Cost-Efficient Resource Provision for Multiple Mobile Users in Fog Computing

Shuaibing Lu*†, Jie Wu†, Yubin Duan†, Ning Wang‡, and Juan Fang*
*Faculty of Information Technology, Beijing University of Technology, Beijing, China
†Center for Networked Computing, Temple University, USA
‡Department of Computer Science, Rowan University, USA
Email: lushuaibing@bjut.edu.cn, {jiewu, yubin.duan}@temple.edu, wangn@rowan.edu, fangjuan@bjut.edu.cn

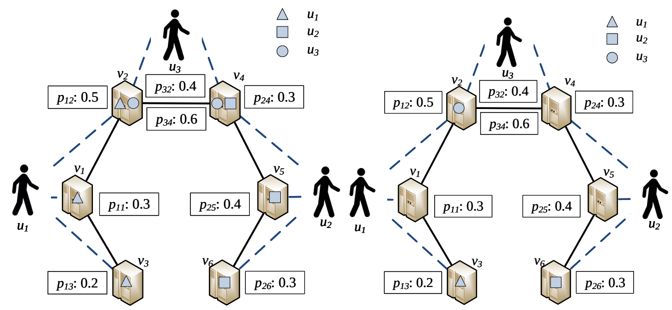
***Abstract*—Fog computing is an emerging paradigm that brings the computing capabilities close to distributed IoT devices, which provides networking services between end devices and traditional cloud data centers. One important mission is to further reduce the monetary cost of fog resources while meeting the ever-growing demand of multiple users. In this paper, we focus on minimizing the total cost for multiple mobile users to provide an efficient resource provisioning scheme in fog computing. The total cost includes two aspects: the replication cost and the transmission cost. We consider two cases for the resource provision problem by focusing on different cost models. First, one simple case where users can only upload one replication is discussed, and an optimal solution is proposed by converting the original problem into one of bipartite graph matching. Then we consider a more complicated case that each user can upload multiple replications on fog nodes in the resource provisioning. For different transmission cost models, the transmission cost is related to the distance of each pair of fog nodes. This problem is proven to be NP-hard. We first propose a non-adaptive algorithm which is proved to be bounded by $\frac{2}{3}W + \frac{1}{3}OPT$. Another $3+\epsilon$-approximation algorithm is proposed based on local search, which has better performance with higher complexity. Extensive simulations also prove the efficiency of our schemes.**

***Index Terms*—Fog computing, multiple users, resource provision, mobility, cost efficiency.**

I. Introduction

Fog computing, which is defined as a distributed computing infrastructure containing a bunch of high-performance physical machines whose computing, storage, and networking services are well connected with each other, is an emerging computing paradigm that brings the computing capabilities close to distributed IoT devices [1]. Due to the repaid generation of an unprecedented volume and variety of data, the demand for high-quality mobile services has been increasing, and how to realise the resource provision that eases the monetary cost for mobile users becomes the key issue of the fog computing. To date, the smartphone penetration in the U.S. has reached 80%. As predicted by Cisco, the average number of connected devices per person will reach 6.58 in 2020 [2].

Due to the increasing number of connected devices on the fog nodes, poor resource provision will result in high cost and heavily unbalanced loads among fog nodes. In this paper, we focus on the resource provision problem for mobile users under the capacity constraints of fog nodes, while realizing the cost efficiency of network operators in fog computing. Our objective is to find a feasible provision scheme that minimizes the total monetary cost for the users under the



(a) Transmission cost minimization. (b) Replication cost minimization.

Fig. 1. Illustration of two extreme assignments.

capacity constraints, and the cost is divided into two parts: replication cost and transmission cost. The replication cost is the offloading cost of users that place replications on the fog nodes. The transmission cost is the bandwidth cost associated with data movement, which only occurs when there is no replication on the real location of the user. We assume that there is no capacity limitation on communication resource, and thus the transmission cost is measured by the shortest distance between real location and other fog nodes.

A. Motivation and Challenges

We give an example that motivates our work in this paper. Some assumptions and notations are not explicitly stated and will be explained in the later section. As shown in Fig 1, there are five fog nodes v_1 to v_5, and there are three users u_1, u_2 and u_3. We define the daily route of each user as the activity track. According to the activity tracks of these three users, these six fog nodes are divided into three sets $U_1 = \{v_1, v_2, v_3\}$, $U_2 = \{v_2, v_4\}$ and $U_3 = \{v_4, v_5, v_6\}$. The probability of user k is represented by the frequency of u_k that stays at that fog node. We use p_{kj} to denote the probability of user k at the j^{th} fog node. For example, $p_{11} = 0.3$, $p_{12} = 0.5$, and $p_{13} = 0.2$, which means that the probability of user u_1 appearing at three fog nodes is respectively 0.5, 0.3, and 0.2.

As shown in Fig. 1(a), one extreme assignment is the solution with minimal transmission cost for users u_1, u_2 and u_3. We offload the replications of the workload on all fog nodes that users are connected to in the graph. For this case, users can do the operation on any fog node where they may stay, and the total cost of the user is the sum of the replication costs. However, the total cost under this solution is the maximum among all possible assignments if the replication cost is extremely high. Another extreme assignment

TABLE I
NOTATIONS

G	Substrate topology of fog nodes, where $G = \{V, E\}$.
V	Set of fog nodes in G, where $V = \{v_i\}$, $1 \leq i \leq n$.
v_i	The i^{th} fog node in V.
E	Set of connections of fog nodes in G, where $E = \{e_{ij}\}$.
e_{ij}	The connection between fog nodes v_i and v_j in E.
U	Set of users, where $U = \{u_k\}$, $1 \leq k \leq n$.
u_k	The k^{th} user in U.
G_k	The activity track of u_k, $G_k = \{V_k, E_k\}$, where $G_k \subseteq G$.
p_{ki}	The probability of u_k on v_i.
r_{ki}	The replication cost of u_k on v_i.
$f_R(u_k)$	Replication cost of u_k, where $f_R(u_k) = \sum_{i \in X_k} r_{ki}$.
$\sigma(v_i)$	The closest fog node of v_i
t_{ij}	The transmission cost between v_i and v_j, where $v_j = \sigma(v_i)$.
$f_T(u_k)$	Transmission cost of u_k, $f_T(u_k) = \sum_{i \in \{V_k \backslash X_k\}} p_{ki} \cdot t_{ij}$.
$f_C(u_k)$	The total cost of u_k.
X_k	Set of fog nodes with replications of u_k, where $X_k \subseteq V$.

is to minimize the replication cost, which means reducing the number of replications on the fog nodes. In Fig. 1(b), we only offload one replication on one fog node for each user. Take u_2 as an example: if the real location of u_2 is v_4, the total cost will be the sum of the replication cost and the transmission cost between v_4 and v_6. Thus, when the transmission cost is large, the total cost of users is extremely high for this solution.

This problem is non-trivial mainly due to the following challenges: (i). Each user has its own activity track. Therefore, to describe the trajectory and formulate the impact of replications on user's total cost is non-trivial. (ii). Users can move between fog nodes in its own fog node set, and the probability of staying at each fog node is different. How to find a feasible provision of the users that can realize the cost minimization within the limited replicate copies is non-trivial. (iii). We need to balance the trade-off between the replication cost and the transmission cost, while guaranteeing the users demand with lowest costs. There is a trade-off between the replication cost and the transmission cost, which means that more replications on fog nodes can reduce the transmission cost of users while increasing the replication cost and the chance of exceeding the capacity constraints.

B. Contributions and Paper Organization

In this paper, we focus on the resource provision problem for multiple users under the capacity constraints while realizing the cost efficiency of network operators in fog computing. Our contributions can be summarized as follows:

- We consider the resource provision problem for users with cost minimization and model the identified problem by considering two types of cost: replication cost and transmission cost. We first discuss a simple case where users can only upload one replication, and one optimal solution is proposed. We then consider a more complicated case, which means that each user can upload multiple replications on fog nodes in the provisioning process. Then the problem will be transformed to finding the number of replications and their optimal provision.
- Then we prove that the problem under the different transmission cost model is NP-hard. We first propose a non-adaptive algorithm which is proved to be bounded by $\frac{2}{3}W + \frac{1}{3}OPT$. Another $3 + \epsilon$-approximation algorithm is proposed based on local search, which has better performance with higher complexity.
- We conduct various simulations to compare our joint optimization methods with several state-of-the-art ones based on two real datasets: one is the published dataset of Mobike company constructed by 16680 users and another is the Microsoft GPS trajectory dataset constructed by 182 users. The results are shown from different perspectives to provide conclusions.

II. RELATED WORK

The concept of fog computing was introduced by Cisco Systems, and it is used to extend the cloud computing paradigm to the edge of the network, thus enabling a new breed of applications and services. There are a lot of extensions and application scenarios on fog computing, three of which are highlighted in [3], including connected vehicle, smart grid, and wireless sensor networks.

Although numerous novel architectures for fog computing [4–6] have been proposed, the resource provision problem in such systems remains a critical challenge. A bundle of existing research in this area are on provisioning fog resources to computational tasks offloaded from mobile devices. Yu et al. [4] studied joint application placement and data routing to support all data streams with both bandwidth and delay guarantees, which consider IoT applications that receive continuous data streams from multiple sources in the network. Skarlat et al. [5] have presented a conceptual framework for fog resource provisioning, which can provide delay-sensitive utilization of available fog-based computational resources with existing constraints. Chen et al. [6] focused on game-theoretical mechanisms for offloading decision making in the presence of multiple users, taking into account the energy consumption and the delay. Similar to the fog computing, Wang et al. [7] focused on the service entity replication for social virtual reality applications in edge computing. Zhang et al. [8] studied the reconfiguration in edge clouds and proposed an efficient online algorithm for configuration updating. These works have considered the performance guarantee for each mobile user while ignoring the interdependent relationship of multiple users between different costs.

Quite a few works have been carried out on the effect of cost efficiency in fog computing networks for mobile users. Arkian et al. [9] have proposed a fog computing based scheme on supporting crowd sensing applications, which jointly investigates data consumer association, task distribution, and virtual machine placement for cost-efficient limited provisioning resources. Pham et al. [10] formulated the task scheduling problem in a cloud-fog computing system and have proposed a heuristic-based algorithm, whose major objective is to achieve the balance between the makespan and the monetary cost of cloud resources. Gu et al. [11] are motivated to integrate fog computation and medical cyber-physical systems and then linearize it into a mixed integer linear programming problem. Most of these works have focused on the cost-efficient resource provision placement problem by only considering the fixed distribution of users; however, they ignored the mobility of multiple users between fog infrastructures. In this paper, we focus on the resource provision problem for multiple mobile users in fog computing. Our objective is to find an appropriate assignment of the workload for users that minimizes the cost and satisfies the constraints on the computation and communication resources.

III. Model and Problem Formulation

A. Fog Node

Given a substrate distribution of the fog nodes which is modeled as a weighted undirected graph G with a set of fog nodes V and a set of connections E, i.e., $G = \{V, E\}$, let $V = \{v_i\}$ denote the set of fog nodes, and v_i be the i^{th} fog node in V. Let $E = \{e_{ij}\}$ denote the set of connections between fog nodes, and e_{ij} be the connection between fog node i and j in E. Let $|e_{ij}|$ be the weight of the connection e_{ij}, which denotes the distance between fog nodes v_i and v_j.

B. Users Model

We use a set $U = \{u_k\}$ to denote the users who have mobilities. Each user with mobile devices is moving around in the fog computing system. We define the daily route of each user as the activity track, and each user u_k has its own activity track which is constructed by a set of fog nodes G_k, where $G_k \subseteq G$. The number of times that the user appears at that fog node is represented by the frequency of the user who stays at that point, and we use p_{ki} to denote the probability of user u_k at fog node v_i.

C. Problem Formulation

In this subsection, we formulate the resource provision problem for multiple users in fog computing. Our goal is to find an appropriate scheme which takes the set of users' workloads as input and decides the amounts and locations of the replications for users' workloads in the fog computing system accordingly such that the total cost is minimized. Let $f_C(u_k)$ denote the total cost of k^{th} user u_k. We consider two aspects: the replication cost $f_R(u_k)$ and transmission cost $f_T(u_k)$. The replication cost $f_R(u_k)$ of user u_k is associated with the creation of fog applications, which is fixed and determined by the fog computing platform based on the requirements [12]. We use X to denote the resource provision of fog nodes for user set U, and X_k to denote the set of chosen fog nodes that offload the replications of user u_k. Let r_{ki} denote the cost for user u_k that offloads its replication on fog node v_i. Each user can offload its replications onto several fog nodes in set G_k, and the replication cost of u_k can be calculated as $f_R(u_k) = \sum_{i \in X_k} r_{ki}$, which is the sum of replication with chosen fog nodes in X_k. Let $f_T(u_k)$ be the transmission cost, which is the bandwidth cost associated with data movement. It only occurs when there is no replication on the real location of u_k. We assume that the transmission cost is positively proportional to the distance between every pair of fog nodes, which is comprehensive yet practical. We assume that there is no capacity limitation on communication resource, and thus the transmission cost is measured by the shortest path between real location and other fog nodes. Let σ be the best assignment of users' workload replications, where $U \rightarrow X$ for the customers. $\sigma(v_i)$ denotes the fog node closest to the current location of u_k that holds the replication. It satisfies $t_{\sigma(i)i} = \min_{j \in X_k} p_{ij} \cdot t_{ij}$, where t_{ij} is the transmission cost between fog nodes i and j, and $v_j = \sigma(v_i)$. Since p_{ki} is the probability of user k at the i^{th} fog node, the transmission cost is the mean value that is transmitted from the other fog nodes in X_k, $f_T(u_k) = \sum_{i \in V_k \setminus X_k} p_{ki} \cdot t_{ij}$. Let $f_C(u_k)$ denote the total cost of user k, where $f_C(u_k) = f_R(u_k) + f_T(u_k)$.

$$\text{minimize} \quad \sum_{k \in U} f_C(u_k) \tag{1}$$

$$\text{subject to} \quad f_C(u_k) = f_R(u_k) + f_T(u_k) \tag{2}$$

$$f_R(u_k) = \sum_{i \in X_k} r_{ki} \tag{3}$$

$$f_T(u_k) = \sum_{j \in V_k / X_k} p_{ki} \cdot t_{ij} \tag{4}$$

$$0 \leq p_{ki} \leq 1, \sum_{i \in V_k} p_{ki} = 1 \tag{5}$$

Equation 1 shows the objective of minimizing the total provision cost for the users. Equation 2 shows that the total provision cost is divided into two parts: replication cost and transmission cost. Equation 3 is the constraint on the total replication cost, which is the sum of the replication cost of fog nodes that offload the workload of user k. Since the replication cost r_{ki} of u_k is the same, we use X_k to denote the set of the chosen fog nodes that offload the replications of user k, and $|X_k|$ to denote the number of fog nodes in set X_k. Equation 4 is the constraint on the transmission cost of the fog nodes, which means the resource provision on each fog node will be transmitted to the real user's location. It is the sum of the production of the probability of user u_k at fog node v_i and the shortest distance among the rest of the fog nodes. Equation 5 is the constraint on the value of the probabilities.

IV. Provision with Single Replication (PSR)

In this section, we discuss a simple scenario that each user can only offload single replication on fog nodes during the resource provision. In this scenario, we consider three cases, in which the capacities of fog nodes are limited to 1, a constant value $\eta > 1$, and ∞. For three cases, we have optimal solutions.

We first consider one simplest case where the capacities of fog nodes are limited to 1, which means that each fog node can only serve one replication of users. We transform our original problem into a bipartite graph, with the set of users in the left and the set of fog nodes on the right. The connections between the user and fog nodes are the activity track of each user, and the weight w_{ki} is the reward when user u_k chooses the i^{th} fog node v_i as the location of the replication. The total cost $f_C(u_k)$ is the sum of the product of the probability p_{ki} of user k at the fog node v_i and the shortest distance between fog node v_i and the rest of possible fog nodes, i.e., $f_C(u_k) = r_{ki} + \sum_{i \in V_k / v_i} p_{ki} \cdot t_{\sigma(i)i}$. Let α be a constantly expected reward of the user that is much larger than the total cost $f_C(u_k)$, where $\alpha \gg f_C(u_k)$. We define the value of weight w_{ki} by using the constant reward α minus the total cost, where $w_{ki} = \alpha - f_C(u_k)$. Then, our problem is converted from minimizing the total cost to finding the maximum weight matching of the bipartite graph.

As shown in Algorithm 1, we take the topology of fog nodes G, and the set of users U as our inputs. The output is the provisioning scheme X for the set of users U. In lines 1 to 3, we first calculate the value of weight w_{ki} according to

Algorithm 1 Provision with Single Replication (PSR)

Input: Topology G, set of users U;
Output: Provision Scheme X of U;

1: **for** user $k=1$ to $k=|U|$ in U **do**
2: **for** fog node $i=1$ to $i=|V_k|$ in G_k **do**
3: Calculate w_{ij} according to the activity track G_k of each user;
4: Construct a bipartite graph with respect to w_{ki};
5: Obtain the maximum weight matching;
6: **return** Provision Scheme X of U;

the activity track G_k of each user. In line 4, we construct a bipartite graph with respect to w_{ki}. We transform our problem from minimizing the total cost to finding the maximum weight matching of the bipartite graph. In line 5, we calculate the maximum weight matching by using the kuhn-munkras algorithm, and the provisioning scheme X of U is returned in line 6. For the case that the capacity of each fog node is limited by a constant $\eta > 1$, the optimal solution still can be calculated by applying the bipartite matching. The difference is that each fog node should be duplicated by $\eta - 1$ times. Then the problem is converted to the case that the capacity of each fog node is 1. When the capacity of each fog node is unlimited, i.e., capacity of each fog node is ∞, the optimal solution can be calculated by a greedy approach. That is, each user offloads his/her workload to the fog node with the smallest cost that contains the transmission and replication cost w_{ki}. The time complexity of PSR algorithm is $O(|U|\cdot|V|\cdot(|U|+|V|)^4)$, where $|U|$ is the number of users in set U, and $|V|$ is number of fog nodes in the activity tracking set of one user.

V. Multiple Replications with Different Transmission Cost Model (D-PMR)

In this section, we discuss a more complicated and realistic scenario where users can offload multiple replications on several fog nodes during the resource provision, and the transmission cost is related to the distance of fog nodes. In this scenario, we assume that each fog node has unlimited capacity. This problem is defined as a D-PMR problem. We first prove that the D-PMR problem is NP-hard. Then, we propose two efficient algorithms. One is a non-adaptive algorithm based on submodular function, and another one is an approximation algorithm based on local search.

A. NP-hard

***Theorem** 1:* The D-PMR problem is NP-hard.

Proof: We conduct the proof via a polynomial-time reduction from the weighted set covering problem, which is known to be NP-hard [13]. A set covering problem is to find a minimum-weight cover $\mathcal{C} \subseteq \mathcal{F}$ whose members cover all elements of $\mathcal{X}$, i.e., $\mathcal{X} = \bigcup_{S\in\mathcal{C}} S$, An instance $(\mathcal{X},\mathcal{F})$ of the set-covering problem consists of a finite set $\mathcal{X}$ and a family $\mathcal{F}$, and every element of X belongs to at least one subset in $\mathcal{F}$: $X=\bigcup_{S\in\mathcal{F}} S$. Each set S in family $\mathcal{F}$ has an associated weight w, and the weight of a cover $\mathcal{C}$ is $\sum_{S\in\mathcal{C}} w$ [13]. The reduction from the set covering problem to our D-PMR problem can be built by treating users in set U as a finite set $\mathcal{X}$, and we reduce all provision schemes of user replications that offload on corresponding fog nodes as a family of $X=\bigcup_{X_k\subseteq U} X_k$ as $\mathcal{F}$. It means which users are offloading replications on the corresponding fog nodes. X_k covers the users, which is treated as S. The cost of X_k is $f_C(u_k)=\sum_{i\in X_k} r_{ki}+\sum_{j\in V_k/X_k} t_{ij}$, which can be treated as w. Our D-PMR problem with minimum cost is to find a minimum-cost subset $X\subseteq U$, whose replications on fog nodes serve all users of U, i.e., $\sum_{X_k\in X} f_C(u_k)$. Since the weighted set covering problem is NP-hard, our D-PMR problem with the minimum cost is NP-hard. ■

B. Non-Adaptive Algorithm based on Submodular

In this subsection, we propose a non-adaptive algorithm with bound $\frac{2}{3}W+\frac{1}{3}OPT$. Before presenting the algorithm, we first carry out a transformation by introducing a new reward function for each user, i.e., $f(Z_k)=W_k-C(Z_k)$. We use a set Z_k to denote the chosen fog nodes which hold the replications of one unique user k in the system, where $Z_k\subseteq X_k$. The cost of setting Z_k is $C(Z_k)=\sum_{i\in Z_k} r_i+\sum_{i\in\{X_k\setminus Z_k\}} p_i\cdot t_{\sigma(i)i}$, where $t_{\sigma(i)i}$ is the minimum transmission cost between fog node $\sigma(i)\in Z_k$ and i. W_k is a constant which denotes the expected cost of users. Then, our original problem of minimizing the total cost is transformed into maximizing the reward of users instead. We first prove that our new function $f(Z_k)$ is a submodular function. In order to simplify the proof in Theorem 2, we use $D_i^{Z_k}$ to denote the minimum transmission cost between fog node k and set Z_k, i.e., $D_i^{Z_k}=t_{\sigma(i)i}|_{\sigma(i)\in Z_k}$.

***Theorem** 2:* The reward function $f(Z_k)=W_k-C(Z_k)$ of the D-PMR problem is a submodular function.

Proof: We first define the reward of one user as $f(Z_k)=W_k-C(Z_k)$, where W_k is the expected cost of the user. Suppose X_k is the provision set of user k, and Y_k is a subset of X_k, where $Y_k\subseteq X_k$. We choose any fog node in graph G which is denoted as b to be the newly added one. If the reward function of the D-PMR problem is a submodular function, we need to prove that $f(X_k\cup b)-f(X_k)\le f(Y_k\cup b)-f(Y_k)$. Since this equation can be converted to $C(X_k)-C(X_k\cup b)\le C(Y_k)-C(Y_k\cup b)$, we can prove it. For the left part of subset X_k, we convert it into

$$\textstyle\sum_{i\in Z_k-X_k} p_i\cdot D_i^{X_k}-\sum_{i\in Z_k-(X_k\cup b)} p_i\cdot D_i^{X_k\cup b} \quad (6)$$

$$=p_b\cdot D_b^{X_k}+\textstyle\sum_{i\in Z_k-(X_k\cup b)} p_i\cdot(D_i^{X_k}-D_i^{X_k\cup b}) \quad (7)$$

For the right part of subset Y_k, we convert it into

$$\textstyle\sum_{i\in Z_k-Y_k} p_k\cdot D_i^{Y_k}-\sum_{i\in Z_k-(Y_k\cup b)} p_i\cdot D_i^{Y\cup b} \quad (8)$$

$$=p_b\cdot D_b^{Y_k}+\textstyle\sum_{i\in Z_k-(Y_k\cup b)} p_i\cdot(D_k^{Y_k}-D_k^{Y_k\cup b}) \quad (9)$$

We compare Equations 7 and 9 by doing subtraction, then we divide it into two parts, which are Equations 10 and 11.

$$p_b\cdot D_b^{X_k}-p_b\cdot D_b^{Y_k} \quad (10)$$

$$\sum_{i\in Z_k-(X_k\cup b)} p_i\cdot(D_i^{X_k}-D_i^{X_k\cup b})-\sum_{i\in Z_k-(Y_k\cup b)} p_i\cdot(D_i^{Y_k}-D_i^{Y_k\cup b}) \quad (11)$$

Since $Y_k\subseteq X_k$, $D_b^{X_k}\le D_b^{Y_k}$, then we have $D_b^{X_k}-D_b^{Y_k}\le 0$ of Equation 10. For Equation 11, we convert the first part into

$$\textstyle\sum_{i\in Z_k-(Y_k\cup b)} p_i\cdot(D_i^{Y_k}-D_i^{Y\cup b}) \quad (12)$$

$$=\textstyle\sum_{i\in Z_k-(X_k\cup b)} p_k\cdot(D_i^{Y_k}-D_i^{Y_k\cup b})+\sum_{X_k-Y_k} p_k\cdot(D_i^{Y_k}-D_i^{Y_k\cup b}) \quad (13)$$

We convert Equations 11 by adding 12. Since $\sum_{X_k-Y_k} p_i \cdot (D_i^{Y_k} - D_i^{Y_k \cup b}) \geq 0$, we only consider the value of $\sum_{i \in Z-(X_k \cup b)} p_i \cdot (D_i^{X_k} - D_i^{X_k \cup b}) - \sum_{i \in Z_k-(X_k \cup b)} p_i \cdot (D_i^{Y_k} - D_i^{Y_k \cup b})$. We compare the values of $D_i^{X_k} - D_i^{X_k \cup b}$ and $D_i^{Y_k} - D_i^{Y_k \cup b}$. Since $D_i^{X_k}$ is the minimum value of shortest paths between fog node i and the fog nodes in set X_k, where $D_i^{X_k} = \min\{D_i^{v_j}|_{v_j \in X_k}\}$, there exist two cases after adding fog node b: one is that the shortest path between b and i is larger than or equal to D_i^X, i.e., $D_i^b \geq D_i^{X_k}$. Another one is that the shortest path between b and i is smaller than $D_i^{X_k}$, i.e., $D_i^b < D_i^{X_k}$. In the first case, if the shortest path between b and i is larger than $D_i^{X_k}$ ($D_i^{X_k} < D_i^b$), there will be no decrease of $D_i^{X_k \cup b}$, and $D_i^{X_k} - D_i^{X_k \cup b} = 0$. However, for $D_i^{Y_k} - D_i^{Y_k \cup b}$, if the shortest path between b and i is smaller than $D_i^{Y_k}$, $D_i^{Y_k} - D_i^{Y_k \cup b} > 0$. Since $D_i^{X_k} - D_i^{X_k \cup b} = 0$, we can have that $D_i^{X_k} - D_i^{X_k \cup b}$ is smaller than $D_i^{Y_k} - D_i^{Y_k \cup b}$. If the shortest path between b and i is larger than $D_i^{Y_k}$, and $D_i^{Y_k} - D_i^{Y_k \cup b} = 0$, then we have that $D_i^{X_k} - D_i^{X_k \cup b}$ is equal to $D_i^{Y_k} - D_i^{Y_k \cup b}$. Thus, we can have that $D_i^{X_k} - D_i^{X_k \cup b} \leq D_i^{Y_k} - D_i^{Y_k \cup b}$. In the second case, if the shortest path between b and i is smaller than $D_i^{X_k}$ ($D_i^{X_k} > D_i^b$), there will be a decrease after adding b, $D_i^{X_k} - D_i^{X_k \cup b} > 0$. Then we have that, $D_i^{X_k} - D_i^{X_k \cup b} < D_i^{X_k} - D_i^b$. Since $D_i^{X_k} \leq D_i^{Y_k}$, we have $D_i^{X_k} - D_i^{X_k \cup b} < D_i^{X_k} - D_i^b \leq D_i^{Y_k} - D_i^b$. For subset Y_k, we have $D_i^b \geq D_i^{Y_k \cup b}$. Then we have $D_i^{X_k} - D_i^{X_k \cup b} \leq D_i^{Y_k} - D_i^{Y_k \cup b}$. Therefore, we can have that the reward function $f(Z_k)$ is a submodular function. ■

Theorem 3: The reward function $f(Z_k) = W_k - C(Z_k)$ is a non-monotone function.

Proof: As the reward W_k is the expected cost of the user, which is a constant, we only need to prove the cost function $C(Z_k)$ is non-monotone. Since $C(Z_k) = R(Z_k) + T(Z_k)$, we have $C(Z_k) - C(Z_k+1) = R(Z_k) + T(Z_k) - (R(Z_k+1) + T(Z_k+1)) = (T(Z_k) - T(Z_k+1)) - r_k$. $|Z_k|$ is the number of replications of one user, where $|Z_k| < |Z_k| + 1$. Because the reduction of the transmission cost after adding one replication depends on the distribution of fog nodes and the tracking activity of users, the relationship of $C(Z_k)$ and $C(Z_k+1)$ is uncertain. (i). If the reduction of the transmission cost is larger than the replication cost r_k, we will have $(T(Z) - T(Z+1)) - r_k > 0$. Then we have $C(Z_k) > C(Z_k+1)$, $f(Z_k)$ is a non-monotone function. (ii). If the reduction of the transmission cost is smaller than the replication cost R, we will have $(T(Z_k) - T(Z_k+1)) - r_k < 0$. Then we have $C(Z_k) < C(Z_k+1)$, $f(Z_k)$ is a monotone function. Since the relationship between replication cost and the reduction of the transmission cost is uncertain, the reward function $f(Z_k) = W_k - C(Z_k)$ is a non-monotone function. ■

Since the $f(Z_k) = W_k - C(Z_k)$ is a non-monotone asymmetric submodular function ($f(Z_k) \neq f(X_k \setminus Z_k)$ for $\forall Z_k \subseteq X_k$), we introduce a non-adaptive algorithm for our D-PMR problem with bound $\frac{2}{3}W + \frac{1}{3}OPT$. Before introducing the algorithm, we first introduce a new definition.

Definition 1 (additive error): Let $X_k = V_k(1/2)$ denote a uniformly random subset of V_k of user u_k. For each element x, define $\omega(x) = \mathbf{E}[f(X_k \bigcup \{x\}) - f(X_k \setminus \{x\})]$.

Algorithm 2 Non-Adaptive for D-PMR (NA-D-PMR)

Input: Topology G, set of users U;
Output: Provision Scheme X of U;

1: **for** each user u_k in U **do**
2: **for** each fog node v_i in V_k **do**
3: Use random sampling to find the estimated value $\widetilde{\omega}(v_j)$ for each fog node v_i in V_k;
4: Independently, sample a random set with probability $p_{ki} = 1/2$, where $X_k = V_k(1/2)$;
5: With probability $8/9$, return X_k;
6: With probability $1/9$, return $X_k' = \{v_j \in V_k : \widetilde{\omega}(v_i) > 0\}$;
7: **return** Provision Scheme X of U;

The expectation $\mathbf{E}[f(X_k)]$ for any distribution of X_k can be estimated by random sampling up to an additive error. Let $\widetilde{\omega}(x)$ be the estimated value of $\omega(x)$ based on the sample mean by using point estimation. Based on that, we propose Algorithm 2 which is inspired by a non-adaptive scheme in [14]. We use it in each iteration for multiple users in set U from lines 1-7. We use the topology of fog nodes G and the set of users U as our input. The provision scheme X of U is the output. In line 3, we first use random sampling to find the estimated value $\widetilde{\omega}(v_i)$ for each fog node v_i in V_k. We use $s_k = |V_k|^5$ to be the number of random samples of each user, and $|V_k|$ to denote the total number of fog nodes based on u_k's tracking activity. In line 4, we sample a random set with probability $p_{ki} = 1/2$ independently, where $X_k = V_k(1/2)$. In lines 5 and 6, for each user, we have a probability of $8/9$ returning X_k, and a probability of $1/9$ returning $X_k' = \{v_i \in V_k : \widetilde{\omega}(v_i) > 0\}$. In line 7, we return the provision scheme.

Theorem 4: The NA-D-PMR algorithm is bounded by $\frac{1}{3}(2W + OPT)$, and the time complexity is $O(|U| \cdot |V|^6)$.

Proof: As $f(Z_k)$ is the reward of users k, we use $f^*(Z_k)$ to denote the optimal reward. Since the non-adaptive method we introduce is a $\frac{1}{3}$-approximation algorithm [14] for reward function $f(Z_k)$, we have $f(Z_k) \geq \frac{1}{3} f^*(Z_k)$. Let $|U|$ be the number of users in set U, the total reward will be $\sum_{k=1}^{|U|} f(Z_k)$. The optimal solution of set U is denoted as OPT, where $OPT = \sum_{k=1}^{|U|} C^*(Z)$. Then we have $|U|W - \sum_{k=1}^{|U|} C(Z) \geq \frac{1}{3}(|U|W - \sum_{k=1}^{|U|} C^*(Z))$. With simple mathematical transformation, we have $\sum_{k=1}^{|U|} C(Z_k) \leq \frac{2}{3}|U|W_k + \frac{1}{3}OPT$. Let $W = |U|W_k$, as the number of users $|U|$ is limited and W_k is a constant larger than the cost value $|Z_k|r_k$, the value of W is also a constant larger than $|U||Z_k|r_k$. Therefore, we have $\sum_{k=1}^{|U|} C(Z_k) \leq \frac{2}{3}W + \frac{1}{3}OPT$. For each user, since the number of chosen random samples is larger than or equal to $|V|^5$, the time complexity of NA-D-PMR algorithm is $O(|U| \cdot |V|^6)$. ■

Since the function $f(Z) = W - C(Z) \geq 0$, we have $f^*(Z) = W - C^*(Z) \geq 0$, and the value of $W \geq C^*(Z)$. When the value of W is equal to the optimal value of the total cost $C^*(Z)$, we have $C(Z) \leq C^*(Z)$, which means that $C(Z)$ is the optimal solution. However, if the value of W is much larger than $C^*(Z)$, our bound will be very loose.

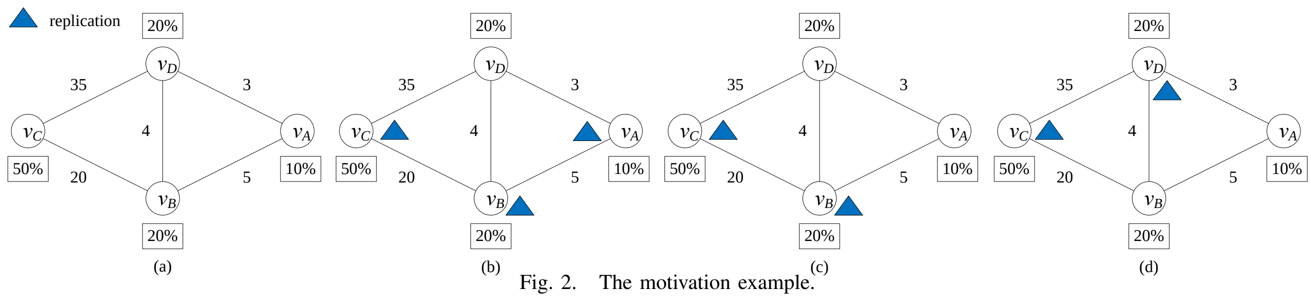


Fig. 2. The motivation example.

C. Approximation Algorithm based on Local Search

In this subsection, we propose an $3+\epsilon$ approximation algorithm based on local search [15]. The insight of Algorithm 3 is based on Theorem 2. We iteratively increase the number of replications until the reduced value of transmission cost is less than single replication cost.

As shown in Algorithm 3, we use the topology of fog nodes G and the set of users U as the inputs. The output is the provision scheme X of users U. We try to find a feasible solution for each user u_k in each iteration from lines 2 to 9. In line 3, we randomly choose a fog node V_i in F. We first add fog nodes one-by-one. If the increment $C(X \bigcup \{v_j|_{j\in Y}\}) - C(X) < 0$, fog node y will add to set X, and this process will be terminated when there is no increment, as shown in lines 4 and 5. We start to check whether there exist fog nodes that can be removed from the obtained set X in line 6. For each fog node v_i in X, if the cost decreases after removing it, $C(X \setminus \{v_i|_{i\in X}\}) - C(X) < 0$, we update the set $X = X \setminus \{v_i|_{i\in X}\}$. Then we start to check whether there exists an exchanging. For each fog node in set X, we check the value of cost after exchanging one arbitrary fog node in the remaining set Y. If $C(X \setminus \{v_i|_{i\in X}\} \bigcup \{v_j|_{j\in Y}\}) - C(X) < 0$, we update the set $X = X \setminus \{v_i|_{i\in X}\} \bigcup \{v_j|_{j\in Y}\}$. Since we directly imply the local search steps based on [15], we have LS-D-PMR is a $3+\epsilon$ approximation algorithm. Detailed proof is presented in [15]. The LS-D-PMR converges to M, where M is the number of fog nodes that are chosen in set G_k, i.e., $M \le 2^{|V|}$. The time complexity of LS-D-PMR is $O(|U| \cdot |V| \cdot M)$.

We use an example in Fig 2 to illustrate our Algorithm 3. As shown in Fig 2(a), we assume that there are four fog nodes v_A, v_B, v_C, and v_D, and the probability that the user stays on each fog node is 10%, 20%, 50%, and 20%, respectively. The weight on each link is the transmission cost between each pair of fog nodes. We use one blue triangle to denote one replication, and the cost of each replication is 3. Firstly, we start to place the replication by adding the fog node iteratively. We assume that the initial cost is very high, $C(X) = 100$. We choose fog node v_A to add to the set $X = X \bigcup v_A$, and calculate the cost $C(X) = 3+0+1+12.5+0.6 = 17.1$. Since $17.1 - 100 < 0$, we update $X = X \bigcup v_A$. After that we choose fog node v_B, and we calculate the cost $C(X) = 3\times2+0+0+10+0.6 = 16.6$. Since $16.1-17.1 < 0$, we update $X = X \bigcup \{v_A, v_B\}$. After that we choose fog node v_C, and we calculate the cost $C(X) = 3\times3+0+0+0+0.6 = 9.6$. Since $9.6 - 16.6 < 0$, we update $X = X \bigcup \{v_A, v_B, v_C\}$. Then we choose fog node v_D, and we find that the cost $C(X) = 3\times4+0+0+0+0 = 12$, which is larger than 9.6, i.e., $12-9.6 > 0$. We stop adding fog nodes. Secondly, we start to remove. Since the fog node set with replications has already been updated to $X = X \bigcup \{v_A, v_B, v_C\}$, we first remove v_A, and we calculate the cost $C(X) = 3\times2+0+0+0.5+0.8 = 7.3$. Since $7.3 - 9.6 < 0$, we update $X = X \setminus \{v_A\} = \{v_B, v_C\}$. After that, we choose fog node v_B, and we find that the cost $C(X) = 3+2.5+4+0+7 = 16.5$, which is larger than 7.3, i.e., $16.5 - 7.3 > 0$. We stop removing fog nodes, and the set is still $X = \{v_B, v_C\}$. Thirdly, we start to exchange. Since $X = \{v_B, v_C\}$, we have $Y = \{v_A, v_D\}$. We first exchange v_B with v_A, $X = X \setminus \{v_B\} \bigcup \{v_A\} = \{v_A, v_C\}$, and we calculate the cost $C(X) = 3\times2+0+0+1+0.6 = 7.6$. We find that the cost $7.6 > 7.3$, we exchange v_B with v_D, $X{=}X \setminus \{v_B\} \bigcup \{v_D\} = \{v_D, v_C\}$. The cost $C(X){=}3\times2+0+0.3+0.8 = 7.1$, $7.1-7.3 < 0$, then we update the set $X = \{v_D, v_C\}$. So, our final result is to place two replications on v_C and v_D.

Algorithm 3 Local Search for D-PMR (LS-D-PMR)

Input: Topology G, set of users U;
Output: Provision Scheme X of U;

1: Initialize sets $X = \emptyset$, $Y = G$;
2: **for** each user u_k in U **do**
3: **repeat**
4: **for** each fog node v_i in V_k **do**
5: **if** $C(X \bigcup \{v_j|_{j\in Y}\}) - C(X) < 0$ **then**
6: Set $\Delta = C(X \bigcup \{v_j|_{j\in Y}\}) - C(X)$;
7: Update set $X = X \bigcup \{v_j|_{j\in Y}\}$;
8: **for** each fog node v_i in X **do**
9: **if** $C(X \setminus \{v_i|_{i\in X}\}) - C(X) < 0$ **then**
10: Set $\Delta = C(X \setminus \{v_i|_{i\in X}\}) - C(X)$;
11: Update set $X = X \setminus \{v_i|_{i\in X}\}$;
12: **for** each fog node V_i in X, V_j in Y **do**
13: **if** $C(X \setminus \{v_i|_{i\in X}\} \bigcup \{v_j|_{j\in Y}\}) - C(X) < 0$ **then**
14: Set $\Delta = C(X \setminus \{v_i|_{i\in X}\} \bigcup \{v_j|_{j\in Y}\}) - C(X)$;
15: Update set $X = X \setminus \{v_i|_{i\in X}\} \bigcup \{v_j|_{j\in Y}\}$;
16: **until** $\Delta > 0$
17: **return** Provision Scheme X of U;

VI. Experiments

In this section, we conduct our experiments on two real datasets, Mobike Dataset [16] and Microsoft GPS trajectory dataset [17] to study the resource provision problem for multiple users in fog computing networks. The results are shown from different perspectives to provide insightful conclusions.

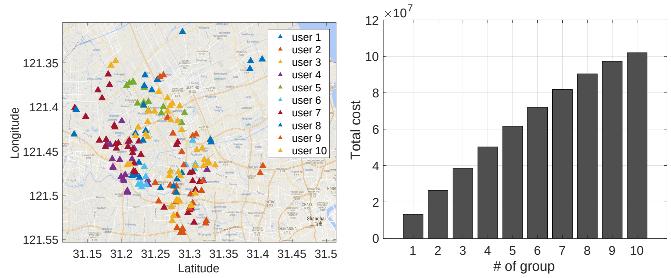
A. Basic Setting

1) Mobike Dataset: We used the published data of Mobike company from the open data platform Soda to construct our real dataset [16]. We used a set of four-month-long history trip data from $08/01/2016$ to $09/01/2016$ with 102361 records. We analyzed this dataset according to the user ids, and obtained the trajectories of 16680 users. Due to the redundant information in the dataset, we extracted several of these parameters to form a new database. The record includes the user id, start location, and end location, respectively. According to the price analysis in [12], we used 10 to denote the unit replication ($1GB$) price of users. We divided them into 10 groups, and each group was constructed by 1668 users. We first analyze the interesting points of users. Due to the large number of users, it was difficult to clearly mark each user's interesting points on the map, so we randomly chose 10 users and marked out the interesting points in Figure 3(a). According to the result shown in Figure 3(a), we found that each user's activity track is different, however, there existed some overlapping points of some users. In order to clearly represent the nodes that overlap between users, we depict the user's trajectory on two-dimensional coordinates on the map of Shanghai with different colour points.

2) Microsoft GPS trajectory dataset: We used the published GPS trajectory dataset which has been collected in the Geolife project of Microsoft Research Asia by 182 users in a period of over five years (from April 2007 to August 2012) [17, 18]. This dataset recorded a broad range of users' outdoor movements, including not only life routines like go home and go to work but also some entertainments and sports activities, such as shopping, sightseeing, dining, hiking, and cycling[19]. Based on the dataset provided by GeoLife GPS Trajectories, [20] clustering the significant locations for 50 users. We extracted several of these parameters of these 50 users to form a new database, which only contains the locations (latitude and longitude). As shown in Figure 5(a), we mark out 10 users' interesting points. Since users' movements were recorded for three years, the user's interest points are widely distributed (crossing different provinces). Most of the points are concentrated in Beijing, we intercept some areas and analyze them, as shown in Figure 5(a).

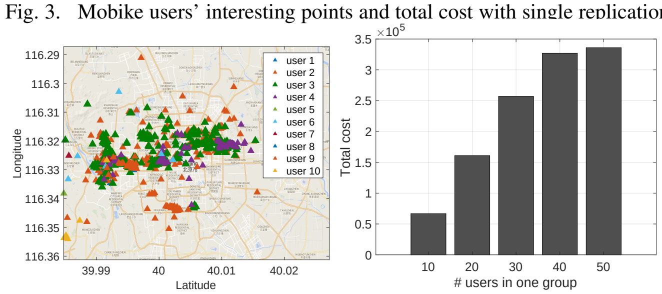
B. Results-Single replication

We first ran the PRS algorithm based on the Mobike dataset. Since the dataset is constructed by the trajectories of 16680 users, we divide them into 10 groups. Each latter group is the union of all the previous groups. The experiment results on the total cost of single replication are shown in Figures 3(b), and we have the following observations. (i). With the scaling number of users, the total cost increases. As shown in Figure 3(b), the total number of users in one group is increased by 1668 compared with the previous one. As the number of users increases, the range of activities of users scales and so does the number of interesting points. In order to satisfy the user's demands, the number of placed fog nodes also increases, which leads to an increase in total cost. (ii). Although the increase in the number of users is linear, the



(a) Mobike users' interesting points. (b) Total cost with single replicaion.

Fig. 3. Mobike users' interesting points and total cost with single replication.



(a) Phone users' interesting points. (b) Total cost with single replicaion.

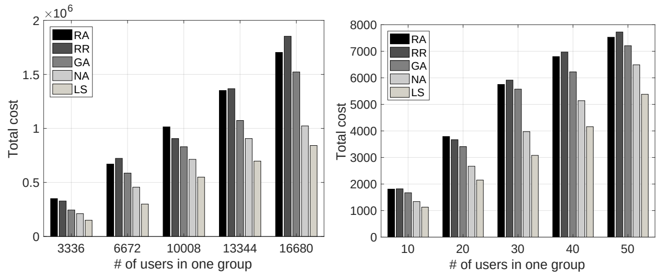
Fig. 4. Phone users' interesting points and total cost with single replication.

increase in total cost is non-linear. Since the group of users was chosen randomly, different users had different trajectories. Users whose trajectories cover more fog nodes will have larger cost. Therefore, due to the uncertainty of users' trajectories in the process of increase, the total cost increases non-linearly.

Then, we ran the PRS algorithm based on the Microsoft GPS trajectory dataset. According to the records of 50 users from the Microsoft GPS trajectory database, we divided it into 5 groups which contain 10 users' trajectories for each. The experiment results in the total cost of single replication are shown in Figure 4(a), and we have the following observations. The total cost obtained by different groups of users is different. Since groups of users were divided randomly, different groups of users had different trajectories. The number of fog nodes and the transmission costs are very different. Users whose trajectories cover more fog nodes will have a larger total cost, as shown in Figure 4(b). We analyzed the average length of trajectories for the users in each group, and we found that the total number of fog nodes in group 1 is larger than group 3, so the corresponding total cost is larger, as shown in Figure 4(b).

C. Results-Multiple replications

Then, we consider the provision with multiple replications of different transmission cost. We compare the proposed algorithms with three baseline approaches: Random Adding algorithm (RA) (the replications on fog nodes of each user are added randomly), Random Removing algorithm (RR) (the replications on fog nodes of each user are removed randomly from its tracking activity), and Greedy Adding algorithm (GA) (the replications on fog nodes of each user are greedily added by the probabilities). The experiment results in the total cost are shown in Figure 5, and we have the following observations. (i). The performances of two random algorithms are nearly the same. As shown in the black and dark gray columns, the performances of the two algorithms are basically the same when the numbers of users in one group are 3336 and 13344 in Mobike dataset (the number of users in one group is 10



(a) mobike users. (b) phone users.

Fig. 5. Total cost in PMR with different transmission cost model.

in Microsoft GPS trajectory dataset). In Figure 5(a), there are a number of 10008 and 16680 users who have better performance under the RA, and 6672 users who have better performance under the RR. In Figure 5(b), there are 20 users who have better performance under the RA, the groups of 30, 40 and 50 users have better performance under the RR. (ii). The greedy algorithm GA has better performance than two random algorithms RA and RR. As shown in the third column in Figure 5, the total cost under the GA is smaller. However, since the GA only considers the probabilities of fog nodes for each user, the results obtained for some groups (groups 13344 and 16680 in Mobike dataset, groups 30, 40 and 50 in Microsoft GPS trajectory dataset) will greatly deviate from the optimal solution. (iii). The total costs of the NA algorithm for some group of users are fluctuating, which means that some of them are much larger than LS algorithm. The reason is that the total cost of the NA algorithm is based on a probability result, as shown in Figure 5. (iv). The total cost under each algorithm is growing with the increasing number of users in one group, as shown in Figure 5(a) and Figure 5(b). (v). The performance of the LS is the best of these five algorithms, as shown in Figure 5. When the trajectories of the group of users are small (groups 3335 and 6672 in Mobike dataset, groups 10 and 20 in Microsoft GPS trajectory dataset), LS can have better performance due to the reduction of the number of iterations.

VII. Conclusion

In this paper, we focus on the resource provision problem for multiple mobile users in cost-efficient fog computing. We aim at minimizing the total cost of users, which includes the replication cost and transmission cost. Two cases are considered. First, we consider a simple case that each user can only upload one replication, and an optimal solution is proposed. Then, we consider a more realistic and complicated case that each user can upload multiple replications on several fog nodes. We prove that the resource provision problem under this scenario is NP-hard. Two algorithms are proposed. One is a non-adaptive algorithm which is bounded by $\frac{2}{3}W + \frac{1}{3}OPT$. Another one is a $3+\epsilon$-approximation algorithm based on local search. The performances of the proposed algorithms are confirmed by extensive experiments based on two real datasets. Extensive simulations show the efficiency and effectiveness of our algorithms.

Acknowledgement

This research was supported in part by Beijing Natural Science Foundation 4192007, and NSF grants CNS 1757533, CNS 1629746, CNS 1564128, CNS 1449860, CNS 1461932, CNS 1460971, and IIP 1439672.

References

[1] X. Sun and N. Ansari, “Edgeiot: Mobile edge computing for the internet of things,” *IEEE Communications Magazine*, vol. 54, no. 12, pp. 22–29, 2016.

[2] T. H. Luan, L. Gao, Z. Li, Y. Xiang, G. Wei, and L. Sun, “Fog computing: Focusing on mobile users at the edge,” *arXiv preprint arXiv:1502.01815*, 2015.

[3] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, “Fog computing and its role in the internet of things,” in *Proceedings of the first edition of MCC workshop on Mobile cloud computing*. ACM, 2012, pp. 13–16.

[4] X. Z. Ruozhou Yu, Guoliang Xue, “Application provisioning in fog computing-enabled internet-of-things: A network perspective,” in *IEEE INFOCOM*. IEEE, 2018, pp. 783–791.

[5] O. Skarlat, S. Schulte, M. Borkowski, and P. Leitner, “Resource provisioning for iot services in the fog,” in *2016 IEEE 9th international conference on service-oriented computing and applications (SOCA)*. IEEE, 2016, pp. 32–39.

[6] Y. Xiao and M. Krunz, “Qoe and power efficiency tradeoff for fog computing networks with fog node cooperation,” in *IEEE INFOCOM*. IEEE, 2017, pp. 1–9.

[7] L. Wang, L. Jiao, T. He, J. Li, and M. Mühlhäuser, “Service entity placement for social virtual reality applications in edge computing,” in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, 2018, pp. 468–476.

[8] J. Zhang, P. Lu, Z. Li, and J. Gan, “Distributed trip selection game for public bike system with crowdsourcing,” in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 2717–2725.

[9] H. R. Arkian, A. Diyanat, and A. Pourkhalili, “Mist: Fog-based data analytics scheme with cost-efficient resource provisioning for iot crowdsensing applications,” *Journal of Network and Computer Applications*, vol. 82, pp. 152–165, 2017.

[10] X.-Q. Pham and E.-N. Huh, “Towards task scheduling in a cloud-fog computing system,” in *Network Operations and Management Symposium (APNOMS), 2016 18th Asia-Pacific*. IEEE, 2016, pp. 1–4.

[11] L. Gu, D. Zeng, S. Guo, A. Barnawi, and Y. Xiang, “Cost efficient resource management in fog computing supported medical cyber-physical system,” *IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 1, pp. 108–119, 2017.

[12] A. Brogi, S. Forti, and A. Ibrahim, “Deploying fog applications: How much does it cost, by the way?” in *CLOSER*, 2018, pp. 68–77.

[13] T. H. Cormen, *Introduction to algorithms*. MIT press, 2009.

[14] U. Feige, V. S. Mirrokni, and J. Vondrak, “Maximizing non-monotone submodular functions,” *SIAM Journal on Computing*, vol. 40, no. 4, pp. 1133–1153, 2011.

[15] D. B. Shmoys, É. Tardos, and K. Aardal, “Approximation algorithms for facility location problems,” in *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*. ACM, 1997, pp. 265–274.

[16] Mobike Shanghai user data. [Online]. Available: http://shanghai.sodachallenges.com/data.html

[17] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, “Mining interesting locations and travel sequences from gps trajectories,” in *Proceedings of the 18th international conference on World wide web*. ACM, 2009, pp. 791–800.

[18] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, “Understanding mobility based on gps data,” in *Proceedings of the 10th international conference on Ubiquitous computing*. ACM, 2008, pp. 312–321.

[19] Y. Zheng, H. Fu, X. Xie, W.-Y. Ma, and Q. Li, “Geolife gps trajectory dataset-user guide,” *Microsoft Research*, 2011.

[20] [Online]. Available: https://github.com/om13shukla/ADB-Project