New Cross-layer QoS-based Scheduling Algorithm in LTE System

MOHAMED A. ABD EL-GAWAD
Network Planning Dep.
National Telecomm. Institute
Cairo, EGYPT
m.algawad@gmail.com

MOHSEN M. TATAWY
Network Planning Dep.
National Telecomm. Institute
Cairo, EGYPT
mohsen@nti.sci.eg

MOHAMED S. EL-MAHALLAWY
Comm. & Electronics Dep.
Arab Academy for Science, Technology & Maritime Transport
Cairo, EGYPT
mahallawy@aast.edu

Abstract: - Long Term Evolution (LTE) represents an emerging and promising technology for providing broadband ubiquitous internet access and integrating a wide variety of communication services such as high speed data, video and multimedia traffic as well as voice signals. One of the promising approaches to achieve that is the adaptive orthogonal frequency division multiplexing (AOFDM). In AOFDM there are a number of adaptive sub-carriers allocation techniques which lead to improve the system performance by assigning sub-channels to the users according to their channel status. In this paper we propose an algorithm takes in its allocation strategy not only the channel status but also quality of service (QoS) aspects like traffic priority and transmission delay. The proposed algorithm performance has been valued and compared in term of throughput with the previous allocation algorithms.

Key-Words: - Adaptive Allocation, Cross-layer Scheduling, LTE, OFDM, QoS, Throughput

1 Introduction

To face the ever-growing demand for packet-based mobile broadband systems, the Third-Generation Partnership Project (3GPP) has introduced LTE specifications as the next step of the current 3.5G cellular networks [1]. LTE is a leading OFDMA-based, mobile broadband technology, supported by a new core network being developed by 3GPP. LTE has been designed to provide interoperability and service continuity with existing Universal Mobile Telecommunication System (UMTS) networks, allowing UMTS operators to capitalize on existing UMTS/High Speed Packet Access (HSPA) and future HSPA+ investments. It offers high spectral efficiency, low latency and high data rates.

OFDM is a special form of multi-carrier transmission technique in which a single high data rate stream is divided into multiple low rate data streams. These data streams are then modulated using sub-carriers which are orthogonal to each other. In this way the symbol rate on each subchannel is greatly reduced, and hence the effect of inter-symbol interference (ISI) due to channel dispersion in time caused by multipath delay spread is reduced. Also guard intervals can be inserted between OFDM symbols to reduce ISI further. The orthogonality between sub-carriers can be maintained, even though the signal passes through time dispersive channel by cyclically extending the OFDM symbol into guard interval [2]-[5].

The main advantages of OFDM are its multipath delay spread tolerance and efficient spectral usage by allowing the overlap in the frequency domain. Another significant advantage is that the modulation and demodulation can be done using inverse Fast Fourier Transformation (IFFT) and fast Fourier Transformation (FFT) operations, which are computationally efficient [2], [3].

In OFDMA systems, the resources can be distinguished both in frequency and time dimension: different users can be assigned sub-carriers belonging to different frequencies or to different OFDM symbol times within the frame. The sub-channelization unit is the slot, which is a set of contiguous sub-carriers in the time-frequency domain. The available resources within the OFDMA frame are composed by a grid of slots; the sub-channel assigned to a certain user is composed by a set of slots [6], [7].

An adaptive sub-carriers allocation based on different channel conditions belonging to different users, it is possible to increase the system performance, by exploiting the multi-user diversity: the more attenuated frequencies for an user may result the better frequencies for another one. In order to obtain the best performance, adaptive sub-
carrier allocation aims to assign each user the less attenuated frequencies of that user’s channel response. The algorithms which demonstrated in this paper are based on the estimation of the user channel capacity by the assignment of a certain slot. Hence, the slots can be assigned to the users so as to reach a certain channel capacity distribution among the users.

The channel capacity of a carrier in the OFDMA system can be expressed as:

\[ C_p = B_p \log_2 \left( 1 + \frac{E}{I_{N_0,B_p}} \right) = B_p \log_2 \left( 1 + \frac{E}{N_0} \right), \]

(1)

Where \( B_p \) is the sub-carrier bandwidth (corresponding to the total system bandwidth divided by the number of sub-carriers), \( E \) is the symbol energy, \( T_s \) is the symbol time and \( N_0 \) is the power spectral density of AWGN channel [6], [8].

This paper is organised as follow: the system model is described in section 2. The adaptive allocation algorithms are described in section 3. The simulation results are demonstrated in section 4. Finally the conclusion is made in section 5.

2 System Model

An LTE downlink time division duplexing (TDD) OFDM system is considered, where the base station can acquire the channel state information (CSI) through the uplink dedicated pilots from all mobile stations at the beginning of each time slot. With cross-layer design the QoS information can be obtained by the traffic controller at the media access control (MAC) layer. The sub-carrier controller at the physical layer uses the CSI and the QoS information which is transferred from the traffic controller to do the sub-carriers allocation process and the results are fed back to the traffic controller to do the date scheduling process [9].

Let us consider a system with \( K \) users and an OFDMA frame of \( A \) slots along the frequency dimension. Each user has two types of traffic, one of them represents real time traffic and the other represents non-real time traffic, each type has different priority and packet delay budget. Furthermore, let’s consider:

- \( k \): user index, with \( k = 0, \ldots, K - 1 \);
- \( i \): slot index along frequency domain, with \( i = 0, \ldots, A - 1 \);
- \( \alpha_{i,k} \): Rayliegh distributed multipath attenuation coefficient within the slots of frequency index \( i \), for the \( k \)-th user;
- \( SNR_{i,k} = \frac{P}{N_0} \alpha_{i,k}^2 \): effective signal to noise ratio within the slots of frequency index \( i \), for the \( k \)-th user;
- \( j \): traffic type index;
- \( lc_j \): packet delay budget for each traffic type;
- \( lm_{k,j} \): time of delivered data for each user over each type of traffic, normal time distribution is assumed;
- \( p_j \): priority value for each type of traffic;

Ideal channel estimation is assumed so the allocation algorithm knows the exact value of \( \alpha_{i,k} \) and \( \frac{P}{N_0} \) [6].

The achieved capacity for user \( k \)-th by the assignment of a slot \( i \) is:

\[ C_{slot_{i,k}} = B_p \log_2 (1 + SNR_{i,k}). \]

The maximum capacity for the \( k \)-th user is the total capacity if all slots are assigned to him:

\[ C_{max,k} = \sum_{i=0}^{A-1} C_{slot_{i,k}}. \]

The goal of an allocation algorithm is determining which slots should be allocated to each user based on the allocation criteria, so the allocation can be represented with a matrix \( X \) whose dimensions are \( K \) rows per \( A \) columns, that holds in each position \( X_{i,k} \) the index of the user whom the slot \( i \) is assigned.

3 Adaptive Allocation Algorithms

3.1 Fair Allocation

The fair allocation algorithm is derived by the work in [8] and approximates a fair capacity distribution among the users (\( K \)), where each available slot is allocated to the user whose capacity is the minimum one among the capacities of the others. In other words the allocation (\( \hat{X} \)) is done to maximize the minimum value among all the capacities \( C_k \).

\[ \hat{X} = \arg\max_X \left( \min_k C_k \right), \]

(4)

A suboptimal solution is searched through an iterative algorithm. Let’s define:
• $S(i)$: the slot of frequency index $i$;
• $S$: the set of free slots;
• $S_k$: the set of slots assigned to the $k$-th user;
• $R_k$: the capacity assigned to the $k$-th user.

The fair algorithm is defined as follow:

1) **Initialization:**
   a) $R_k \leftarrow 0$ for $k = 0, ..., K - 1$;
   b) $S_k \leftarrow \emptyset$ for $k = 0, ..., K - 1$;
   c) $S \leftarrow \{s(i)\}$ where $i = 0, ..., A - 1$;

2) for $k = 0, ..., K - 1$:
   a) find a slot $s(i) \in S$ so that $SNR_{ik} \geq SNR_{n,k}$ for each frequency index $n$ for which at least a free slot $S$ exist;
   b) $S_k \leftarrow S_k \cup \{S(i)\}$;
   c) $S \leftarrow S - \{S(i)\}$;
   d) $R_k \leftarrow R_k + Cslot_{i,k}$;

3) **While** $S \neq \emptyset$:
   a) find the user $k$ so that $R_k \leq R_u$ for each user $u$;
   b) find a slot $s(i) \in S$ so that $SNR_{ik} \geq SNR_{n,k}$ for each frequency index $n$ for which at least a free slot $S$ exist;
   c) $S_k \leftarrow S_k \cup \{S(i)\}$;
   d) $S \leftarrow S - \{S(i)\}$;
   e) $R_k \leftarrow R_k + Cslot_{i,k}$.

In this allocation algorithm, the selected user is the one who has the minimum ratio between the total capacities allocated to that user and the maximum obtainable capacity [4].

### 3.2 Proportional Allocation

In the fair allocation algorithm, the users with the best channel conditions obtain a lower number of resources relative to the users with the worst channel conditions, who need a lot of slots to achieve the same amount of capacity; so the channel capacity is not fully utilized. Proportional allocation [6], [10] solves this issue by making the allocated capacities respect to the maximum capacity value for each user as in relation (3).

$$C_0 : C_1 : ... : C_{K-1} = C_{\text{max,0}} : C_{\text{max,1}} : ... : C_{\text{max,K-1}}, \quad (5)$$

To achieve the proportional allocation, the algorithm has to be modified as follow:

1) **Initialization:**
   a) Compute $C_{\text{max,k}}$ for $k = 0, ..., K - 1$ accordingly to (5);
   b) $R_k \leftarrow 0$ for $k = 0, ..., K - 1$;
   c) $S_k \leftarrow \emptyset$ for $k = 0, ..., K - 1$;
   d) $S \leftarrow \{s(i)\}$ where $i = 0, ..., A - 1$;

2) for $k = 0, ..., K - 1$:
   a) find a slot $s(i) \in S$ so that $SNR_{ik} \geq SNR_{n,k}$ for each frequency index $n$ for which at least a free slot $S$ exist;
   b) $S_k \leftarrow S_k \cup \{S(i)\}$;
   c) $S \leftarrow S - \{S(i)\}$;
   d) $R_k \leftarrow R_k + Cslot_{i,k}$.

### 3.3 The QoS-based Algorithm

In all the previous discussed algorithms, the sub-carrier controller at the physical layer uses the CSI to distribute the resources among the users and the main difference among them is the distribution criteria. The enhancement which is added to the proposed algorithm not only the distribution criteria but also the information which is used in the distribution process. In this algorithm the sub-carrier controller combines the CSI and the QoS information which is transferred from the traffic controller in the MAC layer using the defined weight in (8) to distribute the resources among the system users.

The user selection in this cross-layer algorithm is based on the lowest weight which represents the lowest capacities which is assigned before, the lowest time to serve the user traffic $(t_{ij} - l_{mk})$ and the highest traffic priority $P_j$ which depends on the type of the user traffic. By the same way like the previous algorithm, the user who is selected takes the slot which achieves the highest capacity to that user.

$$Weight_{kj} = C_{g,k,i} \frac{(t_{ij} - l_{mk})}{P_j}, \quad (8)$$

So the algorithm has to be modified as follow:

1) **Initialization:**
   a) $R_k \leftarrow 0$ for $k = 0, ..., K - 1$;
   b) $S_k \leftarrow \emptyset$ for $k = 0, ..., K - 1$;
   c) $S \leftarrow \{s(i)\}$ where $i = 0, ..., A - 1$;

2) for $k = 0, ..., K - 1$:
a) find a slot $s(i) \in S$ so that $\text{SNR}_1 \geq \text{SNR}_n$ for each frequency index $n$ for which at least a free slot $S$ exist;

b) $S_k \leftarrow S_k \cup \{s(i)\};$

c) $S \leftarrow S - \{s(i)\};$

d) $R_k \leftarrow R_k + C_{\text{slot}_{i,k}};\$

c) While $S \neq \emptyset$:

a) find the user $k$ so that weight$_{kj} \leq$ weight$_{uj}$ for each user $u$ over each traffic index $j$;

b) find a slot $s(i) \in S$ so that $\text{SNR}_{1,k} \geq \text{SNR}_{n,k}$ for each frequency index $n$ for which at least a free slot $S$ exist;

c) $S_k \leftarrow S_k \cup \{s(i)\};$

d) $S \leftarrow S - \{s(i)\};$

e) $R_k \leftarrow R_k + C_{\text{slot}_{i,k}}.$

4 Simulation Results

The simulation results are demonstrating the performance of the proposed cross-layer algorithm for an LTE system with a total bandwidth 10 MHZ which is divided into $N=601$ sub-carriers. The transmitted energy on each sub-carrier is assumed to be constant. Also we assume that the simulation environment has 10 users, each user has two traffic queues, one for conversational voice traffic and the other for TCP-based traffic (e.g., WWW, e-mail). We have two priority values and two packet delay budget values to each type of traffic; (priority=1 & delay budget=300ms) for the TCP-based traffic and (priority=2 & delay budget=100ms) for the conversational voice traffic [12]. We assume the remaining time for each user to finish his data is random distributed under the value of the packet delay budget for each type of traffic.

Figure 1 demonstrates the system throughput versus the SNR values for all adaptive allocation algorithms; it's found that the QoS-based algorithm has 1% in throughput less than the other discussed algorithms.

Figure 2 shows the impact of the number of users on system throughput for the QoS-based algorithm and the other discussed algorithms, with SNR= 20 dB. The advantage of the proposed algorithm over the other discussed algorithms is the throughput differentiation between LTE traffic applications as shown in figure 3, in all allocation strategies discussed before there is no differentiation between the two applications. The QoS-based algorithm differentiates between the real time application voice or video and the non-real time application to provide more throughputs to real time applications.
5 Conclusion
In this paper we propose a cross-layer adaptive allocation algorithm for LTE system based on the estimation of the channel capacity with traffic handling procedure to enhance the system performance and guarantee the required integration between different types of communication services. The QoS-based algorithm has approximately 1% reduction in system throughput comparable with the other discussed algorithms so the proposed algorithm slightly reduce the system throughput but on the same time, it has the best performance with LTE real time applications like the conversational voice or video.

References: