Utility-Based Resource Allocation under Multi-Connectivity in Evolved LTE

Konstantinos Alexandris, Chia-Yu Chang, Kostas Katsalis, Navid Nikaein, Thrasyvoulos Spyropoulos
Communication System Department, EURECOM, France
Email: firstname.lastname@eurecom.fr

***Abstract*—In the current 4G era, the dual connectivity technique utilizes radio resources scheduled by two distinct base stations for a single user equipment to enhance the data throughput. Multi-connectivity, as a natural evolution of dual connectivity, is one of the key 5G techniques to improve both the user performance and overall resource utilization, allowing dynamic user traffic steering across multiple connections of one or more radio access technologies (RATs). However, one of the main challenge in multi-connectivity is to efficiently allocate resources across multiple connections under heterogeneous quality of service (QoS) requirements. In this paper, we examine a resource allocation problem under multi-connectivity in an evolved LTE network and propose a utility proportional fairness (UPF) resource allocation that supports QoS in terms of requested rates. We evaluate the proposed policy with the proportional fairness (PF) resource allocation through extensive simulations and characterize performance gain from both the user and network perspectives under different conditions.**

I. Introduction

Toward the development of 5G vision, it is expected that the mobile broadband service will be enhanced to provide a consistent user experience [1]. Considering the current cellular technologies, cell edge users and those experiencing high interference suffer from poor service, even when coordinated signal processing is applied. To this end, the multi-connectivity is considered as an efficient approach in which simultaneous connections to several technologies or bands [2].

The multi-connectivity concept is characterized by effective resource utilization for seamless user experience [3], enhancement among capacity, coverage and mobility [4], and acting as a quick fail-over method [5]. In general, the multiple connections can be applied among multiple radio access technologies (RATs) [6] or within a single RAT which is viewed as establishing multiple connections to different base station (BS). Take the dual connectivity (DC) in LTE as an example, it is a simplified case with two connections in a single-RAT that enables each user equipment (UE) to receive data simultaneously from two distinct BSs in uplink (UL) and/or downlink (DL).

Despite its appealing, a significant challenge of multi-connectivity is presented in [7] related to the efficient radio resource utilization. The resource utilization in multi-connectivity is crucial to enhance the user performance and to deal with the increasing demand of traffic. In general, two types of application traffic exist, (i) user-to-user such as content/video dissemination, peer-to-peer gaming and public safety, and (ii) user-to-network and network-to-user such as social networking and video-on-demand. Nevertheless, another challenge is related to satisfy the quality of service (QoS) requirement of each traffic flow and optimize network throughput across multiple connections.

This paper addresses the problem of resource allocation under multi-connectivity case with user-to-user traffic and QoS requirement. While in the literature, a resource utility proportional fairness allocation criterion is proposed [8], to the best of our knowledge, none of the previous work consider the QoS under the multi-connectivity case. To this end, the contributions of this work are summarized as follows:

1) We introduce the proportional fairness resource allocation problem that aims to maximize network aggregated throughput for multi-connectivity and compare with legacy single-connected case under different scenarios.
2) We propose an utility-based resource allocation that considers the QoS and analyze the performance gain with aforementioned proportional fairness resource allocation.
3) Finally, we investigate the impact of utility function on QoS satisfaction in terms of its shape.

The rest of the paper is organized as follows. Section II provides the system model and assumptions used through this paper. Section III formulates the optimization problem. Simulation results are discussed in Section IV. Finally, Section V presents the concluding remarks and future work.

II. System model

In this section, we describe the system model and the assumptions used in this work. We consider an area $\mathcal{L} \subset \mathbb{R}^2$ served by a set of BSs $\mathcal{B} = \{b_1, \cdots, b_{|\mathcal{B}|}\}$ and a set of UEs $\mathcal{U} = \{u_1, \cdots, u_{|\mathcal{U}|}\}$ is distributed in this area. For instance, in Fig. 1, an example is presented with $\mathcal{B} = \{b_1, b_2, b_3\}$ and $\mathcal{U} = \{u_1, u_2, u_3, u_4\}$. In following, we examine our model in more detail[1]:

A. Air-interface model

A.1-Mobility: In this work, the user mobility is assumed, implying that all location related parameters are changing in time. For simplicity, we drop the time index t in following.

A.2-Signal to interference plus noise ratio (SINR): The SINR of the received signal from the j-th BS (b_j) to the i-

[1] Additionally, bold symbol denotes column vector; $(\cdot)^T$ denotes transpose; $\mathbb{1}_N$ represents a $N \times 1$ all-ones column vector; $|A|$ is the cardinality of a set A; $\|\cdot\|$ denotes the Euclidean norm.

th UE (u_i) per Physical Resource Block (PRB) along the DL direction is denoted as:

$$\text{SINR}^{\text{D}}_{b_j,u_i} = \frac{\text{RSRP}^{\text{D}}_{b_j,u_i}}{\sum\limits_{b_k \neq b_j} p^{\text{D}}_{b_k,u_i} \text{RSRP}^{\text{D}}_{b_k,u_i} + W^{\text{D}}_{b_j} N_0}. \qquad (1)$$

Respectively, the SINR from the i-th UE to the j-th BS in UL direction is denoted as $\text{SINR}^{\text{U}}_{u_i,b_j}$.

The Reference Signal Received Power (RSRP) is as $\text{RSRP}^{\text{D}}_{b_j,u_i} = L^{\text{D}}_{b_j,u_i} P^{\text{D}}_{b_j,u_i} G^{\text{D}}_{b_j,u_i}$ that includes the path loss and shadowing $L^{\text{D}}_{b_j,u_i}$ from the j-th BS to the i-th UE in DL ($L^{\text{U}}_{u_i,b_j}$ for UL), the transmitted power of the j-th BS to the i-th UE $P^{\text{D}}_{b_j,u_i}$ ($P^{\text{U}}_{u_i,b_j}$ for UL), and the combined antenna gain of the j-th BS and the i-th UE $G^{\text{D}}_{b_j,u_i}$ in the DL ($G^{\text{U}}_{u_i,b_j}$ for UL). The N_0 stands for the thermal noise density in dBm per Hz and $W^{\text{D}}_{b_j}$ is the j-th BS DL bandwidth per PRB in Hz, such that their product is the aggregated noise power per PRB for DL ($W^{\text{U}}_{b_j}$ for UL). In addition, high frequency fluctuations (i.e., Rayleigh fading) are assumed to be filtered and equalized.

Further, the RSRP from other BSs $(b_k \neq b_j)$ in the denominator is assumed to be dependent on the *PRB overlapping probability* as $p^{\text{D}}_{b_k,u_i}$ for the i-th UE. In general, we assume the PRB allocation at each BS is uniformly distributed across all PRBs, so the *PRB overlapping probability* is defined as the summation of allocated PRB to all other UEs $(u_q \neq u_i)$ in percentage ($p^{\text{U}}_{u_i,b_k}$ for UL):

$$p^{\text{D}}_{b_k,u_i} = \sum_{u_l \in \mathcal{U}} \sum_{u_q \neq u_i} x^{\star\text{D}}_{b_k,(u_l,u_q)}, \qquad (2)$$

where $x^{\star\text{D}}_{b_k,(u_l,u_q)}$ is the percentage of allocated PRBs by the k-th BS to user-to-user traffic of user pair (u_l, u_q) along the DL that will be elaborated in **B.3** ($x^{\star\text{U}}_{b_k,(u_q,u_l)}$ for UL).

A.3-Physical data rate: The j-th BS can deliver a maximum physical data transmission rate $R^{\text{D}}_{b_j,u_i}$ to a UE. The physical data rate along the DL in bits per second (bps) is given in Eq. (3) based on the Shannon capacity formula:

$$R^{\text{D}}_{b_j,u_i} = B^{\text{D}}_{b_j} W^{\text{D}}_{b_j} \log_2 \left(1 + \text{SINR}^{\text{D}}_{b_j,u_i}\right), \qquad (3)$$

where $B^{\text{D}}_{b_j}$ is the total number of DL PRBs of the j-th BS ($B^{\text{U}}_{b_j}$ for UL). In respect, the physical data rate from the i-th UE to the j-th BS in UL direction is as $R^{\text{U}}_{u_i,b_j}$.

A.4-Power Control: The open-loop power control is applied in UL and each UE compensates the path loss $L^{\text{U}}_{u_i,b_j}$ and shadowing effects based on the power control parameters (i.e., α, P_0). In that sense, the transmitted power of each PRB from the i-th UE to the j-th BS is:

$$P^{\text{U}}_{u_i,b_j} = \min\left(P^{\max}_{u_i}, P_0 + \alpha \cdot L^{\text{U}}_{u_i,b_j}\right), \qquad (4)$$

where $P^{\max}_{u_i}$ is the maximum transmitted power of the i-th UE. However, in the DL, no power control algorithm is applied and the transmitted power from each BS to all UEs is denoted as $P^{\text{D}}_{b_j,u_i} = P^{\text{D}}_{b_j}$.

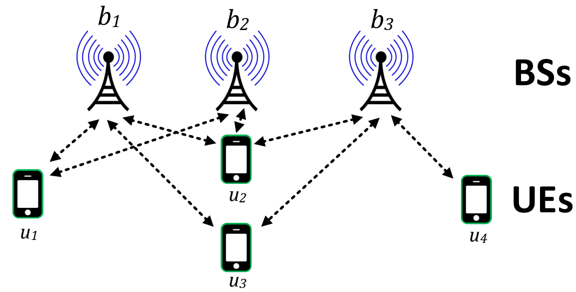


Fig. 1: Multi-connectivity example

B. Connection and Traffic model

B.1-Multi-connectivity: Under multi-connectivity, users can be associated to and communicate with more than one BSs at the same time. We assume the multi-connectivity capability exists in both DL and UL for all UEs and a UE can be connected to a BS if both DL and UL SINR is above a pre-defined threshold, i.e., SINR_{th}. Hence, we define a set $\mathcal{E} \triangleq \{(u_i, b_j), (b_j, u_i) : \min\left(\text{SINR}^{\text{U}}_{u_i,b_j}, \text{SINR}^{\text{D}}_{b_j,u_i}\right) > \text{SINR}_{\text{th}}\}$ that represents all possible connections between UEs and BSs. In contrast, no connection can be established when the condition $\min\left(\text{SINR}^{\text{U}}_{u_i,b_j}, \text{SINR}^{\text{D}}_{b_j,u_i}\right) > \text{SINR}_{\text{th}}$ is not hold for the i-th UE and the j-th BS.

B.2-Local-routing: In principle, routing in the backhaul network is necessary even for the user-to-user traffic served by the same BS. To alleviate the backhaul traffic load, the concept of local-routing is applied [9] in which the user-to-user traffic is routed directly via intermediate BS, thus offloading the core and backhaul network, and reducing the number of hops taken by IP packets to reach the destination [10]. In this case, traffic is not routed via the core network, e.g., LTE evolved packet core (EPC), but only via the intermediate BS [11]. In this work, we focus on the user-to-user traffic flow that can be local-routed, i.e., both user are connected to at least one common BS[2]. Such traffic flow is representative of the public safety (e.g., isolated BSs [10]) and close community application (e.g., community-based video sharing) scenarios where content is shared locally among UEs. To avoid complexity, the backhaul-routed case is out of the scope of this work and will be further surveyed in the future[3].

B.3-Active users pairs: Based on the minimum SINR requirement defined in **B.1**, a set that comprises all active user pairs served by the j-th BS is defined as $\mathcal{C}_{b_j} \triangleq \{(u_i, u_q) : (u_i, b_j), (b_j, u_q) \in \mathcal{E}, u_i \neq u_q\}$ and the user pair $(u_i, u_q) \in \mathcal{C}_{b_j}$ can have user-to-user traffic routed locally via BS b_j. Consequently, a set $\mathcal{C} \triangleq \bigcup_{b_j \in \mathcal{B}} \mathcal{C}_{b_j}$ is formed as the union of all active user pairs. Lastly, two sets are further defined for each user: $\mathcal{D}_{u_i} \triangleq \{u_q : (u_i, u_q) \in \mathcal{C}\}$ comprises all destined UEs from the i-th UE and $\mathcal{S}_{u_i} \triangleq \{u_q : (u_q, u_i) \in \mathcal{C}\}$ comprises all source UEs that can transport traffic to the i-th UE.

B.4-Traffic flow requested rate: It corresponds to the requested rate $\hat{R}_{u_i,u_q}$ determined by the application/service

[2] In Fig.1, the traffic from u_1 to u_2 can be local routed via b_1 or b_2. [3] In Fig.1, the traffic from u_1 to u_4 is not considered

running on the top using an end-to-end established connection of user pair (u_i, u_q), that can go through any intermediate BS via local-routing[4].

III. Problem setup

Based on aforementioned system model, we formulate the optimization problem in this section.

A. Utility function

Our objective here is to allocate optimally the resource to user-to-user traffic based on the applied utility function. We define $x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)} \in \mathcal{X}$ as the percentage of allocated PRB to total PRBs in decimal form along UL/DL direction to transport user-to-user traffic of user pair (u_i, u_q) through BS b_j. It is noted that $x^{\mathrm{U}}_{b_j,(u_i,u_q)} = x^{\mathrm{D}}_{b_j,(u_i,u_q)} = 0$, if $(u_i, u_q) \notin \mathcal{C}_{b_j}$, i.e., no resource is allocated if such user pair can not be local routed through BS b_j. In the following, we introduce two different utility functions: one provides proportional fairness and the other one extends the proportional fairness by taking QoS into consideration.

1) Proportional Fairness (PF): We exploit the logarithmic utility function similar to the one in [12] that maximizes the network aggregated throughput and further include multiple connections to achieve "proportional fairness" naturally. Such "proportionally fairness" characteristic implies that if we increase the allocated data rate of a user pair from the optimal solution, then there must be at least one other user pair will be allocated an inferior data rate that is decreased in a proportion larger than the increased proportion [13]. Such utility function is given as:

$$\Phi(x) = \log(x), \tag{5}$$

2) Utility Proportional Fairness (UPF): The former utility function applies the proportional fairness allocation without considering any QoS requirement. To support QoS in the utility function, we introduce the sigmoid function $S\left(x, \gamma, \hat{R}\right)$ that is used to form the utility function $\Phi(\cdot)$ as:

$$\Phi(x) = \log\left(S\left(x, \gamma, \hat{R}\right)\right) = \log\left(\frac{1}{1 + e^{-\gamma\left(x - \hat{R}\right)}}\right), \tag{6}$$

where γ is a parameter that impacts the shape of sigmoid function, x and $\hat{R}$ respectively corresponds to the allocated data rate and requested data rate of user-to-user traffic of each user pair as introduced in **B.4** of Sec. II. In Fig. 2, the family of sigmoid function with different γ is compared with the linear increment function and step increment function under $\hat{R}$ = 5 Mbps. The sigmoid functions family guarantees that if the allocated rate (i.e., x) is less than the requested rate (i.e, $\hat{R}$), then the priority of this request is increased (i.e., monotonic increment in its slope). However, if the allocated rate is more than the requested one, then the priority of such request is decreased (i.e., monotonic decrement in its slope) [14]. Further, the value of γ impacts the shape of sigmoid function to be more like step function or linear function. It reflects

[4] In Fig. 1, requested rate $\hat{R}_{u_1,u_2}$ and $\hat{R}_{u_2,u_1}$ can go through b_1 and b_2,

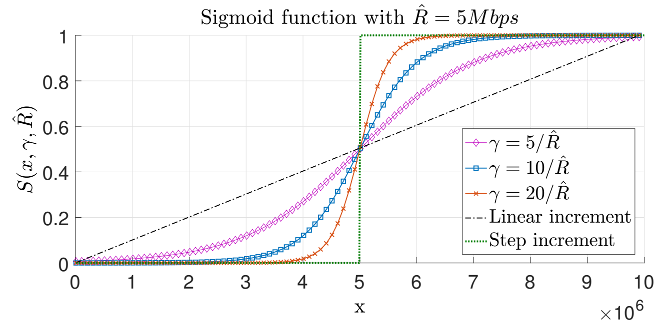


Fig. 2: Sigmoid function family with different γ

the trade-off between the network throughput maximization provided by the "linear" characteristic (better in the network perspective) and the QoS satisfaction by the "non-linear" characteristic (better in the user perspective). Finally, the utility function $\Phi(x)$ takes the logarithm of the sigmoid function that preserves the requested rate of each user pair [15].

Under the multi-connectivity regime assumption (see **B.1** of Sec. II), we can introduce our final objective function based on any utility function $\Phi(\cdot)$ presented in Eq. (5) or Eq. (6) as following:

$$U\left(\mathbf{x}^{\mathrm{U}}_{u_i,u_q}, \mathbf{x}^{\mathrm{D}}_{u_i,u_q}\right) \triangleq \Phi\left(\sum_{b_j \in \mathcal{B}} Q\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)}\right)\right), \tag{7}$$

where $\mathbf{x}^{\mathrm{U}}_{u_i,u_q} \triangleq \left[x^{\mathrm{U}}_{b_1,(u_i,u_q)}, \cdots, x^{\mathrm{U}}_{b_{|B|},(u_i,u_q)}\right]^T$, $\mathbf{x}^{\mathrm{D}}_{u_i,u_q} \triangleq \left[x^{\mathrm{D}}_{b_1,(u_i,u_q)}, \cdots, x^{\mathrm{D}}_{b_{|B|},(u_i,u_q)}\right]^T$ and

$$Q\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)}\right) \triangleq \min\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)} R^{\mathrm{U}}_{u_i,b_j},\; x^{\mathrm{D}}_{b_j,(u_i,u_q)} R^{\mathrm{D}}_{b_j,u_q}\right) \tag{8}$$

stands for the allocated rate for user pair (u_i, u_q). It is noted that the minimum operation is used to only take the bottleneck direction (DL or UL) into account due to the same characteristic is observed in both directions for a single user-to-user traffic[5] of each user pair. Finally, the argument of $\Phi(\cdot)$ in Eq. (7) takes the aggregated allocated rate of user pair (u_i, u_q) over all common BSs (i.e., All $b_j \in \mathcal{B}$).

B. Problem formulation

Based on the proposed system model, assumptions and utility functions, we present our problem formulation. The problem falls into the category of the network utility maximization for resource allocation and is given in Eq. (9). A detailed explanation of the proposed optimization problem follows, presenting both the objective function and constraints.

Objective function: The objective is to allocate the resource $x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)} \in \mathcal{X}$ to each user pair in order to maximize the aggregated network utility function over all user pairs that exchange traffic. The two aforementioned utility functions can be utilized.

[5] Neither the user-to-network traffic nor the network-to-user traffic applies such minimum operation.

$$
\begin{aligned}
\max_{\mathcal{X}} \quad & \sum_{(u_i,u_q)\in\mathcal{C}} U\left(\mathbf{x}^{\mathrm{U}}_{u_i,u_q}, \mathbf{x}^{\mathrm{D}}_{u_i,u_q}\right) \\
\text{s.t.} \quad & \sum_{(u_i,u_q)\in\mathcal{C}_{b_j}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} \le 1,\ \forall\, b_j, \\
& \sum_{(u_i,u_q)\in\mathcal{C}_{b_j}} x^{\mathrm{D}}_{b_j,(u_i,u_q)} \le 1,\ \forall\, b_j, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{D}_{u_i}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} B^{\mathrm{U}}_{b_j} \le B^{\mathrm{U}}_{u_i},\ \forall\, u_i, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{S}_{u_i}} x^{\mathrm{D}}_{b_j,(u_q,u_i)} B^{\mathrm{D}}_{b_j} \le B^{\mathrm{D}}_{u_i},\ \forall\, u_i, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{D}_{u_i}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} P^{\mathrm{U}}_{u_i,b_j} B^{\mathrm{U}}_{b_j} \le P^{\max}_{u_i},\ \forall\, u_i. \quad (9)
\end{aligned}
$$

Constraints: Here, we present a detailed description for each one of them:

1) The first two constraints ensure that the number of allocated PRBs (expressed as the percentage in decimal form on total PRBs) at each BS b_j to all UEs will not exceed the total number of PRBs in both UL/DL.
2) The third and fourth constraints assure that the number of allocated PRBs to each UE among all connected BSs will not exceed the maximum number of allocated PRBs, $B^{\mathrm{U}}_{u_i}, B^{\mathrm{D}}_{u_i}$, in UL/DL of the i-th UE, respectively. Specifically, these constraints take into account all user-to-user traffic related to the i-th UE as $u_q \in \mathcal{D}_{u_i}$ of all destinate UEs in UL and $u_q \in \mathcal{S}_{u_i}$ of all source UEs in DL through any intermediate BS $(b_i \in \mathcal{B})$. Further, these constraints lie on the user capability defined as UE-category in 3GPP TS36.306.
3) Finally, the last constraint is related to the power control mechanism introduced in **A.4** of Sec. II. It restricts the total transmitted power from the i-th user to all connected BSs to be within its power limitation as $P^{\max}_{u_i}$.

Finally, we note that the objective function contains the minimum operation that is concave but non-differentiable. Thus, we need to transform the objective function to a differentiable concave one in order to conclude to a unique tractable global optimal. The following paragraph describes such procedure.

C. Problem transformation

To deal with the minimum operation, we replace $Q\left(x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)}\right)$ in Eq. (8) with the auxiliary variable $z_{b_j,(u_i,u_q)} \in \mathcal{Z}$ and add two extra constraints $z_{b_j,(u_i,u_q)} \le x^{\mathrm{U}}_{b_j,(u_i,u_q)} R_{u_i,b_j}$ and $z_{b_j,(u_i,u_q)} \le x^{\mathrm{D}}_{b_j,(u_i,u_q)} R_{b_j,u_q}$. Then, we define $U\left(\mathbf{z}_{u_i,u_q}\right) \triangleq \Phi\left(\sum_{b_j\in\mathcal{B}} z_{b_j,(u_i,u_q)}\right)$, $\mathbf{z}_{u_i,u_q} \triangleq \left[z_{b_1,(u_i,u_q)}, \cdots, z_{b_{|\mathcal{B}|},(u_i,u_q)}\right]^T$ and the transformed problem in Eq. (10). The objective function in the transformed optimization problem is strictly concave as the sum of $U\left(\mathbf{z}_{u_i,u_q}\right)$ is strictly concave as proved in the Lemma 1.

$$
\begin{aligned}
\max_{\mathcal{X},\mathcal{Z}} \quad & \sum_{(u_i,u_q)\in\mathcal{C}} U\left(\mathbf{z}_{u_i,u_q}\right) \\
\text{s.t.} \quad & z_{b_j,(u_i,u_q)} \le x^{\mathrm{U}}_{b_j,(u_i,u_q)} R^{\mathrm{U}}_{u_i,b_j},\ \forall\, b_j, \\
& z_{b_j,(u_i,u_q)} \le x^{\mathrm{D}}_{b_j,(u_i,u_q)} R^{\mathrm{D}}_{b_j,u_q},\ \forall\, b_j, \\
& \sum_{(u_i,u_q)\in\mathcal{C}_{b_j}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} \le 1,\ \forall\, b_j, \\
& \sum_{(u_i,u_q)\in\mathcal{C}_{b_j}} x^{\mathrm{D}}_{b_j,(u_i,u_q)} \le 1,\ \forall\, b_j, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{D}_{u_i}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} B^{\mathrm{U}}_{b_j} \le B^{\mathrm{U}}_{u_i},\ \forall\, u_i, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{S}_{u_i}} x^{\mathrm{D}}_{b_j,(u_q,u_i)} B^{\mathrm{D}}_{b_j} \le B^{\mathrm{D}}_{u_i},\ \forall\, u_i, \\
& \sum_{b_j\in\mathcal{B}} \sum_{u_q\in\mathcal{D}_{u_i}} x^{\mathrm{U}}_{b_j,(u_i,u_q)} P^{\mathrm{U}}_{u_i,b_j} B^{\mathrm{U}}_{b_j} \le P^{\max}_{u_i},\ \forall\, u_i \quad (10)
\end{aligned}
$$

Lemma 1. *The utility function $U\left(\mathbf{z}_{u_i,u_q}\right)$ in the transformed optimization problem is strictly concave.*

Proof. It is noticed that $U\left(\mathbf{z}_{u_i,u_q}\right)$ function can be written as:

$$
U(\mathbf{z}_{u_i,u_q}) = \Phi\left(\mathbb{1}^T_{|\mathcal{B}|}\mathbf{z}_{u_i,u_q}\right), \quad (11)
$$

and any utility function $\Phi(\cdot)$ is strictly concave. The first one is the logarithmic in Eq. (5) and the second one in Eq. (6) is proved to be strictly concave as the logarithm of the sigmoid function based on Lemma III.1. in [15]. It is known that any composition of a concave function with an affine function is concave [16]. Thus, as expressed in Eq. (11), we conclude that $U(\mathbf{z}_{u_i,u_q})$ is also concave. □

Combined with the linear constraints, the transformed optimization problem is a convex optimization problem with a unique tractable optimal solution.

IV. Simulation Results

In this section, the performance evaluation results are presented for the optimization problem described in Sec. III. Simulation parameters applied to UEs, BSs and network planning are mostly taken from 3GPP (TR36.814, TR36.942, TR25.942) and NGMN documents [17], and some important parameters are listed in TABLE I. Moreover, all assumptions introduced in Sec. II are also held. We use the interior point algorithm to iteratively solve the linear-constrained convex optimization problem of both utility functions: PF and UPF.

A. Comparison: Single-connectivity and Multi-connectivity

Firstly, we compare the performance of legacy single-connectivity case with the multi-connectivity one using the same PF utility function. The considered legacy single-connected case associates each UE to only one BS and only allows each UE to communicate with the associated BS in both UL/DL directions in terms of the best received RSRP. In TABLE II, we compare them in terms of the average

TABLE I: Simulation parameters

Parameter	Value
LTE mode	FDD, SISO
Carrier frequency	DL: 2.19 GHz; UL: 2.0 GHz
Total PRBs of each BS	100 (20 MHz BW)
Maximum PRBs of each UE	100
Number of BSs	3
Initial UE distribution	Uniform of each BS
UE speed	Selected from $[3, 30, 120]$ km/h as [17]
UE direction	Uniform distributed in $[0, 360]$ degree
UE traffic model	Full buffer
SINR threshold	-10 dB
Power control parameters	$P_0 =$ -58 dBm, $\alpha = 0.8$
Maximum transmission power	23 dBm
Thermal noise density	-174 dBm/Hz
Requested rate distribution	Fixed

TABLE II: Comparison of Single/Multi-connectivity

UE number in BS $b_1/b_2/b_3$	Performance metric	Single-connected	Multi-connected
Under-loaded case: 2/2/2	Connected BS	1	2.07
	Connected UE pairs	6	17.52
	Aggregated user rate	0.99 Mbps	20.04 Mbps
Uneven-loaded case: 6/2/2	Connected BS	1	1.34
	Connected UE pairs	34	49.95
	Aggregated user rate	11.68 Mbps	46.73 Mbps
Over-loaded case: 6/6/6	Connected BS	1	1.45
	Connected UE pairs	90	162.22
	Aggregated user rate	55.04 Mbps	57.01 Mbps

TABLE III: QoS metrics comparison of PF and UPF

Metric	Requested rate	PF problem	UPF problem
Satisfaction ratio	0.1Mbps	68.72%	91.39%
	0.5Mbps	42.07%	58.03%
	1Mbps	25.15%	35.33%
	5Mbps	6.42%	23.10%
	10Mbps	< 1%	<1%
Unsatisfied normalized error	0.1Mbps	0.2122	0.0625
	0.5Mbps	0.4131	0.2317
	1Mbps	0.5483	0.3906
	5Mbps	0.7949	0.6607
	10Mbps	0.8833	0.8441

number of connected BS, number of connected user pairs, and aggregated user rate for three representative scenarios, namely under-loaded, uneven-loaded and over-loaded networks.

It can be seen from the table that the number of connected BS per UE is increased with multi-connectivity especially for under-loaded scenario. This is because of the induced inter-cell interference is minor for under-loaded scenario. Moreover, we observe a higher number of connected user pair in multi-connectivity as each UE is able to transmit and receive traffic to a larger set of UEs through the local routing across different BSs. This advantage becomes significant of over-loaded scenario that allows more traffic diversity among users. When comparing the aggregated user rate of all user pairs, we notice that the performance gain is significantly higher in under-loaded scenario followed by the uneven-loaded scenario. This gain is due to schedule UEs across all available BSs, that is indeed one of the expected merit of multi-connectivity. It has to be noted that additional performance gain can be achieved through opportunistic scheduling in time-varying channel of all scenarios. To sum up, the multi-connectivity not only has advantage in user perspective (i.e., more UEs can be reached through multiple BSs) but also in the network perspective (i.e., larger aggregated user rate) in different loading scenarios.

B. Performance analysis of PF & UPF in multi-connectivity

Then, we present the results of both utility functions with a fixed $\gamma = 10/\hat{R}$ in a scenario with 4 UEs that are initially distributed of each BS. Firstly, we define the final allocated rate of the user pair (u_i, u_q) among all intermediate BSs as

$$\zeta_{u_i,u_q} \triangleq \sum_{b_j \in \mathcal{B}} Q\left(x^{\star\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\star\mathrm{D}}_{b_j,(u_i,u_q)}\right), \tag{12}$$

where $x^{\star\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\star\mathrm{D}}_{b_j,(u_i,u_q)}$ are the optimization results of control variables $x^{\mathrm{U}}_{b_j,(u_i,u_q)}, x^{\mathrm{D}}_{b_j,(u_i,u_q)}$. Further, in a quantitative comparison on QoS, we define two different metrics: (i) Satisfaction ratio that represents the ratio of user pairs which are satisfied with the allocated rate in (13), and (ii) Unsatisfied normalized error that shows the normalized Euclidean distance between the allocated rate ζ_{u_i,u_q} and the requested rate $\hat{R}_{u_i,u_q}$ when a user pair (u_i, u_q) is unsatisfied in (14).

$$M_{u_i,u_q} = \text{Prob}\left\{\zeta_{u_i,u_q} \geq \hat{R}_{u_i,u_q}\right\}. \tag{13}$$

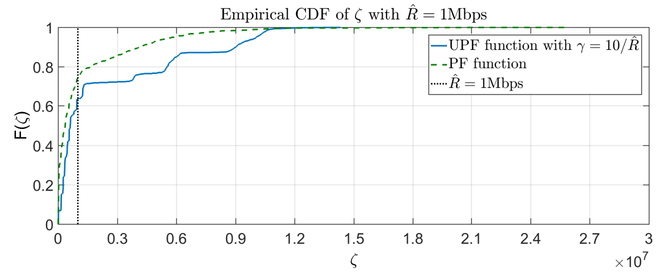
$$E_{u_i,u_q} = \begin{cases} \left\| \frac{\zeta_{u_i,u_q} - \hat{R}_{u_i,u_q}}{\hat{R}_{u_i,u_q}} \right\|, & \text{if } \zeta_{u_i,u_q} < \hat{R}_{u_i,u_q}, \\ 0, & \text{o/w.} \end{cases} \tag{14}$$

Table III shows the results of two QoS metrics using both PF and UPF utility functions with five different requested rate $\hat{R} = \hat{R}_{u_i,u_q},\ \forall (u_i, u_q) \in \mathcal{C}$. In terms of the satisfaction ratio, the UPF is much better than the PF one except in $\hat{R}$ = 10 Mbps case in which both utility functions satisfy less than 1% of user pairs. Further, UPF reduces the unsatisfied normalized error by allocating resources as close as possible to the requested rate. We can see that even the QoS requirement cannot be satisfied for some user pairs mostly in overloaded scenarios, the UPF still provides less error to the requested rate.

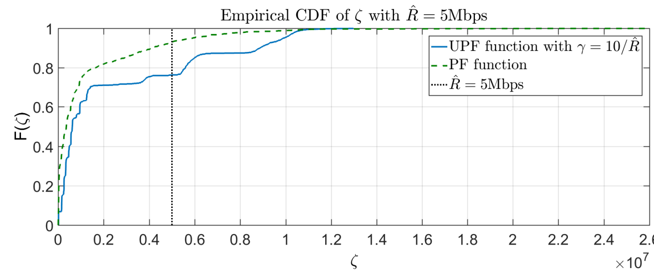
In more qualitative comparison, in Fig. 3, the CDF plot of the allocated rate to all user pairs with three different requested rates are shown. The ratio of satisfied user pair is higher for the UPF case in Fig. 3(a) and Fig. 3(b) and is almost the same of both UPF and PF in Fig. 3(c) that matches the satisfaction ratio in TABLE III. For instance, $M_{u_i,u_q} = 1 - 0.65 = 0.35$ of the UPF case in Fig. 3(a); however, $M_{u_i,u_q} = 0.25$ of the PF case (CDF is 0.75 at this point and it means 75% of user pairs are unsatisfied). We observe that the PF has the same CDF among different requested rate $\hat{R}$ and possesses a longer tail due to the fact that it only maximizes the network throughput without considering any QoS requirements.

C. Impact of γ on UPF problem

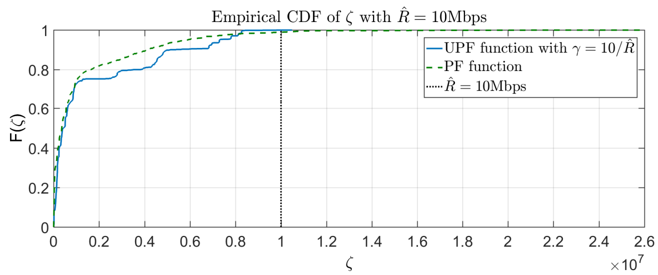
In following, we compare the impact of γ on the UPF utility function in Fig. 4 in terms of the PDF plot of allocated rate ζ_{u_i,u_q} for all user pairs. For simplicity, we only provide the result with $\hat{R} = 1$ Mbps but the same phenomena can be observed for other requested rates. Firstly, we observe that the tail of PDF plot is longer with smaller γ, i.e., the tail is the longest of the three when $\gamma = 5/\hat{R}$. That is because the sigmoid function with smaller γ tends to be more linear that



(a) $\hat{R}$ = 1 Mbps



(b) $\hat{R}$ = 5 Mbps



(c) $\hat{R}$ = 10 Mbps

Fig. 3: CDF plot of PF/UPF utility functions with different $\hat{R}$

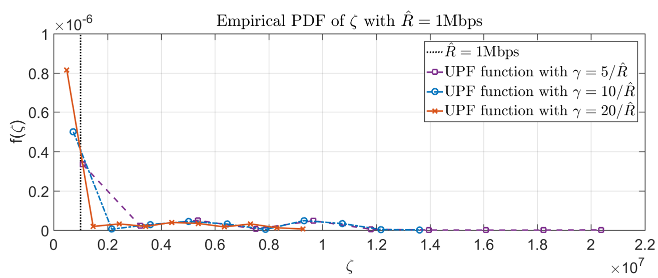


Fig. 4: PDF plot of UPF utility function for several γ

can be satisfied more even though the allocated rate exceeds the requested rate as shown in Fig. 2. The latter approach is better from the network perspective.

Moreover, we observe a significant amount of user pairs are with smaller allocated rate ($\zeta_{u_i,u_q} < 0.4$ Mbps) when γ is large, $\gamma = 20/\hat{R}$, compared to the case when γ is small, $\gamma = 5/\hat{R}$. This is also due to the shape of the sigmoid function in Fig. 2 in which the function with larger γ is more like the step function and prefers to serve the user pair that is close to the QoS requirement. In that sense, some user pairs that are struggle to achieve the requested rate due to the poor SINR condition will be allocated with a smaller data rate or even be deactivated (e.g., $\zeta_{u_i,u_q} = 0$). The latter approach is better from the user perspective.

In summary, the multi-connectivity technique brings benefits in both network and user perspective. Moreover, the UPF takes into account the requested rate in its objective function and is able to satisfy the QoS requirements. Further, the resource allocation policy can be adjusted via changing the value of γ of the sigmoid function in UPF.

V. Conclusion & Future Work

This paper examines a resource allocation problem under multi-connectivity in an evolved LTE network. A utility proportional fairness resource allocation is proposed as an extension to the proportional fairness one, which takes into account the QoS requirement in terms of requested rates. Simulation results reveal that the multi-connectivity can boost the aggregated data rate of user-to-user traffic in under-loaded and uneven-loaded scenarios when compared with the single-connectivity case. In addition, UPF is able to fulfill the requested rate and increase the satisfaction ratio when there are available radio resources among multiple connections. Lastly, the shape of UPF function can be changed in accordance to either user or network perspectives. In future, we plan to extend this work by considering backhaul routing for user-to-network and network-to-user traffic.

Acknowledgments

Research and development leading to these results has received funding from the European Framework Program under H2020 grant agreement 671639 for the COHERENT project.

References

[1] NGMN, "NGMN 5G white paper," 2015.
[2] E. Dahlman *et al.*, "5G wireless access: requirements and realization," *IEEE Communications Magazine*, 2014.
[3] I. Chih-Lin *et al.*, "New paradigm of 5G wireless Internet," *IEEE Journal on Selected Areas in Communications*, 2016.
[4] Nokia, "5G Masterplan - Five keys to create the new communications era," White Paper, 2016.
[5] Ericsson, "5G radio access," White Paper, 2014.
[6] S. Chandrashekar *et al.*, "5G multi-RAT multi-connectivity architecture," in *IEEE ICC workshops*, 2016.
[7] F. B. Tesema *et al.*, "Mobility modeling and performance evaluation of multi-connectivity in 5G intra-frequency networks," in *IEEE Globecom Workshops*, 2015.
[8] W.-H. Wang *et al.*, "Application-oriented flow control: fundamentals, algorithms and fairness," *IEEE/ACM Transactions on Networking*, 2006.
[9] 3GPP, "Feasibility study for proximity services," TR 22.803, 2013.
[10] R. Favraud *et al.*, "Towards moving public safety networks," *IEEE Communications Magazine*, 2016.
[11] A. Laya *et al.*, "Device-to-device communications and small cells: enabling spectrum reuse for dense networks," *IEEE Wireless Communications*, 2014.
[12] Q. Ye *et al.*, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, 2013.
[13] S. Shakkottai *et al.*, "Network optimization and control," *Foundations and Trends in Networking*, 2008.
[14] L. Chen *et al.*, "Utility-based resource allocation for mixed traffic in wireless networks," in *IEEE INFOCOM Workshops*, 2011.
[15] A. Abdel-Hadi *et al.*, "A utility proportional fairness approach for resource allocation in 4G-LTE," in *IEEE ICNC*, 2014.
[16] S. Boyd *et al.*, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
[17] NGMN, "Next generation mobile networks radio access performance evaluation methodology," Tech. Rep., 2008.