# Traffic-aware Carrier Allocation with Aggregation for Load Balancing

Haeyoung Lee, Seiamak Vahid, and Klaus Moessner Institute of Communcation Systems (ICS), University of Surrey, U.K. Email: {Haeyoung.Lee, S.Vahid, K.Moessner}@surrey.ac.uk

Abstract—We consider the resource allocation with aggregation of multiple bands including unlicensed band for heterogeneous traffic. While the mobile data traffic including high volume of video traffic is expected to increase significantly, an efficient management of radio resources from multiple bands is required to guarantee the quality of service (QoS) of different traffic types. In this context, we formulate an optimal resource allocation by using different utility functions for heterogeneous traffic and the two-step resource allocation algorithm including resource grouping has been proposed. Simulation results demonstrate that the proposed algorithm enhances the connection robustness and shows good performance in terms of higher utility value of inelastic traffic even at high traffic loads by steering elastic traffic to unlicensed band.

Index Terms—Resource allocation, carrier aggregation, load balancing, traffic steering

## I. INTRODUCTION

Mobile data traffic is expected to grow annually around 45%, resulting in a ten-fold increase in total data traffic by the end of 2021 compared to 2015. Especially, almost 70% of traffic will be from video by 2021, which means 55% growth annually [1]. In order to support the tremendous growth of data traffic, the use of multiple spectrum bands has been considered, starting on a single band and adding a second and even a third band (e.g., 800 MHz, 1.8GHz, 2.1GHz) [2]. Carrier aggregation (CA) has been standardized in 3GPP and Rel-13 targets aggregation of up to 32 component carriers (CCs) [3]. Along with licensed spectrum, the use of unlicensed spectrum has been investigated, i.e., LTE for Unlicensed (LTE-U) or Licensed Assisted Access (LAA)-LTE [4] and LTE-WLAN Aggregation (LWA) [5]. By using carrier aggregation between the licensed bands and the unlicensed band, the network capacity and data rates can be boosted [6]. However, considering different requirements on quality of service (QoS) of heterogeneous traffic types and different characteristics of each band including unlicensed spectrum, the multi-band traffic management by using carrier aggregation becomes a challenging issue [7].

In the literature, there has been a lot of research efforts on developing mechanisms in support of QoS of different traffic types. In order to measure the degree to which service requirements are satisfied, the approach using utility functions has been investigated widely [8]. Radio resource allocation algorithms based on designed utility functions have been proposed in [9] [10]. Additionally, in [11], the use of multiple carriers are considered and a rate allocation algorithm for multi-carriers is proposed for different traffic types. In [12], unlicensed band utilization is additionally considered. While aforementioned works have been conducted to support QoS of different traffic, to the best of our knowledge, how to efficiently utilize multiple bands with carrier aggregation for heterogeneous traffic has not been fully scrutinized.

In this paper, we consider the problem of resource allocation together with aggregation of multiple bands including unlicensed spectrum for support of various traffic types. For efficient use of multiple bands, the resource grouping concept is introduced. The resource allocation problem is formulated as a two-step optimization problem. Then, by applying the Hungarian method, the optimal primary carriers are allocated. While the secondary carriers allocation problem is formulated as a convex optimization problem, secondary carriers are allocated from multiple bands considering the level of traffic loads and traffic types. At high traffic loads, the proposed twostep resource allocation algorithm based on resource grouping intends to prioritize inelastic traffic to access the licensed spectrum and to utilize unlicensed spectrum for elastic traffic. From the simulation study, it is shown that the proposed algorithm improves the connection robustness by around 176% and the satisfaction level of inelastic traffic's QoS by around 117% even at high traffic loads.

The organization of the paper is as follows. Section II describes the system model and the utility functions of different traffic types. In Section III, the optimization resource allocation problem is formulated. Then, in Section IV, the optimization problem is converted into two-step problem for primary and secondary carriers for a practical approach. The proposed resource allocation with carrier aggregation based on resource grouping is presented. In Section V, the performance of the proposed algorithm is evaluated by simulation and we conclude this paper in Section VI.

# II. SYSTEM AND CHANNEL MODEL

## A. System Model

We consider the downlink transmission of the single-cell OFDM system consisting of one base station (BS) and a set of users (UEs),  $\mathcal{K} = \{1, \dots, K\}$  as depicted in Fig. 1. While the mobile system has its dedicated licensed spectrum over multiple bands, the system can also access unlicensed spectrum as shown in Fig. 2. The set of carriers is labelled as  $\mathcal{N} = \{1, \dots, N\}$ , from the licensed and unlicensed bands. By using carrier aggregation, multiple carriers can be integrated



Fig. 1. Downlink transmission with carrier aggregation to multiple users of heterogeneous traffic types

for the transmission channel. The aggregated channel can include a primary carrier (PC) and multiple secondary carriers (SCs) [4]. The primary carrier provides reliable connection whilst secondary carriers can provide higher data rates [6].

The traffic to be transmitted is classified into two categories, inelastic and elastic traffic [13]. While inelastic traffic is generated by the real-time applications such as VoIP (voice over IP), elastic traffic is generated by applications including FTP (file transfer protocol) and HTTP (hyper text transfer protocol). The set of users  $\mathcal{K}$  is composed of a set of users of inelastic traffic  $\mathcal{K}_{In}$  and a set of users of elastic traffic  $\mathcal{K}_{El}$ .

# B. Channel Model

The signal transmission is on a frame basis, with each frame consisting of multiple OFDM symbols. The received OFDM symbol is  $y_{ij} = h_{ij} \cdot x_{ij} + n_{ij}$ , where  $x_{ij}$  is the transmitted OFDM symbol from the BS to UE *i* on carrier *j* where  $i \in \mathcal{K}$  and  $j \in \mathcal{N}$ . All carriers between the BS and UEs are assumed to be independent and identically distributed (i.i.d.) Rayleigh fading. The channel coefficient,  $h_{ij}$  is complex Gaussian random variable with zero mean and variance  $\sigma_{h_{ij}}^2$ , i.e.,  $h_{ij} \sim C\mathcal{N}(0, \sigma_{h_{ij}}^2)$ . The channel state information (CSI) is assumed to be perfectly known at the receiver. The received signal to noise ratio (SNR) is represented as

$$\rho_{ij} = \frac{|h_{ij}|^2 P_{ij}}{\sigma^2} = \frac{g_{ij} P_{ij}}{\sigma^2},$$
 (1)

where  $g_{ij} = |h_{ij}|^2$ , for all *i*, *j*.  $g_{ij}$  is a Chi-square distributed random variable with 2*M* degrees of freedom where *M* denotes the number of receiver antennas along with multiantenna techniques (e.g., maximal ratio combining). The channel gains are assumed to remain unchanged within each frame. The transmit power,  $P_{ij} = P_{\max}/N$  for equal power allocation with a given  $P_{\max}$ . Notice in (1) that  $\sigma^2$  is the variance of the complex-valued zero-mean additive white Gaussian noise (AWGN). Then, the data rate is expressed as

$$r_{ij} = W_j \cdot \log_2 \left( 1 + \rho_{ij} \right), \tag{2}$$

where  $W_j$  is the bandwidth of carrier *j*. Consider the BS can use multiple carriers by aggregation to transmit to UEs. Then, the data rate to UE *i*,  $r_i$  is given by

$$r_i = \sum_{j \in \mathcal{N}} \alpha_{ij} \cdot r_{ij},\tag{3}$$



Fig. 2. An example of available heterogeneous spectrum over multiple bands

where  $\alpha_{ij}$  is the carrier allocation indicator, i.e.,  $\alpha_{ij} = 1$  indicates that carrier *j* is allocated to UE *i* and  $\alpha_{ij} = 0$ , otherwise. Given the channel inputs, the scheduler at the BS allocates different carriers to different users.

#### C. Applications Utility Functions

We consider utility functions incorporating required data rates to express the satisfaction level of required QoS. The application utility function of UE *i*,  $U_i(r_i)$  is represented by sigmoidal-like function or logarithmic function [11] [13]. These functions have the following properties: 1)  $U_i(0) = 0$ and  $U(r_i)$  is an increasing function of  $r_i$ , 2)  $U_i(r_i)$  is twice continuously differentiable in  $r_i$ . For inelastic traffic UEs, we use the normalized sigmoidal-like utility function, that is

$$U_i(r_i) = c \left( \frac{1}{1 + e^{-a_i(r_i - b_i)}} - d_i \right),$$
 (4)

where  $c_i = (1 + e^{a_i b_i})/e^{a_i b_i}$  and  $d_i = 1/e^{a_i b_i}$  so it satisfies U(0) = 0 and  $U(\infty) = 1$ . Considering the characteristics of traffic types including required QoS, the utility parameters  $a_i$  and  $b_i$  can be set.

The normalized logarithmic utility function is utilized for elastic traffic UEs, that can be represented as

$$U_{i}(r_{i}) = \frac{\log(1 + k_{i}r_{i})}{\log(1 + k_{i}r_{i}^{\max})},$$
(5)

where  $r_i^{\max}$  is the required rate for the user to achieve 100% utilization and  $k_i$  is the slope of the curve that varies based on the user application (i.e., the increasing rate of utility percentage with the allocated rate  $r_i$ ). It satisfies  $U_i(0) = 0$  and  $U_i(r_i^{\max}) = 1$ . From (3),  $U_i(r_i)$  in (4) and (5) can be represented as the multi-variable functions  $U_i(\sum \alpha_{ij}r_{ij})$ .

## **III. PROBLEM FORMULATION**

To allocate multiple carrier resources to multiple UEs, we consider the general utility proportional fairness (PF) objective function formulation in (6).

$$P: \max_{\alpha_{ij}} \prod_{i \in \mathcal{K}} U_i \bigg( \sum_{j \in \mathcal{N}} \alpha_{ij} r_{ij} \bigg), \tag{6}$$

s.t. 
$$\sum_{i \in \mathcal{K}} \alpha_{ij} \le 1, \ \forall j \in \mathcal{N},$$
 (7)

$$\alpha_{ij} = \{0, 1\}, \ \forall i \in \mathcal{K}, \forall j \in \mathcal{N},$$
(8)

$$\sum_{j \in \mathcal{N}} \alpha_{ij} r_{ij} \ge R_i^{\min}, \ \forall i \in \mathcal{K}.$$
(9)

The objective of problem P in (6) is to allocate carriers for UEs to maximize the total system utility while ensuring proportional fairness between UEs (i.e., the product of the utilities of all UEs). The function (6) ensures non-zero resource allocation for all UEs [14]. While  $\alpha_{ij}$  is the allocation indicator as shown in (8), each carrier is exclusively allocated to one user, thus (7) is imposed as the constraint. For UE i, the minimum rate requirement  $R_i^{\min}$  is considered to guarantee a QoS and its constraint is added in (9).

# IV. THE PROPOSED TWO-STEP RESOURCE ALLOCATION WITH RESOURCE GROUPING

By solving the resource allocation problem in (6)-(9), the aggregated channel can be selected for each UE. However, under practical scenarios, for example, in LTE-Advanced, when a UE first establishes or re-establishes radio resource control (RRC) connection with the BS, only one carrier, called as primary component carrier (PCC) is configured (for initial control and data traffic). Then, depending on traffic load and QoS requirements, one or more additional carriers called as secondary component carriers (SCCs) are configured. Therefore the PCC should be robust, and is typically chosen such that it provides the most ubiquitous coverage and/or best signal quality [6]. Considering this, the primary carrier can be allocated from licensed bands and then, the secondary carriers can be allocated from licensed and/or unlicensed bands [4]. In this paper, we introduce resource grouping, whereby available carriers are categorized into three groups considering the channel characteristics and spectrum regime. An example of resource grouping is presented in Fig. 2. Considering a given maximum allowed number of users  $N_{\rm max}$ , the same number of carriers in lower band (i.e., better channel quality) are grouped  $(\mathcal{G}_P)$  for primary carrier while the rest of carriers are grouped  $(\mathcal{G}_S)$  for secondary carriers. Considering the spectrum regime,  $\mathcal{G}_S$  are separated with two groups,  $\mathcal{G}_{S1}$  of carriers in licensed bands and  $\mathcal{G}_{S2}$  in unlicensed bands. With resource grouping, we intend to enhance connection robustness by allocating primary carriers of good channel quality and improve the QoS of inelastic traffic by steering elastic traffic to unlicensed bands. Similarly to [12], we adopt a new two-step resource allocation for primary and secondary carriers.

## A. Resource Allocation of Primary Carriers

While carriers in  $\mathcal{G}_P$  are for the primary carrier, the primary carrier allocation problem can be formulated as follows.

$$P1: \max_{\alpha_{ij}} \sum_{i \in \mathcal{K}} \log \left[ U_i (\sum_{j \in \mathcal{G}_P} \alpha_{ij} r_{ij}) \right]$$
(10)

s.t. 
$$\sum_{i\in\mathcal{K}}\alpha_{ij}=1, \ \forall j\in\mathcal{G}_P,$$
 (11)

$$\sum_{j \in \mathcal{G}_P} \alpha_{ij} \le 1, \; \forall i \in \mathcal{K}, \tag{12}$$

$$\alpha_{ij} \in \{0,1\}, \ \forall i \in \mathcal{K}, j \in \mathcal{G}_P, \tag{13}$$

$$\sum_{j \in \mathcal{G}_P} \alpha_{ij} r_{ij} \ge R_i^{\min}, \ \forall i \in \mathcal{K}.$$
 (14)

Note that  $\arg \max \prod U_i(\sum \alpha_{ij}r_{ij})$  in (6) is equivalent to  $\arg \max \sum \log [U_i(\sum \alpha_{ij}r_{ij})]$ . Thus, the objective function (10) and constraints in (12)- (14) are the same with (6), (7)-(9)of the problem P. (11) is additionally imposed since at most one carrier can be allocated to UE for the primary carrier.

Since P1 in (10)-(14) is formulated as the one-to-one matching problem (between 'UE' and 'carrier') [15], we apply the Hungarian algorithm [16] which is well-known to solve the one-to-one matching problem in polynomial time optimally. We add the step to verify whether the constraint (14) is satisfied or not.

# B. Resource Allocation of Secondary Carriers

After allocating the primary carriers, secondary carriers are allocated depending on the QoS requirements and spectrum availability. Based on (6), the resource allocation problem for secondary carriers is formulated as follows.

$$P2: \max_{\alpha_{ij}} \sum_{i \in \mathcal{K}} \log \left[ U_i (r_i^{pc} + \sum_{j \in \mathcal{G}_n} \alpha_{ij} r_{ij}) \right]$$
(15)

s.t. 
$$\sum_{i \in \mathcal{K}} \alpha_{ij} \le 1, j \in \mathcal{G}_n,$$
 (16)

$$\alpha_{ij} \in \{0, 1\}, \forall i \in \mathcal{K}, \forall j \in \mathcal{G}_n, \quad (17)$$

$$\mathcal{G}_n \in \{\mathcal{G}_S, \mathcal{G}_{S1}, \mathcal{G}_{S2}\}.$$
(18)

In (15),  $r_i^{pc}$  is the achievable data rate of UE i from the

allocated primary carrier, i.e.,  $r_i^{pc} = \sum_{j \in \mathcal{G}_P} \alpha_{ij} r_{ij}$ . While primary carriers are chosen from  $\mathcal{G}_P$ , secondary carriers are chosen from  $\mathcal{G}_{S1}$  and/or  $\mathcal{G}_{S2}$  as described in (18) depending on the traffic types and traffic load in the system. At the low traffic load, (i.e., the traffic load  $L_C$  is lower than the threshold  $L_{TH}$ ), secondary carriers for all K UEs can be selected in  $\mathcal{G}_{S1}$ . As the traffic loads increases and exceeds the given threshold ( $L_C > L_{TH}$ ), UEs of inelastic traffic (in  $\mathcal{K}_{In}$ ) are prioritized to access carriers in licensed bands ( $\mathcal{G}_{S1}$ ) and UEs of elastic traffic (in  $\mathcal{K}_{El}$ ) are allowed to use remained carriers in  $\mathcal{G}_{S1}$  and carriers in  $\mathcal{G}_{S2}$ . When the amount of inelastic traffic is too large to be served by licensed bands, unlicensed band will be utilized for the inelastic traffic as well as elastic traffic. It is to support QoS satisfaction of inelastic traffic even at high traffic loads while exploiting unlicensed band efficiently. Since the channel quality of unlicensed bands is heavily affected by other devices, the unlicensed band is intended to be utilized by elastic traffic most of the time. Once the group of secondary carriers for each UE is decided from  $\mathcal{G}_{S1}$  and/or  $\mathcal{G}_{S2}$ , secondary carriers allocation is carried out by solving the optimization problem P2 in (15).

Similarly with the problem P in (6), the problem P2 is a combinatorial one due to the binary variable  $\alpha_{ii}$ , which makes the problem intractable for large system. One approach to solve it is to relax the binary constraint and allow any value in range of [0,1]. Then, the objective function in (15) is logconvex, since it is a non-negative sum of log-convex functions [11]. Since the constraints are linear and the function is defined on a convex set, the problem can be solved by standard convex

Algorithm 1 The proposed carrier allocation algorithm

**Input:** 1)  $\mathcal{K}$ : the set of users with  $\mathcal{K}_{In}$  and  $\mathcal{K}_{El}$ , 2)  $\mathcal{N}$ : the set of carriers with  $\mathcal{G}_P, \mathcal{G}_{S1}$  and  $\mathcal{G}_{S2}, 3$ )  $\mathcal{W}_N$ : the set of carrier bandwidth, 4)  $P_{\text{max}}$ , 5)  $C_{K \times N}$ : the set of carrier noise ratio, where  $c_{ij} = g_{ij}/\sigma_n^2$ , 6)  $\mathcal{R}^{\min} = \{R_1^{\min}, \cdots, R_K^{\min}\}, 7\}$   $L_{TH}$ : traffic load threshold. **Output:**  $\mathcal{A}_{K \times N} = \{\alpha_{11}, \cdots, \alpha_{KN}\}$ : allocation indicator. 1: Calculate  $\mathcal{R}$  (for  $r_{ij}$ ) and  $\mathcal{D}$  where  $d_{ij} = \log U_i(r_{ij})$ . 2: procedure AllocatePrimaryCarrier() Initialize  $\mathcal{A}_P = \{0\}_{K \times |\mathcal{G}_P|}, flag = N.$ 3: Eliminate the elements of  $\mathcal{D}$  ( $d_{ij} = 0$ ) if  $r_{ij} < R_i^{\min}$ . 4: 5: Make a square matrix  $\mathcal{D}$  with  $v_{\min} = \min(\mathcal{D})$ . procedure  $\operatorname{HUNGARIAN}(\mathcal{D})$ 6: Subtract each entries from  $v_{\max} = \max(\bar{\mathcal{D}})$ . 7. 8: Reduce the matrix by both column/row subtractions. 9: while flag ! = Yes do Find the min. lines number,  $l_{\min}$  to cover zero entries. 10: if  $l_{\min} == K$  flag = Yes. 11:

else Find the min. entry not covered by any line,  $v_{\min}$ . 12: 13: Subtract  $v_{\min}$  from all uncovered elements. 14:

Add  $v_{\min}$  to all elements covered by two lines.

15: Save  $\mathcal{A}_P(1:K) \leftarrow \operatorname{find}(\mathcal{D}(1:K,:) == 0).$ 

Save  $\mathcal{R}_{pc}(1:K) \leftarrow \mathcal{A}_{P} * \mathcal{R}_{\mathcal{G}_{P}}$ . 16:

**return**  $\mathcal{A}_P, \mathcal{R}_{pc} \triangleright$  Selected Primary Carriers and rates 17: 18: **procedure** MONITORTRAFFICLOAD()

Measure current traffic load  $L_C$  and compare with threshold. 19 if  $L_C > L_{TH}$  return HighTraffic; 20: 21: else return LowTraffic; procedure ALLOCATESECONDARYCARRIERS() 22:  $con = [0 \le \alpha_{ij} \le 1; \sum_{i} \alpha_{ij} \le 1];$   $objSig = -\sum_{i} \log(c_i(\frac{\sum_{i} \alpha_{ij} \le 1];}{1 + e^{-a_i(r_i^{p_c} + \sum \alpha_{ij}r_{ij} - b_i)}})$   $obiLog = -\sum_{i} \log(\frac{\log(1 + k_i(r_i^{p_c} + \sum \alpha_{ij}r_{ij}))}{1 + e^{-a_i(r_i^{p_c} + \sum \alpha_{ij}r_{ij})}});$ 23: 24: 2

5: 
$$\operatorname{objLog} = -\sum_{i} \log(\frac{\log(1+k_{i}+2) + ij)}{\log(1+k_{i}+2)}$$

26: **if** mode == HighTraffic

 $\mathcal{A}_{S(In)} = \text{fmincon}(\mathcal{R}, \mathcal{R}_{pc}, \mathcal{K}_{In}, \mathcal{G}_{S1}, \text{con,objSig});$ 27:

28:  $\mathcal{A}_{S(rK)} = \text{fmincon}(\mathcal{R}, \mathcal{R}_{pc}, r\mathcal{K}, r\mathcal{G}_{S}, \text{con,objSig,objLog});$  $\triangleright$  rK: remained UEs, rG<sub>S</sub>: remained carriers in G<sub>S</sub> 29:  $\begin{aligned} \mathcal{A}_{S} &= \mathcal{A}_{S(In)} + \mathcal{A}_{S(rK)}; \\ \textbf{else} \ \mathcal{A}_{S} &= \text{fmincon}(\mathcal{R}, \mathcal{K}, \mathcal{R}^{pc}, \mathcal{G}_{S1}, \text{con,objSig,objLog}); \end{aligned}$ 30: 31:

▷ Selected secondary carriers return  $A_S$ 32:

33: $\mathcal{A} = [\mathcal{A}_P \ \mathcal{A}_S]$	▷ Primary & Secondary	carriers allocation

optimization methods such as the interior-point method. An efficient implementation of the interior-point techniques is available [17] and we use it to compute the optimal solution. The proposed two-step algorithm for allocation of primary and secondary carriers is described in Algorithm 1.

## V. SIMULATION RESULTS

We evaluate the performance of the proposed two-step resource allocation via Matlab simulation and investigate the impacts of two-step allocation and resource grouping on the system performance. The initial configuration parameters are shown in Table I. Four different utility functions are considered as explained in (4) and (5): two for inelastic traffic with  $a_i = \{5,3\}$  and  $b_i = \{10,20\}$ , and two for inelastic traffic with  $k_i = \{0.5, 15\}$  and  $r_i^{\text{max}} = 100$ . While we consider total 20 carriers in three different bands including one unlicensed band, the average channel gain of different bands are assumed different. With setting of maximum number of UE,  $K_{\text{max}} = 8$ , carriers are grouped into three groups. Note that  $K_{\max}$  is equal

TABLE I SIMULATION PARAMETERS

Parameters	Values	
Utility Params. for UEs in $\mathcal{K}_{In}$	$a_i = \{5, 3\}, b_i = \{10, 20\}$	
Utility Params. for UEs in $\mathcal{K}_{El}$	$k_i = \{15, 0.5\}, r_i^{\max} = 100$	
The number of users, $K \& \max$	varying in [1, 8] & $K_{\text{max}} = 8$	
The number of bands	2 in licensed, 1 in unlicensed	
The number of carriers $N$ & per bands	20 & [8, 4, 8] carriers	
Average Channel Gain [dB] per bands	[30, 20, 0]	
Resource grouping	$ \mathcal{G}_P  =  \mathcal{G}_{S2}  = 8,  \mathcal{G}_{S1}  = 4$	

to the number of carriers in  $\mathcal{G}_p$ . For simplicity, the transmit power  $p_{ij}$  and the bandwidth  $W_j$  is assumed to be normalized.

To compare the performance of the proposed algorithm, the one-step allocation algorithm in (6)-(9) (similarly to [11]) and the two-step allocation algorithm in (10)-(18) (similarly to [12]) with only grouping of  $\mathcal{G}_P$  and  $\mathcal{G}_S$  are implemented as the reference. The one-step, two-step and proposed algorithms are referred to as 'One-step', 'Two-step' and 'Proposed', and labelled by 'One', 'Two' and 'Pro', respectively.

Firstly, we focus on the effect of two-step allocation with resource grouping for primary carrier. Fig. 3 shows performance of One-step and Two-step in terms of the average utility value and the average data rate of UEs at high traffic loads. With the bar graph, it is shown *Two-step* produces higher ( $\geq 176\%$ ) average data rates from primary carriers  $(r_i^{pc})$  compared to One-step while the average data rate  $(r_i)$  of One-step is around 5% higher than that of Two-step. By allocating primary carriers from  $\mathcal{G}_P$  (the set of carriers in lower band), *Two-step* could establish more robust connection. From two dashed line in Fig. 3, it is observed that the average utility value of inelastic UEs from Two-step is higher than one of One-step regardless of the number of inelastic/elastic UEs. While inelastic UEs requires less amount of data rates for full QoS satisfaction



Fig. 3. Comparison of One-step and Two-step algorithm's performance: 1) Average utility of inelastic UEs (dashed line) and all UEs (solid line), 2) Average data rate from primary carrier (dark colored bar) and Average rate of UEs (overall bar)



Fig. 4. Comparison of three algorithm's performance: 1) Average utility of inelastic/elastic UEs (solid/dashed line) and 2) Average data rate obtained from unlicensed spectrum use for inelastic/elastic UEs (dark/light colored bar)

(i.e.,  $R_i = 10$  or 20 for  $U_i(R_i) = 1$ ) compared to elastic UEs  $(R_i = 100$  for  $U_i(R_i) = 1$ ), it is analyzed that better data rate from primary carrier allocation results in the benefit to QoS improvement of inelastic traffic. Additionally, while the impact of inelastic traffic's utility on the average utility of all UEs gets bigger with increase of the number of inelastic UEs, the gain of the average utility of all UEs from *Two-step* increases up to 111% against one of *One-step*.

Fig. 4 shows the performance of three algorithms. By comparing the inelastic UEs' data rate obtained from unlicensed band (dark colored bar) of two algorithms, One-step and *Two-step*, it is analyzed that primary carrier allocation from  $\mathcal{G}_P$  results in less use of unlicensed spectrum for inelastic traffic. However, due to the limited resource, as the number of inelastic UEs increases, it is observed that more carriers from unlicensed band become to be utilized for inelastic UEs in both algorithms, resulting in higher data rate from unlicensed spectrum. However, in *Proposed*, the average rate obtained by unlicensed spectrum for elastic traffic is increased (114%)while the rate for inelastic traffic is decreased (14%) at the high traffic loads  $(L_C > L_{TH})$ . In this simulation, the threshold of high traffic is set to 60% of maximum number of users. Thus, when the number of UEs becomes 5, the status of traffic load is regarded as high. Since Proposed prioritizes inelastic traffics to use carriers in licensed spectrum  $\mathcal{G}_{S1}$  for secondary carriers, it leads to steer elastic traffic to unlicensed band at the high load and results in increase of the utility value of inelastic UEs ('A') and decrease of elastic UEs' utility ('B').

While *Proposed* is shown to be superior in terms of inelastic traffic's QoS and robust connection setup, it has the limitation. In *Proposed* with *resource grouping*, carriers in  $\mathcal{G}_P$  are reserved for primary carrier allocation regardless of the traffic load. At the low traffic, some carriers in  $\mathcal{G}_P$  can be reserved without utilization, leading to resource waste. However, in *One-step*, all carriers are utilized fully, thus higher average data rate per UE can be achieved (although the graph is not included here due to a lack of space). In addition, from Fig. 4, it is shown that *Proposed* satisfies inelastic traffic' QoS better than other two algorithms by using resource grouping at high traffic loads. However, the gain in inelastic traffic's QoS support is achieved at the cost of decrease in elastic traffic's QoS. Thus, the resource grouping for different spectrum bands should be carefully decided by considering the QoS requirement of inelastic/elastic traffic, dynamic change of traffic loads and available spectrum.

## VI. CONCLUSIONS

Investigation on resource allocation with traffic steering for service differentiation was provided. For a practical approach, the two-step resource allocation problem was formulated with different utility functions of heterogeneous traffic and the allocation algorithm including resource grouping was proposed. By the simulation results, the proposed approach was shown to be able to support a more robust connection and better QoS to inelastic traffic UEs by steering elastic traffic to unlicensed bands. While fixed resource grouping is considered in this work, future research will consider the flexibility in resource grouping so that the resource allocation can be operated dynamically depending on the varying situation including spectrum availability and the loads of inelastic/elastic traffic.

## ACKNOWLEDGMENT

This work is performed as part of SPEED-5G (https://speed-5g.eu) project supported by the European Unions Seventh Programme for research, technological development and demonstration under grant agreement no. 671705.

# REFERENCES

- [1] Ericsson, "Mobility report on the pulse of the networked society," *White Paper*, Nov. 2015.
- [2] Ofcom, "Measuring mobile braodband performance in the UK," *Research Document*, pp. 1–70, Nov. 2014.
- [3] Nokia Corporation, NTT Docomo Inc., and Nokia Networks, "LTE Carrier Aggregation Enhancement Beyond 5 Carriers," 3GPP Work Item Description (RP-142286), Dec. 2014.
- [4] Nokia Networks, "LTE for Unlicensed Spectrum," 2014. [Online]. Available: info.networks.nokia.com/LTE-for-unlicensed-spectrum-LP.html
- [5] V. Erceg, F. Baboescu, and T. Berham, "Discussion on LTE-WLAN Aggregation (LWA) and LTE WLAN Radio Level Integration with IPsec tunnel (LWIP)," *IEEE 802.11-16/0437r1*, March. 2016.
- [6] H. Lee, S. Vahid, and K. Moessner, "A Survey of Radio Resource Management for Spectrum Aggregation in LTE-Advanced," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 745–760, Second 2014.
- [7] Nokia Networks, "LTE-Advanced Carrier Aggregation Optimization," White Paper, 2014.
- [8] M. Katoozian and H. Yanikomeroglu, "Utility-based adaptive radio resource allocation in OFDM wireless networks with traffic prioritization," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 66–71, Jan. 2009.
- [9] R. Balakrishnan and B. Canberk, "Traffic-Aware QoS Provisioning and Admission Control in OFDMA Hybrid Small Cells," *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 802–810, Feb. 2014.
- [10] L. Tan, Z. Zhu, F. Ge, and N. Xiong, "Utility maximization resource allocation in wireless networks: Methods and algorithms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 7, pp. 1018–1034, July 2015.
- [11] A. Abdelhadi and C. Clancy, "An optimal resource allocation with joint carrier aggregation in 4G-LTE," in *Proc. Int'l Conf. Computing, Netw. Commun. (ICNC)*, California, USA, Feb 2015, pp. 138–142.

- [12] J. Liu, X. Tao, X. Zhou, and Q. Cui, "Utility based resource allocation algorithm with carrier aggregation on unlicensed band," in Proc. Wireless *and Optical Commun. Conf. (WOCC)*, 2015, pp. 180–184. [13] H. Shajaiah, A. Abdelhadi, and T. C. Clancy, "Towards an application-
- aware resource scheduling with carrier aggregation in cellular systems,"
- *IEEE Commun. Lett.*, vol. 20, no. 1, pp. 129–132, Jan. 2016.
  [14] A. Abdel-Hadi and C. Clancy, "A utility proportional fairness approach for resource allocation in 4G-LTE," in *Proc. Int'l Conf. Computing*, 12016. Netw. Commun. (ICNC), Hawaii, USA, Feb 2014, pp. 1034-1040.
- [15] T. M. Nguyen, V. N. Ha, and L. B. Le, "Resource allocation optimization in multi-user multi-cell massive mimo networks considering pilot contamination," IEEE Access, vol. 3, pp. 1272-1287, Aug. 2015.
- [16] D. Jungnickel, Graphs, Networks and Algorithms. Graphs, Networks and Algorithms, 2008.
- [17] V. Venkatasubramanian, "Opportunistic resource allocation and relaying methods for quality of service in the downlink of future cellular wireless networks," Ph.D. Thesis, Victoria Univ., Australia, pp. 1-139, Feb. 2011.