Joint Carrier Aggregation and Packet Scheduling in LTE-Advanced Networks

Xiaolin Cheng†, Gagan Gupta, Prasant Mohapatra†
†Department of Computer Science, University of California at Davis, CA 95616
xlcheng@ucdavis.edu, gagan.gupta@iitdalumni.com, pmohapatra@ucdavis.edu

Abstract*—LTE is the next generation of all-IP mobile communication system designed and developed by 3GPP. It offers unprecedented data transmission speed and low latency to support a variety of applications and services. However, compared to wireline networks, efficient QoS provisioning for diversified applications in wireless access networks such as LTE is challenging due to unreliable and resource-constrained radio interface. In this paper, we investigate an important problem of downlink resource allocation in recently enhanced LTE-Advanced systems where a newly added feature carrier aggregation provides more flexibility in radio resource management in addition to the existing resource block level packet scheduling. The resource allocation problem can be formulated as a complex combinatorial problem with multiple constraints and is solved every time slot. We decompose this highly complex optimization problem and construct a two-tier resource allocation framework which incorporates dynamic component carrier assignment and backlog based scheduling schemes with intelligent link adaptation. An efficient algorithm is developed to dynamically allocate component carriers to users to achieve load balancing. We also present novel backlog based scheduling policies and weighted-CQI based link adaptation scheme to obtain significantly better throughput and delay fairness. Performance of the proposed schemes is evaluated against the static round-robin component carrier assignment, the well-known proportional fairness scheduling rule and the existing link adaptation scheme. Extensive simulation results demonstrate that our schemes offer both better throughput and delay performance as well as user fairness.

I. Introduction

The major attraction of the next generation (4G) wireless systems such as Third Generation Partnership Project (3GPP) Long Term Evolution (LTE) is the availability of high data rate and low end-to-end latency. LTE has been designed and developed by 3GPP [1] as an all-IP network with simplified radio access and core network architecture. On air interface, LTE features OFMDA and MIMO technologies, which significantly improve the data transmission speed along with lower latency. LTE can provide a peak rate of 300Mbps, a significant increase in spectrum efficiency compared to previous cellular systems. In the core network, an all-IP infrastructure with support of IMS (IP Multimedia System) facilitates the addition of a variety of services and applications, including data, voice, video and location. However, it is challenging to provide the required level of quality of service (QoS) and maintain the designed system performance due to the constraints of limited radio resources, unreliable radio propagation channel, and high user demands.

Resource allocation and scheduling have been important aspects of QoS support in wireless networks, and have attracted an increasing attention from the research community. Wireless scheduling has two particular characteristics which distinguish it from conventional wireline scheduling [2]: (1) The radio channel is inherently unreliable and error-prone. Also, errors can be bursty in nature which results into unsuccessful packets transmissions during the burst. The implication is that a good scheduling algorithm needs to be adaptive to changing channel quality. (2) Channel state varies randomly in time on both slow and fast time scale. In fast channel variations (due to fast fading), states of different channels can asynchronously switch from “good” to “bad” within a few milliseconds and vice-versa. A good scheduling algorithm should exploit this condition by giving higher preference to a user whose channel is currently better. Slow channel variations mean that the average channel state depends on user’s location and other factors. Therefore some users may have to demand more radio resources than others even if their data rate requirements are the same.

In OFDMA based LTE networks, data is transmitted on a large number of parallel, narrow-band sub-carriers. During each time slot, multiple users can be allocated a set of sub-carriers (termed resource blocks in LTE) to enable concurrent transmissions. Thus efficient scheduling of radio resources is crucial to achieve high network performance.

In a resource allocation period, each resource block is associated with different channel quality in terms of SINR (signal-to-interference and noise ratio), which is sent back on a feedback channel known as CQI (Channel Quality Indicator) from mobile terminals. Based on the CQI, appropriate modulation and coding schemes can be applied to improve the transmission reliability and rate (known as AMC or adaptive modulation and coding). In this way, the available channel rate or the size of transport block on resource blocks can be determined.

Carrier aggregation introduced in LTE-Advanced (Release 10) [1] is designed to expand the transmission bandwidth by combining multiple carriers in the same or different bands. With this new feature, resource allocation module will have more flexibility when scheduling radio resources for users. However, different assignments of carriers will have significant impact on network performance. For instance, distributing users across available carriers efficiently is important to achieve improved network load balancing and higher utilization. Carriers on different bands may have diversified channel characteristics such as path loss patterns and fading variations, so carrier allocation will be a key part of RF planning.

In this paper, we study the radio resource management (RRM) problem in LTE-Advanced systems by jointly considering both carrier aggregation and packet scheduling. The joint resource allocation is formulated as a complex combinatorial problem. We propose efficient algorithms to address the issues presented in the problem. Specifically, we propose a dynamic component carrier allocation algorithm which utilizes aggregated head-of-line delay of users associated with a carrier as the metric. The algorithm focuses on the user backlog accumulated on a carrier to achieve load balancing as well as better throughput and delay performance. For the resource block level allocation, we propose the backlog based scheduling schemes which are channel-aware and targeted on reducing user backlog (queue length or head-of-line delay). The backlog based schemes have been studied from a mostly theoretical point of view for restricted cases and unrealistic assumptions on arrival processes and channel characteristics. Our goal is to study and propose efficient backlog based scheduling algorithms with realistic settings and scenarios in LTE networks. The proposed resource allocation schemes are enhanced using an intelligent link adaptation mechanism which takes the weighted CQI of all allocated resource blocks for a user as input and determines the MCS (Modulation and Coding Scheme) mode. The weighted-CQI based link adaptation helps in achieving correct balance between transmission rate and error rate.

The rest of the paper is organized as follows. In Section II, we describe the detailed system model. In Section III, we formulate the combinatorial optimization problem for downlink resource allocation. Section IV proposes our dynamic carrier allocation and backlog based scheduling schemes and intelligent link adaptation mechanism. Performance of the proposed schemes is evaluated in Section V. Related work is discussed in Section VI. Section VII concludes the paper.

II. SYSTEM MODEL

A. LTE Downlink Model

The LTE downlink transmission scheme [3] provides scalable bandwidth from 1.4MHz to 20MHz. It is based on conventional OFDM using a cyclic prefix, with a sub-carrier spacing $\triangle f$=15kHz and a cyclic-prefix (CP) duration $T_{CP} \approx 4.7/16.7 \mu s$ (short/long CP). A radio frame is $10ms$ in duration, and divided into 10 equally sized sub-frames (each being $1ms$ long). Each sub-frame is called an *Transmission Time Interval* (TTI), and further divided into 2 slots ($0.5ms$ each). Each slot has 6 or 7 OFDM symbols. The basic transmission parameters are specified in more detail in Table I. The sub-carrier spacing is constant regardless of the transmission bandwidth. Multiple sub-frames can be concatenated into longer TTI to improve support for lower data rates and QoS optimization.

The transmitted downlink signal consists of N_{BW} sub-carriers for a duration of T_{slot}. It can be represented by a resource grid as depicted in Figure 1. Each box within the

TABLE I
LTE DOWNLINK PHYSICAL PARAMETERS

Bandwidth (MHz)	1.4	3	5	10	15	20
Radio frame duration	10 ms					
Sub-carrier spacing	15 KHz					
RB bandwidth	180 KHz					
Number of RBs	6	15	25	50	75	100

grid represents a single sub-carrier for one symbol period and is referred to as a *resource element*. A *physical resource block* (PRB or RB) is defined as one slot in the time domain ($0.5ms$) and 12 consecutive sub-carriers (180KHz) in the frequency domain. A resource block is the smallest element of resource allocation assigned by the base station scheduler. The resource block model is demonstrated in Figure 1.

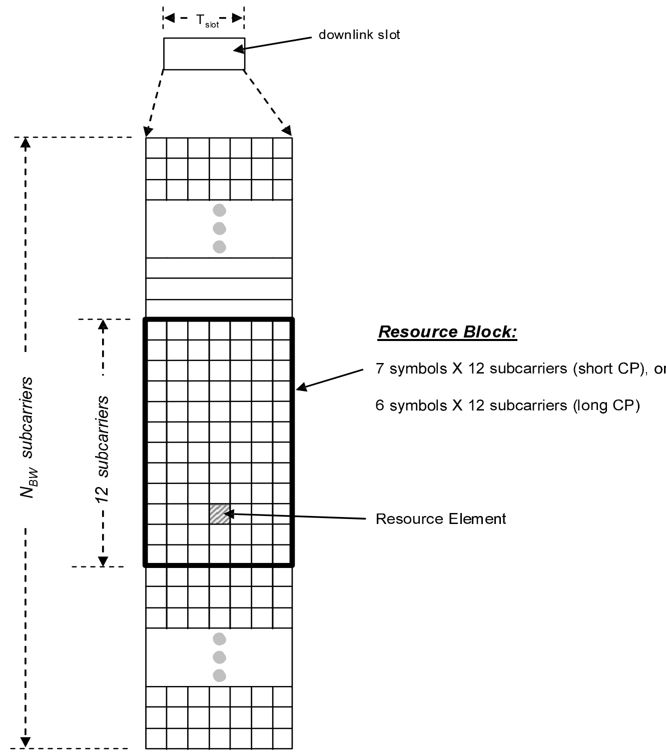

Fig. 1. LTE Downlink Resource Grid

B. Carrier Aggregation

It is unlikely that spectral efficiency of LTE can be improved much beyond its current performance limit to meet the requirement of 1Gbps set by IMT-Advanced. Therefore the only way to achieve significantly higher data rates is to increase the channel bandwidth. Carrier aggregation (CA) [1] is one of the most notable key features of LTE-Advanced which has been standardized in 3GPP as part of LTE Release 10. It allows scalable expansion of effective bandwidth delivered to a user terminal through concurrent utilization of radio resources across multiple carriers. These carriers may be of different bandwidths, and may be in the same or different bands to provide maximum flexibility in utilizing the scarce radio spectrum available to operators. In the carrier

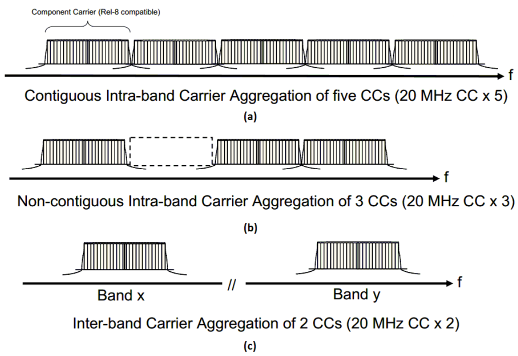

Fig. 2. Types of Carrier Aggregation

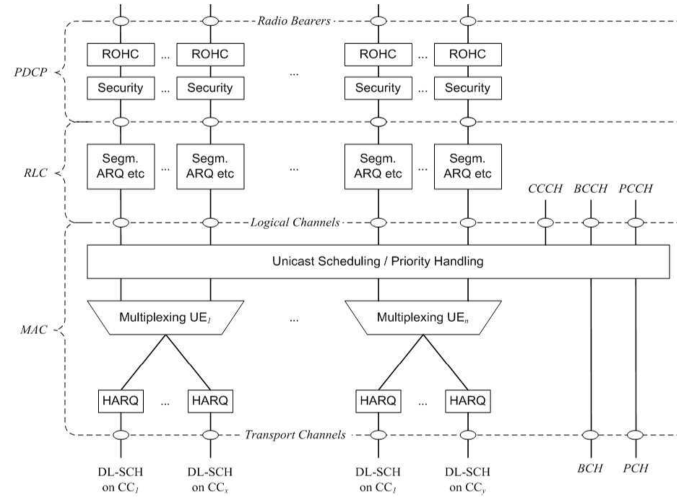

Fig. 3. Downlink Protocol Stack with Carrier Aggregation

aggregation, up to 5 component carriers can be aggregated. The term "component carrier" (CC) in this context refers to any of the bandwidths defined in LTE. The creation of wider bandwidths has three options: intra-band contiguous aggregation, intra-band non-contiguous aggregation, or inter-band non-contiguous aggregation.

Intra-band contiguous CA provides a contiguous bandwidth wider than 20MHz (Figure 2(a)). It may be a unlikely scenario given current frequency allocations of operators, but can be common when new spectrum bands are allocated in the future. Intra-band non-contiguous CA combines multiple CCs belonging to the same band but in a non-contiguous manner (Figure 2(b)). This option can be used in scenarios where spectrum allocation is non-contiguous within a single band. Inter-band non-contiguous CA (Figure 2(c)) is the most general mode where multiple CCs belong to different bands. It can potentially improve mobility robustness by exploiting different radio propagation characteristics of different bands, but may require additional complexity in the RF front-end.

Data aggregation schemes across component carriers can be achieved at the MAC layer where the transmission blocks (TBs) from different component carriers can be built in a way that is transparent to the upper layers (RLC and PDCP) (Figure 3) [1]. Each component carrier has its own transmission configuration parameters as well as an independent HARQ entity in order to support more flexible and efficient data transmissions and guaranteed backward compatibility at the expense of complex control signaling.

C. Channel Models

1) Friis Path Loss Model: The major attenuation of signal strength on a wireless link comes from the large-scale path which can be modeled using the Friis transmission equation [4][5]. The ratio of the power available at the input of the receiving antenna P_r to the output power of the transmitting antenna P_t,

$$P_r = P_t + G_t + G_r + 20 \cdot \log_{10}\left(\frac{\lambda}{4\pi R}\right)$$

where G_t and G_r are the antenna gains of the transmitting and receiving antennas respectively, λ is the wavelength, and R is the distance between the antennas. It is clear that frequency plays an important role in the path loss model: signals in a band with higher frequency attenuate faster, so the band will have a smaller coverage. This fact is well known but its impact on component carrier assignment is significant.

2) Rayleigh Fading Model: Fading adds fast variations upon path loss. Rayleigh fading [6] is a statistical model for modeling the effect of multi-path propagation on a radio signal. Rayleigh fading model assumes that the magnitude of a signal that has passed through such a transmission channel will vary or fade randomly according to a Rayleigh distribution. The MATLAB functions `rayleighchan` and `pwelch` are used to generate the fading traces for simulation [5] [7].

3) PHY Error Model: The PHY error model of NS3 simulator used in the paper is designed according to the standard link-to-system mapping (LSM) techniques [5] [7]. It is aligned with the standard system simulation methodology of OFDMA radio transmission technology. It is based on the mapping of link layer performance obtained using link level simulators to system simulators. In particular, the link layer simulator is used for generating the performance of a single link from a PHY layer perspective (usually in terms of block error rate (BLER) based on SINR), under specific static conditions. NS3 uses the Vienna LTE Simulator [8] to extract the link layer performance and build the PHY error model.

D. Link Adaptation

Adaptive modulation and coding (AMC) has been adopted in LTE to enhance the system throughput. It has been widely proven to be a powerful technique for improving the spectral and energy efficiency and increasing the data rate over a fading channel [9]. AMC is one of the most important techniques of link adaptation. Its objective is to maximize the data rate by adjusting transmission parameters according to channel variation. AMC operates on channel conditions identified by the channel quality indicator (CQI) obtained using SINR estimation.

By using AMC, the combination of different constellation of modulation and different rate of error-control codes are chosen based on the time-varying channel conditions. When channel quality is good (high CQI), AMC schemes with larger constellation sizes and higher channel coding rate can be applied to effectively achieve high transmission rate. When channel quality is poor (low CQI), transmission rate is reduced to ensure transmission quality. Thus it can realize optimal allocation of system resources and maximize throughput.

E. Scheduler in LTE

Since LTE is based on OFDM, it is possible to distribute available transmission radio resources in the frequency domain to different mobile terminals. The resource allocation can be changed dynamically. The MAC scheduler in the eNodeB (base station) assigns both uplink and downlink radio resources. The scheduling decision covers not only the resource block assignment but also the modulation and coding scheme to be used and the antenna techniques.

3GPP does not specify the MAC scheduler, but leaves its design and implementation to vendors. This added flexibility changes the system performance significantly with the use of different scheduling algorithms. Depending on the implementation, the scheduler can base its scheduling decision on the QoS class and the queuing delay of data, on the instantaneous channel conditions, or on fairness indicators. The channel conditions in a wide-band system vary in both time domain and frequency domain. With sufficiently detailed channel-quality information (such as CQI) to eNodeB, the scheduler can perform channel-dependent scheduling in the time and frequency domains, and improve the system capacity.

Carrier Aggregation offers another level of flexibility for the scheduler to allocate the radio resources. This functionality is to configure a set of CCs for each user. The CC assignment is an important apparatus for optimizing system performance, as well as limiting the power consumption of users. The power a UE requires increases with the number of CCs it is assigned (more bandwidth and signals to process). For optimal system performance, it is desirable to have approximately equal load on different CCs.

III. Problem Formulation

We consider a single-cell (one base station or eNodeB) scenario where the downlink bandwidth is configured with C component carriers which totally have M resource blocks (RBs) (similar notations to ones used in [10]). The base station serves N active users. We denote the set of component carriers as $\mathcal{C} = \{c|c = 1, 2, \cdots, C\}$, and all RBs as $\mathcal{M} = \{m|m = 1, 2, \cdots, M\}$, and the set of all users by $\mathcal{N} = \{n|n = 1, 2, \cdots, N\}$. C_n of C component carriers may be configured for user n. During each scheduling slot, the base station can allocate M_n RBs (RBs are not necessarily contiguous) from C_n component carriers to user n. Each RB is assigned to at most one user. We denote the power set of $\mathcal{M}$ as $\mathcal{P}$. Each p in $\mathcal{P}$ identifies a set of RBs. We have $x_i^p = 1$ if and only if p is allocated to user i. In the framework defined by the problem, two levels of resource allocations are involved: component carrier assignment and resource block scheduling. The former runs at a coarse granularity, and the major objective is to load balance across multiple component carriers. The latter runs at each scheduling interval (one TTI) in order to be channel adaptive and achieve better throughput and delay performance. It is evident that two levels are closely related and impact each other.

The resource allocation problem of multiple component carriers and resource blocks within a carrier for multiple users can be formulated as a combinatorial optimization problem. The objective is to maximize the total size of transport blocks (TBS) built for each user at each scheduling interval. The variables are different assignments of component carriers and schedules of resource blocks for users. The constraints are all available component carriers and resource blocks, available AMC modes, and the delay requirements for users. The component carrier allocation is implicitly embedded in the problem formulation which is done at a configurable interval.

Given a set of component carriers $\mathcal{C}$, resource blocks $\mathcal{M}$, a set of active users $\mathcal{N}$ and a set of available AMC schemes $\mathcal{A}$ during a scheduling period in a cell:

$$\textbf{Maximize:}\ \ \textbf{Min}_{i\in\mathcal{N}} \sum_{k\in\mathcal{C}} TBS^k_{UE(i)} \tag{1}$$

subject to:

$$\forall\ \text{RB}\ m \in \mathcal{M} : \sum_{m\in p, i\in\mathcal{N}} x_i^p \le 1 \tag{2}$$

$$\forall\ \text{user}\ i \in \mathcal{N} : \sum_{p\in\mathcal{P}} x_i^p \le 1 \tag{3}$$

$$\forall\ \text{user}\ i \in \mathcal{N}, p \in \mathcal{P} : x_i^p \in \{0,1\} \tag{4}$$

$$\forall\ \text{RB}\ m \in \mathcal{M} : \text{AMC}_{RB_m} \in \mathcal{A} \tag{5}$$

$$\forall\ \text{user}\ i \in \mathcal{N} : \text{delay}_i \le \text{DELAYBOUND}_i \tag{6}$$

Constraint (2) requires each RB be assigned to at most one user. Constraint (3) ensures each user get no more than one set of RBs.

IV. Joint Carrier Aggregation and Packet Scheduling

A. Dynamic Component Carrier Assignment

Carrier aggregation in LTE-Advanced will significantly boost the transmission speed on air interface. Efficient assignment of component carriers (CCs) to UEs to achieve higher network utilization and performance is an important issue that needs to be carefully addressed. In the existing carrier aggregation schemes, CCs are allocated statically when UEs attach to the network and the allocation does not change with time. In our proposed approach, we take channel conditions, network load and other impacting factors into account, and dynamically allocate component carriers for UEs.

In our scheme, the metric to determine the load of a component carrier is the aggregated value of queuing head-of-line (HOL) delay of UEs which are allocated the carrier. This metric can properly quantify and indicate the loading of a particular CC and implicitly factor in the channel conditions. It can help achieve network load balancing and high utilization.

In our scheme, we assume that there are M CCs and each UE can be allocated at most $N(< M)$ CCs. eNodeB periodically runs the dynamic assignment algorithm to reallocate CCs to UEs. The detailed procedure is presented in Algorithm 1. The frequency of running CC reallocation is a configurable parameter on eNodeB.

Algorithm 1 Dynamic CC Reallocation

Require: N out of M to be assigned to each of K UEs.
Clear the CC assignments for all UEs.
Reset AggregHOLPerCC[1...M].
Calculate HOL delay of all UEs.
Sort HOL delay in a descending order into HOL$(1...K)$.
for UE(k), $k = 1...K$ from HOL$(1...K)$ **do**
  for $i = 1...N$ **do**
    LeastHOL = INF;
    LeastHOLFlag = FALSE;
    for $j = 1...M$ **do**
      if AggregHOLPerCC[j] == 0 **then**
        Assign CC j to UE(k);
        AggregHOLPerCC[j] = HOL(k);
        LeastHOLFlag = TRUE;
        break.
      else
        if (LeastHOL $>$ AggregHOLPerCC[j]) and (CC j not assigned to UE(k)) **then**
          LeastHOL = AggregHOLPerCC[j];
          $LeastHOLCCId$ = j.
        end if
      end if
    end for
    if LeastHOLFlag == FALSE **then**
      Assign CC $LeastHOLCCId$ to UE(k);
      AggregHOLPerCC[$LeastHOLCCId$] = AggregHOLPerCC[$LeastHOLCCId$] + LeasetHOL.
    end if
  end for
end for

B. Backlog Based Resource Block Scheduling

At a finer granularity, scheduling resource blocks is another important issue. A number of backlog based policies (e.g.[11][12]) have been proposed in literature for multi-channel systems and shown to be optimal under specific restriction on arrival processes and channel conditions. However, their empirical performance in realistic scenarios (LTE systems in particular) is not known. We study and apply two types of backlog based algorithms (Algorithms 2 and 3): Queue Side Greedy and Server Side Greedy. Here queue implies the packet queue at eNodeB for UEs and server means the available resource blocks.

In both algorithms, backlog can be in terms of either head-of-line (HOL) delay or queue length. In some cases, good queue length performance does not necessarily translate to good delay performance. Simply maintaining low queue length is insufficient in order to guarantee low waiting-time in the queue. For example, a packet that is present in a queue with low queue length may have to wait for a long time to get served if fewer packets are offered to this queue for several time-slots. In our simulations, we will focus the Algorithm 2 (QSG) which has better performance.

Algorithm 2 Queue Side Greedy (QSG) RB Scheduling

Require: M resource blocks to be allocated to N UEs.
for $k = 1...M$ **do**
  Choose the UE i with the largest backlog;
  For UE i, choose the resource block j with the best channel rate;
  Allocate the resource block j to UE i and mark it as allocated;
  Update the backlog for UE i.
end for

Algorithm 3 Server Side Greedy (SSG) RB Scheduling

Require: M resource blocks to be allocated to N UEs.
for resource blocks $i = 1...M$ **do**
  Choose the UE j with the largest metric:
  Metric = backlog$\times$channel rate of resource block i for UE j;
  Allocate the resource block i to UE j;
  Update the backlog for UE j.
end for

In both algorithms, updating the backlog information for UE after one iteration is the key which is not done in conventional scheduling schemes. Backlog is a good indicator for the system load and the priority of UE to be scheduled. Both throughput and delay is taken into account with this metric.

C. Intelligent Link Adaptation

Channel awareness is indispensable for any efficient resource allocation schemes. In LTE, CQI is an important mechanism to obtain the feedback on channel quality and achieve link adaptation. In the existing implementation, the default link adaptation approach is that after all resource blocks are allocated to UEs by the scheduler, the resource blocks for a UE are built into one single transport block and modulated with MCS corresponding to the lowest CQI of the resource blocks of this UE. This approach is safe and conservative, but it may be inefficient and potentially lower the number of bytes that could be transmitted with the transport block. When a resource block is assigned to UE, the scheduler considers its CQI individually which may not be optimal in terms of all CQIs of resource blocks assigned to this UE. A function $\mathcal{F}$ can be designed to take a set of CQIs as input and output a single CQI which can be applied to all allocated resource blocks and achieve an optimal MCS and transport block size. Function $\mathcal{F}$ can be sophisticated or simple (like the default one: Min CQI).

Our basic idea is to incorporate the CQI tradeoff when allocating resource blocks and do the scheduling and link adaptation collectively (Algorithm 4). When a resource block with a particular CQI value is available to a UE, we check if the CQI value of this resource block can collectively enhance the overall transport block (TB) size with the resource blocks already allocated to this UE. If this CQI value is very low, then it will degrade the MCS mode of the transport block built with other resource blocks. In such a case, this resource block may be scheduled to other UE which may have better channel condition (better CQI) on it. In our case, function $\mathcal{F}$ is the weighted CQI average. Once the scheduling during a TTI is finished, the final MCS mode built into the transport block is also determined by the weighted average CQI.

Algorithm 4 Intelligent Link Adaptation

```
UE(i) allocated resource blocks PRB(j) (j = 1...K);
The next resource block PRB(K + 1) allocated to UE(i);
oldAvgCQI = Σ_{j=1}^{K} (CQI_{PRB_j})^2 / Σ_{j=1}^{K} CQI_{PRB_j};
oldMCS = GetMCSFromCQI(oldAvgCQI);
oldTBSize = GetTBSizeFromMCS(oldMCS);
newAvgCQI = Σ_{j=1}^{K+1} (CQI_{PRB_j})^2 / Σ_{j=1}^{K+1} CQI_{PRB_j};
newMCS = GetMCSFromCQI(newAvgCQI);
newTBSize = GetTBSizeFromMCS(newMCS);
if newTBSize > oldTBSize then
    Accept PRB(K + 1) and allocate to UE(i);
else
    Reject PRB(K + 1) and leave it to other UEs.
end if
```

V. Performance Evaluation

In this section we examine the performance of our proposed resource allocation schemes. We use the NS3 [5] as the system simulator. Table II lists the basic LTE parameters settings used in the simulations.

TABLE II
NS3 Parameter Settings

Parameter	Setting
Number of eNodeBs	1
Number of UEs	10
Antenna Model	1x1 Isotropic
Path loss model	Friis transmission equation
Fading model	Rayleigh fading trace
PHY error model	LSM based mapping
Bandwidths	3MHz, 5MHz, 10MHz, 20MHz
Component Carrier Bands	0.7GHz, 0.9GHz, 1.8GHz, 2.025GHz, 2.6GHz
eNodeB TX power	30dBm
UE TX power	10dBm
Noise spectrum density	-174 dBm/Hz
Traffic	Video over UDP

A. Comparison with Static Component Carrier Assignment

The LTE module in NS3 only supports Release 8 LTE, so we extend the model and add the carrier aggregation feature. We compare our dynamic CC assignment scheme with the static round-robin allocation which assigns CCs for UE when it attaches to the network in a round-robin fashion. Our simulations are conducted in a grid scenario with one eNodeB and ten UEs as shown in Figure 4. eNodeB is colocated with UE1. The space between UEs are 1500m. A UDP video session with a rate of 1000KB/s runs on the downlink between eNodeB and each UE. The total simulation time is 10 seconds. Five CCs are available and each UE can have at most three CCs assigned. The bandwidth is 5MHz (25 PRBs), so one UE can have a maximum aggregated 15MHz bandwidth (75 PRBs).

Fig. 4. A 10-UE Grid Scenario

In Figure 5 the aggregated throughput is compared: dynamic reallocation achieves higher aggregated throughput which translates to better network utilization. It is evident in Figure 6 that the average delay performance of dynamic scheme is better than the static assignment. The average delay keeps increasing with the static round-robin assignment.

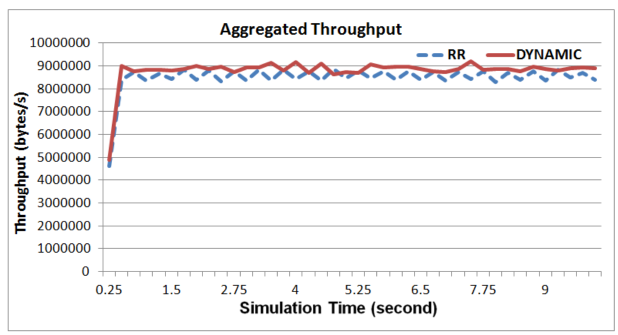

Fig. 5. Aggregated Throughput of 10 UEs

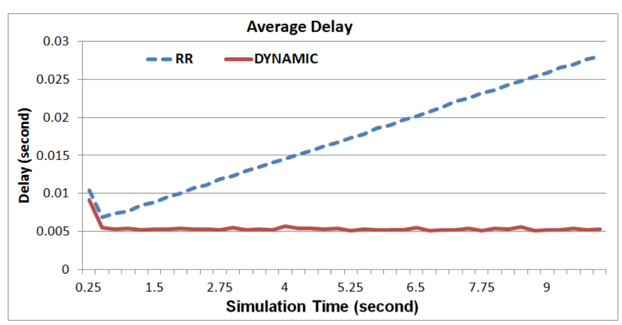

Fig. 6. Average Delay of 10 UEs

When we take a closer look, the problem is from UE6 whose queue is unstable and incurs a large delay. We further compare the throughput and delay of UE6 in Figures 7 and 8 respectively. Clearly both throughput and delay performance of the dynamic scheme is much better than the static approach:

the delay is low and stable; the throughput is much higher. The round-robin allocation assigns CCs 2.6GHz, 2.025GHz and 1.8GHz to UE6 which equivalently makes it a cell-edge user: bands with larger frequencies have smaller coverage. Different CCs may have different frequencies, hence different path loss characteristics and coverage, so proper CC assignment is very important to achieve high network utilization and user fairness. The advantage of our scheme is to dynamically reallocate CCs according to network and user conditions which gives much better network performance and user fairness.

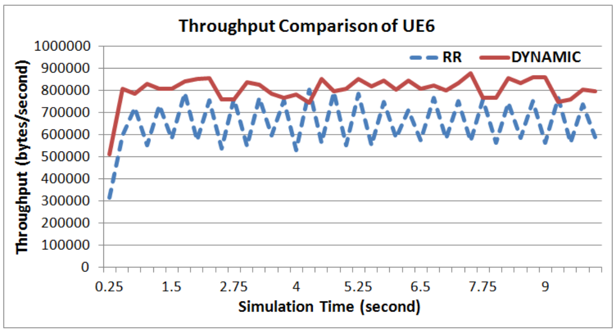

Fig. 7. Throughput of UE6

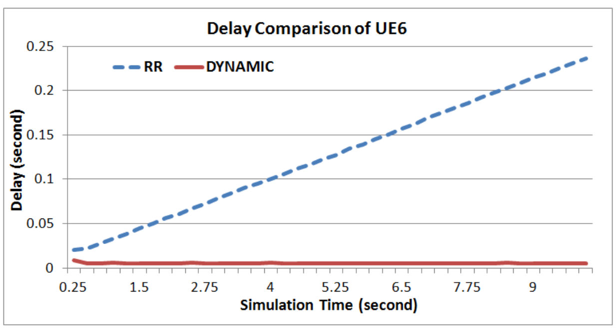

Fig. 8. Delay of UE6

B. Comparison with Proportional Fairness Scheduling Rule

In this section, we compare the backlog based scheduler (QSG) with the well-known Proportional Fairness scheduling rule [13] which is the default scheduler in many eNodeB implementations.

The Proportional Fairness scheduling rule, as the name suggests, attempts a "proportionally fair" allocation of radio resources to different users. Let $\mu_i(t)$ be the state of the channel of user i at time t, i.e. the actual rate supported by the channel. This rate is constant over one slot. Let $\overline{\mu}_i(t)$ denote the moving average rate of user i at time t. Then the rule can be defined as:

$$j = \arg\max_i \frac{\mu_i(t)}{\overline{\mu}_i(t)}$$

Compared to the Max-Rate rule [14] where a user with the highest instantaneous rate or equivalently the best channel condition is scheduled, the proportional fairness rule balances the user requests by considering their historical rates. The scheduler provides fairness in the sense that at any time users with lower moving average rates ("starved" in the past) will have higher priorities to be allocated radio resources. However, it is noted that the proportional fairness rule does not account for packet delay and can result in poor delay performance.

We use the same topology as shown in Figure 4. The horizontal and vertical distance between UEs is 400m. A UDP video session with a rate 300KB/s runs on the downlink between eNodeB and each UE. The simulation time is 10 seconds. The aggregated throughput and average delay across 10 UEs are presented in Figures 9 and 10 respectively. Backlog scheduler offers slightly higher throughput but much better delay performance. In Figures 11 and 12 the average throughput and delay of each UE are compared. Backlog scheduler has lower delay for each UE and better fairness among UEs is achieved. 7 out of 10 UEs have higher or equal average throughput with the backlog scheduler, and other 3 UEs have comparable throughput performance. The throughput of the cell-edge UE 10 is about 14% higher than the PF scheduler, and its delay is substantially lower than the PF scheduler.

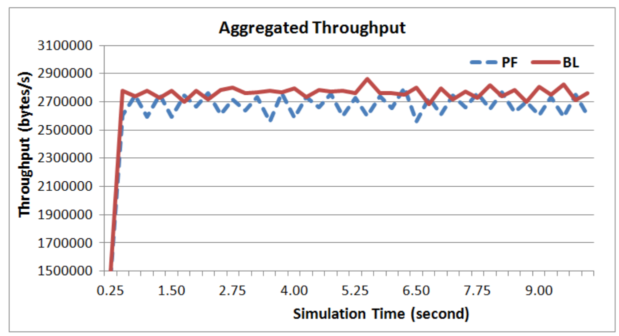

Fig. 9. Aggregated Throughput of 10 UEs

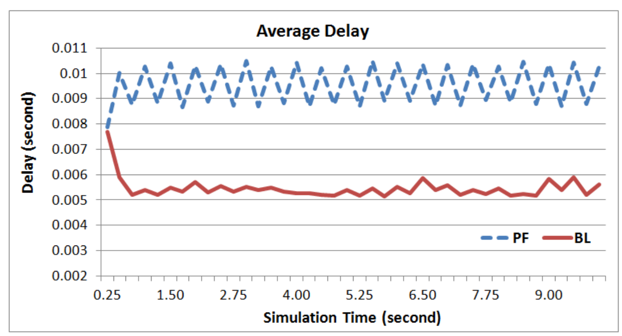

Fig. 10. Average Delay of 10 UEs

C. Comparison with Min-CQI Based Link Adaptation

In this section, we compare our weighted-average-CQI based link adaptation scheme with the minimum CQI based approach. In Figure 4, the space between UEs is 1000m. The bandwidth is 10MHz (50 PRBs). A 200KBps video session runs on the downlink from eNodeB to each UE. To better understand the impact of CQI selection, we also study the Max CQI where the MCS mode is determined by the maximum CQI of the allocated resource blocks for a UE. Min CQI and Max CQI are two extreme cases: most conservative and most

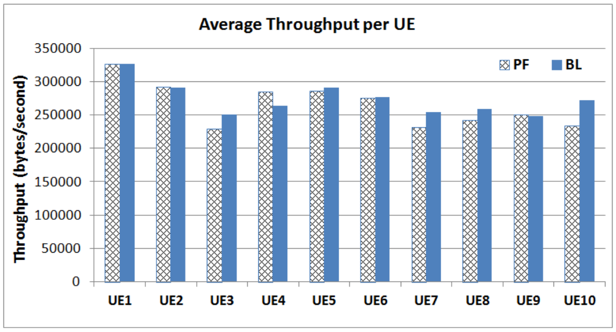

Fig. 11. Individual Average Throughput of 10 UEs

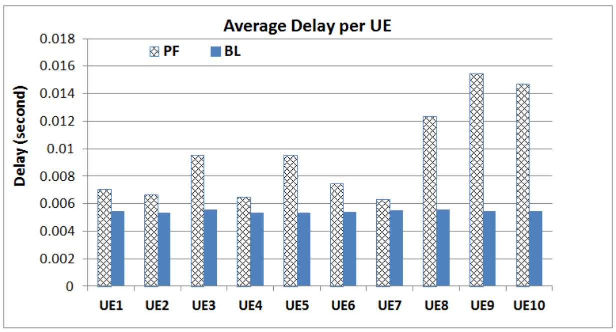

Fig. 12. Individual Average Delay of 10 UEs

optimistic. Higher CQI will indicate higher MCS mode which in turn generates higher bits per symbol. However, higher MCS mode will incur larger BER (Bit Error Rate). This shows that the tradeoff between the transmission rate and the error rate needs to be carefully considered. Min CQI favors low BER but the delay is large while Max CQI has good delay performance but the throughput is very low.

TABLE III
COMPARISON OF DIFFERENT LINK ADAPTATION SCHEMES

	Min CQI	Max CQI	Weighted CQI
Aggregated Throughput (KBps)	1664	1088	1654
Average Delay (ms)	16.7	12.8	14.5

TABLE IV
COMPARISON BETWEEN MIN-CQI AND WEIGHTED-CQI

	Min CQI	Weighted CQI
Aggregated Throughput (KBps)	2245	2190
Average Delay (ms)	31.6	9.6

In our approach, we average the CQIs across all allocated resource blocks in a weighted manner. Hence the high CQIs are weighted more but the tradeoff is also taken into account. Table III presents the results of three link adaptation schemes. Although the aggregated throughput degrades slightly (0.6% less), the delay performance is much better (13% less). It is expected that wider bandwidth may have more CQI variations, so the performance gain would be higher. In another scenario where the downlink bandwidth is 20MHz (100 PRBs) and the rate is 250KBps, Table IV shows the comparison: the throughput degradation is 2.4% but the delay is substantially lower. In Figure 13 we further compare the per UE average delay: UE10 has a constantly increasing delay (the average over the simulation duration is 230ms and off the chart).

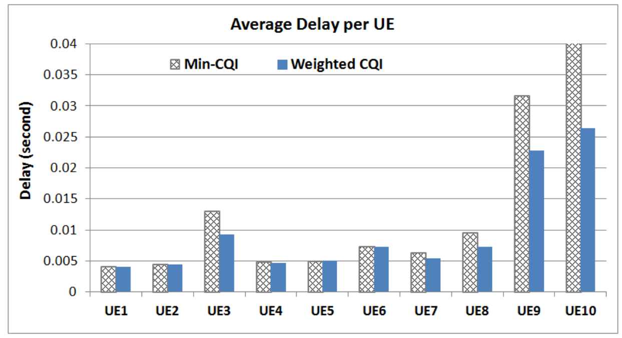

Fig. 13. Individual Average Delay of 10 UEs

D. Overall Performance

In this section, we incorporate the proposed schemes (PROPOSED in the legend) and evaluate their performance collectively in a general scenario. The baseline is the proportional fairness scheduler with the static round-robin component carrier assignment (DEFAULT in the legend). The scenario is one eNodeB with ten UEs. The UEs are uniformly distributed within a 1000m circle centered at the eNodeB, and move under the random waypoint mobility model with a speed uniformly distributed in the range of [0, 15m/s] (from pedestrian to low-speed urban vehicle). Three 3MHz bands are allocated for each UE. The downlink rate for each UE is 600KBps. We compare the aggregated throughput and average delay of UEs in Figures 14 and 15 respectively. The aggregated throughput offered by the proposed schemes is higher, so the network utilization is better. For the delay performance, the default approach produces a constantly increasing delay (the queues are unstable). The standard deviation of the average delay of 10 UEs by our schemes is 0.006ms while it is 3.478ms by the default approach, so the delay fairness is also better.

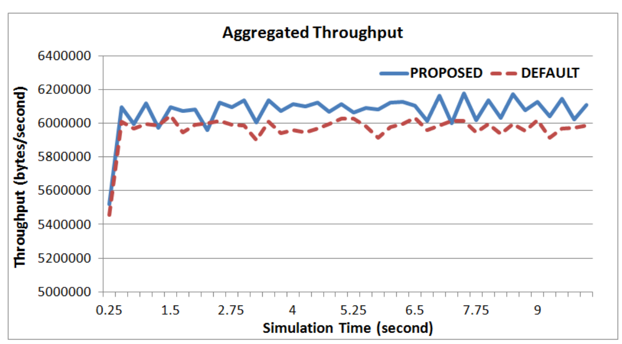

Fig. 14. Aggregated Throughput of 10 Random UEs

VI. RELATED WORK

Resource allocation in OFDMA based LTE networks has been one of the important problems of RRM in wireless networks and attracting increasing research interests recently.

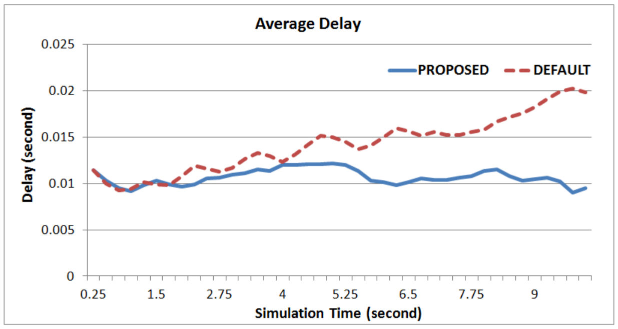

Fig. 15. Average Delay of 10 Random UEs

From the theoretical perspective, scheduling algorithms in multi-channel networks were proposed in [11][12][15]. However, in realistic scenarios, those algorithms may need significant extensions. In [10][16], the authors presented the hardness results for LTE frequency-domain scheduling. Under certain circumstances, it has been proven to be NP-hard.

Joint component carrier allocation and packets scheduling were reported in [17]. The carrier assignment algorithm was simplistic and static. At the subcarrier level, it used the PF scheduler. In [18], the authors focused on the uplink carrier allocation in favor of cell-edge users, and no delay performance was considered. Various schemes have been proposed to address the packet scheduling problem in wireless channel. In addition to the Max-Rate rule and the PF rule discussed in the simulations, two queue driven rules were presented in [19] (the EXP rule) and [20] (the LOG rule) with different focus. As the queues grow linearly, the EXP rule schedules in a manner that emphasizes queue-balancing at the cost of total weighted service rate while the LOG rule schedules in a manner that de-emphasizes queue-balancing in favor of increasing the total weighted service rate (being opportunistic). Both rules are single-user schedulers. Their scheduling metrics bundle two components – channel rate and queue-length, so the scheduling decision may not be optimal individually for both components in multi-subcarrier wideband systems for multi-user scheduling.

Most of the existing CQI schemes [21][22][23] focused on the accurate and timely feedback mechanisms which are important but did not consider the CQI adaptation when scheduling PRBs and constructing transport blocks (TBs). The tradeoff between CQI selection and corresponding MCS mode (TB size) has a significant impact on network performance.

VII. Conclusion

In this paper, we studied the downlink radio resource allocation problem for LTE-Advanced systems which includes both component carrier assignment and resource block scheduling. We proposed a dynamic carrier assignment algorithm and backlog based resource block scheduling policies to achieve load balancing across component carriers and better throughput and delay performance for users compared with the existing schemes. In addition, we presented an intelligent link adaptation scheme based on weighted average CQI which can be plugged into the scheduling algorithms to further enhance the delay performance. The effectiveness of the proposed schemes is evidenced with extensive simulations under realistic models and settings. The proposed schemes consist of a two-tier (carrier and subcarrier) scheduling framework which can be extended and enhanced with more sophisticated metrics and algorithms in future research.

References

[1] 3GPP TS 36.300, "E-UTRA and E-UTRAN Overall Descrption Stage 2 (Release10)," Sep. 2011, v10.5.0.

[2] S. Shakkottai and A. L. Stolyar, "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR," in *in Proceedings of 17th International Teletraffic Congress (ITC-17*, Brazil, Dec. 2001.

[3] 3GPP TS 36.211, "E-UTRA Physical Channels and Modulation (Release9)," March 2010, v9.1.0.

[4] http://http://en.wikipedia.org/wiki/Friis_transmission_equation.

[5] http://www.nsnam.org/.

[6] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall PTR, 2001.

[7] http://iptechwiki.cttc.es/LTE-EPC_Network_Simulator_(LENA).

[8] http://www.nt.tuwien.ac.at/about-us/staff/josep-colom-ikuno/lte-simulators/.

[9] M. Alouini and A. Goldsmith, "Adaptive Modulation over Nakagami Fading Channels," *Wireless Personal Communications*, May 2000.

[10] H. Yang, F. Ren, C. Lin, and J. Zhang, "Frequency-Domain Packet Scheduling for 3GPP LTE Uplink," in *Proc. IEEE INFOCOM*, San Diego, CA, March 2010, pp. 1–9.

[11] S. Bodas, S. Shakkottai, L. Ying, and R. Srikant, "Low-complexity scheduling algorithms for multi-channel downlink wireless networks," in *INFOCOM*, San Diego, CA, March 2010.

[12] ——, "Scheduling for small delay in multi-rate multi-channel wireless networks," in *INFOCOM*, Shanghai, China, April 2011.

[13] A. Jalali, R. Padovani, and R. Pankaj, "Data Throughput of CDMA-HDR a High Efficiency-High Data Rate Personal Communication Wireress System," in *Proc. IEEE VTC*, Tokyo , Japan, May 2000.

[14] B. S. Tsybakov, "File Transmission over Wireless Fast Fading Downlink," *IEEE Transactions on Information Theory*, August 2002.

[15] M. Sharma and X. Lin, "OFDM Downlink Scheduling for Delay-Optimality: Many-Channel Many-Source Asymptotics with General Arrival Processes," in *ITA Workshop*, San Diego, CA, Feb. 2011.

[16] S.-B. Lee, S. Choudhury, A. Khoshnevis, S. Xu, and S. Lu, "Downlink MIMO with Frequency-Domain Packet Scheduling for 3GPP LTE," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, April 2009.

[17] Y. Wang, K. Pedersen, T. Soensen, and P. Mogensen, "Carrier Load Balancing and Packet Scheduling for Multi-Carrier Systems," *IEEE Trans. On Wireless Comm.*, May 2010.

[18] R. Sivaraj, K. Z. A. Pande, K. Govindan, and P. Mohapatra, "Edge-prioritized Channel- and Traffic-aware Uplink Carrier Aggregation in LTE-Advanced Systems," in *IEEE Symp. on WoWMoM*, San Francisco, CA, June 2012.

[19] S. Shakkottai and A. L. Stolyar, "Scheduling for Multiple Flows Sharing a Time-Varying Channel: The Exponential Rule," *American Mathematical Society Translations, Series 2*, vol. 207, 2002.

[20] B. Sadiq, S. J. Baek, and G. de Veciana, "Delay-optimal opportunistic scheduling and approximations: The log rule," in *IEEE INFOCOM*, Rio de Janeiro, Brazil, April 2009, pp. 1692–1700.

[21] N. Kolehmainen, J. Puttonen, P. Kela, T. Ristaniemi, T. Henttonen, and M. Moisio, "Channel Quality Indication Reporting Schemes for UTRAN Long Term Evolution Downlink," in *IEEE VTC 2008*, Singapore.

[22] X. Chen, H. Yi, H. Luo, H. Yu, and H. Wang, "A Novel CQI Calculation Scheme in LTE/LTE-A Systems," in *WCSP 2011*, Nanjing, China.

[23] S. Donthi and N. Mehta, "Performance Analysis of Subband-Level Channel Quality Indicator Feedback Scheme of LTE," Chennai, India, NCC 2010.