

Toward 5G High Utilizations: A survey on OFDMA-based Resource Allocation Techniques in Next-Generation Broadband Wireless Access Networks

Ibrahim Al-Surmi^{1,*}, Ali Mohammed Mansoor², Aqeel Abdullah Ahmed³

¹Faculty of Engineering and IT, Yemen and Gulf University for Science & Technology, Sana'a, Republic of Yemen.

²Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia.

³Faculty of Multimedia and Creative Technology, International University of Technology Twintech, Sana'a, Republic of Yemen.

Abstract

LTE is marketed as 4G standard for wireless communication services. LTE network technologies provide a high speed triple play services like data, voice, and video with QoS provision. The resource allocation techniques primary concerns are the QoS support that aims to satisfy the diverse service requirements as well as to guarantee the required data rates from the available resources. In this context different network architecture for broadband wireless data network like LTE made OFDMA a practical access technology for wireless multimedia services. Therefore, this survey investigates the techniques that aim to support QoS for real-time and non real-time services in OFDMA-based packet scheduling and adaptive multiuser frequency-time domain resource allocation. Moreover, this survey addressed the different aspects of resource allocation algorithms design to be contributed to the emerging 5G next generation networks.

Received on 05 September 2020; accepted on 23 December 2020; published on 10 February 2021

Keywords: Wireless communication, Packet scheduling, Resource allocation, LTE, QoS

Copyright © 2021 I. Al-Surmi *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.10-2-2021.168713

1. Introduction

Among the versatile scenarios supported by future 5th Generation mobile communication (5G), high mobility is one of the most important and extreme use cases, which attracts much interests [1, 3]. The IEEE 802.16 is one of the latest wireless networks of IEEE standards technologies [1]. IEEE 802.16 adopted the Long-Term Evolution (LTE). It is presented to compete with Broadband Wireless Access (BWA) such as Universal Mobile Telecommunications System (UMTS), Code Division Multiple Access (CDMA) and High-Speed Uplink/Downlink Packet Access (HSU/DPA).

The standard of IEEE 802.16 defines Medium Access Control (MAC) and Physical (PHY) layers of the air interface for a Broadband Wireless Access (BWA). The Physical layer radio interfaces have been defined under Line-of-Sight (LOS) in an early release of the IEEE 802.16 standard in 2001 and deployed in 10-66 GHz frequency band. Moreover, the advancement has been issued and documented in 2004 named as the IEEE 802.16d [2]. It is capable to operate under Non-Line-of Sight (NLOS) with spectrum ranging from 2-11 GHz. Additionally, further enhancements in 2006, rebelled as IEEE 802.16e [3]. Furthermore, the recent version has been introduced in May 2009 called 802.16m [4], to comply with the IMT-advanced next generation mobile networks standard and this version merges the earlier versions to provide advance air interface.

In addition, Figure 1 shows the access protocol as identified in IEEE 802.16 profiles. The MAC layer

*Corresponding author. I. Al-Surmi Email: iaa_a@yahoo.com
A. M. Mansoor Email: ali.mansoor@um.edu.my
Aqeel A. A. Email: aqeel_je@iutt.edu.ye

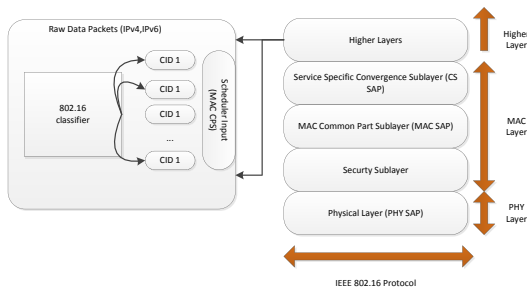


Figure 1. IEEE 802.16 Layer Protocols

consists of three types of sub-layers first, service Convergence Sub-layer (CS), second, MAC Common Part Sub-layer (MAC-CPS) and the Security Sub-layer. The CS sub-layer is the central interface for the higher layer, which is capable of interacting with transport layer by the mapping functions. Moreover, an adaptive CS can be provided for each protocol involving the Asynchronous Transfer Mode (ATM) and Internet Protocol version 4 (IPv4) services.

The MAC-CPS involves in several roles such as QoS administration, system access control, bandwidth allocation service, and system maintenance. The security sub-layer within a shared wireless networks transmits the connection in protected process through supervision of the authentication, provision of secure key exchange, and protects the data with encryption policies.

Moreover, the standard has defined two operation modes to cope with the wireless broadband solutions, referred as Point-to-Multipoint (PMP) mode as well as mesh mode, as showed in Figure 2. This paper considers the PMP mode with frame-base's precedence of DL sub-frame over the UL sub-frame for the Downlink (DL) and Uplink (UL) transmission. In this mode, BS is responsible to transmit the service to multiple Subscriber Stations (SSs) within its range as a one-hop structure. Moreover, the communication between SSs is direct with the BS, and not among them. Furthermore, the data transmission route from BS to SS is known as DL direction, whereas the route from the SS to the BS is known as UL direction. The BS plays important roles which operate as centralized data transmissions to the active connected SS and present as a gateway to the peripheral networks.

On the other hand, OFDMA make its effort for users to operate with smaller power amplifiers. OFDMA has been adopted for various standards such as IEEE 802.11n [5] and 4G generation cellular systems such as IEEE 802.16m and 3GPP Long Term Evolution (LTE) [6]. Therefore, there is an open issues relate to

how towards fulfil the QoS requirement and service differentiation in IEEE 802.16 networks.

The bandwidth allocation algorithms are the key mechanisms in LTE that are used to support the required QoS. The mechanisms must be designed to maximize efficient utilization of spectrum and of systems resources more effectively. Therefore, scheduling is the problem of determining bandwidth priority that will be given to the users.

This survey focuses on bandwidth scheduling algorithms for QoS provisioning in LTE networks especially in the MAC and Physical layers. In order to answer the need of QoS provisioning, this paper describes categorization of LTE QoS provisioning procedures that operate on physical and MAC layers.

2. Scheduling Algorithms Categorizations in LTE

LTE is the most commonly used for the IEEE 802.16 implementation. It is one of the candidatures applicable emerging technologies for BWA because of its compatibility with 4G-all-IP wireless networks as well as its cost effectiveness. LTE is a viable alternative to the cable modem and DSL technologies due to its high resource utilization, easy implementation and low cost. In addition, LTE attractive for areas without infrastructure as well as it offers higher data rate over a metropolitan area up to 70 km with a variety of QoS requirements. Therefore, efficient scheduling algorithm is particularly very important of their capability to preserve bandwidth and improve system usage throughput, besides QoS provisioning for diverse user requirements [4].

A comprehensive survey of various LTE scheduling schemes for both point-to-multipoint and multi-hop relay systems performed in [7]. The authors discussed researches on the main important aspects component of LTE MAC layer that helps to provide QoS

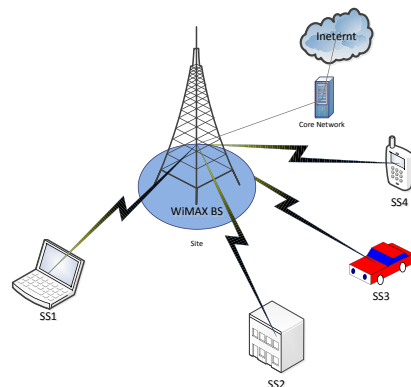


Figure 2. PMP Operation Mode

guarantee to various service classes in order to assist network designers, researchers, and Internet service providers in designing effective practical schedulers for single-hop or multi-hop relay systems. In this section, we discuss the bandwidth algorithms in both two layer LTE MAC and PHY layers in both DL and UL directions. Figure ?? provides more explanations about the scheduling categorizations. Next subsections elaborate in more detail about their mechanisms and the related algorithms. This figure also identified in colour the related significant classifications of this paper.

2.1. Scheduling Service Types in LTE

The priorities of traffic is handled in schedulers after the classifications are performed to differentiate the active connections, and the entire admitted connections are attached with a specific QoS requirements to specific service class queues. Moreover, these scheduling services particularly identify the bandwidth required to allocate both UL and DL traffic. Explicitly, the IEEE 802.16 standard stipulates the following scheduling service classes as depicted in Table 1. There are two major types of bearers:

Guaranteed Bit Rate (GBR): the GBR bearers are used for real-time services, such as conversational voice and video. A GBR bearer has a smallest amount of bandwidth that is reserved by the network, and always consumes resources in a radio base station regardless of whether it is used or not. GBR bearers, if implemented properly, should not experience packet loss due to congestion on the radio link or the IP network. Moreover, GBR bearers will be defined with the lower latency and jitter tolerances that are typically required by real-time services.

non-guaranteed bit rate(Non-GBR): the Non-GBR bearers do not have specific network bandwidth allocation. The Non-GBR bearers are intended for best-effort services, such as file downloads, email, and Internet browsing. These bearers will experience packet loss when a network is congested. A maximum bit rate for non-GBR bearers is not specified on a per-bearer basis. Nevertheless, An Aggregate Maximum Bit Rate (AMBR) will be specified on a per-subscriber basis for all non-GBR bearers.

2.2. LTE MAC layer QoS scheduling algorithms

LTE MAC layer is connection-oriented, in which the end points user uses an initial protocol to establish an end-to-end connection before any data is sent as well as an active user can undergo unidirectional communication. This affords a structure for requesting bandwidth, linking QoS parameter and flow information status, transmitting data and facilitating other procedures related to service terms establishment.

On the other hand, service type connections are individually differentiated by a Connection Identifier (CID). The information contained in CID is attached to each MAC Protocol Data Unit (PDU) within the MAC header. In addition, the data exchanged amongst the BS and its SSs over their MAC layer is the PDU responsibility. The BS gets the information of their connections and the SSs queues status by their bandwidth request messages. While the BS grants an aggregate uplink bandwidth to individual SS that sequentially redistribute among its connections.

Moreover, the process of transmitting and receiving might be performed in both Frequency Division Duplex (FDD) and Time Division Duplex (TDD) methods, which used separately by the DL and UL sub-frames. In addition, both sub-frames that use FDD take place on separate frequency bands, and may be synchronized in time. The SSs can either transmit and receive simultaneously, or transmit and receive at non-overlapping time intervals as full-duplex and half-duplex approaches, respectively.

Moreover, both sub-frames that uses TDD utilize the same frequency, however their transmissions happen at different times. On the other hand, Time Division Multiple Access (TDMA) permits the SSs to share a wireless channel in the UL sub-frame. Nevertheless, a Downlink MAP (DL-MAP) and Uplink MAP (UL-MAP) information are broadcasting by the DL sub-frame, and the data is transmitted in bursts using TDMA.

Furthermore, the core function of QoS provisioning is the service flow which is a set of packets that are identified by their QoS parameters, such as traffic priority, tolerated jitter, minimum reserved traffic rate, maximum latency, etc. These QoS parameters are associated with each queue and vary in respect with the corresponding service flows to guarantee a certain QoS in the system.

The scheduling algorithms in the MAC layer guarantee the QoS for different multimedia service classes. Moreover, these algorithms are employed to control the traffic or connection that will be served before others within the same queue. The priority must adhere to the procedures dictated by the scheduling algorithms that fulfil with QoS requirements. Certainly, the scheduler plays as a supplier to the connected SSs with the required resources.

In addition, the main scheduling algorithms' concerns are utilizing the full bandwidth, boosting fairness among SSs, and maintaining the required QoS in the whole network. These scheduling algorithms are performed both in the BS and SSs sides.

On the other hand, this paper discuss scheduling algorithms that support QoS MAC layer in order to assure efficient scheduling algorithms, specific aspects must be considered and evaluated in the architecture of the scheduling policy. Therefore, designing an

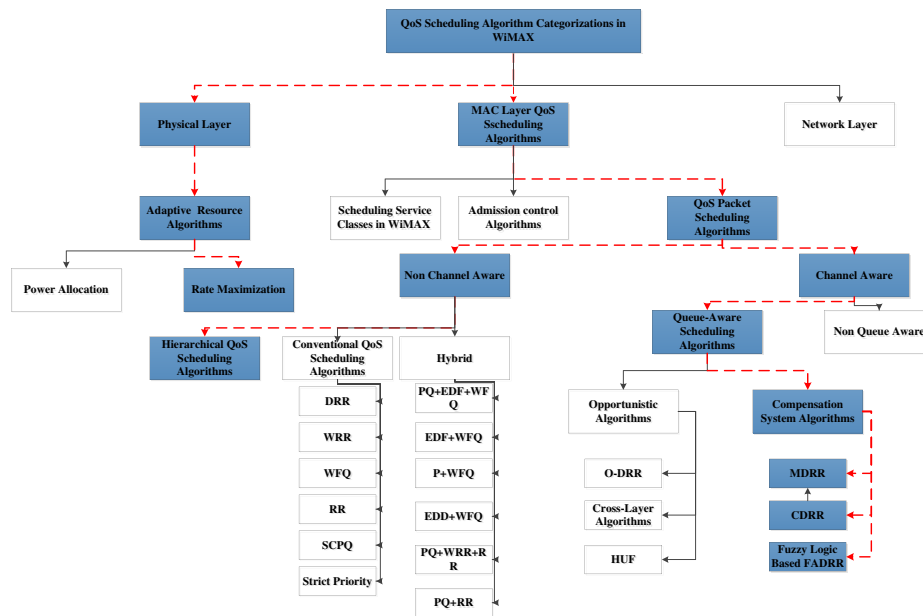


Figure 3. QoS Scheduling Algorithms Categorizations in LTE

Table 1. Scheduling Service Types in LTE

QCI	Resource Type	Priority	Packet Delay Budget (NOTE 1)	Packet Error Loss Rate (NOTE 2)	Example Services
1 (NOTE 3)	GBR	2	100 ms	10^{-2}	Conversational Voice
2 (NOTE 3)	GBR	4	150 ms	10^{-3}	Conversational Video (Live Streaming)
3 (NOTE 3)	GBR	3	50 ms	10^{-3}	Real Time Gaming
4 (NOTE 3)	GBR	5	300 ms	10^{-6}	Non-Conversational Video (Buffered Streaming)
5 (NOTE 3)	Non-GBR	1	100 ms	10^{-6}	IMS Signalling
6 (NOTE 4)	Non-GBR	6	300 ms	10^{-6}	Video (Buffered Streaming), TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7 (NOTE 3)	Non-GBR	7	100 ms	10^{-3}	Voice, Video (Live Streaming), Interactive Gaming
8 (NOTE 5)	Non-GBR	8	300 ms	10^{-6}	Video (Buffered Streaming), TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9 (NOTE 6)	Non-GBR	9	300 ms	10^{-6}	Video (Buffered Streaming), TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)

efficient scheduling algorithm must consider the QoS parameters of different service classes as shown in Figure ??.

Fairness: In addition to supporting QoS requirements for each user, the allocation of available bandwidth should be fairly distributed among the active

users. Therefore, boosting fairness is an important target that poses challenges to the design of scheduling algorithm.

Channel Utilization: This represents the time slots that are utilized to transmit the requested packets. Moreover, the design of scheduler must be able to steer clear of resources wastage which consequences from

allocating excessive resources for the connection that do not have plenty data to send.

Complexity: The scheduling algorithm must not include complex portions that will increase the complexity when dealing with different types of services constraints because it designed to be implemented in a simple manner.

Scalability: It defined as the capability to cope with increased number of SSs in the network. Hence, the scheduling algorithm should be constant even as the number of SSs increases.

The optimal selection of a scheduling algorithm relies on a combination of architecture mentioned above and its ease of implementation. Moreover, scheduling algorithms should handle multiple connections in order to guarantee the assigned throughputs, delay limits and loss rates.

In general, the main responsibility of a scheduling algorithm is to allocate the next service. Such an allocation procedure mainly depends on the QoS requirements of each service class. Therefore, designing scheduler algorithm at the BS would be preferable in order to guarantee superior performance. Moreover, current scheduling structures need to be modified that utilize QoS parameters to cope with the particular characteristics, for example the type of service class and the LTE frame structure.

The QoS scheduling algorithms are broadly classified as either a channel-aware scheme or non channel-aware scheme [8]. The former scheme concerned with the information change in channel status and uses this information in the scheduler, while the latter scheme ignores the channel status information. This paper revises scheduling algorithms and discusses the state-of-the-art which has recently been proposed for LTE's QoS provisioning.

Non Channel-Aware Scheduling Algorithms. Non Channel-Aware scheduling algorithm keeps away from the use of channel state condition like power allocation, channel fault, and loss of data rates. These scheduling algorithms are capable to guarantee the QoS requirements for various service classes, in essence the limitations of the delay and throughput. In addition, jitter is one of the QoS parameters, until now the proposed algorithms cannot fully guarantee jitter limits. In [9] the research compared different approaches for LTE network. Moreover, comparison of the scheduling algorithms and the QoS disciplines are classified and discussed as follows.

A. Conventional QoS Scheduling Algorithms

The conventional scheduling algorithms are concerned on a classical structure. This structure is simple and flexible to fulfil with both hybrid

and hierarchical structures. Constantly, these algorithms intended initially for wired networks, but are later implemented to manage the QoS requirements by majority of broadband services like LTE. Table 2 shows a comparison of well-known conventional QoS scheduling algorithm.

1. Round Robin (RR)

Round Robin (RR) algorithm is one of the simplest conventional scheduling algorithm which equally allocates the resources in sequence to all active users [10]. Nevertheless, the algorithm must make sure adequate resources either for specified packet or bytes towards ensuring fairness among active users. Hence, a huge packet size will take on the entire bandwidth which violates the system fairness.

Moreover, RR algorithm might be reserve resources for a connection with no traffic to transmit. So, an adjustment is required to avoid idle users and allocate for active users. However, in such algorithms more considerations are given to compute the data rate levels at a certain period, and provide overall fairness between the entire participating connections [11].

2. Weighted Round Robin (WRR)

The Weighted Round Robin (WRR) [12] was developed to differentiate traffic in different queues in order to allow several streams to get served. The WRR algorithm works similar to RR algorithm except it gives a weight to each queue. This weight is proportional to overall available shared bandwidth used in the system. Consequently, the amount of packets that must be de-queued fluctuates based on the weight given to these queues. Hence, the mechanism of diversity of weights enforces the priority between active queues, and consequently the connected users. Yet, the drawback of WRR algorithm is unfair allocation of data rates for those connections with low priority queues.

3. Deficit Round Robin (DRR)

The Deficit Round Robin (DRR) algorithm mitigates the unfairness associated with of RR and WRR algorithms [13]. Moreover, it inherits the low complexity mechanism of $\mathcal{O}(1)$ designed for RR algorithm. The DRR algorithm is associated with a Deficit Counter (DC) which is originally established to distinguish the quantum of every

queue in the system. This quantum is a number of bytes assigned to be used by each queue at any time it allowed to be de-queued.

Moreover, the associated DC is periodically increased by one quantum on each round at any time the scheduler performs the de-queuing process, unless the queue is empty. Furthermore, these packets are de-queued if the total quantum added to the residual of previous deficit counter is larger than packet size. Otherwise, this individual queue remains idle until the next scheduler visit.

On the other hand, when these packets are entirely de-queued, at that time any residual values in the DC will be set to zero, hence, this process will lead to unfairness when that value is neglected and left without utilization.

Cicconetti, et al. [14] adopt the latency awareness of these conventional algorithms in order to cope with characteristics of IEEE 802.16 schedulers. This decision is motivated by the fact that both rtPS and nrtPS scheduling services demand a basic QoS requirement such as minimum reserved traffic rate. For instance, for real-time services like VoIP, DRR algorithm remains the most appropriate selection amongst conventional scheduling algorithms.

4. Weighted Fair Queuing (WFQ)

Weighted Fair Queuing (WFQ) is designed in [15] for a packet estimation of the Generalized Processor Sharing (GPS) algorithm. The GPS algorithm schedules the packet individually after splitting it into bits. In this process the WFQ algorithm achieves a superior performance in comparison to the WRR algorithm when handling packets with variable size. The limitation of the WFQ algorithm happens when ignoring packet starting time and it still de-queues the packets that have not begun service within the GPS algorithm. Modifications of WFQ has been proposed in several literatures, one of the most effective scheduling algorithm is Worst-case Fair Weighted Fair Queuing (WF2Q) described in [16]. This algorithm retains the delay constraints and accomplishes worst-case fairness.

Moreover, the WF2Q utilizes the same characteristics of WFQ through virtual time concept. In addition, the WF2Q selects packets with the smallest ending time in the Head-of-Line (HoL) instead of the lowest virtual finishing time of all packets in the queue.

5. Self-Clocked Fair Queuing (SCFQ)

Self-Clocked Fair Queuing (SCFQ) is a fair queuing algorithm designed for LTE services as proposed by [17]. SCFQ expresses fairness in an independent manner. It adopts virtual time as the key of allocation opportunity. Hence, organizing packets in each queue take place upon the finishing time, in which the higher priority is given to the packets with smaller completion time to be transmitted in a First-in-First-out (FIFO) mode. This system is superior in providing fairness between diverse traffic streams. Nevertheless, the complexity characteristic of the virtual time method results in a challenge to be applied for LTE services. For instance, the virtual clock acts as reference for all traffic streams; hence it cannot be reset until all the streams are idle which rarely happens and can lead to numerical overflow problem. Another drawback is that the selected packets must be sorted in the queue, which inevitably introduces insignificant delay due to the swapping process.

6. Strict-Priority (SP)

The selection order as mentioned in Strict-Priority algorithm [18], is based on the priority of weight order. First, the packets are categorized based on the QoS classes and then allocated into different priority queues. In addition, the algorithm services the highest priority queue until it is empty, and it moves to the next highest priority queue.

Thus, this SP algorithm may not be effective in LTE network due to no compensation for inadequate bandwidth. Moreover, this algorithm is only suitable for low bandwidth serial lines that does not automatically adapt to changing network requirements.

After all, the algorithm process may result in bandwidth starvation for the low priority QoS classes when the packets may not get forwarded and no guarantee is offered to even one flow.

B. Hybrid Schemes

The hybrid scheme can be applied to overcome the drawbacks related with conventional algorithm. The main purpose with hybrid scheme is to select suitable scheduling algorithm that prioritize the flows by assign bandwidth adequate to diverse connections. This type of scheduling is accomplished by implementing various algorithms. Moreover, combined simple algorithms

Table 2. Conventional QoS Scheduling Algorithms

Scheduling Algorithm	Contributions	Drawbacks
RR	Very simple	Unfairness and cannot guarantee QoS requirements
WRR	Simple and can support throughput requirements	Unfairness
DRR	Simple, support variable packet size, and have lower complexity	Not fair on a short time scale
SP	Simple and can meet the delay guarantee	Lower throughput, violates low traffic type
WFQ	Weight mechanism guarantee throughput, delay, and fairness	Complex
SCFQ	Adopts virtual time that guarantee throughput and delay , providing fairness	Complex

such as, DRR or WRR will result in superior performance. Thus, as proposed in [19] hybrid scheduling algorithms are recommended for UL direction for diverse LTE service type.

Tarchi, et al. [20] proposes a hybrid structure algorithm for UL direction, where UGS is scheduled by Packet Based Round Robin (PBRR) that is suitable for stable allocation of bandwidth. Moreover, the Earlier Deadline First (EDF) is proposed for the rtPS service type that is suitable for real-time delay sensitive services. On contrary, the WFQ algorithm is proposed for non real-time services like nrtPS and BE that is capable of reserving suitable weights for classes' service.

Moreover, Esmailpour, et al. [21] suggested a dynamic QoS-based bandwidth allocation (DQBA) in order to support heterogeneous traffic in LTE networks along with diverse QoS requirements. This scheme aims to maximize the system capacity based on traffic characteristics and service demand by dynamically adjusts bandwidth allocation for ongoing and new arrival connection.

Other researchers, Vinay, et al. [22], presented an enhancement method by combination between EDF and WFQ in order to serve real-time application. In addition, Wongthavarawat, et al. [23] proposed a constant allocation for real-time services using EDF, that used WFQ and equal sharing to schedule connections of diverse classes' service. In fact, the hybrid algorithms lead to superior complexity either by their structure implementation or huge resource consumptions.

objective is to provide particular needs to several service classes. The bandwidth is allocated in a certain way at the first hierarchy level to associated service classes and usually inherits conventional schemes to schedule different connections within each service class.

Wongthavarawat, et al [24], were the earliest work on hierarchical algorithm has been presented for resource assignments in LTE systems called Uplink Packet Scheduling (UPS). Figure 4 shown this UPS, in the first-tier, the overall available bandwidth is allocated in a strict priority fashion. The highest priority is dedicated to UGS service class, followed by rtPS, nrtPS and BE, respectively. However, fairness among these service classes is not guaranteed, mainly when massive packet size is being served.

In the second-tier, several methods are applied to handle QoS parameters for each service class. This algorithm allocates a fixed number of packets to the connections belonging to UGS service class. The rtPS connections are scheduled using EDF that allocate the bandwidth first for packets with earliest deadline. The nrtPS service class is

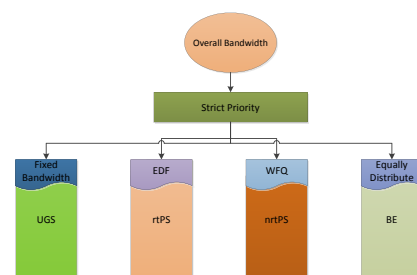


Figure 4. Hierarchical Structure

C. Hierarchical Structures

The hierarchical schedulers are composite of scheduling techniques in diverse levels. The

allocated the required resources based on WFQ. The remaining bandwidth is assigned equally for every other connections belongs to the BE service class.

A number of bandwidth allocation schemes have been proposed to tackle the drawbacks of class services in IEEE 802.16 [25, 26]. Moreover, a survey of certain hierarchical algorithms for mobile LTE is discussed by B. Li, et al. [27].

Chen, et al. [28] proposed a two-tier hierarchical scheduling. The former tier, Deficit Fair Priority Queuing (DFPQ) assigns the total available bandwidth for DL and UL services. Moreover, DFPQ aims to ensuring fairness to different service class and to improve total system throughput. Furthermore, DFPQ dynamically modify DL and UL bandwidth in order to improve system utilization.

Safa, et al. [29] presented an enhancement of DFPQ algorithm in order to improve the performance of real-time services. This is done by considering a deadline constraint. Nevertheless, the DFPQ cannot guarantee QoS requirements for real-time services such as VoIP and video streaming. Even though this can minimise the latency of real-time service, but this also degrades the opportunity for non real-time services such as nrtPS and BE.

Msadaa, et al. [30] proposed a new QoS scheduling architecture based on DFPQ. However this method gives more bandwidth to rtPS at the expenses of Best Effort (BE) traffic.

Shang, et al. [31], presented a hierarchical model for UL packet scheduling for LTE system. This algorithm is based on soft-QoS and hard-QoS structures. The soft-QoS is for rtPS and nrtPS in which their QoS parameters fluctuate between the minimum bandwidth requirements and maximum bandwidth required, while the UGS service class is served under hard-QoS. This structure is effectively capable of allocating the available bandwidth among BE and other service classes along with ensuring fairness.

Sun, et al. [32] proposed two QoS schedulers located at BS and SS sides. Both schedulers give higher priority at BS and SS sides. For rtPS services it is guaranteed the required bandwidth by the calculation of the deadline based on the arrival time. However a fixed priority assigned to nrtPs and BE starved their connection when more rtPS connections existed in the network.

Meng, et al. [33], proposed Adaptive Proportional Fairness (APF) scheduling algorithm. It is designed to extend the PF scheduling algorithm

to real-time services and satisfies various QoS requirements. The APF tries to differentiate the delay performance of each queue based on the Grant per Type-of- Service (GPTS) principle.

Nie, et al. [34], presented a QoS priority and fairness scheduling scheme for DL traffic. It guarantees the delay requirements of UGS, ertPS and rtPS service classes, which considered as two-level scheduling scheme that maximizes the BE traffic throughput. The first level adapted a strict priority between service classes. Moreover, UGS is treated in the second level by assigning a fixed rate, and then APF allocates bandwidth for rtPS and ertPS service classes. Furthermore, the Proportional Fairness algorithm (PF) is applied to treat the non real-time application for nrtPS and BE service classes.

On the other hand, Wongthavarawat, et al. [2] defined three levels of priority. Prioritizing additional rtPS flow may solve the problem of interrupting rtPS packets, but this may lead to a situation where rtPS flows gain arbitrarily large free access at the expense of BE and nrtPS flows. To overcome this shortcoming, Customize Deficit Round Robin (CDRR) with the technique of additional queue has been considered in [35], where it considers only real-time packets which are just prior to the deadline. However, using additional queue for rtPS connections cause additional access latency and bandwidth allocation disorder where errors occur and the flow becomes backlogged. This is undesirable especially if the flow is of type high priority.

There are several other scheduling algorithms with different design goals [36, 38, 39]. However, up-to-date trivial algorithms have been performed to create effective scheduler structure with significant performance in the two-tier hierarchical scheduler, which requires further optimization.

Channel-Aware Scheduling Algorithms. This paper considers the Channel State Information (CSI) when the scheduling allocates the data rate for each connection in LTE systems. Therefore, the scheduling algorithms are capable of manipulating the differences in channel quality for each user, and then based on these characteristics, the scheduler will prioritize the allocation processes, besides taking into consideration of QoS provisioning in different service classes. A comprehensive survey and taxonomy of scheduling in IEEE 802.16e LTE networks can be found in [39].

Moreover, this paper provides an overview of various key scheduling algorithms that considered

channel conditions for LTE networks. Channel aware scheduling system can be divided into two main categories, scheduling algorithm based compensation system and opportunistic scheduling [40].

A. Opportunistic Scheduling Algorithm

The opportunistic algorithms are considered a greedy decision when de-queue packets from service class that experience good channel quality. Moreover, the algorithm will defer de-queuing packets with bad channel condition. Hence, in these channel-aware algorithms, the higher bandwidth granted only for connections that experience a good channel condition. By doing so, the utilization of bandwidth is improved. However this can prioritize non real-time connection located near the base station that experience good channel condition over real-time connection.

Rath, et al. [41], proposed the Opportunistic-Deficit Round Robin (O-DRR) scheduler, which is an analytical method for getting an optimal polling interval for UL data traffic via the polling interval mechanism, the BS polls service traffic periodically to make sure that the traffic delays are achieved. The system considering several situations, for instance, the SSs must ensure that the queue should not be empty as well as the receive Signal-to-Noise Ratio (SNR) must exceed the threshold value. Nevertheless, the allocation mechanism of the O-DRR algorithm leads to an additional overhead at the BS because it requires the manipulation of quantum size and a DC for each SS, repeatedly.

Ball, et al. [42] presented a scheduling algorithm for the rtPS. It manipulates a scheduling list that contains all the SSs, which can be served at the next frame. Nevertheless, the algorithm specifies that the SSs that have low transmission quality are suspended temporarily from the transmission list for a period of time. Moreover, this mechanism is repeated periodically for all SSs. In addition, the scheduler grants another suspended period of time if the transmission quality is still low.

Gan, et al. [43] designed a cross layer scheduling algorithm to cope with the features of UL traffic in LTE system. It referred as dynamic Modulation and Coding Scheme (MCS) and Interference Aware Scheduling (DMIA). The main feature for its structure is taking into account the queue status, the status of the channel and the QoS parameters of service type queue. The main algorithm aim is to improve the total throughput, besides sustaining the QoS requirement of diverse

classes. Thus, this algorithm considered as an optimization problem to the current scheduling algorithm.

Niyato, et al. [44], proposed a queue-aware algorithm in the SS side for UL direction. It defined set thresholds, which can be identified for bandwidth allocation for the connected services as follows:

$$\Psi = \{\psi_1, \psi_2, \dots, \psi_b, \psi_{1b,max}\}$$

Where, $\psi_1 \in \{1, \dots, X\}, \psi_b < \psi_{b+1}$ and $b = 1, \dots, b_{max}$

The purpose of these sets is to recognize the required bandwidth to be allocated in the UL sub-frame. In specific, the sum of bandwidth assigned to polling service is considered as a function of amount of PDUs in a queue. Nevertheless, they estimate the period amongst successive thresholds in the set is equal as follows:

$$b(x) = \begin{cases} 0, & x=0 \\ b & \psi_b \leq x < \psi_{b+1} \\ b_{max} & \psi_{b_{max}} \leq x \end{cases} \quad (1)$$

Lin, et al. [45] proposed a latency and modulation aware bandwidth allocation algorithm called Highest Urgency First (HUF). HUF converts the incoming data rates into time slots in order to determine the influence of several MCS. Nevertheless, the MCS diversity does not fully exploit. Furthermore, HUF process force the request that approach its deadline to be discarded.

B. Scheduling Algorithms based on Compensation System

The scheduling algorithms are categorized under the channel aware scheduling algorithms that are able to amend the allocation based on the variability of LTE channel and QoS parameter provisioning. In this paper, we elaborate more descriptions with reference to related studies related to different algorithms such as fuzzy algorithm which used as an intelligent approach to deal with diverse traffic services.

The authors in [46] discussed a resource allocation and scheduling of cloud computing for five major topics, such as, locality-aware task scheduling; reliability-aware scheduling; energy-aware; Software as a Service layer; and work flow scheduling. Moreover, they classified these five topics into different parts such as performance-based and cost-based resource allocation. In addition, they presented a comparative analysis of the

five identified problems with their representative algorithms. On the other hand, a modification to DRR known as Modified Deficit Round Robin (MDRR) was proposed in [47]. They implement a quantum ϕ and DC for each service type queue. The scheduler, in each round, assigns the service type queue by DC value that is in every round added by the ϕ value. Moreover, the scheduler transmits the packets traffic until the DC empties or when the queue length of the packet is greater than the DC, then the scheduler move to the next queue. Nevertheless, real-time packets will experience disruption delay due to heavy traffic in the system. Furthermore, unfairness in resource sharing among non real-time connections will arise when assigning fixed weight to the queues.

Another research as in Laias, et al. [48], they presented a Customized Deficit Round Robin (CDRR). Their algorithm concerned with real-time service through adding a new queue to schedule real-time connections imminent to the deadline. Nevertheless, an extra queue adds further delay which degrades the overall system throughput for non real-time connections such as nrtPS and BE. Moreover, it will violate the packets deadline for real-time connection such as rtPS queues due to the interception of the extra queue for the real-time signal that leads to increase system overhead, which is not desirable mainly for a high real-time traffic.

On the other hand, in this category, the fuzzy logic approaches considered as one of the important scheduling algorithms. The wireless queue scheduling scheme which uses an adaptive fuzzy logic to alleviate the effect of inaccurate state information, achieves better scheduling performance. In addition, a reinforcement learning scheme is adopted to improve the scheduling performance.

Fuzzy logic is able to achieve a near precise decision according to incomplete and inaccurate information. It can be applied to resource allocation and management in wire line and wireless networks. As mentioned in [49], fuzzy logic is used for queues scheduling in order to solve the problem of inaccurate state information in queue scheduling. This is done through using fuzzy inference to make a near optimal solution regardless of in accurateness of status information such as, traffic characteristics, channel condition, and queue status.

In Kumar, et al.[50], the researches use a dynamic fuzzy based priority scheduler intended to provide an implement of the IEEE 802.16e standard that focus on the QoS aspects such as, delay,

throughput, and bandwidth utilization. They proposed a fuzzy based scheduling algorithm to overcome the conventional scheduling algorithms drawbacks that are not meeting the necessary QoS parameters.

IEEE 802.16 standard, as discussed earlier, stipulates the signalling mechanisms of QoS and the scheduling services classes to cope with the various application requirements. Table 3 shows the existing IEEE 802.16 QoS support algorithms that are implemented at MAC Layer, which includes the state-of-the-art algorithms that are executed at the BS. Moreover, for more contribution to this field this paper elaborates more details about algorithms characteristics, as well as either contribution or drawbacks of supporting multimedia traffic in order to make straightforward future research.

2.3. LTE Physical Layer QoS Scheduling Algorithms

IEEE 802.16 provisions a variation of physical layers that are frequencies diverse from 2-66 GHz. The Wireless MAN-SC (Single Carrier) is considered for 10-60 GHz band as it supports low system complexity.

Even though IEEE has standardized this PHY, there are not many platforms employing it due to the nature of this PHY characteristic that involves LOS communication. The purpose for allowing NLOS communication, IEEE 802.16 intended to implement the Orthogonal Frequency Division Multiplexing (OFDM) via frequency range of 2-11 GHz spectrum.

Moreover, this support type of PHY is referred to as IEEE 802.16d that eventually targeted for fixed SSS. Popularly this type of communication is used in TDMA. In arrangement of time division and frequency division multiple accesses in combination with OFDM is called Orthogonal Frequency Division Multiple Access (OFDMA). Further details about these techniques can be found in [3]. Figure 5 depicted OFDMA subcarriers structure. In this paper, we highlight a several of the current mechanism conducted in the area of resource allocation in LTE networks with detailed emphasis on OFDMA-based LTE networks.

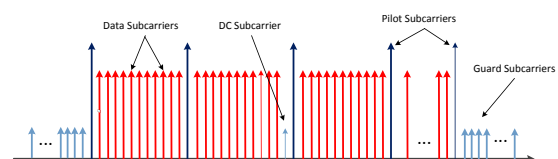


Figure 5. OFDMA Subcarrier Structure

Table 3. Summary of various LTE Scheduling Algorithms

Papers	Objective	Key Idea	Limitations
Tarchi [27]	A hybrid structure algorithm for UL direction	UGS is scheduled by PBRR, EDF for the rtPS and WFQ for nrtPS and BE	It is complex because it depend on different algorithms to handle their QoS requirements
Vinay [29]	Bandwidth allocation for UL traffic to gives better performance for rtPS	Combined EDF and WFQ to serve rtPS	It is complex because it combined two algorithms
Wongthavarawat [31, 32]	Enhance the system throughput and fairness to queues	Constant allocation for rtPS using EDF, whereas WFQ and Equal Sharing are used to schedule nrtPS.	Hybrid algorithms lead to higher complexity
Chen [35]	Ensuring fairness to different service class and to improve throughput	A two-tier hierarchical scheduling for downlink(DL) and uplink(UL) services	Unable to guarantee QoS requirements for rtPS services
Msadaa [36]	Minimise the latency of real-time connection	Guarantee QoS requirements for real-time services	Gives more bandwidth to rtPS at the expenses of BE traffic
Safa [37]	To improves the performance of real-time services by considering a deadline constraint	Modify uplink and downlink bandwidth dynamically	degraded the opportunity for non real-time services such as nrtPS and BE
Shang [38]	Allocating the available bandwidth among BE.	This algorithm is based on the so called soft-QoS and hard-QoS structure	Complex and unfairness for overall system traffic
Sun [39]	Two QoS schedulers located at BS and SS to guaranteed required bandwidth	Gives higher priority to UGS, ertPS and rtPS during the connection setup	Fixed priority assigned to nrtPs and BE starved their connection when more rtPS connections existed in the network
Meng [40]	Designed to extend the PF scheduling algorithm to real-time services and satisfies various QoS requirements	The algorithm tries to differentiate the delay performance of each queue based on GPTS principle	Applied to treat the non real-time application for nrtPS and BE
Nie [41]	Guarantees the delay requirements of UGS, ertPS and rtPS in downlink traffic	QoS priority and fairness scheduling	Not support nrtPS and BE
Rath [49]	To make sure that the traffic delays are achieved	For getting an optimal polling interval for uplink data flow via the polling interval mechanism	The allocation mechanism of the algorithm leads to an additional overhead at the BS

OFDMA. The OFDMA refers to a multiuser OFDM, is recognized as a powerful access technology that is mainly employs the OFDM technique using a modulation technique [51]. The capacity of the system is improved by employing a combination of frequency and multiuser diversities. This is done through allocating a variety of divisions of OFDM subcarriers towards numerous users giving adequate attention to interferences.

In addition, many research has been attempted to improve the optimal subcarrier allocation to the end users, however, there are remaining an open issue related to the allocation problem, with developing problems yet to be conducted [52, 53].

Adaptive Modulation and Coding (AMC) Schemes. In IEEE 802.16e, once higher layer data have been classified into their corresponding service type queues and scheduled by the MAC layer, they are mapped into OFDMA slots by a mapper. A slot is the basic resource unit in an

OFDMA frame structure visualized in a rectangular two-dimensional allocation [3]. In order to ease the resource allocation process in the OFDMA downlink, the subcarriers are grouped into sub-channels by using a mode of permutation.

There are two main classes of permutation modes. The first permutation is Partial Usage of Subcarriers (PUSC) and Full Use of Subcarriers (FUSC) modes. They are diversity permutation schemes that distribute the subcarriers of a sub-channel pseudo-randomly in a wide frequency band. The second permutation mode is Adaptive Modulation and Coding (AMC). In this method a number of carriers adjacent on the frequency spectrum are grouped into a band of AMC sub-channels.

In a multipath fading channel different sub-channels experience different levels of fading. Achievable rates can be maximized by adjusting the modulation and coding rate according to the fading level for each

Table 4. IEEE 802.16 OFDMA PHY Modulation and Coding Schemes

Mode	Modulation	Coding Rate	Information bits/symbol	Receiver SNR (dB)
1	QPSK	1/2	1	5
2	QPSK	3/4	1.5	8
3	16QAM	1/2	2	10.5
4	16QAM	3/4	3	14
5	64QAM	1/2	3	16
6	64QAM	2/3	4	18
7	64QAM	3/4	4.5	20

sub-channel [3]. In each frame, user’s connections are allocated a successive set of subcarriers, forming a rectangular slot. Each allocation is represented in the DL-MAP message by the slot offset and the number of slots in the allocation frame.

On the other hand, the BS combats fading on each slot by adjusting the modulation technique. Moreover, users have experience the same modulation scheme in each rectangular slot and this experience is a same bit rate which depends on the distance away from BS. The AMC objective is to maximize the data rate through adjusting transmission modes to channel variations while maintaining a prescribed BER. Table 4 presents the AMC for IEEE 802.16 OFDMA [11].

LTE OFDMA Frame Structure. IEEE 802.16 standard defines a frame sub-channelization and structure. Figure 6 shows frame sub-channelization mechanisms: Distribute sub-channelization (PUSC) and adjacent sub-channelization (AMC), respectively. Moreover, Figure 7 shows frame structure [56]. Furthermore, the frame constructs of two sub-frames, DL and UL and the burst time can be distinguished by DL-MAP and UL-MAP as referred for their usage in Frame Control Header (FCH). Thus, the users granted their required bandwidth based on these bursts.

Furthermore, in order to suggest LTE scheduler algorithm it must note that, the number of burst per frame must be distinguished, since the number

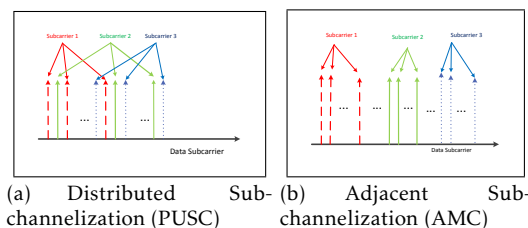


Figure 6. Comparison of Sub-channelization Mechanism - a) Distributed (PUSC) and b) Adjacent (AMC)

of frame burst increases, the overhead is increased proportionately.

Multiuser Diversity. The multiuser diversity scheme is recognized as a wireless system with diverse users sharing a time-varying channel. Moreover, this scheme presented by [54] considered the uplink overall capacity over fading channels with full knowledge of Channel State Information. Their results show the mechanism that manages the capacity for the users that can deliver the maximum available bandwidth. Furthermore, with employments of multiuser characteristics, the overall capacity can be enhanced proportional to the increased number of connections in the network.

Other researchers as in [55] attained similar outcomes for the adjustment of downlink fading channels. Even though fading is typically observed as an impairment that has to be improved, the role of multiuser diversity alters this observation through adjusted fading features. Hence, the multiuser diversity is gradually obtained among diverse users from the presence of channel variations.

Particularly, the achieved diversity can be obtained by exploiting the characteristic of wireless channels, and transmissions scheduling towards users that achieved

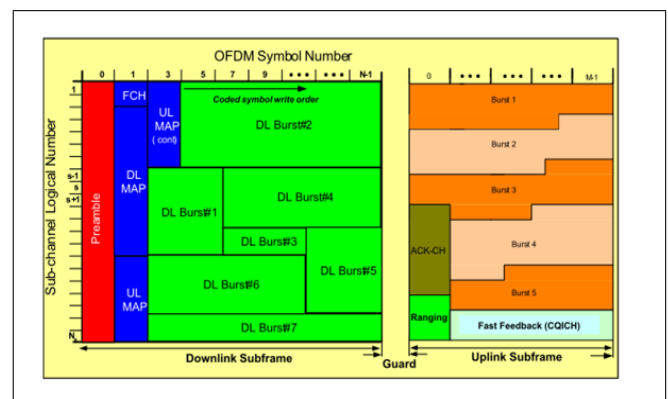


Figure 7. LTE OFDMA Frame Structure [56]

good channel conditions. Hence, the multiuser diversity can significantly increase the spectral efficiency of wireless systems [57] with conventional scheduling mechanisms and CSI feedback.

Generally, the concept of multiuser diversity has attracted a significant consideration, as well as recognised as a key principle in developing well-organized adaptive resource allocation algorithms for wireless communication systems [58].

3. Conclusion

This survey has provided the fundamental information about the design of traffic scheduling algorithms for BWA networks. An overview of BWA networks was presented, with a focus on the IEEE 802.16 standard. In particular, the MAC and PHY layers, as well as the IEEE 802.16 QoS architecture were also discussed.

In addition, a number of essential concepts for designing wireless scheduling algorithms including wireless channel modelling, AMC technique, and multiuser diversity have been described. Furthermore, a detailed about wireless channel poses both challenges and opportunities to traffic scheduling has been discussed. Therefore, a well-designed scheduler should exploit the time-varying nature of wireless channels via the AMC technique and multiuser diversity in order to improve the system performance. This is because overall system utilization starts to be challenged when degradation may happen due to bandwidth allocation disorder. This paper elaborates more detail for MAC QoS scheduling algorithms and discusses their categorizations. In addition, an overview of QoS architecture and the related supported algorithm are presented in detail with their classifications and limitations. Furthermore, this survey investigates the techniques that intend to support QoS for real-time and non real-time services in mobile LTE, especially during packet scheduling and bandwidth allocation that has attracted a significant consideration in developing well-organized adaptive resource allocation and scheduling algorithms for wireless communication systems.

Acknowledgement. This study is supported by the Fundamental Research Grant Scheme (FRGS), Project: FP007-2016 from Ministry of Higher Education, Kuala Lumpur, Malaysia.

References

- [1] IEEE STD P802.16/COR1/D2 (2005) *Draft IEEE Standard for Local and metropolitan area networks Corrigendum to IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems, Corrigendum to IEEE Std 802.16-2004* .
- [2] K. WONGTHAVARAWAT AND A. GANZ (2003) *IEEE 802.16 based last mile broadband wireless military networks with quality of service support,* presented at the Proceedings of the 2003 IEEE conference on Military communications Volume II, Boston, MA.
- [3] AMENDMENT AND CORRIGENDUM TO IEEE STD 802.16-2004 (2006) *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1*, IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005 pp. 0_1-822.
- [4] IEEE P802.16M/D10 (2010) *IEEE Draft Amendment Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Broadband Wireless Access Systems - Advanced Air Interface* pp. 1-1132.
- [5] IEEE STD 802.11v-2011 (AMENDMENT TO IEEE STD 802.11-2007 AS AMENDED BY IEEE STD 802.11k-2008, IEEE STD 802.11r-2008, IEEE STD 802.11y-2008, IEEE STD 802.11w-2009, IEEE STD 802.11n-2009, IEEE STD 802.11p-2010, AND IEEE STD 802.11z-2010) (2011) *IEEE Standard for Information technology- Local and metropolitan area networks- Specific requirements- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: IEEE 802.11 Wireless Network Management* pp. 1-433.
- [6] F. KHANOQ (2009) *LTE for 4G Mobile Broadband: Air Interface Technologies and Performance*
- [7] BAYAN AF, WAN (2012) *"In-depth understanding, analysis, and classification of wimax scheduling state-of-the-art for effective practical scheduler design and development IETE Technical Review.vol 29, pp.449-60.*
- [8] J. RAKESH AND U. DALAL (2010) *A Survey of Mobile LTE IEEE 802.16 m Standard* arXiv preprint arXiv:1005.0976.
- [9] S. MURAWWAT, ET AL. (2012) *An overview of scheduling strategies for PMP mode in IEEE 802.16* Sci. Int.(Lahore), vol. 24, pp. 307-315.
- [10] E. L. HAHNE AND R. G. GALLAGER (1986) *Round robin scheduling for fair flow control in data communication networks* NASA STI/Recon Technical Report N, vol. 86, p. 30047.
- [11] C. SO-IN, ET AL. (2009) *Scheduling in IEEE 802.16 e mobile LTE networks: key issues and a survey* IEEE Journal of Selected Areas in Communications, , vol. 27, pp. 156-171.
- [12] M. KATEVENIS, ET AL. (1991) *Weighted round-robin cell multiplexing in a general-purpose ATM switch chip* IEEE Journal on Selected Areas in Communications, , vol. 9, pp. 1265-1279.
- [13] M. SHREEDHAR AND G. VARGHESE (1996) *Efficient fair queuing using deficit round-robin* Networking IEEE/ACM Transactions on, vol. 4, pp. 375-385.
- [14] C. CICONETTI, ET AL. (2006) *Quality of service support in IEEE 802.16 networks* Network, IEEE, vol. 20, pp. 50-55.
- [15] A. K. PAREKH AND R. G. GALLAGER (1992) *A generalized processor sharing approach to flow control in integrated services networks-the single node case* in INFOCOM'92. Eleventh Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE, pp. 915-924.
- [16] C. JON, ET AL. (1996) *WF2Q: worst-case fair weighted fair queuing* in Proceedings of IEEE INFOCOM.
- [17] S. J. GOLESTANI (1994) *A self-clocked fair queueing scheme for broadband applications* in INFOCOM'94. Networking for Global Communications, 13th Proceedings IEEE, pp.

- 636-646.
- [18] J.-P. GEORGES, ET AL. (2005) *Strict Priority versus Weighted Fair Queuing in Switched Ethernet networks for time critical applications* in Parallel and Distributed Processing Symposium, 2005. Proceedings. 19th IEEE International, pp. 141-141.
- [19] A. LERA, ET AL. (2007) *Channel-aware scheduling for QoS and fairness provisioning in IEEE 802.16/LTE broadband wireless access systems* Network, IEEE, vol. 21, pp. 34-41.
- [20] D. TARCHI, ET AL. (2006) *Quality of service management in IEEE 802.16 wireless metropolitan area networks* IEEE International Conference on Communications, 2006. ICC'06. pp. 1789-1794.
- [21] A. ESMAILPOUR AND N. NASSER (2011) *Dynamic QoS-based bandwidth allocation framework for broadband wireless networks* IEEE Transactions on Vehicular Technology, vol. 60, pp. 2690-2700.
- [22] K. VINAY, ET AL. (2006) *Performance evaluation of end-to-end delay by hybrid scheduling algorithm for QoS in IEEE 802.16 network* IFIP International Conference on in Wireless and Optical Communications Networks, pp.5.
- [23] K. WONGTHAVARAWAT AND A. GANZ (2003) *IEEE 802.16 based last mile broadband wireless military networks with quality of service support* in Military Communications Conference. MILCOM'03. 2003 IEEE, pp. 779-784.
- [24] K. WONGTHAVARAWAT AND A. GANZ (2003) *Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems* International Journal of Communication Systems, vol. 16, pp. 81-96.
- [25] A. SAYENKO, ET AL. (2006) *Ensuring the QoS requirements in 802.16 scheduling* presented at the Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems, Terromolinos, Spain.
- [26] O. IOSIF, ET AL (2010) *Performance analysis of uplink resource allocation in WIMAX* in 8th International Conference on Communications (COMM), pp. 351-354.
- [27] B. LI, ET AL. (2007) *A survey on mobile LTE [wireless broadband access]* Communications Magazine, IEEE, vol. 45, pp. 70-75.
- [28] C. JIANFENG, ET AL. (2005) *A service flow management strategy for IEEE 802.16 broadband wireless access systems in TDD mode* in Communications, IEEE International Conference on ICC, Vol. 5, pp. 3422-3426
- [29] H. SAFA, ET AL. (2007) *New Scheduling Architecture for IEEE 802.16 Wireless Metropolitan Area Network* in Computer Systems and Applications, IEEE/ACS International Conference, pp. 203-210.
- [30] I. C. MSADAA, ET AL. (2007) *An Adaptive QoS Architecture for IEEE 802.16 Broadband Wireless Networks* IEEE International Conference in Mobile Adhoc and Sensor Systems, pp. 1-3.
- [31] Y. SHANG AND S. CHENG) (2005) *An enhanced packet scheduling algorithm for QoS support in IEEE 802.16 wireless network* in Networking and Mobile Computing, ed: Springer, pp. 652-661
- [32] J. SUN, ET AL. (2006) *Quality of service scheduling for 802.16 broadband wireless access systems* IEEE 63rd in Vehicular Technology Conference, pp. 1221-1225.
- [33] X. MENG (2007) *An efficient scheduling for diverse QoS requirements in LTE* University of Waterloo.
- [34] W. NIE, ET AL. (2011) *Packet Scheduling with QoS and Fairness for Downlink Traffic in LTE Networks* JIPS, vol. 7, pp. 261-270.
- [35] E. LAIAS AND I. AWAN (2010) *An interactive QoS framework for fixed LTE networks* Simulation Modelling Practice and Theory, vol. 18, pp. 291-303.
- [36] D.-J. DENG, ET AL. (2009) *Delay constrained uplink scheduling policy for rtPS-ertPS service in IEEE 802.16e BWA systems* Int. J. Commun. Syst., vol. 22, pp. 119-133.
- [37] I. P. HSIEH, ET AL. *A slot-based BS scheduling with maximum latency guarantee and capacity first in 802.16e networks* International Journal of Communication Systems, pp. n/a-n/a.
- [38] L. YI-NENG, ET AL. (2008) *A Latency and Modulation Aware Bandwidth Allocation Algorithm for LTE Base Stations* in Wireless Communications and Networking Conference, IEEE, pp. 1408-1413.
- [39] I. C. MSADAA, ET AL. (2010) *Scheduling and CAC in IEEE 802.16 fixed BWNs: a comprehensive survey and taxonomy* IEEE Communications Surveys & Tutorials, vol. 12, pp. 459-487.
- [40] D. NDIKI, ET AL. (2010) *A Comparative Overview IEEE 802.16 e QoS Scheduling Algorithms* in Second International Conference on Evolving Internet (INTERNET), pp. 74-79.
- [41] H. K. RATH, ET AL. (2006) *NXG02-4: An Opportunistic Uplink Scheduling Scheme to Achieve Bandwidth Fairness and Delay for Multiclass Traffic in Wi-Max (IEEE 802.16) Broadband Wireless Networks*, pp. 1-5.
- [42] C. BALL, ET AL. (2005) *Performance analysis of temporary removal scheduling applied to mobile wimax scenarios in tight frequency reuse* pp. 888-894.
- [43] W. GAN, ET AL. (2009) *A cross-layer designed scheduling algorithm for LTE uplink* 9th International Conference on Electronic Measurement & Instruments, ICEMI'09, pp. 1-122-1-127.
- [44] D. NIYATO AND E. HOSSAIN (2005) *Queue-aware uplink bandwidth allocation for polling services in 802.16 broadband wireless networks* in Global Telecommunications Conference, GLOBECOM'05. IEEE, pp. 5 pp.-3706.
- [45] Y. N. LIN, ET AL. (2009) *Highest Urgency First (HUF): A latency and modulation aware bandwidth allocation algorithm for LTE base stations* Computer Communications, vol. 32, pp. 332-342.
- [46] MA T, CHU Y, ZHAO L, ANKHBAYAR (2014) *Resource Allocation and Scheduling in Cloud Computing: Policy and Algorithm* IETE Technical Review. vol. 31 pp. 4-16.
- [47] DOC. ID 18841 *Understanding and configuring MDRR/WRED on the Cisco 12000 series internet routers* doc. ID 18841 Available: http://www.cisco.com/warp/public/63/mdrr_wred_overview.html
- [48] E. LAIAS AND I. AWAN (2010) *An interactive QoS framework for fixed LTE networks* Simulation Modelling Practice and Theory, vol. 18, pp. 291-303.
- [49] J. CHENG, ET AL. (2008) *wireless Queue Scheduling Based on Adaptive Fuzzy Logic* in Wireless Communications, Networking and Mobile Computing. WiCOM'08. 4th International Conference on, pp. 1-4.
- [50] D. D. N. P. KUMAR, ET AL. (2012) *Fuzzy based Priority Scheduler for LTE with Improved QoS constraints* International Journal of Advanced Networking and

- Applications, vol. 4.
- [51] IEEE UNAPPROVED DRAFT STD P802.16j/D7 (2008) *IEEE Draft Amendment to IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Multihop Relay Specification* IEEE UNAPPROVED DRAFT STD P802.16j/D7
- [52] J. G. ANDREWS, ET AL (2007) *Fundamentals of LTE: Understanding Broadband Wireless Networking (Prentice Hall Communications Engineering and Emerging Technologies Series)* Prentice Hall PTR.
- [53] S. SADR, ET AL. (2009) *Radio Resource Allocation Algorithms for the Downlink of Multiuser OFDM Communication Systems* Communications Surveys & Tutorials, IEEE, vol. 11, pp. 92-106.
- [54] R. KNOPP AND P. A. HUMBLET (1995) *Information capacity and power control in single-cell multiuser communications* in Communications, Gateway to Globalization, IEEE International Conference on ICC '95 Seattle, vol.1, pp. 331-335.
- [55] D. N. TSE (1997) *Optimal power allocation over parallel Gaussian broadcast channels*, in IEEE International Symposium on Information Theory, pp. 27-27.
- [56] S. CHIEOCHAN AND E. HOSSAIN (2009) *Adaptive radio resource allocation in OFDMA systems: a survey of the state-of-the-art approaches* Wireless Communications and Mobile Computing, vol. 9, pp. 513-527.
- [57] M. ERGEN, ET AL. (2003) *QoS aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access systems*, Broadcasting, IEEE Transactions, vol. 49, pp. 362-370.
- [58] ALSAHAG, A. M., ALI, NOORDIN , N. K., & MOHAMAD, H. (2014) *Fair uplink bandwidth allocation and latency guarantee for mobile WiMAX using fuzzy adaptive deficit round robin* Journal of network and computer application, vol. 39, pp. 17-25.