Fair Multi-resource Allocation in Mobile Edge Computing with Multiple Access Points

Erfan Meskar
University of Toronto
Toronto, Ontario, Canada
emeskar@ece.utoronto.ca

Ben Liang
University of Toronto
Toronto, Ontario, Canada
liang@ece.utoronto.ca

ABSTRACT

We consider the problem of fair multi-resource allocation for mobile edge computing (MEC) with multiple access points. In MEC, user tasks are uploaded over wireless communication channels to the access points, where they are then processed with multiple types of computing resources. What distinguishes fair multi-resource allocation in the MEC environment from more general cloud computing is that a user may experience different levels of wireless channel quality on different access points, so that the user's channel bandwidth demand is not fixed. Existing resource allocation studies for cloud computing generally consider Pareto Optimality (PO), Envy-Freeness (EF), Sharing Incentive (SI), and Strategy-Proofness (SP) as the most desirable fairness properties. In this work, we show these properties are no longer compatible in MEC, since there exists no resource allocation rule that can satisfy PO+EF+SP or PO+SI+SP. Hence, we propose a resource allocation rule, called Maximum Task Product (MTP), that retains PO, EF, and SI. Extensive simulation driven by Google cluster traces further shows that MTP improves resource utilization while achieving these fairness properties.

ACM Reference Format:
Erfan Meskar and Ben Liang. 2020. Fair Multi-resource Allocation in Mobile Edge Computing with Multiple Access Points. In *The Twenty-first ACM International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (Mobihoc '20), October 11–14, 2020, Boston, MA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3397166.3409144

1 INTRODUCTION

Computation offloading, *i.e.*, uploading and executing jobs on a remote server, is a strategy to overcome the restrictions of resource-constrained mobile devices. Mobile cloud computing facilitates computation offloading by providing access to the vast resource hosted by clouds, *e.g.*, Google Cloud Platform. However, the latency experienced in reaching a distant cloud server through a wide area network limits the effectiveness of mobile cloud computing. In 2014, the European Telecommunications Standards Institute (ETSI) launched a new standardization group on the so-called Mobile/multi-access Edge Computing (MEC) with the purpose of providing cloud computing capabilities within the radio access network [3]. Hence, instead of utilizing the servers in the core network, mobile users can offload their tasks to the MEC servers at the edge of the network. MEC offers a service environment characterized by proximity, low latency, high bandwidth, and personalized mobile applications [9, 12].

MEC users often have direct access to utilize the computing capabilities of multiple MEC-capable access points (APs) to offload their tasks. To execute a task at an AP, the task input data and execution results are sent through the shared wireless communication link, which may have different channel quality for different APs. In addition to this wireless communication resource, user tasks in MEC further require multiple types of computing resources (*e.g.*, CPU cores and memory) provided by the MEC servers [7]. Hence, MEC is characterized by the allocation of multiple communication and computation resources. Different MEC tasks can consume vastly different amounts of these resources. For instance, language translation, face recognition, and augmented reality applications typically have CPU-intensive tasks, graph analytics and data indexing may have memory-bound tasks, and video processing and vehicle-to-infrastructure communication services can bottleneck on the wireless communication link bandwidth. Such high level of diversity in resource demands significantly complicates resource allocation in MEC. It is challenging to ensure efficient resource utilization and fairness among MEC users that offload different types of tasks. Indeed, it is non-trivial to even define fairness in the multi-resource environment.

Yet, developing a fair resource allocation mechanism is of immense significance to the quality of experience in MEC. To evaluate the fairness of an allocation rule in the multi-resource computing environment, it is common to check whether it satisfies several core properties that are commonly considered as the most desirable [4, 6, 14, 17, 18]:

- *Pareto Optimality (PO)*: increasing a user's utility is impossible without decreasing the utility of another user. This property is critical to high resource utilization, *e.g.*, to avoid the trivial fairness of not allocating any resource to any user.
- *Envy-Freeness (EF)*: no user prefers the allocation of another user. This property avoids the user-perceived unfairness of an allocation rule.
- *Sharing Incentive (SI)*: the utility a user receives is at least as much as the utility it receives from simply splitting the total resources equally. This property ensures that the users are motivated to participate in the resource allocation scheme.
- *Strategy-Proofness (SP)*: the users cannot benefit by lying about their resource demands. This property prevents users from manipulating the scheduler.

Mobihoc '20, October 11–14, 2020, Boston, MA, USA

ACM ISBN 978-1-4503-8015-7/20/10…$15.00
https://doi.org/10.1145/3397166.3409144

To achieve these fairness properties in the MEC environment is more challenging than in those purely computing scenarios in [4, 6, 14, 17, 18]. First, the communication resource and the computing resources are substantially different. In particular, the communication link is external to the computing servers residing within an AP, while existing solutions require that there is at least one server that has non-zero capacity in all resources. More importantly, the wireless channel quality to access different APs usually is different, mainly due to the various distances between the users and the APs. Thus, the link bandwidth demand of a user varies from AP to AP. In contrast, existing solutions require fixed resource demand regardless of computing servers. As explained in Sections 2 and 4, all of the existing fair allocation solutions fall short in the MEC environment.

In this paper, we study the problem of fair resource allocation in the MEC environment with multiple APs. The user tasks require multiple computing resources and communication link bandwidth. Moreover, the users may demand different communication link bandwidth for different APs. Our contributions are as follows:

- We study the performance of existing allocation rules and their direct extensions in MEC and show that they are no longer effective. In particular, Dominant Resource Fairness (DRF) does not ensure fairness when applied to MEC since equalizing the users' dominant share can violate EF. Moreover, allocation rules based on the Kalai-Smorodisky (KS) bargaining solution, *e.g.*, Task Share Fairness (TSF) and Containerized-DRF (C-DRF), do not satisfy EF in MEC, either.
- We study the compatibility of the four core fairness properties in MEC with multiple APs. We show that EF, PO, and SP cannot be satisfied simultaneously. Moreover, SI, PO, and SP cannot be satisfied simultaneously, either. Hence, an allocation rule that utilizes resources efficiently (*i.e.*, satisfying PO) can at best satisfy EF and SI.
- We propose a non-wasteful multi-resource allocation rule for MEC with multiple APs, termed Maximum Task Product (MTP). It is based on constrained maximization of the product of number of tasks over all jobs. We show that it satisfies PO, EF, and SI, which is the best result possible as explained above.
- The performance and efficiency of MTP are further evaluated and compared with existing allocation rules via trace-driven simulation with Google cluster data. MTP is shown to promote efficient resource utilization and achieve superior job completion time while maintaining user fairness.

The organization of this paper is as follows. In Sec. 2, we summarize the existing multi-resource fair allocation solutions. After describing the system model and allocation properties in Sec. 3, we analyze the extended version of the existing allocation rules to show how they fail in MEC in Sec. 4. In Sec. 5, first we present two impossibility theorems to outline the limitations to multi-resource fair allocation in MEC, and then we propose MTP for multi-APs MEC and prove its multi-resource fairness properties. We further evaluate the performance of MTP via trace-driven simulations in Sec. 6 and give concluding remarks in Sec. 7.

2 RELATED WORK

When a single resource, such as the communication bandwidth, is shared among multiple users, max-min fairness satisfies PO, EF, SI, and SP [6]. However, MEC is a multi-channel multi-rate environment. In general wireless communication systems, the proportional fairness (PF) scheduling scheme has been used to provide balance between throughput and fairness. Zhang *et al.* studied the application of PF scheduling in the multi-channel multi-rate environment in [10] and [22]. They show that PF satisfies PO and leads to equal equivalent airtime (*i.e.*, the weighted sum of the airtime of a user on every channel with the weights being the "shadow price"). Several algorithms were proposed to achieve PF allocation in multi-channel multi-rate environments [23–25]. While these works address the multi-rate property of the environment, they all limit the discussion to a single type of resource (*i.e.*, communication channel). However, a resource allocation mechanism in MEC must assign computing resources (*e.g.*, CPU and memory) as well as communication link bandwidth.

The problem of multi-resource allocation was studied by Ghodsi *et al.* [6] in the cloud computing environment. They proposed Dominant Resource Fairness (DRF), an allocation mechanism that describes a notion of fairness when allocating multiple types of resources. DRF computes the share of demanded resources for each user and finds each user's dominant share. It then applies max-min fairness across users' dominant shares. Ghodsi *et al.* [6] proved that DRF meets all four of the desirable properties (*i.e.*, PO, EF, SI, and SP) when all resources are pooled into a single server and tasks are infinitesimally divisible. Parkes *et al.* [14] extended DRF and studied the problem of indivisible tasks. They proved that no mechanism satisfies PO, SI, and SP in that case.

While DRF and several subsequent works address the demand heterogeneity of multiple resources, they all limit the discussion to a simplified model where all resources are concentrated into one server. However, in cloud computing and MEC environments, server heterogeneity presents challenges to developing a fair resource allocation mechanism. In systems with multiple heterogeneous servers, applying DRF per server may lead to allocation with arbitrarily low resource utilization [18]. Instead of allocating resources separately in each server, DRF for Heterogeneous servers (DRFH) jointly considers resource allocation across all servers [18]. It defines the global dominant share for a user based on the aggregate of all the resources and then computes the max-min optimal allocation regarding the global dominant shares. DRFH satisfies PO, EF, and SP [18].

In [17] and [4], although not explicitly stated as such, the proposed allocation rules can be considered as variants of the Kalai-Smorodisky (KS) bargaining solution. Wang *et al.* [17] proposed Task Share Fairness (TSF), and showed that it satisfies the four properties even in the existence of task placement constraints and heterogeneous servers. Unlike previous mechanisms, Friedman *et al.* [4] directly allocated containers, which are isolated bundles of resources. This differs from the model in [6, 17, 18] as users cannot combine bundles. They proved that in both single-server and multi-server systems, no deterministic mechanism allocating containers to users could satisfy PO, SI, and SP simultaneously. Instead, Friedman *et al.* [4] propose Containerized-DRF (C-DRF), a randomized

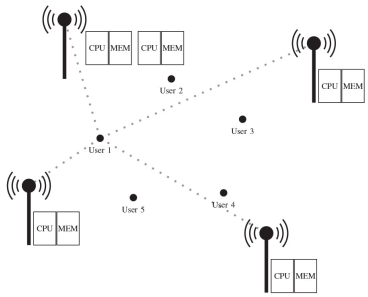


Figure 1: Illustrative example of system model.

mechanism, which satisfies all of the desired properties on average (in an ex-ante sense) in multiple servers with indivisible jobs.

Although DRFH, C-DRF, and TSF consider resource allocation across heterogeneous servers, they cannot be directly applied to MEC. These mechanisms require a server in which every type of resource is contained. In MEC, however, in addition to computing resources, the wireless communication link is the resource that exists outside the computing servers (*i.e.*, shared by the servers). The problem of multi-resource fair allocation in an MEC environment with a single AP was studied in [13], where DRF with an External Resource (DRF-ER) was proposed and shown to satisfy PO, EF, and SP. However, DRF-ER was proposed for a single AP, and it fails to capture the impact of multi-rate channels when there are multiple APs. More specifically, we are interested in MEC where the users' communication bandwidth demand differs from AP to AP. DRF and its follow-up work are not suitable for such environments.

In Sec. 4, we will provide a more detailed discussion on the failure of existing solutions when applied to MEC, after we present the mathematical model of the system in the next section.

3 SYSTEM MODEL AND ALLOCATION PROPERTIES

We consider an MEC environment with a set of APs, denoted by $\mathcal{E}$, each equipped with edge computing servers, as illustrated in Figure 1. A set of users access the edge computing services over the shared communication links. We denote the set of users by $\mathcal{J}$ and the set of computing servers on AP e by $\mathcal{S}_e$. Let $\mathcal{R}$ be the set of resources in the servers (*e.g.*, CPU and memory). For server $s \in \mathcal{S}_e$, its capacity for resource r is denoted by $c_{e,s,r}$, and its capacity profile is $\mathbf{c}_{e,s} = \left(c_{e,s,r}\right)_{r\in\mathcal{R}}$. Without loss of generality, we normalize the aggregate capacity of any computational resource $r \in \mathcal{R}$ to one (*i.e.*, $\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}} c_{e,s,r} = 1$).[1] We denote the MEC computational capacity profile in AP e by $\mathbf{c}_e = \left(\mathbf{c}_{e,s}\right)_{s\in\mathcal{S}_e}$ and the MEC computational capacity profile by $\mathbf{c} = (\mathbf{c}_e)_{e\in\mathcal{E}}$.

The wireless communication link of each AP is a dedicated resource that exists outside of the computing servers. All users share this link when they upload their tasks to the servers at the AP. Let the link bandwidth of AP e be c_e^{BW}, and $\hat{\mathcal{R}}$ be the augmented set of resources which is constructed by adding the external resource (*i.e.*, bandwidth) to the set of computational resources $\mathcal{R}$. Without loss of generality, we normalize the link bandwidth of AP e to one (*i.e.*, $c_e^{\mathrm{BW}} = 1$). However, we emphasize here that the actual physical link bandwidth, when measured in Hz, of different APs in our model usually have different values.

Users in the MEC environment require computing resources and wireless communication bit rate in a customized proportion. For any resource $r \in \mathcal{R}$, user j requires $d_{j,r}$ share of the aggregate capacity of resource r, and bit rate ρ_j per task. To achieve this bit rate, user j requires $d_{j,e}^{\mathrm{BW}}$ of the link bandwidth of AP e. User j may experience different level of wireless channel quality on each AP. For instance, consider the example in Figure 1 where the distances from User 1 to different APs are different. Hence, $d_{j,e}^{\mathrm{BW}}$ can be different for each $e \in \mathcal{E}$, which is a distinguishing characteristic of multi-AP MEC that renders existing solutions ineffective (see Sec.4). We denote the demand profile of user j by $\mathbf{d}_j = \left(\left(d_{j,r}\right)_{r\in\mathcal{R}}, \left(d_{j,e}^{\mathrm{BW}}\right)_{e\in\mathcal{E}}\right)$.

Let $A_{j,e}^{\mathrm{BW}}$ be the share of bandwidth of AP e that is allocated to user j, and $A_{j,e,s,r}$ be the share of computational resource r that is allocated to user j in server s of AP e. We denote the resource allocation profile of user j by $\boldsymbol{A}_j$ where

$$\boldsymbol{A}_j = \left(\left(A_{j,e,s,r}\right)_{\{r\in\mathcal{R},\, s\in\mathcal{S}_e,\, e\in\mathcal{E}\}}, \left(A_{j,e}^{\mathrm{BW}}\right)_{e\in\mathcal{E}}\right).$$

Given some resource allocation profile $\boldsymbol{A}_j$, the number of tasks that user j can execute in server $s \in \mathcal{S}_e$ of AP $e \in \mathcal{E}$ is no more than $\min_{r\in\mathcal{R}}\left\{\frac{A_{j,e,,s,r}}{d_{j,r}}\right\}$. Hence, the number of tasks that this user can execute in AP $e \in \mathcal{E}$ is no more than $\sum_{s\in\mathcal{S}_e}\min_{r\in\mathcal{R}}\left\{\frac{A_{j,e,,s,r}}{d_{j,r}}\right\}$. Moreover, the number of tasks that user j can execute in AP $e \in \mathcal{E}$ is bounded by $\frac{A_{j,e}^{\mathrm{BW}}}{d_{j,e}^{\mathrm{BW}}}$. Therefore, the number of tasks that this user can execute in AP $e \in \mathcal{E}$ is $\min\left\{\frac{A_{j,e}^{\mathrm{BW}}}{d_{j,e}^{\mathrm{BW}}}, \sum_{s\in\mathcal{S}_e}\min_{r\in\mathcal{R}}\left\{\frac{A_{j,e,s,r}}{d_{j,r}}\right\}\right\}$. Thus, the total number of tasks that user j can execute is given by the following utility function.

$$u_j\left(\boldsymbol{A}_j\right) = \sum_{e\in\mathcal{E}}\min\left\{\frac{A_{j,e}^{\mathrm{BW}}}{d_{j,e}^{\mathrm{BW}}}, \sum_{s\in\mathcal{S}_e}\min_{r\in\mathcal{R}}\left\{\frac{A_{j,e,s,r}}{d_{j,r}}\right\}\right\}. \quad (1)$$

An MEC environment E is a tuple

$$E = \left(\mathcal{E}, (\mathcal{S}_e)_{e\in\mathcal{E}}, \hat{\mathcal{R}}, \mathbf{c}, \left(c_e^{\mathrm{BW}}\right)_{e\in\mathcal{E}}, \mathcal{J}, \mathbf{u}\right),$$

where $\mathbf{u} = \left(u_j\right)_{j\in\mathcal{J}}$ is the set of utility functions, which we term the preference profile. Let $x_{j,e,s}$ denote the number of tasks that user j can execute in server s of AP e. Let $\mathbf{x}_j$ and $\mathbf{x}$ denote user j's and all users' task allocation profiles where $\mathbf{x}_j = \left(\left(x_{j,e,s}\right)_{s\in\mathcal{S}_e}\right)_{e\in\mathcal{E}}$, and $\mathbf{x} = \left(\mathbf{x}_j\right)_{j\in\mathcal{J}}$. Given the environment E, the set of feasible task allocation is defined as

$$\mathcal{X}(E) = \Bigg\{\mathbf{x} \mid x_{j,e,s} \geq 0, \sum_{j\in\mathcal{J}}\sum_{s\in\mathcal{S}_e} x_{j,e,s}\, d_{j,e}^{\mathrm{BW}} \leq 1,$$

$$\sum_{j\in\mathcal{J}} x_{j,e,s}\, d_{u,r} \leq c_{e,s,r},\ j \in \mathcal{J}, e \in \mathcal{E}, s \in \mathcal{S}_e, r \in \mathcal{R}\Bigg\}. \quad (2)$$

Definition 1 (Resource Allocation Rule). An allocation rule is a pair of functions (f, F) that specifies for each environment E a non-wasteful allocation, where f and F specify the number of

[1] We can achieve this by scaling the resource capacity and users' demand.

allocated tasks and the amount of allocated resource, respectively. We denote the number of tasks allocated to user j in server s of AP e by $f_{j,e,s}(E)$. The share of computational resource r that is allocated to user j in server s of AP e is denoted by $F_{j,e,s,r}(E)$, and the share of bandwidth of AP e that is allocated to user j is denoted by $F^{\mathrm{BW}}_{j,e}(E)$.

The resource allocation profile of user j derived by rule (f, F) is denoted by

$$\boldsymbol{F}_j(E) = \left(\left(F_{j,e,s,r}(E)\right)_{\{r\in\mathcal{R},\, s\in\mathcal{S}_e,\, e\in\mathcal{E}\}}, \left(F^{\mathrm{BW}}_{j,e}(E)\right)_{e\in\mathcal{E}} \right).$$

Definition 2 (Non-wasteful Resource Allocation Rule). To avoid waste, redundant resources should not be allocated. Allocation rule (f, F) is non-wasteful if

$$F_{j,e,s,r}(E) = f_{j,e,s}(E)\, d_{j,r}, \tag{3}$$

$$F^{\mathrm{BW}}_{j,e}(E) = \sum_{s\in\mathcal{S}_e} f_{j,e,s}(E)\, d^{\mathrm{BW}}_{j,e}. \tag{4}$$

for any $j \in \mathcal{J}$, $e \in \mathcal{E}$, $s \in \mathcal{S}_e$, and $r \in \mathcal{R}$. Thus, allocating resource and task are equivalent in a non-wasteful allocation rule.

We are interested in non-wasteful allocation rules with efficient MEC resource utilization. To that end, we define Pareto optimality as follows:

Definition 3 (Pareto Optimality (PO)). Allocation rule (f, F) is Pareto optimum, if for any environment E, there exists no feasible allocation $\mathbf{y} \in \mathcal{X}(E)$ such that $\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e} y_{i,e,s} \geq \sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e} f_{i,e,s}(E)$ for all $i \in \mathcal{J}$, and $\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e} y_{j,e,s} > \sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e} f_{j,e,s}(E)$ for some $j \in \mathcal{J}$.

The fairness-related properties, *i.e.*, envy-freeness and sharing incentive, are very familiar in the literature (see Sec. 1). First, we define envy-freeness for the MEC environment which embodies the basic meaning of fairness in the sense that no user envies the allocation of another.

Definition 4 (Envy-Freeness (EF)). Allocation rule (f, F) is envy-free, if for any environment E, no user prefers the allocation of another user, *i.e.*, $u_j(\boldsymbol{F}_i(E)) \leq u_j(\boldsymbol{F}_j(E))$ for any environment E and users $i, j \in \mathcal{J}$.

In a multi-resource system in which all resources are pooled into a single server, an allocation satisfies sharing incentive if no user receives utility less than what it can receive under equal division of resources [6]. However, there can be many possible equal division of resources when there are more than one server in the system. In this paper, we focus on the case when each server is equally divided among all users.

Definition 5 (Sharing Incentive (SI)). Allocation rule (f, F) satisfies sharing incentive, if for any environment E, the total number of tasks each user is allowed to process is at least as much as the total number of tasks the user would be allowed to process if each server were equally divided among the users.

Finally, to prevent users from gaming the allocation mechanism we consider strategy-proofness.

Definition 6 (Strategy-Proofness (SP)). Let

$$E = \left(\mathcal{E}, (\mathcal{S}_e)_{e\in\mathcal{E}}, \hat{\mathcal{R}}, \mathbf{c}, \left(c^{\mathrm{BW}}_e\right)_{e\in\mathcal{E}}, \mathcal{J}, \mathbf{u}\right)$$

be an arbitrary environment in which user j reports its true demand, *i.e.*, $\mathbf{d}_{j,r}$, for all resources in $\hat{\mathcal{R}}$. Moreover, let

$$E' = \left(\mathcal{E}, (\mathcal{S}_e)_{e\in\mathcal{E}}, \hat{\mathcal{R}}, \mathbf{c}, \left(c^{\mathrm{BW}}_e\right)_{e\in\mathcal{E}}, \mathcal{J}, \mathbf{u}'\right)$$

be the environment in which user j reports fake demand $\mathbf{d}'_j$, where $\mathbf{d}'_j \neq \mathbf{d}_j$. Allocation rule (f, F) satisfies SP if no user benefits from reporting fake demand, *i.e.*, $u_j\left(\boldsymbol{F}_j(E')\right) \leq u_j\left(\boldsymbol{F}_j(E)\right)$.

To satisfy these fairness properties, existing allocation rules in the fair multi-resource allocation literature all make the assumption that a user demands a fixed amount of a resource across different servers [4, 6, 13, 14, 17, 18]. However, in the multi-AP MEC environment, a user's channel bandwidth demand is not fixed, and in Sec. 4 we will show that existing allocation rules are no longer effective. Our objective is to develop a multi-resource fair allocation scheme for this unique MEC environment, which retains the most important fairness properties. However, in Sec. 5 we show that it is impossible to attain all four properties altogether. Hence, we then focus on designing an allocation rule that is efficient and fair, *i.e.*, one that satisfies PO, EF, and SI.

4 FAILURE OF EXISTING ALLOCATION RULES

In this section, we study the existing allocation rules in the literature. We show that directly extending these allocation rules to MEC cannot satisfy the required properties even in some simple cases.

4.1 Equal Division

We start with the simplest allocation rule, namely Equal Division (EQ), which equally divides all the resources among the users. EQ is guaranteed to satisfy EF and SI since all users receive the same allocation. Moreover, EQ does not depend on the users' reported demands, so it satisfies SP as well. However, it is easy to see that EQ generally is not Pareto optimum. In Sec. 6, we show resources are utilized poorly under EQ in realistic scenarios.

4.2 Dominant Resource Fairness

Ghodsi *et al.* [6] proposed DRF, an allocation mechanism that describes a notion of fairness when allocating multiple types of resources in a single computing server. DRF and its extensions (*e.g.*, DRFH [18] and DRF-ER [13]) find the most demanding resource of a user (*i.e.*, dominant resource) and compute the share of dominant resource for each user (*i.e.*, dominant share). Max-min fairness is applied across the users' dominant shares. Ghodsi *et al.* proved that DRF meets the four desirable properties (*i.e.*, PO, EF, SI, and SP) when tasks are infinitesimally divisible.

The following example shows that applying max-min fairness across the users' dominant shares is not enough to satisfy PO or EF in MEC with multiple APs, even in the simplified case with demands on only the bandwidth resource. Consider an MEC environment that consists of two users (*i.e.* user i and user j) with an unlimited number of tasks and two APs, a and b. Let the users' demands be

$d_{i,a}^{\text{BW}} = 1$, $d_{i,b}^{\text{BW}} = 1/10$, and $d_{j,a}^{\text{BW}} = 1/10$, $d_{j,b}^{\text{BW}} = 1$. There is an infinite number of resource allocation realizations with the max-min fair dominant shares (*i.e.*, 1/2 for each user). Unfortunately, DRF is unable to differentiate between these realizations. For instance, consider the following resource allocation realizations. The first assigns AP a to user i and AP b to user j, and the second assigns AP b to user i and AP a to user j. Both of these realizations satisfy max-min fairness of dominant shares. However, the first one is neither Pareto optimum nor envy-free, while the second one is Pareto optimum and envy-free.

4.3 Relative Task Fairness

In the case of cloud computing environemnts with a single server and divisible jobs, DRF can be interpreted as the KS bargaining solution [5, 8]. Notice that user j's dominant share (*i.e.*, $x_j \max_{r \in \mathcal{R}} \{d_{j,r}\}$) is equal to the ratio of allocated jobs to user j divided by the number of potential jobs for user j if it had the entire server to itself (*i.e.*, $\frac{x_j}{1/\max_{r \in \mathcal{R}}\{d_{j,r}\}}$). Hence, DRF applies max-min fairness across the users' relative number of tasks, which is the KS bargaining solution.

This interpretation is adopted in C-DRF and TSF to allocate resources in cloud computing environments with heterogeneous servers [4, 17]. C-DRF specializes in indivisible tasks. In the case of divisible tasks with placement constraints, Wang *et al.* show that the KS bargaining solution derived by TSF satisfies PO, EF, SI, and SP. However, the following example shows that the KS bargaining solution fails to satisfy EF or SP in MEC environments, even in the simplified case with demands on only the bandwidth resource.

Consider an MEC environment with two APs, namely a and b. Let the link bandwidth demands of users i, j, and k be as follows:

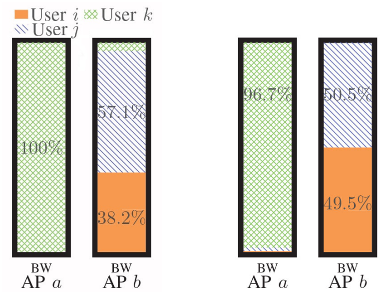
$$d_{i,a}^{\text{BW}} = 1, \quad d_{j,a}^{\text{BW}} = 6/20, \quad d_{k,a}^{\text{BW}} = 13/20,$$
$$d_{i,b}^{\text{BW}} = 1/160, \quad d_{j,b}^{\text{BW}} = 3/20, \quad d_{k,b}^{\text{BW}} = 7/20.$$

Figure 2(a) shows users' share derived from a resource allocation rule based on the KS bargaining solution, using TSF as an example. This allocation rule satisfies PO (the proof is similar to the proof for DRF) and SI (maximizing the equalized relative number of tasks guarantees each user's utility to be better than that of the equal share). However, this resource allocation is not envy-free since user i benefits from exchanging allocation with user j.

More generally, in Sec. 5, Theorem 2 states that there exists no allocation rule that satisfies PO, SI , and SP altogether in MEC with multiple APs. Hence, the resource allocation rule based on the KS bargaining solution (*e.g.*, either C-DRF or TSF) cannot satisfy SP. Figure 2(b) depicts the KS bargaining solution when user i misreports its demand on AP b as $d_{i,b}^{\prime\text{BW}} = 0.5$. This figure shows that the KS bargaining solution is not immune to the strategic behavior since user i could benefit from reporting a fake demand.

5 FAIR MULTI-RESOURCE ALLOCATION WITH MAXIMUM TASK PRODUCT

In MEC with multiple APs, user demand for communication bandwidth is not fixed and depends on the AP. In this section, we first prove the impossibility of satisfying the four properties, PO, EF, SI, and SP, altogether. To be more specific, a resource allocation rule that satisfies PO and EF (or PO and SI) violates SP, necessarily. Hence, we propose a non-wasteful fair multi-resource allocation rule, termed Maximum Task Product, and prove that it satisfies PO, EF, and SI for any number of APs.



(a) TSF with true demands

(b) TSF with fake demands

Figure 2: TSF allocation. (a) All users report their real demand. User i benefits from exchanging allocation with user j. (b) User i increases its number of tasks by reporting fake demand.

5.1 Impossibility Results

First, we present the following three lemmas that concern an envy-free and Pareto optimal allocation in a simplified MEC environment with two APs. These lemmas will be used in to prove Theorems 1 and 2 where we address general MEC environment with more APs.

Lemma 1. *Consider an MEC environment E with $\mathcal{E} = \{e_1, e_2\}$, $\hat{\mathcal{R}} = \{BW\}$, and $\mathcal{J} = \{i, j\}$. If*

$$\frac{d_{i,e_2}^{BW}}{d_{i,e_1}^{BW}} < \frac{d_{j,e_2}^{BW}}{d_{j,e_1}^{BW}} < 1, \tag{5}$$

an allocation rule (f, F) that satisfies PO and EF must have

$$F_{i,e_1}^{BW}(E) = 1 - F_{j,e_1}^{BW}(E) = 0, \quad F_{i,e_2}^{BW}(E) = 1 - F_{j,e_2}^{BW}(E) = \beta,$$

where β is restricted by the following constraint:

$$\frac{1}{2}\left(1 + \frac{d_{i,e_2}^{BW}}{d_{i,e_1}^{BW}}\right) \le \beta \le \frac{1}{2}\left(1 + \frac{d_{j,e_2}^{BW}}{d_{j,e_1}^{BW}}\right). \tag{6}$$

Proof. The allocation rule (f, F) that satisfies PO must utilize the bandwidth resource entirely. Hence, we have $F_{i,e_1}^{\text{BW}}(E) = 1 - F_{j,e_1}^{\text{BW}}(E) = \alpha$ and $F_{i,e_2}^{\text{BW}}(E) = 1 - F_{j,e_2}^{\text{BW}}(E) = \beta$, for some $0 \le \alpha, \beta \le 1$. We study the four possible cases for α and β, and show that only the case ($\alpha = 0, 0 < \beta < 1$) with constraint (6) can satisfy both PO and EF.

Case 1.I ($\alpha = 0, \beta = 0$). In this case, user i envies user j. Hence, (f, F) cannot satisfy EF in this case.

Case 1.II ($0 \le \alpha \le 1, \beta = 1$). It is easy to show that user j envies user i, for all $0 \le \alpha \le 1$. Thus, (f, F) cannot satisfy EF in this case.

Case 1.III ($0 < \alpha \le 1, 0 \le \beta < 1$). We construct the new allocation profiles $\boldsymbol{A}_i$ and $\boldsymbol{A}_j$ from $\boldsymbol{F}_i$ and $\boldsymbol{F}_j$ by having user i giving away ϵ

on e_1 and recieving $\epsilon\frac{d^{\mathrm{BW}}_{i,e_2}}{d^{\mathrm{BW}}_{i,e_1}}$ on e_2, where $\epsilon = \min\left\{\alpha, \frac{d^{\mathrm{BW}}_{i,e_1}}{d^{\mathrm{BW}}_{i,e_2}}(1-\beta)\right\}$. It is easy to check that $\boldsymbol{A}_i$ and $\boldsymbol{A}_j$ are feasible allocation profiles. Then, we have

$$u_i(\boldsymbol{A}_i) = \frac{\alpha-\epsilon}{d^{\mathrm{BW}}_{i,e_1}} + \frac{\beta+\epsilon\frac{d^{\mathrm{BW}}_{i,e_2}}{d^{\mathrm{BW}}_{i,e_1}}}{d^{\mathrm{BW}}_{i,e_2}} = u_i(\boldsymbol{F}_i(E)),$$

$$u_j(\boldsymbol{A}_j) = \frac{1-\alpha+\epsilon}{d^{\mathrm{BW}}_{j,e_1}} + \frac{1-\beta-\epsilon\frac{d^{\mathrm{BW}}_{i,e_2}}{d^{\mathrm{BW}}_{i,e_1}}}{d^{\mathrm{BW}}_{j,e_2}} = u_j(\boldsymbol{F}_j(E)) + \frac{\epsilon}{d^{\mathrm{BW}}_{j,e_1}} - \frac{\epsilon\frac{d^{\mathrm{BW}}_{i,e_2}}{d^{\mathrm{BW}}_{i,e_1}}}{d^{\mathrm{BW}}_{j,e_2}}. \tag{7}$$

Equations (5) and (7) imply that $u_j(\boldsymbol{A}_j) > u_j(\boldsymbol{F}_j(E))$. Hence, (f, F) cannot satisy PO in this case.

Case 1.IV ($\alpha = 0, 0 < \beta < 1$)**.** Consider a feasible allocation $A^{\mathrm{BW}}_{i,e_1} = 1 - A^{\mathrm{BW}}_{j,e_1} = \theta$ and $A^{\mathrm{BW}}_{i,e_2} = 1 - A^{\mathrm{BW}}_{j,e_2} = \gamma$, where $u_i(\boldsymbol{A}_i) \geq u_i(\boldsymbol{F}_i(E))$ and $u_j(\boldsymbol{A}_j) \geq u_j(\boldsymbol{F}_j(E))$.

$$u_i(\boldsymbol{A}_i) \geq u_i(\boldsymbol{F}_i(E)) \Rightarrow \frac{\theta}{d^{\mathrm{BW}}_{i,e_1}} + \frac{\gamma}{d^{\mathrm{BW}}_{i,e_2}} \geq \frac{\beta}{d^{\mathrm{BW}}_{i,e_2}}$$
$$\Rightarrow \theta\frac{d^{\mathrm{BW}}_{i,e_2}}{d^{\mathrm{BW}}_{i,e_1}} \geq \beta - \gamma \tag{8}$$

$$u_j(\boldsymbol{A}_j) \geq u_j(\boldsymbol{F}_j(E)) \Rightarrow \frac{1-\theta}{d^{\mathrm{BW}}_{j,e_1}} + \frac{1-\gamma}{d^{\mathrm{BW}}_{j,e_2}} \geq \frac{1}{d^{\mathrm{BW}}_{j,e_1}} + \frac{1-\beta}{d^{\mathrm{BW}}_{j,e_2}}$$
$$\Rightarrow \beta - \gamma \geq \theta\frac{d^{\mathrm{BW}}_{j,e_2}}{d^{\mathrm{BW}}_{j,e_1}} \tag{9}$$

Equations (5), (8), and (9) are at odds unless $\theta = 0$ and $\gamma = \beta$. Consequently, $\boldsymbol{A}_i = \boldsymbol{F}_i(E)$ and $\boldsymbol{A}_j = \boldsymbol{F}_j(E)$. Hence, $u_i(\boldsymbol{A}_i) \geq u_i(\boldsymbol{F}_i(E))$ and $u_j(\boldsymbol{A}_j) \geq u_j(\boldsymbol{F}_j(E))$ imply $u_i(\boldsymbol{A}_i) = u_i(\boldsymbol{F}_i(E))$ and $u_j(\boldsymbol{A}_j) = u_j(\boldsymbol{F}_j(E))$, for all $0 < \beta < 1$. Thus, (f, F) satisfies PO in this case. Moreover, to satisfy EF we must have

$$u_i(\boldsymbol{F}_i(E)) \geq u_i(\boldsymbol{F}_j(E)) \Rightarrow \frac{\beta}{d^{\mathrm{BW}}_{i,e_2}} \geq \frac{1}{d^{\mathrm{BW}}_{i,e_1}} + \frac{1-\beta}{d^{\mathrm{BW}}_{i,e_2}}$$
$$\Rightarrow \beta \geq \frac{1}{2}\left(\frac{d^{\mathrm{BW}}_{i,e_2}}{d^{\mathrm{BW}}_{i,e_1}} + 1\right), \tag{10}$$

$$u_j(\boldsymbol{F}_j(E)) \geq u_j(\boldsymbol{F}_i(E)) \Rightarrow \frac{1}{d^{\mathrm{BW}}_{j,e_1}} + \frac{1-\beta}{d^{\mathrm{BW}}_{j,e_2}} \geq \frac{\beta}{d^{\mathrm{BW}}_{j,e_2}}$$
$$\Rightarrow \frac{1}{2}\left(\frac{d^{\mathrm{BW}}_{j,e_2}}{d^{\mathrm{BW}}_{j,e_1}} + 1\right) \geq \beta. \tag{11}$$

Equations (10) and (11) imply (6). □

Lemma 2. *Consider an MEC environment E with $\mathcal{E} = \{e_1, e_2\}$, $\hat{\mathcal{R}} = \{BW\}$, and $\mathcal{J} = \{i, j\}$. If*

$$1 < \frac{d^{BW}_{i,e_2}}{d^{BW}_{i,e_1}} < \frac{d^{BW}_{j,e_2}}{d^{BW}_{j,e_1}}, \tag{12}$$

an allocation rule (f, F) that satisfies EF and PO must have

$$F^{BW}_{i,e_1}(E) = 1 - F^{BW}_{j,e_1}(E) = \alpha, \quad F^{BW}_{i,e_2}(E) = 1 - F^{BW}_{j,e_2}(E) = 1,$$

where α is restricted by the following constraint:

$$\frac{1}{2}\left(1 - \frac{d^{BW}_{i,e_1}}{d^{BW}_{i,e_2}}\right) \leq \alpha \leq \frac{1}{2}\left(1 - \frac{d^{BW}_{j,e_1}}{d^{BW}_{j,e_2}}\right). \tag{13}$$

Proof. Lemma 1 is equivalent to Lemma 2 if we swap users and APs and set $\alpha = 1 - \beta$. □

Lemma 3. *Consider an MEC environment E with $\mathcal{E} = \{e_1, e_2\}$, $\hat{\mathcal{R}} = \{BW\}$, and $\mathcal{J} = \{i, j\}$. If*

$$\frac{d^{BW}_{i,e_2}}{d^{BW}_{i,e_1}} < 1 < \frac{d^{BW}_{j,e_2}}{d^{BW}_{j,e_1}}, \tag{14}$$

an allocation rule (f, F) that satisfies PO and EF must have

$$F^{BW}_{i,e_1}(E) = 1 - F^{BW}_{j,e_1}(E) = 0, \quad F^{BW}_{i,e_2}(E) = 1 - F^{BW}_{j,e_2}(E) = \beta, \tag{15}$$

or

$$F^{BW}_{i,e_1}(E) = 1 - F^{BW}_{j,e_1}(E) = \alpha, \quad F^{BW}_{i,e_2}(E) = 1 - F^{BW}_{j,e_2}(E) = 1, \tag{16}$$

where α and β are restricted by the following constraints:

$$\frac{1}{2}\left(1 + \frac{d^{BW}_{i,e_2}}{d^{BW}_{i,e_1}}\right) \leq \beta \leq 1, \quad 0 \leq \alpha \leq \frac{1}{2}\left(1 - \frac{d^{BW}_{j,e_1}}{d^{BW}_{j,e_2}}\right) \tag{17}$$

Proof. The proof outline of Lemma 3 is similar to that of Lemma 1 and is omitted due to space constraint. □

Now we can prove that PO, EF, and SP are not compatible in an MEC environment with multiple APs.

Theorem 1. There exists no allocation rule (f, F) that can satisfy PO, EF, and SP altogether for MEC environments with $|\mathcal{E}| > 1$.

Proof. It suffices to construct a counter-example for MEC environments for each of $|\mathcal{E}| > 1$. Let the allocation rule (f, F) be PO and EF. We first consider the case for $|\mathcal{E}| = 2$ and construct an environment such that there is always a user who can benefit from lying. Consider an MEC environment E with $\mathcal{E} = \{e_1, e_2\}$, $\hat{\mathcal{R}} = \{\mathrm{BW}\}$, and $\mathcal{J} = \{i, j\}$, where (14) holds. Lemma 3 suggests two possible scenarios.

Scenario 1.I (Allocations are given by (15), and β is constrained by (17))**.** The number of tasks that user i can execute is $\frac{\beta}{d^{\mathrm{BW}}_{i,e_2}}$. Suppose user i reports fake demand profile $\boldsymbol{d}'_i$ such that $1 < \frac{d'^{\mathrm{BW}}_{i,e_2}}{d'^{\mathrm{BW}}_{i,e_1}} < \frac{d^{\mathrm{BW}}_{j,e_2}}{d^{\mathrm{BW}}_{j,e_1}}$. We denote the new environment in which user i reports the fake demand $\boldsymbol{d}'_i$ by E'. By Lemma 2, $u_i(\boldsymbol{F}_i(E')) = \frac{\alpha}{d^{\mathrm{BW}}_{i,e_1}} + \frac{1}{d^{\mathrm{BW}}_{i,e_2}}$, where

$$\frac{1}{2}\left(1 - \frac{d'^{\mathrm{BW}}_{i,e_1}}{d'^{\mathrm{BW}}_{i,e_2}}\right) \leq \alpha \leq \frac{1}{2}\left(1 - \frac{d^{\mathrm{BW}}_{j,e_1}}{d^{\mathrm{BW}}_{j,e_2}}\right).$$

Since $\alpha > 0$, user i could benefit from lying.

Scenario 1.II (Allocations are given by (16), and α is constrained by (17))**.** The number of tasks that user j can execute is $\frac{1-\alpha}{d^{\mathrm{BW}}_{j,e_1}}$. Suppose user j reports fake demand profile $\boldsymbol{d}'_j$ such that $\frac{d^{\mathrm{BW}}_{i,e_2}}{d^{\mathrm{BW}}_{i,e_1}} < \frac{d'^{\mathrm{BW}}_{j,e_2}}{d'^{\mathrm{BW}}_{j,e_1}} < 1$.

We denote the new environment in which user i reports the fake demand $\boldsymbol{d}'_j$ by E'. By Lemma 1, $u_j\left(\boldsymbol{F}_j(E')\right) = \frac{1}{d^{\text{BW}}_{j,e_1}} + \frac{1-\beta}{d^{\text{BW}}_{j,e_2}}$, where

$$\frac{1}{2}\left(1+\frac{d^{\text{BW}}_{i,e_2}}{d^{\text{BW}}_{i,e_1}}\right) \le \beta \le \frac{1}{2}\left(1+\frac{d'^{\text{BW}}_{j,e_2}}{d'^{\text{BW}}_{j,e_1}}\right).$$

Since $\beta < 1$, user j could benefit from lying.

Consequently, PO, EF, and SP are not compatible for MEC environments with $|\mathcal{E}| = 2$.

For the case of more than two APs, consider as counter example an MEC environment E with $|\mathcal{E}| > 2$, $\hat{\mathcal{R}} = \{\text{BW}\}$, and $\mathcal{J} = \{i, j\}$. Furthermore, consider the scenario where the set of access points is partitioned into two arbitrary non-empty disjoint sets $\mathcal{E}_1$ and $\mathcal{E}_2$ (*i.e.*, $\mathcal{E}_1 \cup \mathcal{E}_2 = \mathcal{E}$ and $\mathcal{E}_1 \cap \mathcal{E}_2 = \emptyset$) and a user have equal demands on all APs in a partition, *i.e.*, for any AP in $\mathcal{E}_1$, the demands of user i and j are $d^{\text{BW}}_{i,\mathcal{E}_1}$ and $d^{\text{BW}}_{j,\mathcal{E}_1}$, respectively, and for any AP in $\mathcal{E}_2$, the demands of user i and j are $d^{\text{BW}}_{i,\mathcal{E}_2}$ and $d^{\text{BW}}_{i,\mathcal{E}_2}$, respectively. This example is similar to an environment of two APs where the demands of users for APs 1 and 2 are normalized by $|\mathcal{E}_1|$ and $|\mathcal{E}_1|$, respectively. Let the allocation rule (f, F) be PO and EF. Analogous to the case of $|\mathcal{E}| = 2$ above, we can show that there exists a user who can benefit from lying when

$$\frac{d^{\text{BW}}_{i,a_2}/|\mathcal{E}_2|}{d^{\text{BW}}_{i,a_1}/|\mathcal{E}_1|} < 1 < \frac{d^{\text{BW}}_{j,a_2}/|\mathcal{E}_2|}{d^{\text{BW}}_{j,a_1}/|\mathcal{E}_1|},$$

thus completing the construction of a counter-example for $|\mathcal{E}| > 2$. □

Theorem 2. There exists no allocation rule (f, F) that can satisfy PO, SI, and SP altogether for MEC environments with $|\mathcal{E}| > 1$.

Proof. We will reuse the counter examples in the proof of Theorem 1. First we show that SI implies EF for any MEC environemnt E with $|\mathcal{E}| > 1$, $\hat{\mathcal{R}} = \{\text{BW}\}$, and $\mathcal{J} = \{i, j\}$. Suppose the allocation rule (f, F) satisfies SI. Then

$$u_i\left(\boldsymbol{F}_i(E)\right) \ge \sum_{e\in\mathcal{E}} \frac{1/2}{d^{\text{BW}}_{i,e}}. \tag{18}$$

Moreover, given any allocation rule (f, F) and the environment E as described above, we have

$$u_i\left(\boldsymbol{F}_i(E)\right) + u_i\left(\boldsymbol{F}_j(E)\right) \le \sum_{e\in\mathcal{E}} \frac{1}{d^{\text{BW}}_{i,e}}. \tag{19}$$

Equations (18) and (19) indicate that user i cannot envy user j's allocation. Similarly, user j does not envy user i's allocation. Hence, the non-wasteful allocation rule (f, F) satisfies EF. Since we have already shown in Theorem 1 that PO, EF, and SP are incompatible, there cannot exist any allocation rule that satisfies PO, SI, and SP altogether for this couter-example. □

5.2 MTP and Properties

Theorems 1 and 2 indicate that one has to be less strict with the allocation properties. In this section, we present a non-wasteful fair multi-resource allocation for MEC environments, termed MTP, which satisfies PO, EF, and SI.

We represent this allocation rule by (λ, Λ), where λ, and Λ correspond to f and F in Definitions 1 and 2, respectively. It is based on the following convex optimization problem, which can be solved efficiently by well-known numerical methods.

$$\underset{x_{j,e,s}}{\text{maximize}} \quad \sum_{j\in\mathcal{J}} \log \sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} x_{j,e,s} \tag{20a}$$

subject to

$$\sum_{j\in\mathcal{J}} x_{j,e,s}\, d_{j,r} \le c_{e,s,r}, \qquad e\in\mathcal{E},\ s\in\mathcal{S}_e,\ r\in\mathcal{R}, \tag{20b}$$

$$\sum_{j\in\mathcal{J}} \sum_{s\in\mathcal{S}_e} x_{j,e,s}\, d^{\text{BW}}_{j,e} \le c^{\text{BW}}_e, \quad e\in\mathcal{E}. \tag{20c}$$

Let $x^*_{j,e,s}$ be the optimizer of problem (20). Then the number of tasks allocated by function λ is set to

$$\lambda_{j,e,s}(E) = x^*_{j,e,s},$$

and the resource allocation function Λ follows (3) and (4). The term MTP reflects the fact that the sum of log terms in (20a) is equivalent to the product of number of tasks over all jobs. We note here that although (20a) has the form of PF, as explained in Sec. 2, existing solutions based on PF cannot be applied to multi-resource schedule in the MEC environment. In the following, we show that the proposed non-wasteful allocation rule (λ, Λ) satisfies PO, EF, and SI.

Theorem 3. The non-wasteful allocation rule (λ, Λ) is Pareto optimal.

Proof. Let us assume by way of contradiction that there exists an environment E such that the allocation derived by rule (λ, Λ) does not satisfy PO. Then, there exists some task allocation profile $\boldsymbol{y} \in \mathcal{X}(E)$ that Pareto dominates $\boldsymbol{\lambda}(E)$. Therefore,

$$\sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} y_{j,e,s} \ge \sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} \lambda_{j,e,s}(E),$$

for all $j \in \mathcal{J}$, and there exists some $i \in \mathcal{J}$ such that,

$$\sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} y_{i,e,s} > \sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} \lambda_{i,e,s}(E).$$

Consequently,

$$\begin{aligned}\sum_{j\in\mathcal{J}} \log \sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} y_{j,e,s} &> \sum_{j\in\mathcal{J}} \log \sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} \lambda_{j,e,s}(E)\\ &= \sum_{j\in\mathcal{J}} \log \sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} x^*_{j,e,s}.\end{aligned}$$

This contradicts the fact that $x^*_{j,e,s}$ is the optimal solution to problem (20). □

Theorem 4. The non-wasteful allocation rule (λ, Λ) is envy-free.

Proof. By way of contradiction, suppose user u envies user v. Then, for all $r \in \mathcal{R}$

$$\sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} d_{v,r}\lambda_{v,e,s}(E) > \sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} d_{u,r}\lambda_{u,e,s}(E), \tag{21}$$

and

$$\sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} \frac{d^{\text{BW}}_{v,e}}{d^{\text{BW}}_{u,e}}\lambda_{v,e,s}(E) > \sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} \lambda_{u,e,s}(E). \tag{22}$$

Since problem (20) is convex with affine constraints, $\lambda_{j,e,s}(E) = x^*_{j,e,s}$ is optimal if and only if there exists a set of multipliers, μ, γ^{BW}, and ρ, such that the KKT conditions are satisfied [1]. Therefore, for any $e' \in \mathcal{E}$ and $s' \in \mathcal{S}_{e'}$, we have

$$\frac{1}{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\lambda_{u,e,s}(E)} = \sum_{r\in\mathcal{R}}\mu_{e',s',r}d_{u,r} + \gamma^{\text{BW}}_{e'}d^{\text{BW}}_{u,e'} - \rho_{u,e',s'}. \tag{23}$$

This implies that

$$\begin{aligned}1 &\le \sum_{r\in\mathcal{R}}\mu_{e',s',r}\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}d_{u,r}\lambda_{u,e,s}(E) \\ &+ \gamma^{\text{BW}}_{e'}\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}d^{\text{BW}}_{u,e'}\lambda_{u,e,s}(E) \\ &\overset{(a)}{<} \sum_{r\in\mathcal{R}}\mu_{e',s',r}\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}d_{v,r}\lambda_{v,e,s}(E) \\ &+ \gamma^{\text{BW}}_{e'}\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}d^{\text{BW}}_{v,e'}\lambda_{v,e,s}(E) \\ &- \gamma^{\text{BW}}_{e'}\left[\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}d^{\text{BW}}_{v,e'}\lambda_{v,e,s}(E) - \sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}d^{\text{BW}}_{u,e'}\lambda_{u,e,s}(E)\right],\end{aligned} \tag{24}$$

where (a) is due to (21).

Let $\mathcal{E}^+_v$ and $\mathcal{S}^+_{v,e}$ be

$$\mathcal{E}^+_v = \{e\in\mathcal{E} \mid \sum_{s\in\mathcal{S}_e}\lambda_{v,e,s}(E) > 0\},$$
$$\mathcal{S}^+_{v,e} = \{s\in\mathcal{S}_e \mid \lambda_{v,e,s}(E) > 0\}.$$

For any $e^+ \in \mathcal{E}^+_v$ and $s^+ \in \mathcal{S}^+_{v,e^+}$, $\lambda_{v,e^+,s^+}(E) > 0$. Hence,

$$\begin{aligned}&\sum_{r\in\mathcal{R}}\mu_{e^+,s^+,r}\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}d_{v,r}\lambda_{v,e,s}(E) + \\ &\gamma^{\text{BW}}_{e^+}\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}d^{\text{BW}}_{v,e^+}\lambda_{v,e,s}(E) = 1.\end{aligned} \tag{25}$$

Equations (24) and (25) imply that

$$\sum_{e\in\mathcal{E}}\sum_{e\in\mathcal{S}_e}d^{\text{BW}}_{v,e^+}\lambda_{v,e,s}(E) - \sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}d^{\text{BW}}_{u,e^+}\lambda_{u,e,s}(E) < 0.$$

Thus,

$$\frac{d^{\text{BW}}_{v,e^+}}{d^{\text{BW}}_{u,e^+}} < \frac{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\lambda_{u,e,s}(E)}{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\lambda_{v,e,s}(E)}.$$

From this, we have

$$\begin{aligned}\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}\frac{d^{\text{BW}}_{v,e}}{d^{\text{BW}}_{u,e}}\lambda_{v,e,s}(E) &= \sum_{e^+\in\mathcal{E}^+_v}\sum_{s^+\in\mathcal{S}^+_{v,e^+}}\frac{d^{\text{BW}}_{v,e^+}}{d^{\text{BW}}_{u,e^+}}\lambda_{v,e^+,s^+}(E) \\ &< \frac{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\lambda_{u,e,s}(E)}{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\lambda_{v,e,s}(E)}\sum_{e^+\in\mathcal{E}^+_v}\sum_{s^+\in\mathcal{S}^+_{v,e^+}}\lambda_{v,e^+,s^+} \\ &= \frac{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\lambda_{u,e,s}(E)}{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\lambda_{v,e,s}(E)}\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}\lambda_{v,e,s}.\end{aligned}$$

Hence,

$$\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}\frac{d^{\text{BW}}_{v,e}}{d^{\text{BW}}_{u,e}}\lambda_{v,e,s}(E) < \sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}\lambda_{u,e,s}(E).$$

This contradicts (22), which is the supposition that user u envies user v. Thus, the non-wasteful allocation rule (λ, Λ) satisfies EF. □

Now, let us define

$$\psi_{i,e,s} = \min_{r\in\mathcal{R}}\left\{\frac{c_{e,s,r}}{d_{i,r}}\right\} \times \min\left\{\frac{\frac{c^{\text{BW}}_e}{d^{\text{BW}}_{i,e}}}{\sum\limits_{s\in\mathcal{S}_e}\min\limits_{r\in\mathcal{R}}\left\{\frac{c_{e,s,r}}{d_{i,r}}\right\}}, 1\right\}.$$

Note that

$$\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}\psi_{i,e,s} = \sum_{e\in\mathcal{E}}\min\left\{\frac{c^{\text{BW}}_e}{d^{\text{BW}}_{i,e}}, \sum_{s\in\mathcal{S}_e}\min_{r\in\mathcal{R}}\left\{\frac{c_{e,s,r}}{d_{i,r}}\right\}\right\}$$

is the maximum number of tasks that user i can execute by monopolizing all resources in an MEC environment. Next, we show that the non-wasteful allocation rule (λ, Λ) satisfies SI, *i.e.*,

$$\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}\lambda_{i,e,s}(E) \ge \frac{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\psi_{i,e,s}}{|\mathcal{J}|}, \tag{26}$$

for any environment E and any user $i \in \mathcal{J}$.

Theorem 5. The non-wasteful allocation rule (λ, Λ) satisfies SI.

Proof. Equation (23) implies that for any $u \in \mathcal{J}$,

$$\begin{aligned}\frac{\psi_{u,e,s}}{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\lambda_{u,e,s}(E)} &\le \sum_{r\in\mathcal{R}}\mu_{e,s,r}d_{u,r}\psi_{u,e,s} \\ &+ \gamma^{\text{BW}}_e d^{\text{BW}}_{u,e}\psi_{u,e,s} \\ &\le \sum_{r\in\mathcal{R}}\mu_{e,s,r}c_{e,s,r} + \gamma^{\text{BW}}_e d^{\text{BW}}_{u,e}\psi_{u,e,s}.\end{aligned}$$

Thus,

$$\begin{aligned}\frac{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\psi_{u,e,s}}{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\lambda_{u,e,s}(E)} &\le \sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}\sum_{r\in\mathcal{R}}\mu_{e,s,r}c_{e,s,r} + \\ &\sum_{e\in\mathcal{E}}\gamma^{\text{BW}}_e d^{\text{BW}}_{u,e}\sum_{s\in\mathcal{S}_e}\psi_{u,e,s} \\ &\le \sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}\sum_{r\in\mathcal{R}}\mu_{e,s,r}c_{e,s,r} + \sum_{e\in\mathcal{E}}\gamma^{\text{BW}}_e c^{\text{BW}}_e.\end{aligned} \tag{27}$$

For all $u \in \mathcal{J}$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_e$, $\lambda_{u,e,s}(E)\rho_{u,e,s} = 0$. Hence, by multiplying (23) by $\lambda_{u,e,s}(E)$ we have

$$\frac{\lambda_{u,e,s}(E)}{\sum\limits_{e\in\mathcal{E}}\sum\limits_{s\in\mathcal{S}_e}\lambda_{u,e,s}(E)} = \sum_{r\in\mathcal{R}}\mu_{e,s,r}\lambda_{u,e,s}(E)d_{u,r} + \gamma^{\text{BW}}_e\lambda_{u,e,s}(E)d^{\text{BW}}_{u,e}.$$

This implies that

$$1 = \sum_{r\in\mathcal{R}}\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}\mu_{e,s,r}\lambda_{u,e,s}(E)d_{u,r} + \sum_{e\in\mathcal{E}}\gamma^{\text{BW}}_e\sum_{s\in\mathcal{S}_e}\lambda_{u,e,s}(E)d^{\text{BW}}_{u,e}.$$

Therefore,

$$|\mathcal{J}| = \sum_{r\in\mathcal{R}}\sum_{e\in\mathcal{E}}\sum_{s\in\mathcal{S}_e}\mu_{e,s,r}\sum_{u\in\mathcal{J}}\lambda_{u,e,s}(E)d_{u,r} +$$

$$\sum_{e\in\mathcal{E}} \gamma_e^{\mathrm{BW}} \sum_{u\in\mathcal{J}} \left(\sum_{s\in\mathcal{S}_e} \lambda_{u,e,s}(E)\right) d_{u,e}^{\mathrm{BW}}.$$

Hence,

$$|\mathcal{J}| = \sum_{r\in\mathcal{R}} \sum_{e\in\mathcal{E}} \sum_{s\in\mathcal{S}_e} \mu_{e,s,r} c_{e,s,r} + \sum_{e\in\mathcal{E}} \gamma_e^{\mathrm{BW}} c_e^{\mathrm{BW}}. \qquad (28)$$

Equations (27) and (28) imply (26). □

6 TRACE-DRIVEN SIMULATION RESULT

In Sec. 5, we have studied the desirable fairness properties of MTP. In this section, we further evaluate its performance in resource utilization and compare it with TSF and EQ. The former is the best known fair multi-resource allocation rule for multiple servers, and the latter provides strict multi-resource fairness by sacrificing resource utilization. It is worth mentioning that TSF is not directly applicable to our problem, and we extend it by using (1) to find the max-min fair allocation in terms of the users' relative number of tasks. We evaluate the performance of MTP in a realistic setting via large-scale simulation using Google cluster traces [15], which reports the resource usage for computing tasks from Google engineers and services. Other publicly available traces, such as those from Yahoo or Facebook [2], do not provide information on the usage of different resources [16], so we cannot use those traces.

6.1 Experimental Setup

In Google cluster traces, jobs each consisting of multiple tasks are submitted to the servers. The arrival time, duration, and resources demand (CPU and memory) of the tasks are available in the traces. Unfortunately, the wireless communication link bandwidth demand is not provided by any of the publicly available traces. To model the required bandwidth of the tasks, we assumed that $d_{j,\mathrm{CPU}} f_{\mathrm{CPU}} = X\rho_j$, where $d_{j,\mathrm{CPU}}$ is the CPU demand of user j, f_{CPU} is the CPU frequency of the server, ρ_j is the required bit rate of user j and X is a random variable with Gamma distribution [11, 19–21]. In this paper, we adopt the parameter settings of [21] and use shape parameter $\alpha = 4$ and rate parameter $\beta = 200$ to generate X. We consider a general MEC system with frequency division multiple access (FDMA) and estimate $d_{j,e}^{\mathrm{BW}}$, the demanded link bandwidth of user j on AP e, based on ρ_j. The demanded link bandwidth of user j on AP e is uniformly picked between $\frac{0.9\rho_j}{3.5W}$ and $\frac{9\rho_j}{3.5W}$. Note that the link bandwidth demands are normalized by the channels' bandwidth, which is set to 20 MHz. We take the 7-hour computing demand data from the Google traces and simulate their processing on a smaller MEC environment with 5 APs. Table 1 shows the server configuration at each AP. We randomly picked 10 percent of the task submissions of jobs with completion time of less than two hours to create a suitable task arrival data set for this MEC server configuration.

6.2 Resource Utilization and Pareto Optimality

Since it is difficult to directly illustrate Pareto optimality, we use resource utilization as a proxy. Figure 3 compares the resource utilization of MTP, TSF, and EQ. This figure illustrates that MTP and TSF outperform EQ in the utilization of all resources. This is mainly because the latter relinquishes Pareto optimality to satisfy

Table 1: Server configuration. To show the relative capacity of different servers, the resource demand and capacity are re-normalized so that the maximum capacity of each resource is 1.

	Server 1		Server 2	
	CPU core	GB memory	CPU core	GB memory
AP 1	0.25	1	-	-
AP 2	0.25	0.75	0.75	0.25
AP 3	0.25	1	1	1
AP 4	1	1	0.5	0.75
AP 5	0.25	1	0.75	0.5

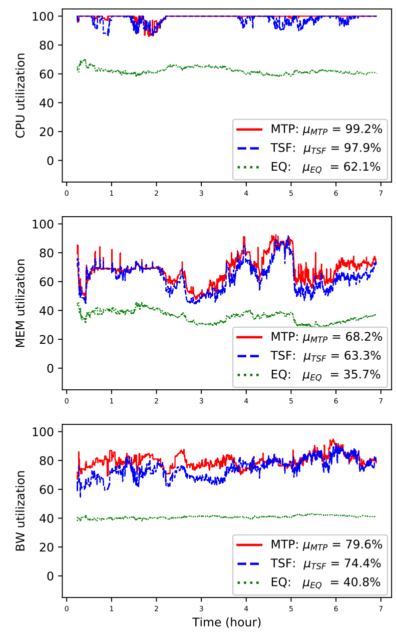


Figure 3: Time series of resource utilization. The average resource utilization is denoted by μ.

strategy-proofness. Although TSF does not satisfy envy-freeness as opposed to MTP, Figure 3 shows that TSF also has no advantage over MTP in terms of resource utilization.

6.3 Fairness Properties

To study the performance of an allocation rule in terms of envy-freeness, we present a notion of enviousness. The enviousness of user i over j, denoted by $\eta_{i,j}$, is defined as the percentage increase in the number of tasks when user i swaps its allocation with user j. The maximum enviousness is $\eta = \max_{i,j\in\mathcal{J}} \{\eta_{i,j}\}$. Figure 4 illustrates the time series of maximum enviousness for the three allocation rules. Since MTP and EQ satisfy EF, η always equal zero. However, we observe that for more than 95% of the simulation duration, EF was violated by TSF. Moreover, the enviousness of users in TSF is substantial.

We remark that the three allocation rules satisfy SI, TSF by maximizing the equalized relative number of tasks which guarantees

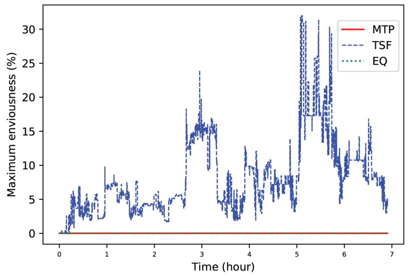


Figure 4: Time series of maximum enviousness.

each user's utility to be better than that of the equal share; EQ by equally sharing all resources; and MTP by Theorem 5. Therefore we omit the experimental results on SI to avoid redundancy.

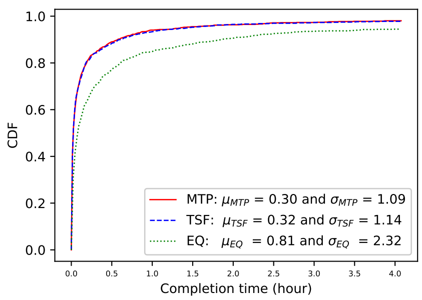


Figure 5: Cumulative distribution function of job completion time. The average and standard deviation are represented by μ and σ, respectively.

6.4 Overall Job Completion Time

Figure 5 compares the cumulative distribution function of the job completion time for the three allocation rules. The job completion time of EQ is 30 minutes greater than that of MTP and TSF on average. Moreover, MTP provides a similar completion time distribution as TSF, even though it guarantees EF while TSF does not.

7 CONCLUSION

In this paper, we consider a system where mobile users run their tasks on MEC servers through multiple MEC-capable APs. Each task requires a specific amount of computing resources and communication data. Since the wireless communication link exists outside the computing servers, we cannot directly apply the conventional multi-resource fair allocation mechanisms. Moreover, a user's demand for the link bandwidth differs among different APs. We show that PO, EF, and SI are no longer compatible with SP in MEC with multiple APs. For this environment, we have proposed the MTP allocation rule, and shown that it satisfies PO, EF, and SI. Simulation driven by Google cluster traces further shows that MTP improves resource utilization while achieving these fairness guarantees.

ACKNOWLEDGMENTS

This work has been supported in part by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] Stephen Boyd and Lieven Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.
[2] Yanpei Chen, Archana Ganapathi, Rean Griffith, and Randy Katz. 2011. The Case for Evaluating MapReduce Performance Using Workload Suites. In *Proc. IEEE International Symposium on Modelling, Analysis, and Simulation of Computer and Telecommunication Systems*.
[3] ETSI. 2014. First Meeting of New Standardization Group on Mobile-edge Computing. (2014). http://www.etsi.org/news-events/news/838-2014-10-news-etsi-announces-first-meeting-of-new-standardization-group-on-mobile-edge-computing
[4] Eric Friedman, Ali Ghodsi, and Christos-Alexandros Psomas. 2014. Strategyproof allocation of discrete jobs on multiple machines. In *Proc. ACM Conf. on Econ. and Comput.*
[5] Eric J. Friedman, Ali Ghodsi, Scott Shenker, and Ion Stoica. 2011. Strategyproofness, Leontief Economies and the Kalai-Smorodinsky Solution. Manuscript.
[6] Ali Ghodsi, Matei Zaharia, Benjamin Hindman, Andy Konwinski, Scott Shenker, and Ion Stoica. 2011. Dominant Resource Fairness: Fair Allocation of Multiple Resource Types. In *Proc. of NSDI*, Vol. 11.
[7] Robert Grandl, Ganesh Ananthanarayanan, Srikanth Kandula, Sriram Rao, and Aditya Akella. 2015. Multi-resource packing for cluster schedulers. *ACM SIGCOMM Computer Communication Review* 44, 4 (2015), 455–466.
[8] Ehud Kalai and Meir Smorodinsky. 1975. Other Solutions to Nash's Bargaining Problem. *Econometrica* 43, 3 (1975), 513–518.
[9] Ben Liang. 2017. Mobile edge computing. *Key Technologies for 5G Wireless Systems* (2017).
[10] Soung Chang Liew and Ying Jun Zhang. 2008. Proportional Fairness in Multi-Channel Multi-Rate Wireless Networks - Part I: The Case of Deterministic Channels with Application to AP Association Problem in Large-Scale WLAN. *IEEE Transactions on Wireless Communications* 7, 9 (Sep. 2008), 3446–3456.
[11] Jacob R. Lorch and Alan Jay Smith. 2001. Improving Dynamic Voltage Scaling Algorithms with PACE. In *Proc. ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*.
[12] Yuyi Mao, Changsheng You, Jun Zhang, Kaibin Huang, and Khaled B. Letaief. 2017. A Survey on Mobile Edge Computing: The Communication Perspective. *IEEE Communications Surveys Tutorials* 19, 4 (Aug. 2017), 2322–2358.
[13] Erfan Meskar and Ben Liang. 2018. Fair multi-resource allocation with external resource for mobile edge computing. In *Proc. IEEE Conf. on Comput. Commun. Workshops (INFOCOM WKSHPS)*.
[14] David C. Parkes, Ariel D. Procaccia, and Nisarg Shah. 2015. Beyond Dominant Resource Fairness: Extensions, Limitations, and Indivisibilities. *ACM Trans. Econ. Comput.* 3, 1, Article 3 (March 2015), 3:1–3:22 pages.
[15] Charles Reiss, Alexey Tumanov, Gregory R. Ganger, Randy H. Katz, and Michael A. Kozuch. 2012. Heterogeneity and dynamicity of clouds at scale: Google trace analysis. In *Proc. ACM Symposium on Cloud Computing*.
[16] Siqi Shen, Vincent van Beek, and Alexandru Iosup. 2015. Statistical Characterization of Business-Critical Workloads Hosted in Cloud Datacenters. In *Proc. IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*.
[17] Wei Wang, Baochun Li, Ben Liang, and Jun Li. 2016. Multi-resource Fair Sharing for Datacenter Jobs with Placement Constraints. In *Proc. International Conference for High Performance Computing, Networking, Storage and Analysis*.
[18] Wei Wang, Ben Liang, and Baochun Li. 2015. Multi-Resource Fair Allocation in Heterogeneous Cloud Computing Systems. *IEEE Transactions on Parallel and Distributed Systems* 26, 10 (Oct. 2015), 2822–2835.
[19] Wanghong Yuan and Klara Nahrstedt. 2003. Energy-efficient Soft Real-time CPU Scheduling for Mobile Multimedia Systems. In *Proc. ACM Symposium on Operating Systems Principles*.
[20] Wanghong Yuan and Klara Nahrstedt. 2006. Energy-efficient CPU Scheduling for Multimedia Applications. *ACM Trans. Comput. Syst.* 24, 3 (Aug. 2006), 292–331.
[21] Weiwen Zhang, Yonggang Wen, Kyle Guan, Dan Kilper, Haiyun Luo, and Dapeng O. Wu. 2013. Energy-Optimal Mobile Cloud Computing under Stochastic Wireless Channel. *IEEE Transactions on Wireless Communications* 12, 9 (Sep. 2013), 4569–4581.
[22] Ying Jun Zhang and Soung Chang Liew. 2008. Proportional Fairness in Multi-Channel Multi-Rate Wireless Networks - Part II: The Case of Time-Varying Channels with Application to OFDM Systems. *IEEE Transactions on Wireless Communications* 7, 9 (Sep. 2008), 3457–3467.
[23] Hui Zhou, Pingyi Fan, and Jie Li. 2009. Cooperative Proportional Fairness Scheduling for Wireless Transmissions. In *Proc. ACM International Conference on Wireless Communications and Mobile Computingy*.
[24] Hui Zhou, Pingyi Fan, and Jie Li. 2011. Global Proportional Fair Scheduling for Networks With Multiple Base Stations. *IEEE Transactions on Vehicular Technology* 60, 4 (May 2011), 1867–1879.
[25] Hui Zhou, Pingyi Fan, Xiang-Gen Xia, and Khaled. B. Letaief. 2010. Achieving Network Wide Proportional Fairness: A Pricing Method. In *Proc. IEEE Wireless Communication and Networking Conference*.