A Transportation Problem based Resource Allocation Scheme for an LTE-Advanced System with Carrier Aggregation

M. Saad ElBamby and Khaled M. F. Elsayed
Department of Electronics and Electrical Communications Engineering, Cairo University
Giza, Egypt 12613
mselbamby@ieee.org, khaled@ieee.org

Abstract—We formulate the downlink cross-carrier resource allocation problem in LTE-Advanced carrier aggregation (CA) systems as an unbalanced transportation problem (TP). The component carriers (CC’s), or their associated sub-bands, represent the supply points whereas the users requesting traffic represent the demand points. The cost of shipping from a supply point to a demand point is set as a function of the proportional fairness (PF) metric. To balance the problem, an artificial supply point is added to represent the shortage in satisfying the demand in case of an overloaded system. Backward compatibility with LTE Release 8 (Rel-8) is guaranteed by adjusting the cost of shipping to Rel-8 users to restrict them to operate on a single CC. The results show that using Vogel approximation method (VAM), which is a sub-optimal efficient method, provides near-optimum results in terms of throughput and fairness and is much faster than the simplex method for optimal solution. The proposed TP-based scheduler is shown to achieve considerable performance improvement over a PF scheduler that handles multiple CC’s in the presence of both Rel-8 and LTE-A users.

Keywords- LTE-Advanced; carrier aggregation; resource allocation; transportation problem; Vogel Approximation

I. INTRODUCTION

In order to meet the increasing demand for high data rate requirements of IMT-Advanced, as defined by the International Telecommunication Union (ITU), the Third Generation Partnership Project (3GPP) has introduced carrier aggregation (CA) as one of the Long Term Evolution-Advanced (LTE-A) features that could serve this goal [1]. With carrier aggregation, multiple carrier chunks, namely component carriers (CC’s), are aggregated so that a user can be scheduled simultaneously on multiple CC’s, thus higher data rates can be achieved.

An important problem that arises in scheduling resources in CA is backward compatibility. Backward compatibility should be maintained such that a legacy Rel-8 user equipment (UE), which naturally only supports a single CC, should still be able to coexist with Rel-10 UE (LTE-A UE), which can be scheduled on the entire aggregated CC’s. Furthermore, the downlink radio resource management (RRM) of LTE-A should have some new functionalities and improvements to support carrier aggregation. For example, load balancing between Rel-8 users and LTE-A users should be considered in a way to enhance both throughput and fairness under different load conditions among the existing CC’s. In addition channel-aware packet scheduling (PS) is required to exploit the channel diversity not only between the users across a given CC but also for a given user across multiple CC’s to provide frequency domain scheduling gain. This adds a new dimension to inspect in the scheduling and resource allocation which is the CC dimension in addition to the time and frequency dimension typically associated with OFDMA-based wireless access systems.

The performance gain of CA over independent carriers is evaluated in [2]. Simulation results for different traffic models show that CA can enhance the throughput, fairness and latency. In [3], different load balancing methods across CC’s are investigated, it shows that using Round Robin (RR) load balancing is better than mobile hashing (MH) to balance Rel-8 users over CC’s whereas LTE-A users are scheduled across the entire CC’s. It also shows that the proposed cross CC packet scheduling (PS) algorithm provides better performance than independent scheduling per CC. An uplink resource allocation framework for CA-based LTE-A systems is presented in [4]. Performance evaluations results show that CA with specific CC selection algorithm improves performance of the average and cell center users’ throughput, particularly in low load traffic conditions. On the other hand, the cell edge user throughput of LTE-A UE’s maintains the same level of performance as Rel-8 UE’s. In [5], the problem of resource allocation in the IEEE 802.16 band-AMC Mode is modeled as an unbalanced transportation problem and solved so that the total power is minimized with proportional rate constraints between the users.

The main contribution of this paper is the formulation of the CA Resource Allocation as a transportation problem and its efficient solution to obtain the resource block allocations for both LTE-A users and Rel-8 users. The cost of shipment in the transportation problem is used to model an allocation that considers the proportional fairness metric of the users in a global way. The formulated transportation problem is converted to a linear programming (LP) problem and solved using both the exact simplex solution and the Vogel Approximation Method (VAM). The VAM solves the problem efficiently and gives near-optimum results as compared with the simplex method but in a much shorter time [6].

The rest of this paper is organized as follows. In section 2 we give an overview of carrier aggregation concept and the system model. In section 3 we briefly discuss the transportation problem and introduce the resource allocation problem formulation as a transportation problem and present the solution methods. Simulation parameters and performance evaluation are presented in section 4. Finally, section 5 concludes the paper.

1 This work is part of the 4G++ project supported by the National Telecom Regulatory Authority of Egypt.
II. PRELIMINARIES

A. Overview of Carrier Aggregation Concept

Carrier aggregation (CA) is one of the LTE-A Release 10 (Rel-10) main features. LTE-A is designed to meet the peak data rates required by IMT-Advanced: 1 Gb/s for the downlink and 500 Mb/s for the uplink [7]. This requires users’ access to a total bandwidth up to 100 MHz. Since the maximum supported bandwidth in LTE is 20 MHz, bandwidth is expanded through aggregating up to five CC’s. These CC’s may be intra-band (located at the same band) or inter-band (located at different bands). The intra-band CC’s may be contiguous or non-contiguous, depending on the spectrum availability. These configurations are depicted in Fig. 1. Moreover, CA is designed to be backward compatible, which means that both LTE Rel-8 and LTE-A user equipment (UE) can be scheduled on the same CC deployed by the Rel-10 eNodeB (eNB).

![Figure 1. Carrier aggregation spectrum configurations: a) intra-band contiguous; b) intra-band non-contiguous; c) inter-band non-contiguous.](image)

B. System Model

We consider a single cell LTE-A system, where we focus on the resource allocation for the downlink direction. The number of CC’s in the system is equal to \( L \). Each CC has \( f \) physical resource blocks (PRB), where the PRB is the smallest allocation unit for the scheduler. We assume a uniform transmitted power of \( P_f / V \) on each PRB, where \( P_f \) is the transmitting power of each CC. CC’s operate in Frequency-Division Duplex (FDD) mode, meaning that downlink and uplink transmission take place in different CC’s. Each UE in the cell is able to connect to a set of CC’s \( F \):

\[
F = \{f_1, f_2, \ldots, f_l\} \quad (1)
\]

\[
l = \begin{cases} 1 & \text{for Rel-8 UE,} \\ L & \text{for LTE-A UE.} \end{cases} \quad (2)
\]

As Rel-8 UE’s are capable of connecting to single CC only, a load balancing scheme is needed to distribute their load across CC’s. Round Robin (RR) is used herein to balance loads among CC’s. It is shown in [3] that RR load balancing provides better performance than the mobile hashing (MH) balancing scheme.

The Exponential Effective SINR Mapping (EESM) model [8] is used to combine the SINR on each subcarrier to obtain the PRBs’ effective SINR, which is:

\[
\text{SINR}_{\text{eff}} = -\ln \left( \frac{1}{N_{SC}} \sum_{i=1}^{N_{SC}} e^{-\text{SINR}_i} \right)
\]

where SINR, is the SINR of subcarrier \( i \) and \( N_{SC} \) is the number of subcarriers on each resource block.

The system is comprised of an enhanced NodeB (eNB) and UUE’s (including Rel-8 users and LTE-A users) randomly dropped in the layout. Each user requests a specified number of bytes (generated in accordance with an arbitrary traffic model) each transmission time interval (TTI) of 1 millisecond, which is the subframe size of LTE. The resource allocation is dynamically performed on a subframe basis. Unsatisfied demand in a certain subframe is accumulated to the added demand of the next subframe. To enable channel-aware resource allocation, exploiting the users’ Channel Quality Indicator (CQI) reports is performed. In a CA system, the user may be allocated resource blocks across different CC’s, therefore to support this functionality the LTE-Advanced specifications provide certain provisions for CQI reporting in CA systems. Since per-PRB CQI reporting over all CC’s would place a large signaling overhead for the UE uplink channel, the LTE-Advanced provides a method to report CQI’s efficiently but at a reduced granularity. Our proposed method explicitly emulates the CQI reporting method in CA systems and exploits it in the proposed solution, and we briefly explain the method in the following subsection.

C. CQI Reporting Mode

The channel aware PS requires feedback of CQI across the CC’s. This gives rise to a high uplink overhead. Reduction of this feedback overhead is done by reporting the CQI in terms of sub-bands (SBs). Each SB contains a number of PRB’s and the UE reports one CQI for all of them. This method of reporting is exploited in the proposed solution as will be discussed later. The mechanism in which UE select and report the CQI of these SBs is standardized by 3GPP [9] and is performed as follows:

1- The UE selects a set of \( M \) preferred SBs each of size \( k \) PRBs from each CC (where \( k \) and \( M \) Depend on the CC bandwidth).

2- The UE reports one CQI value representing the average channel quality of these \( M \) selected SBs.

3- The UE also report a wideband CQI value for each CC, which is calculated assuming transmission on the all SBs.

The remaining SBs are then not associated with any CQI values. A simple method is developed in [10] to construct values for these unclaimed SBs; it assumes that the unclaimed CQI’s have a value, which is below the wideband CQI by an offset \( \Delta \). This offset is calculated such that the average CQI of the selected and unclaimed CQI groups equals the wideband CQI.

![Figure 2. Example of the difference between the actual and constructed CQI.](image)
After this modification, only two values are assigned to the whole carrier, the average best \( M \) value for the best \( M \) SBs, and the wideband value minus the offset value \( D \) for the remaining SBs. Fig. 2 gives an example of the difference between the full (actual) CQI and the constructed one for a 20 MHz CC that contains 100 PRB’s divided into 25 SBs. Best 6 SBs have one value which is the average of their values. The remaining SBs have one value which is calculated according to the above mentioned method.

III. MODELING THE CROSS-CC RESOURCE ALLOCATION AS A TRANSPORTATION PROBLEM

We formulate the cross-CC resource allocation problem as an unbalanced transportation problem where the users’ demand may exceed the capacity of the sources. This essentially models a highly loaded system. The goal of the optimization problem is to efficiently allocate the resources to the users in order to maximize the overall system throughput.

A. Transportation Problem Basics

A transportation problem minimizes the cost of shipping units from supply points to demand points so that the needs of each demand point is satisfied and each supply point serves within its capacity. The main parameters of the transportation problem are listed in Table I. A balanced transportation problem can be efficiently solved as linear programming (LP) optimization problem. The problem may be unbalanced for one of the following two reasons. The first is when the total number of units contained in the supply points is higher than the total number of required units at the demand points. In this case, the problem is balanced by adding an artificial dummy demand point with its demand value as the excess available units in the supply. Shipping to the dummy demand point from any supply will be of zero cost as actually this shipping will not occur.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S )</td>
<td>Number of supply points.</td>
</tr>
<tr>
<td>( D )</td>
<td>Number of demand points.</td>
</tr>
<tr>
<td>( s_i )</td>
<td>Number of available units at supply point ( i ), ( i \in {1, 2, \ldots, S} ).</td>
</tr>
<tr>
<td>( d_j )</td>
<td>Number of needed units at demand point ( j ), ( j \in {1, 2, \ldots, D} ).</td>
</tr>
<tr>
<td>( c_{i,j} )</td>
<td>Cost of transferring one unit from supply point ( i ) to demand point ( j ).</td>
</tr>
<tr>
<td>( x_{i,j} )</td>
<td>Number of assigned units from supply point ( i ) to demand point ( j ).</td>
</tr>
</tbody>
</table>

The second reason that makes the transportation problem unbalanced, is when the total number of units in the supply points is less than those needed at the demand points. The problem is balanced here by adding an artificial dummy supply point with its supply value as the shortage of the required units. The cost of shipping from the dummy supply point to each demand point should reflect the waste in profits that occurs when this demand point gets one unit less than the needed number of units. After solving the problem, units that are shipped from this supply point will represent the deficit in satisfying the demand. In both cases, this addition converts the problem to a balanced transportation problem which is formulated as follows:

\[
\begin{align*}
\min & \quad \sum_{i=1}^{S} \sum_{j=1}^{D} x_{i,j}c_{i,j} \\
\text{subject to} & \quad \sum_{i=1}^{S} s_i = \sum_{j=1}^{D} d_j \\
& \quad \sum_{i=1}^{S} x_{i,j} = d_j, \quad \forall j \\
& \quad \sum_{j=1}^{D} x_{i,j} = s_i, \quad \forall i \\
& \quad x_{i,j} \geq 0, \quad \forall i, \forall j
\end{align*}
\]

where the above parameters are defined in Table I.

B. Mapping of the Resource Allocation Problem as a Transportation Problem

The mapping of supply points herein will depend on the CQI reporting frequency resolution. If the UE is reporting a wideband CQI, the whole CC will be considered as one supply point. If the CQI feedback is reported on a UE-selected SB basis (described in the previous section), then the whole SB will have one CQI value and hence each SB inside a CC will represent a separate supply point containing a number of PRB’s. This is because the cost of shipping a unit from a supply point to a demand point depends on the achieved rate (as will be discussed later), so it is a function of the user’s CQI.

The downlink users’ queues at the eNB with pending traffic demand at a certain subframe will represent the demand points. The demand at each point is represented by the number of PRB’s required to satisfy the user’s pending traffic. The user’s pending traffic demand is expressed in terms of bytes, and therefore needs to be mapped to a number of PRB’s using the user’s average channel state. This is done by calculating the average CQI of the user over all carriers. Then, the demand bits are divided by the corresponding transport block size (TBS) of one PRB at this value of the CQI. This gives the demand in terms of PRB’s. The TBS can be obtained by mapping the CQI index to the corresponding Modulation and Coding Scheme (MCS) level and then, the TBS tables of [9] are used to determine the block size in bits. The mapping of CA problem to a transportation problem is conceptually explained in Fig. 3 which illustrates the demand points being the user queues at the eNB’s requesting a number of PRB’s and the supply points being the CC’s each comprised of a numbers of SB’s each supplying a number of PRB’s to satisfy the demand.

We further define the cost of assigning a resource to a user (which is to minimize the total cost) to relate to the proportional fairness (PF) metric as follows:

\[
C_{i,j} = \frac{R_{i,j}}{R_{j}}
\]

where \( C_{i,j} \) is the cost of assigning one unit from CC/SB \( i \) to user \( j \), \( R_{i,j} \) is the instantaneous rate of user \( j \) in the CC/SB \( i \) and \( R_j = \sum_i R_{i,j} \) is the historical total average rate for user \( j \) in the previous allocations over all CC’s, \( R_{i,j} \) is the average rate of user \( j \) over CC/SB \( i \). This cost metric is the opposite of the proportional
fairness metric so that the higher the metric, the lower the assignment cost.

\[
\begin{array}{c|c|c}
\text{CC}_1 & \text{CC}_2 & \text{Supply Points} \\
\hline
\text{SB}_1 & \text{SB}_2 & \text{PRB} \\
\hline
\end{array}
\]

Figure 3. Mapping of CA problem to a transportation problem

Since the eNB is assumed to be transmitting with full power distributed uniformly across CC’s, the achieved rate is typically a function of the user's CQI only.

A dummy supply will be added to represent the shortage in satisfying the demand if it is greater than the available resources. The cost of assignment from this supply is selected higher than the range of the available costs \( C_{ij} \) to increase the opportunity for users to be allocated to a real supply point.

We assume some users in the system are Rel-8 users, who are capable of connecting to only one CC at a time. These users will be assigned to a certain CC using Round Robin (RR) load balancing method. The cost of shipping to these users from other CC’s is set to a value \( \alpha \). This ensures that those users will be allocated resources only from the selected CC. This high value of \( \alpha \) will cause some cost values to be positive whereas the typical values in the problem are negative. To avoid this, we initially scale the whole costs to higher values, so that the addition of \( \alpha \) will keep all of them negative. The suitable absolute value of \( \alpha \) should be approximately 100 times the average of the available costs \( C_{ij} \), increasing it above this value has no effect on the solution. It is worth mentioning that we are herein interested in obtaining the solution with minimum total cost rather than the value of the total cost itself, thus the exact value of \( \alpha \) is not of a particular importance in the problem as long as it prevents assignment from that CC’s.

C. Solution Methods

The problem is transformed into a linear programming (LP) problem and then solved using the simplex method. The simplex method can be considered a substantial generalization of standard Gauss-Jordan (GJ) elimination in linear algebra. It starts by the pivot operation which is similar to the pivot used in solving systems of linear equations, but restricts the choice of pivot by the use of two rules; the entrance rule that determines the pivot column and the exit rule that determines the pivot row. These rules are explained in [11]. By following these two rules, the optimal solution of the transportation problem is obtained after a finite number of pivots.

However, there are some other methods which can efficiently solve the problem if it is balanced, i.e., total number of available supply units equals the total demand. The Vogel’s Approximation Method (VAM) can be used to obtain a feasible solution that is close to the optimal one. Some experimental research showed that on the average, VAM yielded the optimal solution about 20% of the time and it yielded a solution with around 0.5% loss of optimality about 80% of the time [6]. The VAM method is performed as follows:

1. Construct an assignment matrix whose dimension is \( S \) by \( D \), begin with all cells unallocated.
2. Compute for each row and each column the difference between the lowest and next lowest cost cell in the row or column.
3. Among those rows and columns, select the one with maximum difference.
4. Allocate as much as possible to the \( x_{ij} \) with the lowest cost cell in the selected row or column. Decrease the corresponding supply and demand. Drop the row and/or column whose supply or demand is zero.
5. Make any allocations where only one unallocated cell remains in a row or column. After reducing the corresponding supply and demands and dropping the row and/or column, repeat Step 4 as necessary.
6. Stop if no rows and columns remain. Otherwise return to Step 1 with the reduced problem.

An important property of the TP is the integer solutions property. It states that when all units in the source points and demand points have an integer value, all basic feasible solutions for \( x_{ij} \), including the optimal solution also have integer values.

IV. PERFORMANCE EVALUATION

The proposed solution is simulated using MATLAB. The channel model on the links is the WINNER II C2 model as specified in [12]. The main parameters used in the simulation are summarized in Table II. For each experiment, we generate 50 different channel and user distributions and take the average of the 20 experiments as the reported valued. The traffic demand for the users is generated in terms of a constant bit rate (CBR) application. The rate is changed in order to change the system offered load. We assume there are no packet transmission errors, hence no HARQ operation is applied. The performance of the proposed scheme is compared against an adapted version of Proportional Fair (PF) scheduler to cope with multiple CC’s in a CA setup. This scheduler allocates a resource in a certain CC to the user with the highest PF metric only if the user is connected to this CC. The main metrics are the average user throughput and fairness index expressed by the Jain’s Fairness Index (FI) [13], which is formulated as follows:

\[
FI = \frac{(\sum_i R_i)^2}{U \sum_i R_i^2}
\]

where \( U \) is the number of UE’s in the system. This index measures the degree of fairness in the allocated rates between users and it has value that falls between zero and one with more fairness achieved as we are close to one.
In the first experiment, we increase the offered load and measure the average UE throughput and fairness comparing the performance of the proposed scheme with the adapted PF scheduler. The results show that the average UE throughput increases with increasing the offered loads until certain value then it tends to saturate as the cell capacity is therefore limited by the number of available PRB’s, as shown in Fig. 4 and Fig. 5 for a percentage of LTE-A users of 100% and 50%, respectively. As compared to the PF scheduler, the proposed scheme can achieve about 4% throughput gain at moderate loads if all users are LTE-A and 3% if half of them are LTE-A. At low loads, the users demand is lower than the available resources so they get their entire traffic, and hence no difference between the two schemes is expected. Also there is no noticeable difference in their performance at high loads, as the system capacity is approached. Thus a certain bound cannot be exceeded and the two schemes typically achieve the same throughput. Fig. 6 depicts the average Rel-8 UE throughput and average LTE-A UE throughput versus the offered load for the two schemes. The throughput of LTE-A users is greater as they are scheduled on the entire bandwidth. It is shown that our scheduler achieves better throughput than the PF scheduler in moderate load conditions, especially for Rel-8 users who achieves up to 5% throughput gain. Being served well in the presence of LTE-A users, Rel-8 coexistence in the system doesn’t degrade the overall performance and hence the fairness of the system remains high.

### TABLE II. SIMULATION PARAMETERS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value/description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layout scenario</td>
<td>Typical urban macro-cell [12]</td>
</tr>
<tr>
<td>Carrier aggregation pattern</td>
<td>4 × 10 MHz non-contiguous CC’s at CA band 1</td>
</tr>
<tr>
<td>Number of PRB’s per CC</td>
<td>50 PRB, each contains 12 subcarriers</td>
</tr>
<tr>
<td>Duplex mode</td>
<td>FDD</td>
</tr>
<tr>
<td>Sub-frame duration</td>
<td>1 ms</td>
</tr>
<tr>
<td>FFT size</td>
<td>1024</td>
</tr>
<tr>
<td>Sampling frequency</td>
<td>15.36 MHz</td>
</tr>
<tr>
<td>Antenna configuration</td>
<td>1*1</td>
</tr>
<tr>
<td>Modulation schemes</td>
<td>QPSK, 16-QAM, 64-QAM</td>
</tr>
<tr>
<td>Traffic model</td>
<td>CBR</td>
</tr>
<tr>
<td>Number of UE’s</td>
<td>20</td>
</tr>
<tr>
<td>UE distribution</td>
<td>Uniform over cell area</td>
</tr>
<tr>
<td>CQI reporting</td>
<td>Best-M sub-band reporting</td>
</tr>
<tr>
<td>Simulation time</td>
<td>1000 subframes repeated 50 times</td>
</tr>
</tbody>
</table>

The proposed scheme also achieves better fairness at all load conditions; this is because in the PF scheduler case, the best user (i.e., the user with the highest metric) is scheduled each time in a greedy approach whereas the TP-based scheduler applies the proportional fairness metric in a more global way in which the metrics of all users play a role in obtaining the solution. Therefore, the TP-based scheduler achieves better fairness in the long-term. In both cases, the fairness decreases with increasing load but remains in a favorable range (above 0.9).

In the second experiment, we assess the results of the VAM method as compared with the optimal simplex method. The difference between the two methods in the realizable average UE throughput is shown in Fig. 7. The figure depicts the performance in three different scenarios: when all users are LTE-A, half of the users are LTE-A and the rest being Rel-8 users, and all users are Rel-8. VAM performance is slightly degraded with around 0.5% loss due to the sub-optimality of the scheme. However, the time consumed in the simplex method is nearly 30 times larger than that of the VAM method.
In the third experiment, we assess the advantage of increasing the percentage of LTE-A capable users. Fig. 8 illustrates the throughput performance versus the percentage of LTE-A users at moderate load condition (100 Mbps). It is shown that the system performs better as the percentage of LTE-A users increases; the more the number of LTE-A users the better performance is achieved as more users become able to access more than one CC and this can take advantage of the best channel states for the PRB’s available in the different CC’s, hence, the advantage of CA can be exploited. The gain of the proposed method compared with the PF scheduler varies from 3-4% when VAM is used, VAM loss due to sub-optimality is always less than 1%.

In brief, our proposed scheme obtains considerable performance gain than the traditional PF scheduler in terms of throughput and fairness, either by using the VAM method or the simplex method.

![Graph](image-url)

**Figure 7.** VAM versus simplex method in obtaining average UE throughput

![Graph](image-url)

**Figure 8.** Average UE throughput versus percentage of LTE-A users for different schedulers

V. CONCLUSIONS

In this paper, the problem of downlink cross-carrier resource allocation in an LTE-A setup with carrier-aggregation is formulated as a transportation problem (TP). Simulation results demonstrate that the proposed scheme can achieve good throughput in the presence of both Rel-8 users and LTE-A users. Throughput increases as the number of LTE-A users increase. It is also shown that the proposed TP scheduler with cost metric that is the opposite of the PF metric achieves better performance than the Proportional Fair (PF) scheduler. Both the simplex (optimum) method and Vogel Approximation Method (VAM) solve the TP efficiently. VAM is a sub-optimal but efficient solution that solves TP in much smaller time than the simplex method. Our future work will consider the assessment of loss of performance due to the granularity of the CQI reporting for the different CC’s. We will also consider the multi-cell resource allocation problem in existence of inter-cell interference coordination schemes such as fractional-frequency reuse and soft-frequency reuse.

REFERENCES