A Cross-Layer Design for Dynamic Resource Block Allocation in 3G Long Term Evolution System

Hossam Fattah and Hussein Alnuweiri
Department of Electrical & Computer Engineering
Texas A&M University
Email: hossamf@ieee.org

***Abstract*—3G Long Term Evolution (LTE) has emerged as a comprehensive evolution of the Universal Mobile Telecommunications System (UMTS). LTE system uses resource blocks which are the basic unit of exchanging information in both downlink and uplink. However, resource allocation scheme, which is a crucial component to guarantee quality of service, remains as an open issue. In this paper, we propose a novel cross-layer scheduling algorithm for LTE system that allocates resources, both as resource blocks and a modulation and coding scheme, among users with different traffic load. The algorithm minimizes the overall average packet delay and takes into account queueing theory, modulation and coding scheme supported by each user, channel condition, and available transmit power. Simulation results shows that the proposed algorithm provides substantial reduction in average delay compared to conventional algorithm.**

***Index Terms*—Wireless Networks, LTE, UTRAN, OFDMA, Queueing Analysis.**

I. Introduction

3G LTE standard grew out of the Third Generation Partnership Project (3GPP) as an evolution of UMTS/HSPA to ensure it remains competitive in the future [1] especially with the emergence of the IEEE 802.16 (WiMAX) wireless broadband standard. In late 2004, 3GPP began a project to define the long-term evolution of UMTS cellular technology. The specifications related to this effort are formally known as the evolved UMTS Terrestrial Radio Access Network (E-UTRAN), but are more commonly referred to by the project name LTE. UMTS/HSPA specifications were released in 3GPP Release 5 and 6 and they continue to evolve in Release 7 and 8 under the name HSPA+. 3GPP Release 8 specifies LTE as well as further enhancements to the existing HSPA+ technology.

LTE evolves from the circuit-switched and packet-switched architecture of the UMTS to an all-IP, packet-only system in addition to many other key features such as higher data rate, higher spectral efficiency, scalable bandwidth, multi-antenna configuration, low user-plane latency, full mobility, co-existence with legacy standards, support for TDD/FDD mode of operations, better power-saving mechanism, seamless integration of unicast and enhanced broadcast transmission, and reasonable system and terminal complexity. The LTE physical layer is designed primarily for full duplex operation in a paired spectrum with asymmetrical modulation for downlink and uplink. OFDM is used for the downlink while SC-FDMA is used for the uplink. Although SC-FDMA has many similarities to OFDM, its main advantage is that it has a low power amplifier de-rating requirement, thereby conserving User Equipment (UE) battery life time and extending its range.

LTE uses the concept of a resource block (RB), which is a block of subcarriers both in frequency and time domain [2], [3]. That is, a RB is typically a consecutive number of subcarriers in frequency domain and a number of consecutive OFDM symbols in time domain (i.e., 12 consecutive subcarriers by 7 OFDM symbols) [3]. In FDD mode of operation, a 10 ms radio frame is divided into 20 slots of 0.5 ms each. Each two slots constitute a 1 ms sub-frame. Each sub-frame represents a Transmission Time Interval (TTI) which is the minimum amount of a transmission unit. A number of consecutive RBs in time domain which constitutes one TTI (typically two consecutive RBs [2]) represents the minimum transmission unit[1]. The physical layer interface is a transport block which is a group of RBs with a common Modulation and Coding Scheme (MCS). Each TTI contains a maximum of one transport block per UE [3], [4]. Figure 1 (a) shows the LTE radio frame structure and the resource block structure both in time and frequency domain whereas (b) shows how RBs are organized in each TTI where each TTI has six RBs.

Subcarrier and bit allocation schemes in OFDMA systems are the main challenge that has been the focus of many proposals in [5]–[7] and references therein. Basically, a subcarrier and bit allocation algorithm can be categorized as a static allocation [8] and dynamic allocation [5], [6]. In static allocation, each user occupies a pre-determined number of subcarriers that are fixed from one OFDM symbol to the other. However, in dynamic allocation algorithms, the subcarrier and bit allocation varies from one OFDM symbol to another based on the instantaneous channel condition. Dynamic allocation algorithms achieve higher throughput, lower power consumption, and higher system capacity than static methods due to their adaptive nature.

Several LTE packet scheduling schemes has been proposed. Maximizing throughput is the objective of the proposed algorithms in [9] where RBs is allocated to a user with the best channel condition thus maximizing its throughput. A proportional fairness scheme is also proposed in [9] that aims at maximizing the throughput while achieving fairness by

[1]During the reminder of this paper, we use the term RB to denote two consecutive RBs in time domain that constitute one TTI as in [2]

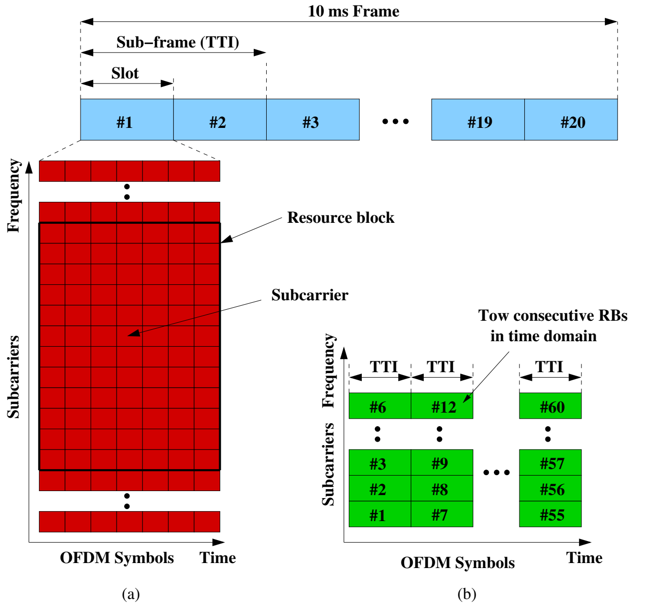

Fig. 1. (a) LTE Radio frame and resource block structure as defined in [3] and (b) TTI and resource block organization where each TTI has six RBs.

taking the user past average throughput into consideration. Proportional fairness is also the focus of the scheme in [10]. Algorithms in [9], [10] focuses on maximizing throughput and achieving fairness which can be desirable features for non real-time traffic but not for real time traffic with Quality of Service (QoS) constrains such as packet delay. A VoIP scheduling scheme is proposed in [11] where the RB is assigned to a user with the longest queue and the best channel condition. A user with a longer queue will always hog the channel thus effecting QoS parameters for other users. The scheme outlined in [12] tries to minimize the number of control channel needed for exchanging scheduling information between UEs and Base Station (BS) by introducing a semi-persistent scheduling scheme that maximizes the number of VoIP users. In such a context, notation of QoS guarantee is not considered. Two algorithms that either allocates RBs to users in a round-robin fashion or to a user with the best channel were proposed in [13] but the first one does not consider the channel condition when scheduling users and the latter ignores users with bad channels thus increasing their delay.

Packet transfer delay, defined as the time between the arrival of a packet at the transmitter and its complete reception at the receiver, is a commonly used measure of performance for non-real time services such as web browsing VoIP, or FTP applications. Although the proposed algorithms focus on maximizing throughput, achieving proportional fairness, or maximizing the number of VoIP users, queue dynamics and application level requirements are ignored which leads to unsatisfactory design from a higher layer system performance perspective especially with the provision of QoS requirements such as packet delay. While achieving fairness is important, it is not the only requirement users need nor it is very easy to determine the relative *weights* to be assigned to users. Delay and packet loss rate are, however, important and measurable quantities that directly impact the Quality of Service (QoS) provided to the user. Delay, in particular, affects the user Quality of Experience (QoE). Moreover, minimizing delay is strongly justified by the fact that even most common multimedia streaming services, such as Skype and Windows Media Services [14], are carried over delay-friendly TCP connections [14]. Therefore, reducing delay for TCP connections results in significant improvement to multimedia streaming services. Reducing delay for non-real time services is an important target for content delivery over wireless networks [15]. Moreover, the proposed schemes [9]–[13] assume users with homogenous traffic load, do not consider capability of a UE in terms of its supportable MCS, and finally, they ignore the fundamental fact that the transmit power allocated to all subcarrier in an OFDM symbol should not exceed a maximum target value.

In this paper, a cross-layer scheduling algorithm is proposed which allocates resources, both RB and MCS, dynamically according to channel condition and traffic load so as to minimize the overall average packet transfer delay, while satisfying BER, queue stability constraints, UE supportable MCS, and available transmit power constraints. The proposed algorithm is shown to provide significant lower average delays. To the best of our belief, this is the first RB allocation algorithm that considers packet delay as a criteria for enhancing QoS while taking into account fundamental and practical constrains as outlined by the LTE specifications [3], [4]. In the proposed algorithm, an optimization problem is formulated, which aims at minimizing the overall average packet transfer delay for all users subject to target BER, available transmit power level, queue stability, UE supportable MCS, transport block constrains. Simulation results show that the proposed algorithm outperforms conventional algorithm in terms of the overall packet delay.

This paper is organized as follows: Section II outlines the system model and problem formulation. Section III covers the proposed Resource Block Allocation (RBA) algorithm. Finally, simulation results and conclusion are provided in Section IV and Section V respectively.

II. System Model and Problem Formulation

Figure 2 shows the system model of an LTE transmitter and receiver with cross-layer design for allocating RBs in a frame to different users. Each OFDM symbol contains a number of subcarriers in the set $S_c = \{1, 2, \ldots, C\}$ where C is the total number of subcarriers in an OFDM symbol. All RBs in the set $S_r = \{1, 2, \ldots, R\}$ are available for allocation to users where R is the total number of RBs in a frame. $S_r^t \in S_r$ is the set representing all RBs in the tth TTI whereas $S_r^n \in S_r$ is the set representing all RBs in a frame allocated to a user n. Therefore $S_r^t \cap S_r^n$ is the set of RBs assigned to user n in the tth TTI. $S_t = \{1, 2, \ldots, T\}$ is set of all TTIs where T is the total number of TTIs in a frame. Each RB is a grid with size equal to $K < |S_c|$ consecutive subcarriers by F consecutive OFDM symbols. Following the LTE standard constraints in [3], [4], all RBs have the same grid size, subcarriers in each RB have the same MCS, and all RBs allocated to a user in an TTI have also the same MCS. Each user n supports a maximum number,

M, of MCSs in the set $S_\Phi^n = \{\Phi_{n,1}, \Phi_{n,2}, \dots, \Phi_{n,M}\}$. $\Phi_{n,m}$ represents the number of data bits per subcarrier (i.e., 1, 2, 4, 6 for BPSK, QPSK, 16-QAM, and 64-QAM respectively).

At the transmitter, there are N users where each user is represented as an $M/D/1$ queue. Data packets are of fixed length and arrive at queue n according to a Poisson process with arrival rate λ_n bits per frame. Queue n packets are serviced with a service rate R_n bits per frame. The average packet transfer delay of a user is given by [16]:

$$\bar{D}_n = \frac{1}{R_n}[1 + \frac{\rho_n}{2(1-\rho_n)}], \quad \forall n \in U, \tag{1}$$

where $\rho_n = \lambda_n / R_n < 1$ and $U = \{1, 2, \dots, N\}$.

Let $h_{n,r}$ denote the channel gain of the rth RB to user n. Channel gain, $h_{n,r}$, can be estimated using the Channel Quality Indicator (CQI) which is reported by the user measurement entity to the BS to provide time and frequency channel quality information for better spectral efficiency and resource allocation. For downlink RBs, users uses the Physical Uplink Control Channel (PUCCH) to convey channel quality information to the BS. BS conveys downlink RBs allocations and MCS assignments to all users using the Physical Downlink Control Channel (PDCCH) [3]. For uplink RBs, the BS estimates the channel quality of the received uplink RBs and conveys uplink RB allocations to users using the PDCCH.

An RB allocation matrix is denoted by $a_{n,r}$ (i.e. $a_{n,r}$ is an N rows by R columns matrix where $a_{n,r}$ = 1 if RB r is allocated to user n and $a_{n,r}$ = 0 if RB r is not allocated to user n). The total service rate for user n can be calculated as:

$$R_n = K \cdot F \cdot \sum_{r=1}^{R} (a_{n,r} \cdot \Phi_{n,b_{n,r}}), \ \forall r \in S_r^n \tag{2}$$

where $b_{n,r}$ is an MCS assignment matrix (N by R) representing the MCS index assigned for each RB (i.e., $b_{n,r}$ = 2 if RB r allocated to user n is assigned MCS $\Phi_{n,2}$).

The transmit power of the rth RB allocated to user n is denoted by $P_{n,r}$. That is, $P_{n,r}$ represents the sum of transmit power allocated to all subcarriers of an OFDM symbol belonging to rth RB allocated to user n. $P_{n,r}$ can be expressed as

$$P_{n,r} = K \cdot \frac{\sigma^2 \Gamma}{h_{n,r}} (2^{\Phi_{n,b_{n,r}}} - 1), \tag{3}$$

where Γ is a constant Signal-to-Noise Ratio (SNR) which is related with the target BER and σ^2 denotes the variance of additive white Gaussian noise. The total transmit power of an OFDM symbol in the tth TTI can be expressed as

$$P_t = \sum_{r=1}^{R} \sum_{n=1}^{N} (a_{n,r} \cdot P_{n,r}), \quad \forall r \in S_r^t. \tag{4}$$

All OFDM symbols in a TTI have the same transmit power due to the fact that each RB has the same MCS and thus the same number of allocated bits to each of its subcarrier. Each TTI can have different transmit power.

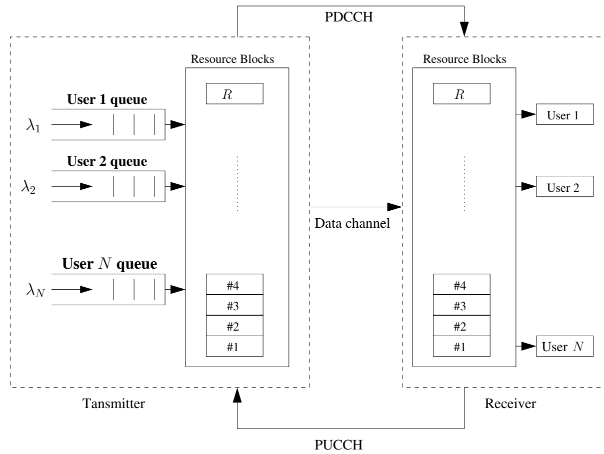

Fig. 2. LTE model for cross-layer design and RB allocation.

The objective of the resource block allocation algorithm, which allocates RBs and assigns MCS to each RB, is to minimize the overall average packet delay for all users while providing the BER requirement, queue stability constraint, power resource limitation, channel condition awareness, supportable MCS, and transport block constrains. This is an optimization problem that can be formulated as follow:

$$(\mathbf{P}) \qquad \min_{\{R_n\}_{n=1}^{N}} \ \bar{D} = \frac{1}{\lambda_1 + \lambda_2 + \cdots + \lambda_N} \sum_{n=1}^{N} \lambda_n \bar{D}_n$$

subject to the following constrains:

$$\begin{aligned}
C1: \quad & \sum_{n=1}^{N} a_{n,r} \le 1, \quad \forall r \in S_r, \\
C2: \quad & P_t \le P_v, \quad \forall t \in S_t, \\
C3: \quad & \Phi_{n,b_{n,r}} \in S_\Phi^n, \quad \forall n \in U, \ \forall r \in S_r^n, \\
C4: \quad & \Phi_{n,b_{n,x}} = \Phi_{n,b_{n,y}} = \cdots = \Phi_{n,b_{n,z}}, \\
& \quad \forall n \in U, \ \forall x, y, z \in S_r^t \cap S_r^n, \ \forall t \in S_t, \\
C5: \quad & R_n > \lambda_n, \quad \forall n \in U,
\end{aligned} \tag{5}$$

where C1 guarantees that each RB can only be used by at most one user; C2 indicates that transmit power on all subcarriers of an OFDM symbol does not exceed the available transmit power P_v; C3 guarantees that an RB allocated to a user has its MCS supportable by that user; C4 ensures that all RBs assigned to a user in a TTI (e.g. a transport block) have the same MCS; and finally C5 guarantees that the service rate of a user is larger than its arrival rate. The problem **P** is a complex combinatorial optimization problem that minimizes the weighted delay for all users. It is hard to find an optimal solution to this problem using known optimization techniques. For such a purpose, we propose a suboptimal solution to this problem as outlined in the next section.

III. Resource Block Allocation (RBA) Algorithm

In this section, we provide an RBA algorithm for solving problem **P**. The RBA has two phase, the first phase is the RB allocation phase and is outlined in Fig. 3 while the second one is the MCS assignment phase and is outlined in Fig. 4.

The RB allocation phase starts by initializing the RBs allocated to each user, $a_{n,r}$, the MCS assigned to each RB, $b_{n,r}$, the set of RBs allocated to user n, S_r^n, a TTI counter, t, the set of users, U, and finally the weight of each user, w_n. Step (1.b) sets each user weight in proportional to its arrival rate (i.e., it is the ratio of a user arrival rate to the minimum arrival rate). The RBA tries to spread RBs allocation to each user among TTIs as much as possible and at the same time select the best channel for a user. The reason for the first criteria is that allocating different RBs in the same TTI for the same user does not utilize the power resource efficiently since they must all have the same MCS (as required by $C4$). In other words, changing the MCS assignment for a user requires changing the MCS for all its allocated RBs in a TTI and this leads to a poor utilization of the power resource. The second criteria stems from the fact that choosing an RB with the best channel condition is important for minimizing the transmit power.

In the second stage, the algorithm allocates one RB to each user and guarantees that each TTI has only one RB allocated to a single user. In the third stage, RB allocations are done in proportional to user weights. First, a user is chosen according to its weight. Then, a TTI that has the minimum number of RBs allocated to that user, is chosen (Step (3.b)). An RB from the selected TTI with the best channel condition to that user is allocated to it (Step (3.c)). The RBA concludes the allocation phase by updating the allocation matrix $a_{n,r}$ and the set of RBs allocated to each user.

The MCS assignment phase starts with initializing the TTI counter, t, as shown in the fourth stage. In the fifth stage, the RBA distributes $R_n = \lambda_n + 1$ bits for each user among its allocated RBs in order to fulfil $C5$. For each user, the RBA continuously selects the RB with the lowest MCS (Step (d.1)), finds the corresponding TTI (Step (d.2)), and increments MCS for all RBs in this TTI belonging to the same user (Step (d.4)) until the user total service rate exceeds its arrival rate.

The six stage aims at assigning MCS for all RBs in a way to minimize the overall average packet delay. This stage keep toggling between TTIs and in each TTI, it increases the MCS for those RBs in that TTI which yields the lowest average delay. In each TTI, the RBA starts by calculating the total service rate for each user as currently assigned to its RBs (Step (b.1)) and the total service rate when the MCS for all its RBs in this TTI are incremented (Step (b.2) and (b.3)). Note that if an RB allocated to a user is already assigned the maximum supportable MCS, Step (b.2) has no effect and the total service rate as in Step (b.3) remains the same as that in Step (b.1). As a result of incrementing the MCS, the RBA calculates the corresponding reduction in weighted delay (Step (b.5)). If the transmit power, as a result of incrementing the MCS, exceeds the available transmit power or the RBs are already assigned the maximum supportable MCS, the user is excluded for any further processing during this TTI (Step (b.6)). If all users in a given TTI are excluded, then all RBs in this TTI are considered

Input: $\lambda_n,\ N,\ S_r, S_t,\ h_{n,r}$
output: $a_{n,r}$
(1) Initialization
  (a) Set $a_{n,r} = 0$, $b_{n,r} = 0$, $S'_r = S_r$, $S_r^n = \phi$, $t = 1$, $U = \{1, 2, \cdots, N\}$;
  (b) Calculate $w_n = \dfrac{\lambda_n}{\arg\min_i \lambda_i}$, $\forall n, \forall i \in U$;
(2) for each $n \in U$,
  (a) Find $r^* = \arg\max_r h_{n,r},\ \forall r \in S_r^t \cap S'_r$;
  (b) Set $a_{n,r^*} = 1$, $S'_r = S'_r - r^*$, $S_{r^*}^n = S_{r^*}^n + r^*$;
  (c) if $t > |S_t|$, Set $t = 1$, else Set $t = t + 1$;
(3) while $S'_r \neq \phi$,
  (a) Find $n^* = \arg\min_n \dfrac{|S_r^n|}{w_n}$;
  (b) Find $t^* = \arg\min_t |S_r^t \cap S_r^{n^*}|,\ \forall t \in S_t,\ S_r^t \neq \phi$;
  (c) Find $r^* = \arg\max_r h_{n^*,r},\ \forall r \in S_r^{t^*} \cap S'_r$;
  (d) Set $a_{n^*,r^*} = 1$, $S'_r = S'_r - r^*$, $S_{r^*}^{n^*} = S_{r^*}^{n^*} + r^*$;

Fig. 3. Pseudo code for the RB allocation phase.

to have reached the maximum possible MCS and the TTI is excluded from any further processing (Step (c)). Step (d) selects a user with the maximum weighted delay reduction, $\Delta\bar{D}_{max}$, and its corresponding power increase ΔP_{min}. In Step (f), a *Cost* function is calculated for all users as follow:

$$C(\Delta\bar{D}_n, \Delta P_n) = \frac{\Delta P_n}{\Delta P_{min}} - \frac{\Delta\bar{D}_n}{\Delta\bar{D}_{max}} \quad \forall n \in U, \qquad (6)$$

where $\Delta\bar{D}_n$ is user n reduction in weighted delay and ΔP_n is the corresponding increase in power. The Cost function $C(.)$ is a function that determines the overall cost between a reduction in weighted delay and the corresponding increase in power. In other words, it aims at finding a *reasonable* cost that balances between reduction in weighted delay and power increase. After calculating $C(.)$ for all users, the algorithm selects the user n^* with the minimum cost function and increments the MCS for its RBs in the given TTI as shown in Step (g) and (h) respectively. Finally, Step (h), increments the TTI counter to move to another TTI. This phase concludes when all RBs in each TTI either reach the maximum supportable MCS or their total transmit power reach the available transmit power.

IV. Simulation Results

In this section, we provide simulation results for the proposed RBA algorithm. We use an LTE OFDM system with a total number of RBs, R, equals to 60 [2]. That is, 6 RBs per each TTI where the total number of TTIs in a frame, T, is 10. The required SNR for a target BER is $8\ dB$. The variance of the additive white Gaussian noise at each receiver is assumed to be $\sigma^2 = 1mW$. The available transmit power is $P_v = 10W$. We assume a fixed channel gain of unity and a variable channel gain that is uniformly distributed within $[0.2\ 1]$. We assume that each user supports MCSs in the set $S_\Phi^n = \{1, 2, 4, 6\}$ corresponds to BPSK, QPSK, 16-QAM, and 64-QAM respectively. Arrival and service rates are in bits per

Input: $\lambda_n,\ N,\ S_r,\ S_t,\ S_\Phi^n,\ h_{n,r},\ \Gamma,\ \sigma^2,\ P_v,\ a_{n,r}$
output: $b_{n,r}$
(4) Initialization
  (a) Set $t = 1$;
(5) for each $n \in U$,
  (a) Set $b_{n,r} = 1,\ \forall r \in S_r^n$;
  (b) Set $R_n = \lambda_n + 1$;
  (c) Set $R_n = R_n - (KF \sum_{r=1}^{R} (a_{n,r}\Phi_{n,b_{n,r}})$;
  (d) while $R_n > 0$,
    (d.1) Find $r^* = \arg\min_r b_{n,r},\ \forall r \in S_r^n$;
    (d.2) Find $t^* = \arg_t r^*,\ t \in S_t$;
    (d.3) Calculate R_n' as in Eq. (2);
    (d.4) Set $b_{n,r} = b_{n,r} + 1,\ \forall r \in S_r^{t^*} \cap S_r^n$;
    (d.5) Calculate R_n'' as in Eq. (2);
    (d.6) Set $R_n = R_n - (R_n'' - R_n')$;
(6) while $S_t \neq \phi$,
  (a) Set $U' = U$;
  (b) for each $n \in U'$
    (b.1) Calculate R_n' and P_t' as in Eq. (2) and (4);
    (b.2) Set $b_{n,r}' = b_{n,r} + 1,\ \forall r \in S_r^t \cap S_r^n$;
    (b.3) Calculate R_n'' and P_t'' as in Eq. (2) and (4);
    (b.4) Calculate $\Delta P_n = P_t'' - P_t'$;
    (b.5) Calculate $\Delta \bar{D}_n$ corresponding to R_n' and R_n'';
    (b.6) if $(P_t'' > P_v)$ or $(R'' = R')$, $U' = U' - n$;
  (c) if $U' = \phi$, Set $S_t = S_t - t$ and GOTO Step (i);
  (d) Find $n^+ = \arg\max_n \Delta\bar{D}_n,\ \ \forall n \in U'$;
  (e) Set $\Delta\bar{D}_{max} = \Delta\bar{D}_{n^+}$ and $\Delta P_{min} = \Delta P_{n^+}$;
  (f) for each $n \in U'$
    (e.1) Calculate $C(\Delta\bar{D}_n, \Delta P_n)$ as in Eq. (6)
  (g) Find $n^* = \arg\min_n C(.),\ \forall n \in U'$;
  (h) Set $b_{n^*,r} = b_{n^*,r} + 1,\ \forall r \in S_r^t \cap S_r^n$;
  (i) if $t > |S_t|$, Set $t = 1$, else Set $t = t + 1$;

Fig. 4. Pseudo code for the MCS assignment phase.

frame and packet sizes are assumed to be fixed and of size 500 bits.

We compared the proposed RBA algorithm to a conventional static allocation algorithm. In the static algorithm, the number of RBs allocated to each user satisfies the rate proportionality where the number of RBs allocated to user n is $\lfloor w_n R / \sum_{i=1}^{N} w_i \rfloor$. The remaining $R - \sum_{n=1}^{N} \lfloor w_n R / \sum_{i=1}^{N} w_i \rfloor$ RBs are allocated to the user with the best channel condition where a user can get at most one unassigned RB. In each TTI, MCS assigned for RBs belonging to each user is incremented until there is no more power available or the user reaches its maximum supportable MCS.

We assume all users are divided into two groups; group A and group B where the arrival rates, λ_A, of users in group A are equal to each others and allowed to vary whereas the arrival rates, λ_B, of users in group B are always fixed and set to 150 bits per frame. Figs. 5 and 6 show the results for a unity channel gain while Figs. 7 and 8 show those results when the channel gain is variable and is the same for both algorithms.

Figure 5 show the overall average packet delay versus arrival rate, λ_A, for a unity channel gain when $N = 8$ and $N = 16$. When $N = 8$, RBA provides lower delay than the static algorithm when $\lambda_A > 2000$ bits/frame. When $N = 16$, delay of the static algorithm becomes very large (i.e. infinity) when $\lambda_A > 2050$ while the RBA sustains finite queueing delay. When the number of users increases to $N = 20$ and $N = 24$, as shown in Fig. 6 (a) and (b) respectively, both algorithms yeild the same average delay. This is mainly because with a channel gain of unity for all RBs, channel condition exploitation is not utilized by the RBA algorithm. In addition, increasing the number of users results in less resource for each user thus less delay distinction between both algorithms.

Figure 7(a) and (b) show the overall average packet delay versus arrival rate, λ_A, for a variable channel gain when $N = 8$ and $N = 16$ respectively. In a variable channel gain scenario, RBA provides significant lower average delay that the static algorithm. As shown in Fig. 7 for $N = 8$ and $N = 16$, RBA provides about 50 ms and 270 ms delay reduction when $\lambda_A = 2000$ and $\lambda_A = 1500$ bits/frame respectively. Not only RBA exploits channel condition but it also allocates RBs and assigns their MCS in a way to reduce the overall average delay. Fig. 8 shows also a significant delay reduction of about 180 ms and 250 ms at $\lambda_A = 1000$ bits/frame when the number of users increased to $N = 20$ and $N = 24$ respectively.

In Figs. 7 and 8, the delay for the static algorithm tends to increase more rapidly than the RBA algorithm (e.g., when $\lambda_A > 1500$ bits/frame in Fig. 7 (a)). The static algorithm does not consider queue arrival rates when assigning MCS to RBs while the RBA keeps adapting to varying arrival rate thus its delay tends to remain steady-still irrespective of the arrival rate. However, when the channel gain is variable, increasing the number of users increases the delay reduction obtained by the RBA algorithm because with a larger number of users, the resources allocated to each user becomes lesser and the delay reduction goal of the RBA becomes more significant.

V. Conclusion

In this paper, we have introduced a novel, practical, and efficient RB allocation and MCS assignment scheme for LTE wireless networks. An optimal problem is formulated and an algorithm that provides a solution to this problem is provided. The algorithm first allocates RBs to different users in proportional to their arrival rates, then, it performs MCS assignment for each RB in a way to minimize the overall average packet delay while taking into account queue dynamics, power limitation, channel condition, and MCS capability of each user. The proposed LA algorithm is shown to provide better lower delay than other algorithms. Although an M/D/1 queueing model and an average packet transfer delay were used in this paper, the algorithm can be extended to more general queueing models and other performance metrics.

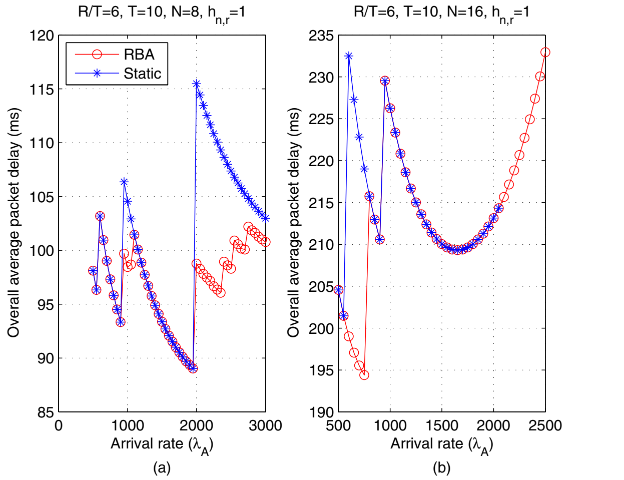

Fig. 5. Delay versus arrival rate for (a) $N = 8$ and (b) $N = 16$ when channel gain is fixed.

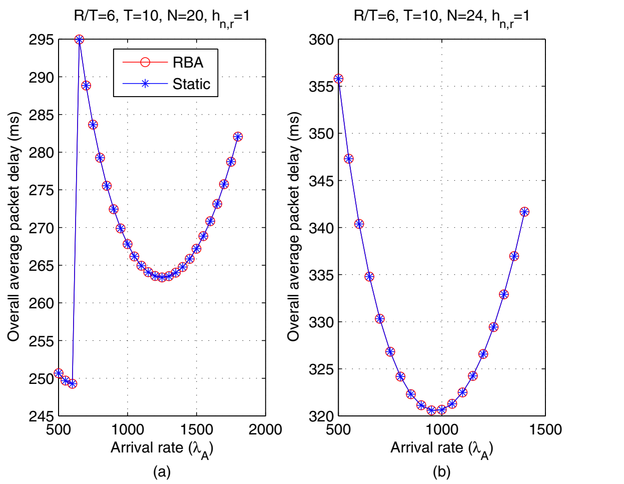

Fig. 6. Delay versus arrival rate for (a) $N = 20$ and (b) $N = 24$ when channel gain is fixed.

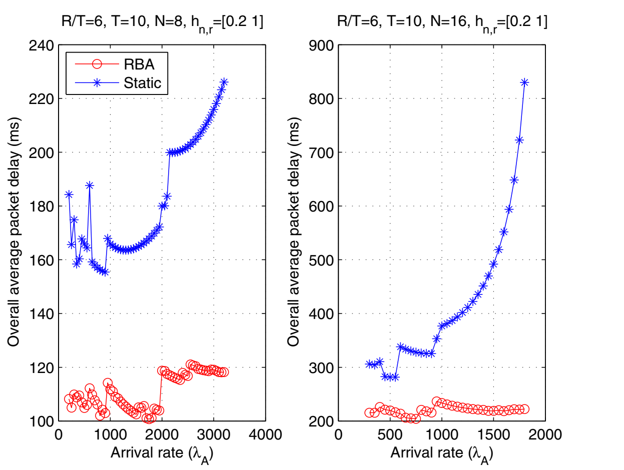

Fig. 7. Delay versus arrival rate for (a) $N = 8$ and (b) $N = 16$ when channel gain is variable.

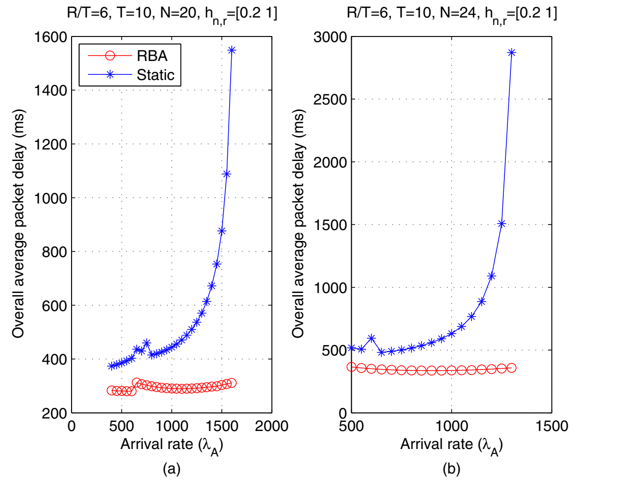

Fig. 8. Delay versus arrival rate for (a) $N = 20$ and (b) $N = 24$ when channel gain is variable.

REFERENCES

[1] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2*, 3GPP Std. TS 36.300 V8.7.0, December 2008.

[2] *Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA)*, 3GPP Std. TR 25.814 V7.1.0, September 2006.

[3] *Evolved Universal Terrestrial Radio Access (E-UTRA);Physical Channels and Modulation*, 3GPP Std. TS 36.211 V8.6.0, March 2009.

[4] *Evolved Universal Terrestrial Radio Access (E-UTRA);Medium Access Control (MAC) protocol specification*, 3GPP Std. TS 36.321 V8.5.0, March 2009.

[5] G. Yu, Z. Zhang, and P. Qiu, "Adaptive subcarrier and bit allocation in OFDMA systems supporting heterogeneous services," *Wireless Personal Communications: An International Journal*, vol. 43, no. 4, pp. 1057–1070, 2007.

[6] J. Shi and A. Hu, "Maximum utility-based resource allocation algorithm in the IEEE 802.16 OFDMA system," *IEEE International Conference on Communications (ICC)*, pp. 311–316, May 2008.

[7] D. Hui, V. Lau, and W. H. Lam, "Cross-layer design for OFDMA wireless systems with heterogeneous delay requirements," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2872–2880, August 2007.

[8] R. Van Nee and R. Prasad, *OFDM Wireless Communications*. Artech House, 2000.

[9] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, "Dynamic packet scheduling performance in UTRA long term evolution downlink," *International Symposium on Wireless Pervasive Computing, (ISWPC)*, pp. 308 – 313, May 2008.

[10] A. Pokhariyal, T. Kolding, and P. Mogensen, "Performance of downlink frequency domain packet scheduling for the UTRAN long term evolution," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1 – 5, September 2006.

[11] S. Choi, K. Jun, Y. Shin, S. Kang, and B. Choi, "MAC scheduling scheme for VoIP traffic service in 3G LTE," *IEEE Vehicular Technology Conference (VTC)*, pp. 1441 – 1445, September 2007.

[12] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, "Principle and performance of semi-persistent scheduling for VoIP in LTE system," *International Conference on Wireless Communications, Networking and Mobile Computing (WiCom)*, pp. 2861 – 2864, September 2007.

[13] J. Sanchez, D. Morales-Jimenez, G. Gomez, and J. Enbrambasaguas, "Physical layer performance of long term evolution cellular technology," *Mobile and Wireless Communications Summit*, pp. 1–5, July 2007.

[14] E. Brosh, S. A. Baset, D. Rubenstein, and H. Schulzrinne, "The delay-friendliness of TCP," in *SIGMETRICS '08: Proceedings of the 2008 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*. ACM, 2008, pp. 49–60.

[15] S. Ganguly, M. Chatterjee, and R. Izmailov, "Non-real-time content scheduling algorithms for wireless data networks," *IEEE Transactions On Computers*, vol. 55, no. 7, pp. 893–905, 2006.

[16] L. Kleinrock, *Queueing Systems*. John Wiley & Sons, Inc., 1975, vol. 1.