Dynamic Resource Pricing and Scalable Cooperation for Mobile Cloud Computing

Xumin Huang, Rong Yu, Jiawen Kang and Jiefei Ding
Guangdong University of Technology, Guangzhou, China
Email: huangxu.min@163.com, yurong@ieee.org, Kjwx886@163.com, Jiefei.Ding@ieee.org

Sabita Maharjan, Stein Gjessing and Yan Zhang
Simula Research Laboratory, Oslo, Norway
Email: sabita@simula.no, steing@ifi.uio.no, yanzhang@ieee.org

***Abstract*—Mobile cloud computing is a new paradigm to improve the quality of mobile services, which has drawn considerable attentions in both industrial and academic fields. In this paper, we consider the resource management and sharing problems for radio and computing resources to support mobile applications in mobile cloud computing. In such an environment, service providers can cooperate to form coalition to share their idle resources with each other. We propose a coalition game model based on two-sided matching theory. The coalition game model efficiently reflects the scalable cooperation among the service providers for sharing their idle resources. As a result, the resources can be better utilized and the quality of service for users can be improved. The simulation results indicate that our scheme can optimize the resource utilization and significantly improve the quality of service of the users.**

I. Introduction

With the rapid development of wireless technology, mobile devices can receive various information and services, which shows high potential for a large number of applications. However, massive mobile applications are typically supported by computing modules on mobile devices. These terminals have limited computing capability. Mobile Cloud Computing (MCC) possesses a huge potential to address such issues. MCC combines wireless access services and cloud computing into mobile environment to improve the performance of mobile applications [1]. With the help of cloud computing, mobile applications can offload some computing modules to be executed on a powerful server in a cloud. As a result, MCC could not only reduce the delay of mobile applications but also lower the energy consumption of mobile devices [2], [3].

However, running a mobile application in an MCC environment requires various resources, mainly including radio and computing resources. These resources should be previously reserved before usage to ensure Quality of Services (QoS). Generally, service providers (SPs) reserve a certain amount of remote radio resource from network providers (NPs) [4]. The SPs also reserve some remote computing resources (e.g., CPU, memory, and storage) from data centers, which are owned by cloud providers (CPs) [5]. Both the resources are used to support different mobile applications. The number of users running mobile applications is limited by the available radio and computing resources of the SPs thus impairing the service quality and user experience [6], [7]. Moreover, multiple applications have different resource requirements. In some cases, after running some applications, an SP may fall short of one kind of resources while it has plenty of other resources, or an SP may lack many resources. As a result, it is an important issue for SPs to design efficient resource management schemes satisfying real-time requirements of MCC applications [8].

There are some previous studies to address above issues [8]–[11]. In [8], some SPs cooperate to create a resource pool to share their radio and computing resources. All the SPs can use the resources from the resource pool when required, thereby increasing resource utilization. The revenue of an SP is based on the contribution to user's demand. However, if some resources of the SPs are idle, they may not obtain maximum revenues. In [10], the authors propose a service decision using semi-Markov decision process to balance the computation loads among multiple cloud domains. The authors in [11] focus on dynamic analysis and develop a provably-efficient dynamic scheduling and pricing algorithm to achieve a higher average profit but with big delay. The aforementioned schemes demonstrate that it is beneficial for SPs to cooperate in the MCC. In fact, every SP desires to autonomously cooperate with its optimal partner to maximize revenue. However, the above schemes do not fully optimize the idle resources of the SPs.

In this paper, we introduce a coalition game model for resource sharing between different SPs in the MCC framework. In our model, resource sharing is conducted in two steps. In the first step, each SP evaluates its revenues and decides to either work alone or join a coalition in the cloud market. In the second step, the SP will either rent resources from others or lease resources to others. The two groups of SPs effectively match their demands by using two-sided matching scheme.

The major contributions of this paper can be summarized as follows: (1) We combine prices with resource-demands in mobile cloud computing environment and adopt a coalition-game model close to realistic business model in the cloud market. (2) We apply two-sided matching theory for forming coalitions. Our scheme can not only improve fairness of transaction but also optimize resource utilization of the SPs. A win-win situation for the SPs is achieved in every coalition. (3) The revenue of an SP not only includes its profit from the users that use its applications, but also charges from the SPs which rent its resources and pays its rent on time. Therefore, our scheme also enhances the revenues from idle resources.

978-1-4673-7211-4/15 $31.00 © 2015 IEEE
DOI 10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.155

Moreover, the resource cooperation is scalable as more SPs and resources can be resiliently added into the cooperation.

The rest of this paper is organized as follows. System model and assumptions are described in Section II. In Section III, we present our problem formulation and analysis. In Section IV, we elaborate our proposed scheme. Extensive simulations are conducted and discussed in Section V. Section VI concludes the paper.

II. System model

In mobile cloud computing environment, a mobile application is typically divided into local computing module running on a mobile device and remote computing module running on a server. As a result, both radio and computing resources are required to run mobile applications. The network provider which provides radio resource (i.e., bandwidth) and the cloud provider which offers computing resource (such as CPU, storage and memory) collaborate to support the mobile applications [9].

The remote resource is the fixed asset of an SP, which is cheaper than resource obtained on an on-demand basis. However, the total number of available resource decides the maximum number of supporting applications. The available resources of SPs are different since the SPs support multiple applications with different resource requirements for some time. The model in this paper is proposed to take advantages of both remote resource reserving and local resource demands. As shown in Fig. 1, different SPs can cooperate with each other to share their available resources if necessary, in order to enhance their revenues and improve the resource utilization. In other words, through cooperation, an SP (e.g., SP_3) can expand its capacity to support applications by renting extra local resources from CPs or NPs. While an another SP (e.g., SP_2) can lease a local access to its remote reserved resource to SP_3. The deal is only based on the condition of mutual benefit. In this way, SP_3 can increase its applications that can be accessed or run to obtain higher QoS. Besides, SP_2 can improve its revenue by utilizing its idle resources. In Section V, we will show that the QoS, resource utilization and revenues increase significantly.

III. Problem Formulation

A. Wireless Network and Data Center

We consider an MCC environment with N mobile service providers. A service provider can negotiate with other service providers and share their reserved radio and computing resources. A service provider i at time t is indexed by SP_i^t, $i = 1, 2, \cdots, N$. Resource management for MCC includes the radio resource and computing resources, which are denoted as $\mathrm{B}_{\mathrm{SP_i}}^t$ and $\mathrm{C}_{\mathrm{SP_i}}^t$, respectively. There are k users denoted as $\mathrm{U} = \{u_1, u_2, u_3, ...u_k\}$ in the MCC environment. Let us denote the set of mobile applications as $\mathrm{A} = \{a_1, a_2, a_3, ...a_k\}$. Besides, a set of radio resource and a set of computing resource are represented as $\mathrm{B} = \{b_1, b_2, b_3, ...b_k\}$ and $\mathrm{C} = \{c_1, c_2, c_3, ...c_k\}$, respectively.

The system model is based on the following assumptions.

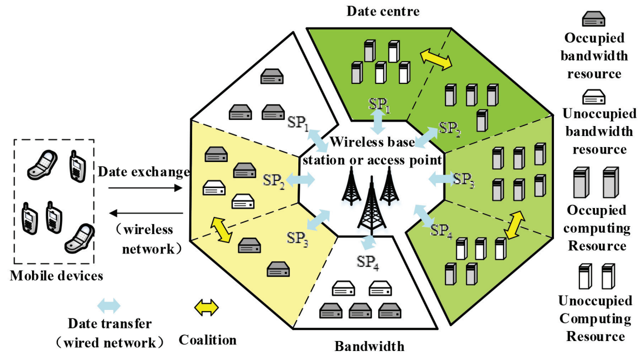


Fig. 1: The mobile cloud computing environment

1) The arrival and departure time of each user are assumed to follow Poisson distribution. The size of base stations and data centers follows the same distribution.
2) The mobile application a_i will consume the radio resource b_i, which is a random value in the range of $[b_{\min,i}, b_{max,i}]$, and has probability density function (PDF) $\rho_b(b_i)$. At the same time, the mobile application a_i will use the computing resource c_i whose value follows arbitrary distribution with PDF $\rho_c(c_i)$ in the range $[c_{\min,i}, c_{max,i}]$ [12].

Based on aforementioned assumptions, the quality of bandwidth service of application a_i , denoted as q_i^b, can be defined as

$$q_i^b = \begin{cases} 0, \quad \text{if } b_i < b_{\min,i}, \\ \int_{b_{\min,i}}^{b_i} \rho_b(b_i) db_i, \quad \text{if } b_{\min,i} < b_i \leq b_{\max,i}, \\ 1, \quad \text{if } b_i > b_{\max,i}. \end{cases} \tag{1}$$

Where, b_i is the radio resource which will be provided by SP_i^t. In a similar way, since SP_i^t can provide computing resource c_i, the quality of computing service, q_i^c, can be defined as

$$q_i^c = \begin{cases} 0, c_i \leq c_{\min,i}, \\ \int_{c_{\min,i}}^{c_i} \rho_c(c_i) dc_i, c_{\min,i} < c_i \leq c_{\max,i}, \\ 1, c_i > c_{\max,i}. \end{cases} \tag{2}$$

Therefore, the QoS requirement of user i is the minimum value between q_i^b and q_i^c, i.e.,

$$\mathrm{QOS}_i = min(q_i^b, q_i^c). \tag{3}$$

Next, we need to evaluate the ability of running application for the demand of each SP.

B. Quantification of Resource

Each SP has a certain amount of remote reserved resource. Radio resource and computing resource can be shared with each other by renting for a short time. Let $\mathrm{DB}_{\mathrm{SP}_i}$ and $\mathrm{DC}_{\mathrm{SP}_i}$ represent the absolute value of the difference between the required amount of the resource and the actual amount of the resource, respectively.

$$\mathrm{DB}_{\mathrm{SP}_i} = \left| \mathrm{B}_{\mathrm{SP_i}} - b_{\max,i} \right|, \tag{4}$$

$$\mathrm{DC}_{\mathrm{SP}_i} = \left| \mathrm{C}_{\mathrm{SP_i}} - c_{\max,i} \right|. \tag{5}$$

If the available radio resource $\mathrm{B_{SP_i}}$ is more than the upper bound of radio resource $b_{\max,i}$, the $\mathrm{DB}_{\mathrm{SP}_i}$ indicates the amount of resource that SP_i^t can offer to rent out. On the contrary, SP_i^t wants to rent the amount of radio resource $\mathrm{DB}_{\mathrm{SP}_i}$ for local usage. In eqn. (5), $\mathrm{DC}_{\mathrm{SP}_i}$ is obtained by the difference between the available computing resource $\mathrm{C_{SP_i}}$ and the upper bound of computing resource $c_{\max,i}$.

C. Utility Function

The utility function of service provider SP_i^t consists of two parts: the satisfaction and the cost. The satisfaction function S_i^t is different for the case of renting extra local resources ($Case1$) and renting out its own resources ($Case2$). But, in both cases, the satisfaction function is convex and starts from zero point. For simplicity, we take radio resource cooperation for example. We define the satisfaction function in $Case1$ as

$$S_i^t = w_{b,i}^t log(1 + x_{b,i}^t), \tag{6}$$

where $0 \leq x_{b,i}^t \leq 1$, and

$$w_{b,i}^t = \frac{\alpha}{1 + e^{-\beta(\mathrm{B_{n,SP_i}} - \mathrm{B_{SP_i}})}}. \tag{7}$$

Here, $x_{b,i}$ represents the amount of leased bandwidth resource of service provider i [13]. The willingness of SP_i^t is denoted by $w_{b,i}^t$ which is an S function curve [14]. The predefined constants α and β are determined by the user preference. Apparently, the larger the available resource $\mathrm{B_{SP_i}}$, the lower is the willingness of renting resource from other SPs. It means that more and more idle resources will be rented out.

In $Case2$, $x_{b,i}$ represents the resource which can be rented out. Thus, it is a negative value. Alternatively, we adopt the satisfaction function in this case as follows:

$$S_i^t = w_{b,i}^t(log(2 + x_{b,i}^t) - 1), where\ -1 \leq x_{b,i}^t \leq 0. \tag{8}$$

Let $\mathrm{P}_{b,C1}^t > 0$ and $\mathrm{P}_{b,C2}^t < 0$ denote the unit price in $Case1$ and $Case2$, respectively. The sign is an indication of the cost or the earning. The unified utility functions in two cases for service provider i in time slot t are then given by

$$u_{b,i}^t(\mathrm{SP}_i^t) = \begin{cases} w_{b,i}^t log(1+x_{b,i}^t) - \mathrm{P}_{b,C1}^t x_{b,i}^t, 0 \leq x_{b,i}^t \leq 1, \\ w_{b,i}^t(log(2+x_{b,i}^t)-1) + \\ \qquad \mathrm{P}_{b,C2}^t x_{b,i}^t, -1 \leq x_{b,i}^t < 0. \end{cases} \tag{9}$$

In time slot t, each SP will choose a proper x_i^t to maximize the utility $u_{b,i}^t(\mathrm{SP}_i)$ according to the given price.

D. Service Providers Classification

The SPs can be classified into two categories: $Case1$ or $Case2$, according to their potential behavior. In this paper, the prices in both sides are set by the NP. The NP only has one price for renting. However, different coalitions can have different cooperative prices. We first analyze the characteristic of utility in $Case1$. Differentiating $\mathrm{u}_{b,i}^t(\mathrm{SP}_i^t)$ with respect to $x_{b,i}^t$, we get

$$\frac{\partial \mathrm{u}_{b,i}^t(\mathrm{SP}_i^t)}{\partial x_{b,i}^t} = \frac{w_{b,i}^t}{(1 + x_{b,i}^t) ln2} - \mathrm{P}_{b,C1}^t, \tag{10}$$

$$\frac{\partial^2 \mathrm{u}_{b,i}^t(\mathrm{SP}_i^t)}{\partial {x_{b,i}^t}^2} = -\frac{w_{b,i}^t}{(1 + x_{b,i}^t)^2 ln2} < 0. \tag{11}$$

Clearly, the utility function for $Case1$ is concave, which indicates that the maximum value of this function exists. Therefore, using first order optimality condition $\frac{\partial \mathrm{u}_{b,i}^t(\mathrm{SP}_i^t)}{\partial x_{b,i}^t} = 0$, we get

$$ {x_{b,i}^t}^* = \frac{w_{b,i}^t}{\mathrm{P}_{b,C1}^t \mathrm{ln2}} - 1, \tag{12}$$

where ${x_{b,i}^t}^*$ is called the *best response*, which maximizes the utility on the condition of price $\mathrm{P}_{b,C1}^t$. Then, we substitute ${x_{b,i}^t}^*$ into eqn. (9) and have

$$ {\mathrm{u}_{b,i}^t}^*(\mathrm{SP}_i^t) = w_{b,i}^t \log(\frac{w_{b,i}^t}{\mathrm{P}_{b,C1}^t \mathrm{ln2}}) - \frac{w_{b,i}^t}{\ln 2} + \mathrm{P}_{b,C1}^t. \tag{13}$$

Therefore, the utility function of service provider SP_i^t can be converted into the *optimal utility* function in terms of $\mathrm{P}_{b,C1}^t$. Then, if we take the first and the second derivative of ${\mathrm{u}_{b,i}^t}^*(\mathrm{SP}_i^t)$ with respect of $x_{b,i}^t$, we obtain

$$\begin{aligned} \frac{\partial {\mathrm{u}_{b,i}^t}^*(\mathrm{SP}_i^t)}{\partial x_{b,i}^t} &= -\frac{w_{b,i}^t}{\mathrm{P}_{b,C1}^t ln2} + 1, \\ \frac{\partial^2 {\mathrm{u}_{b,i}^t}^*(\mathrm{SP}_i^t)}{\partial {x_{b,i}^t}^2} &= \frac{w_{b,i}^t}{{\mathrm{P}_{b,C1}^t}^2 ln2} > 0. \end{aligned} \tag{14}$$

Clearly, the optimal utility function is convex. Using $\frac{\partial \mathrm{u}_{b,i}^t(\mathrm{SP}_i^t)}{\partial \mathrm{P}_{b,C1}^t} = 0$, we can get the price ${\mathrm{P}_{b,C1,i}^t}^*$ (CP) as

$$ {\mathrm{P}_{b,C1,i}^t}^* = \frac{w_{b,i}^t}{\mathrm{ln2}}. \tag{15}$$

The ceiling price ${\mathrm{P}_{b,C1,i}^t}^*$ is interpreted as a threshold for SP_i^t. Under different given price sets by the NP, the SP_i^t may have different best utility cases as follows.

- If $\mathrm{P}_{b,C1}^t < {\mathrm{P}_{b,C1,i}^t}^*$, the best utility is obtained with the best response ${x_{b,i}^t}^* > 0$ and accordingly ${\mathrm{u}_{b,i}^t}^*(\mathrm{SP}_i^t) > 0$. In this case, the SP would like to rent resources.
- If $\mathrm{P}_{b,C1}^t = {\mathrm{P}_{b,C1,i}^t}^*$, the best utility is obtained with the best response ${x_{b,i}^t}^* = 0$ and accordingly ${\mathrm{u}_{b,i}^t}^*(\mathrm{SP}_i^t) = 0$. In this case, the SP does not expect to rent resources.
- If $\mathrm{P}_{b,C1}^t > {\mathrm{P}_{b,C1,i}^t}^*$, the best response is ${x_{b,i}^t}^* < 0$. This violates the limit of ${x_{b,i}^t}^* \geq 0$. In this case, the SP refuses to rent resources.

Following a similar analysis to the utility function of $Case2$, i.e., eqn. (9), we have

$$\frac{\partial \mathrm{u}_{b,i}^t(\mathrm{SP}_i^t)}{\partial x_{b,i}^t} = \frac{w_{b,i}^t}{(2 + x_{b,i}^t) ln2} - \mathrm{P}_{b,s}^t, \tag{16}$$

$$\frac{\partial^2 \mathrm{u}_{b,i}^t(\mathrm{SP}_i^t)}{\partial {x_{b,i}^t}^2} = -\frac{w_{b,i}^t}{(2 + x_{b,i}^t)^2 ln2} < 0. \tag{17}$$

The utility function in $Case2$ is also concave. Thus, we can obtain the best response of $\frac{\partial \mathrm{u}_{b,i}^t(\mathrm{SP}_i^t)}{\partial x_{b,i}^t} = 0$ as

$$x_{b,i}^t{}^* = -(2 + \frac{w_{b,i}^t}{\mathrm{P}_{b,C2}^t \ln 2}). \tag{18}$$

Now substituting eqn. (18) into eqn. (9), we get the optimal utility function in $Case2$ in terms of $\mathrm{P}_{b,s}^t$ as

$$\begin{aligned} u_{b,i}^t * (SP_i^t) = w_{b,i}^t \log(-\tfrac{w_{b,i}^t}{P_{b,C2}^t \ln 2}) - w_{b,i}^t - \tfrac{w_{b,i}^t}{\ln 2} \\ - 2P_{b,C2}^t. \end{aligned} \tag{19}$$

The first order and second order derivatives of $\mathrm{u}_{b,i}^t{}^*(\mathrm{SP}_i^t)$ with respect to the price $\mathrm{P}_{b,C2}^t$ are written as

$$\frac{\partial \mathrm{u}_{b,i}^t{}^*(\mathrm{SP}_i^t)}{\partial \mathrm{P}_{b,C2}^t} = -\frac{w_{b,i}^t}{\mathrm{P}_{b,C2}^t ln2} - 2, \tag{20}$$

$$\frac{\partial^2 \mathrm{u}_{b,i}^t{}^*(\mathrm{SP}_i^t)}{\partial \mathrm{P}_{b,C2}^t{}^2} = \frac{w_{b,i}^t}{\mathrm{P}_{b,C2}^t{}^2 ln2} > 0. \tag{21}$$

From the first order optimality condition $\frac{\partial \mathrm{u}_{b,i}^t{}^*(\mathrm{SP}_i^t)}{\partial \mathrm{P}_{b,C2}^t} = 0$, we obtain the minimum price (MP) $\mathrm{P}_{b,C2,i}^t{}^*$ that the SP is willing to rent out its resource for

$$\mathrm{P}_{b,C2,i}^t{}^* = -\frac{w_{b,i}^t}{2\ln 2}. \tag{22}$$

We also have three situations to estimate the preference of the SP by using the MP, $MP = |\mathrm{P}_{b,C2,i}^t{}^*|$.

- When $\mathrm{P}_{b,C2}^t < \mathrm{P}_{b,C2,i}^t{}^*$, the best utility is obtained with the best response $x_{b,i}^t{}^* > 0$, and accordingly $\mathrm{u}_{b,i}^t{}^*(\mathrm{SP}_i^t) > 0$. In this case, the SP prefers to rent out its resource.
- When $\mathrm{P}_{b,C2}^t = \mathrm{P}_{b,C2,i}^t{}^*$, the best utility is obtained with the best response $x_{b,i}^t{}^* = 0$ and accordingly $\mathrm{u}_{b,i}^t{}^*(\mathrm{SP}_i^t) = 0$. In this case, the SP does not expect to rent out its resource.
- When $\mathrm{P}_{b,C2}^t > \mathrm{P}_{b,C2,i}^t{}^*$, the best utility can be obtained with the best response $x_{b,i}^t{}^* < 0$. This violates positivity requirement of $x_{b,i}^t$. In this case, the SP refuses to rent out its resource.

By comparing eqns. (13) and (19), we can further obtain the critical point of price (PP). For the same service provider, the prices in $Case1$ and $Case2$ are the same, thus, $\mathrm{P}_{b,C2}^t = -\mathrm{P}_{b,C1}^t$. We compare the different *best utility* function for one service provider i, and we have

$$\begin{aligned} w_{b,i}^t \log(\tfrac{w_{b,i}^t}{\mathrm{P}_{b,C1}^t \ln 2}) - \tfrac{w_{b,i}^t}{\ln 2} + \mathrm{P}_{b,C1}^t = w_{b,i}^t \log(\tfrac{w_{b,i}^t}{\mathrm{P}_{b,C1}^t \ln 2}) - \\ w_{b,i}^t - \tfrac{w_{b,i}^t}{\ln 2} + 2\mathrm{P}_{b,C1}^t. \end{aligned} \tag{23}$$

Therefore,

$$\mathrm{P}_{b,C1}^t{}^{**} = w_{b,i}^t. \tag{24}$$

Here, $\mathrm{P}_{b,C1}^t{}^{**}$ denotes PP. It is easy to get the result when $MP < PP < CP$. Therefore, the classification of SP is based on the price given by the NP (NPP), and follows the rules shown by the price zone distribution in Fig. 2 [15].

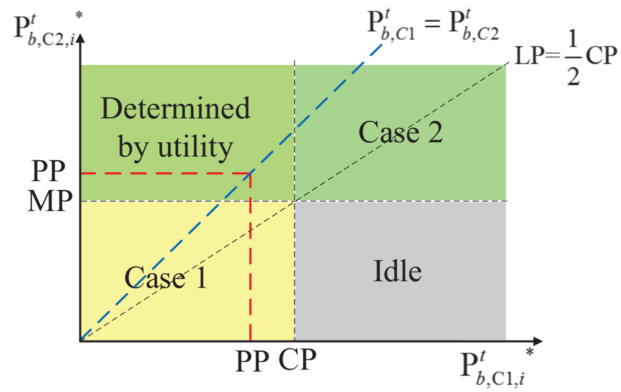


Fig. 2: Classification in each case

- If NPP < PP, the SP prefers to rent extra resource. It is represented by the blue line in the red square in Fig. 2.
- If NPP > PP, the SP prefers to rent out its own resource. It is represented by the blue line out of red square in Fig. 2.
- If NPP = PP, the SP will be self-sufficient. It is represented by the point between blue line and red square in Fig. 2.

The analysis process for computing resources is similar, and thus straightforward.

IV. Discussion on Coalition Result

A. Pareto Optimality

In a coalition game, players prefer joining different coalitions to improve their own utility. There are several possible operations: i) an individual player would like to join a certain coalition if the utility cloud be improved in this coalition; ii) a player of a coalition A would like to leave A but join coalition B if B provides a better utility; iii) a player would like to leave a coalition and work alone if leaving brings a larger utility. We employ a simple but effective mechanism, namely merge-and-split, to derive the coalition game stable formation. In the merge-and-split mechanism, *Pareto optimality* is used as the criterion of the operations of the players.

Definition 1: Consider two sets of coalitions $\mathcal{G}_1 = \{G_1^1, G_2^1, \cdots, G_l^1\}$ and $\mathcal{G}_2 = \{G_1^2, G_2^2, \cdots, G_m^2\}$, which are two different partitions of a same set $G \subset \mathcal{S}$. For a player SP_i, let $u_k(SP_i)$ denote the utility of SP_i in the coalition $\mathcal{G}_k$ ($k = 1, 2$). The coalition $\mathcal{G}_1$ is preferred over $\mathcal{G}_2$ by *Pareto order*, denoted by $\mathcal{G}_1 \triangleright \mathcal{G}_2$, if and only if

$$\begin{aligned} &u_1(SP_i) \geq u_2(SP_i), \forall SP_i \in \mathcal{S}', \\ &\text{with at least an inequality for a player } SP_k. \end{aligned} \tag{25}$$

Following the criterion of Pareto order, the players will be reorganized so that the coalitions are reformed for improving the utilities. This procedure usually takes many rounds. In each round, all the coalitions should be involved so that their utilities are ensured to increase, or at least not to decrease. It means that, the reorganization of coalitions is naturally a global operation. In order to facilitate the procedure, we decouple the global operation by a series of distributed operations using the following two fundamental rules.

- *Merge:* For any set of coalitions $\{G_1, \cdots, G_l\}$, if $\{\bigcup_{j=1}^{l} G_j\} \triangleright \{G_1, \cdots, G_l\}$, then merge $\{G_1, \cdots, G_l\}$ to $\{\bigcup_{j=1}^{l} G_j\}$, denoted by $\{G_1, \cdots, G_l\} \rightarrow \{\bigcup_{j=1}^{l} G_j\}$.
- *Split:* For any coalitions $U_{j=1}^{l} G_j$, if $\{G_1, \cdots, G_l\} \triangleright \{\bigcup_{j=1}^{l} G_j\}$, then split $\{\bigcup_{j=1}^{l} G_j\}$ into $\{G_1, \cdots, G_k\}$, denoted by $\{\bigcup_{j=1}^{l} G_j\} \rightarrow \{G_1, \cdots, G_k\}$.

By using these rules of merge-and-split, the SPs are allowed to negotiate and constitute the local coalitions. So the globally Pareto-optimal collection of coalitions can be consolidated gradually.

B. Conditions of Coalition Formation

In the coalition game, both the members in $Case1$ and in $Case2$ try to maximize their utility values. Therefore, they prefer to the condition which makes them achieve the maximum utility. There are two conditions: working alone or working in a coalition.

1) Condition 1: Working alone.

After coalition game, some SPs may work alone. If, these SPs do not find other SPs to improve their utility since the total number of members in the two sides may be different. Besides, some SPs cannot improve their utility by working in any coalition, such as, the self-sufficient SPs.

2) Condition 2: Resource trading in a coalition.

In this condition, there is at least one SP in $Case1$ SP_A^t and one SP in $Case2$ SP_B^t in one coalition. We suppose SP_A^t and SP_B^t are both their final choice in a coalition, thus, $G_j = \{\mathrm{SP}_A^t, \mathrm{SP}_B^t\}$. Here also we consider radio resource discussion, and the process of computing resource is similar. As the rule of equivalent quantity of trading resource, we use the $\mathrm{DB}_{\mathrm{SP}_A}$ and $\mathrm{DB}_{\mathrm{SP}_B}$ to balance the ratio differences, which are caused by the different quantity of resource. Let $\mathrm{P}_{b,C1,n}^t$ and $\mathrm{P}_{b,C2,n}^t$ denote the negotiated price for renting resources of SP_A^t and SP_B^t, respectively. Here, NP will charge a service fee of $\sigma \mathrm{P}_{b,C2,n}^t (0 < \sigma < 1)$, which could be a very small part of renting price. If $\sigma = 1$, no SP will participate in coalition formation since the loss outweighs the gain. Therefore, eqn. (12) and (18) take the form

$$\begin{cases} x_{b,\mathrm{A}}^t = \frac{w_{b,\mathrm{A}}^t}{\mathrm{P}_{b,C1,n}^t \ln 2} - 1, \\ x_{b,\mathrm{B}}^t = -(\frac{w_{b,\mathrm{B}}^t}{\mathrm{P}_{b,C2,n}^t \ln 2} + 2), \\ \mathrm{DB}_\mathrm{A} x_{b,\mathrm{A}}^t + \mathrm{DB}_\mathrm{B} x_{b,\mathrm{B}}^t = 0, \\ \mathrm{P}_{b,C1,n}^t = -(1+\sigma)\mathrm{P}_{b,C2,n}^t. \end{cases} \tag{26}$$

After solving these equations, we can obtain the quantity of trading resources $x_{b,\mathrm{A}}^t$ and $kx_{b,B}^t$, and the two prices, $\mathrm{P}_{b,C1,n}^t$ and $\mathrm{P}_{b,C2,n}^t$. Moreover, the balance ratio can be denoted as k.

$$x_{b,\mathrm{A}}^t = -kx_{b,B}^t = \frac{(2k+1)w_{b,A}^t}{w_{b,A}^t + k(1+\sigma)w_{b,B}^t} - 1, \tag{27}$$

$$\mathrm{P}_{b,C1,n}^t = \frac{w_{b,A}^t + k(1+\sigma)w_{b,B}^t}{(2k+1)\ln 2}, \tag{28}$$

$$\mathrm{P}_{b,C2,n}^t = -\frac{w_{b,A}^t + k(1+\sigma)w_{b,B}^t}{(2k+1)\ln 2(1+\sigma)}, \tag{29}$$

$$k = \frac{\mathrm{DB_B}}{\mathrm{DB_A}}. \tag{30}$$

C. Stability of the Coalition Formation

In this section, we mainly discuss the stability and convergence of the proposed strategy of coalition formation. We use Pareto-optimal $\mathbb{D}_c$-stable partition to demonstrate the stability of coalition according to [16].

Definition 1: A collection of coalitions $S = \{S_1, ..., S_k\}$ is said to be $\mathbb{D}_c$-stable if it satisfies two conditions:

(a) $i \in \{1, ..., k\}$ and for each partition $\{P_1, ..., P_l\}$ of the coalition G_i: $u(G_i) \geq \sum_{j=1}^{l} u(SP_j)$.

(b) $S \subseteq \{1, ..., k\}$: $\sum_{j \in S} u(G_i) \geq u(\cup_{i \in T} G_i)$.

Theorem 1: The final coalition formation under the proposed strategy can be $\mathbb{D}_c$-stable [17].

Proof: We first consider condition (a). In the final coalition set $\mathcal{G} = \{G_1^1, G_2^1, \cdots, G_l^1\}$, we assume that the SP_i is included in the coalition G_k, i.e., $\mathrm{SP}_i \in G_k$. However, if the SP_i can obtain a higher utility by working alone, or joining other coalitions G_l, condition (a) will be violated. Therefore, according to split-and-merge rules, SP_i will leave from the current coalition G_k. Thus, coalition G_k will not exist. The coalition formations in $\mathcal{G}$ are unstable and can not be the final coalition set. Therefore, condition (a) must be satisfied for any stable coalition generated under the proposed strategy.

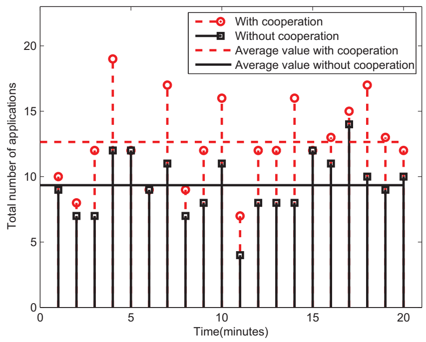
For condition (b), we consider the situation in the same final coalition set $\mathcal{G} = \{G_1^1, G_2^1, \cdots, G_l^1\}$. If coalition G_k can obtain a higher utility, when it combines with other SPs and come into a larger coalition $G_k'(G_k \subseteq G_k')$, the G_k will merge into G_k', such that $u(G_k) < u(G_k')$. $\mathcal{G}$ can not be the final coalition set for the same reason. Thus, for stable formation of the final coalition set, condition (b) needs to be satisfied.

In summary, conditions (a) and (b) will both involve into the final coalition set to ensure the stability of the final result.

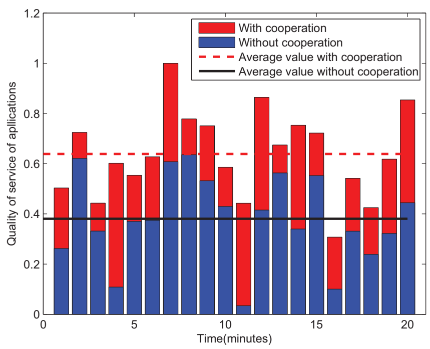
Theorem 2: In the Theorem 1, the final coalition formation is $\mathbb{D}_c$-stable. Therefore, if this partition exists and is stable on the final coalition set, the Pareto optimal solution will be the only one stable solution.

Theorem 3: In the matching process, the coalition could only be formed on the different sides.

Proof: If there exists a coalition on the side of $Case1$, the SP_A and SP_B both are included in coalition G_k, and $\mathrm{P}_{b,C2}^t < \mathrm{P}_{b,C2,B}^{t\ *} < \mathrm{P}_{b,C2,A}^{t\ *}$ according to the rules of classification. Since the coalition will only exist on the mutual benefit condition, SP_A will be the buyer while SP_B will be the seller. The cooperation price $\mathrm{P_{A,B}}$ will satisfy the condition



(a) Total number of applications.



(b) QoS of applications.

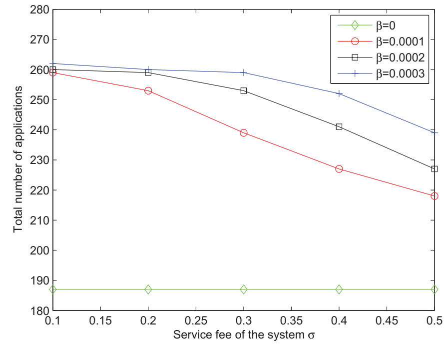
Fig. 3: The comparison of total number of applications and QoS ($\lambda = 1, \mu = 1$).

$\mathrm{P}_{b,C2,B}^{t}{}^{*} < \mathrm{P}_{\mathrm{A,B}} < \mathrm{P}_{b,C2,A}^{t}{}^{*}$. It means that after participating into the coalition, SP_B is still the seller, and SP_A is the buyer. Therefore, the coalition can only exist when it contains the members from both the sides at the same time.

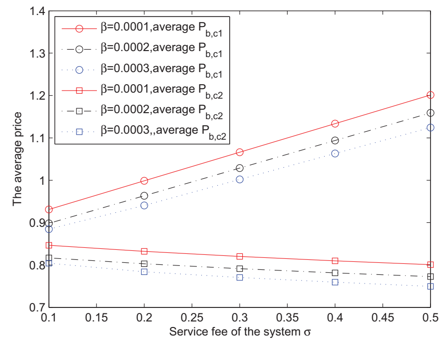
V. Numerical Results

We consider 20 mobile service providers and 10 service areas. The reserved bandwidth of each service provider at the base station is a random value which follows an uniform distribution and ranges from 1 to 10 Mbps. The servers of each service provider at the data center are random values and follow an uniform distribution in the range [1,10]. We consider the set of applications from users, whose bandwidth and server requirements both follow an uniform distribution that ranges from 1 to 4. The arrival and departure time of applications follow Poisson distribution. $\frac{1}{\lambda}$ is the arrival rate of new applications and μ represents the average value of service time [9].

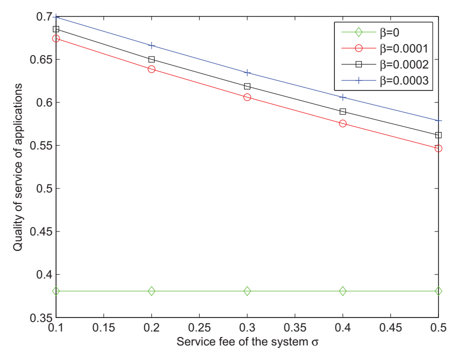
1) Impact of coalition: Fig. 3(a) shows the average number of running applications in every observation time. The solid lines with squares represent the average number of running applications without cooperation during the observation time. The dotted lines with circles are the average number of running applications with cooperation among SPs. Clearly, the average



(a) The impact on total number of applications.



(b) The impact on the average coalition price (take the price of bandwidth for example).



(c) The impact on QoS of applications.

Fig. 4: Performance evaluation with respect to service fee of the system ($\lambda = 1, \mu = 1$).

number of running applications with cooperation is larger than that without cooperation. For instance, at $t = 14$, the average number with cooperation is almost twice of that without cooperation. Considering the whole observation duration, the average value of all running applications without cooperation (dotted line) is about 13, while that without cooperation (solid line) is about 9, which is 44% less than that with cooperation. The figure clearly illustrates that cooperation among SPs can improve resource utilization. Moreover, Fig. 3(b) shows the cooperation also improve all SPs' QoS during working. As a result, Fig. 3 illustrates that our proposed scheme improves the resource utilization and the number of served applications for SPs. And the proposed scheme also increases the QoS of applications.

2) Impact of service price: The service fee, that is announced by the NP and the CP, will have a negative impact on SPs cooperation. It is because that the service fee will add extra cost to the SPs. For simplicity, we consider the service fees of radio and computing resource are of the same values. As mentioned above, the smaller σ brings the lower service fee. Therefore, more and more SPs will participate in coalition formation to improve utility. In Fig. 4(a), the line with rhombus indicates the situation that no one is willing to participate in the coalition formation. Without cooperation between different SPs, the total number of service applications is decided only by the remote reserving resources of each SP. The figure shows that lower service fee during cooperation is conducive to improve resource cooperation.

To further analyze the impact of service fee on coalition formation, we compare the average coalition price with respect to service fee. We use $\beta = [0.003, 0.002, 0.001]$ indicated by in blue, black and red lines in Fig. 4(b), respectively. The lines with circles represent the coalition price for renting extra resources. The lines with squares represent the coalition price for renting one's own idle resources. The difference of coalition price between the two cases will enlarge as the service fee increases. Besides, the coalition price in $Case1$ increases, while the coalition price in $Case2$ decreases. Thus, the extent of coalition formation will decrease, when SPs need to pay higher extra price for local resource and the SPs will earn less.

As shown in Fig. 4(c), the average QoS for each SP is also affected by service fee of the system. The average QoS in the case of no coalition, when $\beta = 0$, is shown by the line with rhombus which is a constant value. In the case of coalition game, the average QoS will decrease as the service fee increases or as the willingness to cooperate is reduced. In summary, on one hand, the service fee will negatively affect coalition formation. On the other hand, if more SPs are willing to take part in the coalition game, SPs can receive high QoS.

VI. Conclusion

In this paper, we introduced a coalition game based model for resource sharing between different service providers in the mobile cloud computing environment. The service providers conform with a virtual resource network which provides it reserved radio and computing resources in order to support the mobile applications. Among service providers, they can cooperate with each other by renting their resource for a short time. The coalition game based scheme promotes scalable cooperation, from which both the SPs and the users can benefit a lot. Further, we have introduced and applied the two-sided matching theory to speed up the coalition formation process. Simulation results indicate that our scheme enhances resource utilization of the SPs and also improves the QoS of the users.

VII. Acknowledgment

The work is partially supported by programs of NSFC under Grant nos. 61422201, 61370159 and U1201253, U1301255, Guangdong Province Natural Science Foundation under Grant no. S2011030002886, High Education Excellent Young Teacher Program of Guangdong Province under grant no. YQ2013057, Science and Technology Program of Guangzhou under grant no. 2014J2200097 (Zhujiang New Star Program), 240079/F20 funded by the Research Council of Norway, and the European Commission FP7 Project CROWN (grant no. PIRSES-GA-2013-627490).

References

[1] M. R. Rahimi, J. Ren, C. H. Liu, A. V. Vasilakos, and N. Venkatasubramanian, "Mobile cloud computing: A survey, state of art and future directions," *MONET*, vol. 19, no. 2, pp. 133–143, 2014.

[2] M. R. Rahimi, N. Venkatasubramanian, S. Mehrotra, and A. V. Vasilakos, "Mapcloud: Mobile applications on an elastic and scalable 2-tier cloud architecture," in *Proc. of the IEEE/ACM Fifth International Conf. on Utility and Cloud Computing*, pp. 83–90, 2012.

[3] F. Xu, F. Liu, H. Jin, and A. V. Vasilakos, "Managing performance overhead of virtual machines in cloud computing: A survey, state of the art, and future directions," *Proc. of the IEEE*, vol. 102, no. 1, pp. 11–31, 2014.

[4] http://www.prepaidmvno.com/category/network/mno/att/.

[5] http://aws.amazon.com/ec2/.GoGrid,http://www.gogrid.com/.

[6] L. Popa, G. Kumar, M. Chowdhury, A. Krishnamurthy, S. Ratnasamy, and I. Stoica, "Faircloud: sharing the network in cloud computing," in *Proc. of the ACM SIGCOMM on Applications, technologies, architectures, and protocols for computer communication*, pp. 187–198, 2012.

[7] L. Mashayekhy, M. M. Nejad, D. Grosu, and A. V. Vasilakos, "Incentive-compatible online mechanisms for resource provisioning and allocation in clouds," in *IEEE 7th International Conf. on Cloud Computing*, pp. 312–319, 2014.

[8] R. Kaewpuang, D. Niyato, P. Wang, and E. Hossain, "A framework for cooperative resource management in mobile cloud computing," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 12, pp. 2685–2700, 2013.

[9] D. Niyato, P. Wang, E. Hossain, W. Saad, and Z. Han, "Game theoretic modeling of cooperation among service providers in mobile cloud computing environments," in *IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 3128–3133, 2012.

[10] H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, "An smdp-based service model for interdomain resource allocation in mobile cloud networks," *IEEE Trans. on Vehicular Technology*, vol. 61, no. 5, pp. 2222–2232, 2012.

[11] S. Ren and M. van der Schaar, "Dynamic scheduling and pricing in wireless cloud computing," *IEEE Trans. on Mobile Computing*, vol. 13, pp. 2283–2292, Oct 2014.

[12] F. Qian, Z. Wang, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "Profiling resource usage for mobile applications: a cross-layer approach," in *Proc. of the 9th international conf. on Mobile systems, applications, and services*, pp. 321–334, ACM, 2011.

[13] Z. Fan, "A distributed demand response algorithm and its application to phev charging in smart grids," *IEEE Trans. on Smart Grid*, vol. 3, no. 3, pp. 1280–1290, 2012.

[14] A. Tversky and D. Kahneman, "The framing of decisions and the psychology of choice," *Science*, vol. 211, no. 4481, pp. 453–458, 1981.

[15] R. Yu, J. Ding, W. Zhong, Y. Liu, and S. Xie, "Phev charging and discharging cooperation in v2g networks: A coalition game approach," *Internet of Things Journal, IEEE*, vol. 1, pp. 578–589, Dec 2014.

[16] N. K. A. T. Chao Wei, Zubair Md. Fadlullah, "Gt-cfs: A game theoretic coalition formulation strategy for reducing power loss in micro grids," *IEEE Trans. on Parallel and Distributed Systems*, SEPTEMBER 2013.

[17] F. P. Kelly, A. K. Maulloo, and D. K. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research society*, pp. 237–252, 1998.