the Technical Program Committee (TPC) Co-Chair of numerous conferences. Dr. Yu is a registered Professional Engineer in the province of Ontario, Canada, and a senior member of the IEEE. He serves as a member of Board of Governors of the IEEE Vehicular Technology Society.