Radio Resource Scheduling in 3GPP LTE: A Review

Davinder Singh[#1], Preeti Singh[#2]

[#]*University Institute of Engineering &Technology, Panjab University, Chandigarh, India*

***Abstract*— Long Term Evolution (LTE) ,proposed by 3rd Generation Partnership Project (3GPP) as a 3.9G technology, represents a very promising answer to the ever rising bandwidth demand of mobile applications. To support vast range of multimedia and internet services at high data rates that too with increased spectral efficiency; LTE incorporates various Radio Resource Management (RRM) procedures. The key to achieve optimal performance of base station is dynamically scheduling limited resources like power and bandwidth to offer the best service for terminals with the lowest cost. In this context, radio Resource allocation strategies play a key role in distributing radio resources among different stations by taking into consideration the channel conditions as well as QoS requirements. The present paper provides review of radio resource allocation strategies present in the literature.**

***Keywords*— LTE, OFDMA, QOS, Downlink Scheduling, Radio-Resource Management**

I. INTRODUCTION

In recent years, demand for data services, such as mobile TV, file sharing, video telephony and social networking have grown very fast. As a 3.9G technology, Long Term Evolution (LTE) is proposed by 3rd Generation Partnership Project (3GPP) to meet these requirements [1]. Worldwide 20 LTE cellular operators with more than 32 million subscribers prefer LTE [2]. The reason is that the higher bandwidth speed and the higher system capacity of LTE attract operators. Furthermore, the latency of end to end in LTE is possible to be kept within 50 ms. LTE achieves above said goals by making use of various the Radio Resource Management (RRM) procedures.

The key to achieve optimal performance of base station is dynamically scheduling limited radio resources like power and bandwidth to offer the best service for terminals with the lowest cost in order to satisfy the end user's QOS requirement. The major node performing radio resource management functions in LTE is evolved NodeB (eNB). The packet scheduler in the MAC layer works at the eNB, and is the prime entity which assigns portions of spectrum shared among users. Since the channel quality is vulnerable to variations in time and frequency domains due to various effects, such as fading, multipath propagation and Doppler Effect, so the packet scheduler makes use of efficient resource allocation strategy to maximize spectral efficiency while keeping low channel drops.

Ensuing text in the paper discusses various resource allocation strategies of LTE present in the literature with focus on LTE air interface and downlink scheduling from eNB to UE.

II. DOWNLINK RESOURCE ALLOCATION IN LTE

At the physical layer, LTE allows flexible bandwidth which varies from 1.4 MHz up to 20 MHz. Radio spectrum access is based on the Orthogonal Freq. Division Multiplexing (OFDM) scheme. The air interface has been designed to use OFDMA for downlink i.e. from eNodeB to UE and Single Carrier – FDMA has been selected for uplink i.e. for transmission from UE to eNodeB.

Radio resources in LTE are apportioned into the time/frequency domain [3]. Along the time domain they are assigned every Transmission Time Interval (TTI). TTI has been reduced to 1ms in LTE in order to support low latency data transfer. The time is divided in frames. Each 10ms Frame is divided into ten 1ms sub-frames i.e. TTIs, with each sub-frame further divided into two 0.5ms Slots. Each slot consists of 7 OFDM symbols with normal cyclic prefix. In the frequency domain, instead, the total bandwidth is divided in sub-channels of 180 kHz, each one with 12 consecutive and equally spaced OFDM sub-carriers. *Resource Block* (RB) which is formed by the intersection between a sub-channel in frequency domain and one TTI in time domain is the smallest allocable resource unit.

Spectrum portions should be distributed every TTI among the users. Packet schedulers work in the time and frequency domain with coarseness of one TTI and one RB respectively. The fastest scheduling is required to be done within 1ms according to the symbol length of RB.

III. PROCEDURE OF DOWNLINK SCHEDULING

The per-RB metrics' comparison that serves as the transmission priority of each user on a specific RB is taken into account for resource allocation for each UE. For example the *k*-th RB is allocated to the *j*-th user if its metric $m_{j;k}$ is the largest one among all i-UEs, i.e., if it satisfies the equation:

$$m_{j;k} = \max_i \{m_{i;k}\} \qquad (1)$$

The whole process of downlink scheduling can be divided in a sequence of operations that are repeated, in general, every TTI (see fig.1):

1) The Evolved Node B prepares the list of flows which can be scheduled in the current TTI .Flows could be formulated only if there are packets to send at MAC layer and UE at receiving end is not in the idle state.

2) Each UE decodes the reference signals, reports CQI (Channel Quality Indicator) to eNB which helps to estimate the downlink channel quality. The eNB can configure if the CQI report would correspond to the whole downlink bandwidth or a part of it which is called sub-band.

3) Then the chosen metric is computed for each flow according to the scheduling strategy using the CQI

information. The sub-channel is assigned to that UE that presents the highest metric.

4) For each scheduled flow, the eNB computes the amount of data that will be transmitted at the MAC layer i.e. the size of transport block during the current TTI. The AMC (Adaptive Modulation and Coding module) at MAC layer selects the best MCS (Modulation and Coding Scheme) that should be used for the data transmission by scheduled users. Link adaptation involves tailoring the modulation order (QPSK, 16-QAM, 64-QAM) and coding rate for each UE in the cell, depending on the downlink channel conditions.

5) Physical Downlink Control Channel (PDCCH) is used to send the information about the users, the assigned Resource Blocks, and the selected MCS to terminals in the form of DCI (Downlink Control Information).

6) Each UE reads the PDCCH payload .If a particular UE has been scheduled; it will try to access the proper PDSCH payload.

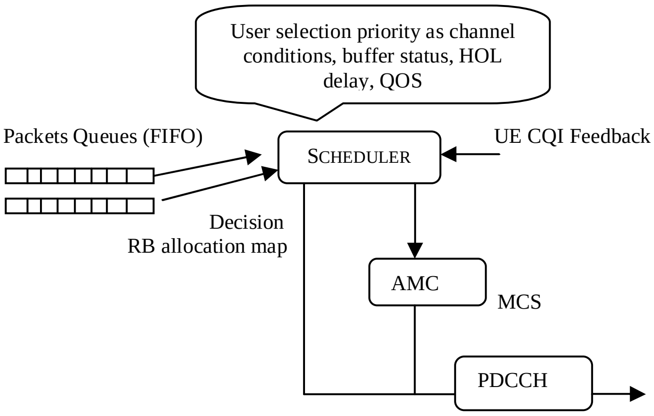

Fig.1. General Model of packet scheduler

The users are prioritized by packet scheduler on the basis of a scheduling algorithm being used. These algorithms while making scheduling decisions, takes into account the instantaneous or average channel conditions, Head of Line (HOL) packet delays, status of receiving buffer or type of service being used [4].

IV. DOWNLINK RESOURCE SCHEDULING ALGORITHMS

Generally, scheduling can be divided into two classes: channel-independent scheduling and channel-dependent scheduling. The performance of channel-independent scheduling can never be optimal in wireless networks due to varying nature of instantaneous channel conditions. On the contrary, channel-dependent scheduling can achieve better performance by allocating resources based on channel conditions with optimal algorithms. Channel dependent schedulers can further be classified on the basis of QOS support as QOS unaware or QOS aware channel dependent scheduling algorithms. Apart from these semi-persistent and energy aware solutions also exist in the literature.

A. Channel Independent Scheduling Strategies

Channel independent strategies were firstly introduced in wired networks and are based on the assumption of time-invariant and error-free transmission media. Being unrealistic for LTE networks, they are typically used in conjugation with channel-dependent strategies to improve system performance.

1) First in First out (FIFO): Though FIFO is the simplest of all possible scheduling disciplines but it is inefficient and unfair. This scheduler serves the packets in the queue in order of arrival and when the queue is full, it drops the packets that are just arriving [5]. The major setback is that it cannot differentiate among connections; therefore all packets experience the same delay, jitter and packet loss irrespective of which packet it is.

The metric of i-th user on the *k*-th RB can be translated from its behaviour as:

$$m_{i,k}^{FIFO} = t - T_i \qquad (2)$$

Where t is the current time and T_i is the time instant when request was issued by i-th user.

2) Round Robin: Round Robin allocates resources to each UE, completely neglecting channel quality or data rate. Initially, the terminals are ordered randomly in a queue. The new terminals are inserted at the end of the queue. The first terminal of this queue is assigned all the available resources by scheduler, and then put it at the rear of the queue. The rest of steps follow the same way, until no terminal applies for resources. Round Robin (RR) metric is similar to the one defined for FIFO. The only difference is that, in this case, T_i refers to the last serving time instant of the user.

On one hand, it seems to be a fair scheduling, since every terminal is given the same amount of resources. On the other hand, it neglects the fact that certain terminals in bad channel conditions need more resources to carry out the same rate, so it is absolutely unfair. This scheme is impractical in LTE because different terminals have different service with different QoS requirements [6].

3) Weighted Fair Queuing: In Weighted Fair scheduling introduced in [7], the packets are grouped into various queues. A weight is assigned to each queue which determines the fraction of the total bandwidth available to the queue. In this case, a specific weight (w_i) is associated to the *i*-th user (or class of users) and then it is used to modify Round-Robin metric as:

$$m_{i,k}^{WFQ} = w_i \cdot m_{i,k}^{RR} \qquad (3)$$

To assure that flows with larger packets are not allocated more bandwidth than flows with smaller packets, it also supports variable-length packets. The Weighted Fair scheduling assigns the bandwidth for each service based on the weight assigned to each queue and not based on the number of packets.

4) Blind Equal Throughput: The Blind equal throughput (BET) is a channel unaware strategy that aims at providing throughput fairness among all the users. To counteract the unfair sharing of the channel capacity, the BET scheduler uses a priority metric which considers past average user throughput as follows [8]:

$$m_{i,k}^{BET} = \frac{1}{\overline{R}^{i}(t-1)} \quad (4)$$

Where $\overline{R}^{i}(t-1)$ is the average throughput of terminal *i* over windows in the past.

The smoothed value of $R^i(t)$ is computed using any weight moving average formula, e.g.,

$$\overline{R}^{i}(t) = \{\left(1-\frac{1}{T}\right)\cdot\overline{R}_{i}(t-1) + \frac{1}{T}\cdot R^{i}(t) \quad (5)$$

Where $R^i(t)$ is the instantaneous value of data rate at time instant t .

It is clear from equation (4) that the BET scheduler prioritizes users with lower average throughput in the past. This implies that users with bad channel conditions are allocated more resources compared to the users with good channel conditions. Thus throughput fairness is achieved but at the cost of spectral efficiency.

5) Largest Weighted Delay First: To avoid packet drops, it is required that each packet has to be received within a certain delay deadline in Guaranteed delay services. It incorporates the information about the specific packet timing, when the packet was created and its deadline while calculating the priority metric. For Real-Time flow, its metric is calculated as [9]:

$$m_{i,k}^{MLWDF} = \alpha_i \cdot D_{HOL,i} \quad \text{where} \quad \alpha_i = -\frac{\log \delta_i}{\tau_i} \quad (6)$$

Where $D_{HOL,i}$ is waiting time of the packet at the head of the line and δ_i represents drop probability and τ_i defines target delay for the *i*-th user.

Similar to Round Robin, neglecting channel conditions leads to poor throughput in LWDF.

B. Channel Dependent/QOS unaware Scheduling Strategies

Channel-dependent scheduling strategies allocate resources with optimal algorithms by taking into consideration the channel conditions. The user channel quality can be estimated from CQI reports which help the scheduler to estimate the channel quality perceived by each UE and serves as an indication of the data rate which can be supported by the downlink channel.

The ensuing text discusses the channel dependent but QOS unaware strategies that exist in the literature:

1) Maximum Throughput: Being a channel dependant scheduling, Max Throughput takes advantage of multiuser diversity to carry out maximum system throughput. First, scheduler analyzes CQI reports from UEs to obtain data rate of an identical sub-channel for different terminals. This information can be used in the priority metric to prioritize users with good channel conditions over users with bad channel conditions. Thus scheduler assigns the resource to the user which can achieve the highest throughput in this sub-channel based on SNR. The priority metric for the MaxT scheduler is given as follows [8]:

$$m_{i,k}^{Max-T} = d_{k}^{i}(t) \quad (7)$$

Where $d_{k}^{i}(t)$ is the expected data-rate for *i*-th user at time *t* on the *k*-th Resource-block. It can be calculated by considering the Shannon expression for the channel capacity as:

$$d_{k}^{i}(t) = \log[1 + SINR_{k}^{i}(t)] \quad (8)$$

MaxT performs unfair resource sharing of the resources since it aims at blind maximization of throughput only.

2) Proportional Fair: The Proportional Fair (PF) algorithm can improve the fairness among users without losing the efficiency in terms of average (or aggregate) throughput. The terminals are ranked according to the priority function which is defined as the ratio of the instantaneous to average throughput.. Then scheduler assigns resources to terminal with highest priority. Repeat the last two steps until all the resources are used up or all the resources requirements of terminals are satisfied [10]-[13].

The PF was designed specifically for the Non-Real Time class and hence does not assure any QoS requirement such as delay, jitter and latency. The preference metric or priority function is obtained by merging the metrics of MaxT and BET and is given as:

$$m_{i,k}^{PF} = m_{i,k}^{Max-T} \cdot m_{i,k}^{BET} = \frac{d_{k}^{i}(t)}{\overline{R}^{i}(t-1)} \quad (9)$$

$d_{k}^{i}(t)$ is the estimation of supported data rate of terminal *i* for the resource block *k*. $\overline{R}^{i}(t-1)$ is the average data rate of terminal *i* over a windows in the past.

T_{PF} is the windows size of average throughput and can be adjusted to maintain fairness. Normally T_{PF} should be limited in a reasonable range so that terminals cannot notice the quality variation of the channels.

If $R^i(t)$ is the instantaneous value of data rate a time instant t ,then if *i*-th terminal is selected

$$\overline{R}^{i}(t) = \{\left(1-\frac{1}{T_{PF}}\right)\overline{R}_{i}(t-1) + \frac{1}{T_{PF}}\cdot R^{i}(t) \quad (10)$$

If *i*-th terminal is not selected, then

$$\overline{R}^{i}(t) = \{\left(1-\frac{1}{T_{PF}}\right)\overline{R}_{i}(t-1) \quad (11)$$

Rabie K. Almatarneh *et. al.* in [12] evaluated the performance of two dimensional (time slot and frequency subcarrier) PF scheduling in OFDMA wireless systems ; both analytically and by simulation. Closed-form expressions for the average throughput and Jain’s fairness index as the performance metrics; have been derived. The algorithm performance is investigated for a broad range of the traffic load and the number of subbands.

In [13], the approach of PF was formulated as an optimization problem in order to maximize the achieved throughput of a LTE system. Here, a multiuser scheduler with

PF is proposed. A suboptimal PF scheduler, which has a much lower complexity at the cost of some throughput degradation, is also proposed. Numerical results show that the proposed PF scheduler provides a superior fairness performance with a modest loss in throughput, as long as the user average SINRs are fairly uniform.

3) Throughput to Average: Throughput to Average (TTA) scheduling algorithm [10] tries to divide the available resources between all users with the priority metric:

$$m_{i,k}^{TTA} = \frac{d_k^i(t)}{d^i(t)} \quad (12)$$

The above metric performs averaging of resources evenly between the users. Here, the achievable throughput in the current TTI is used as normalization factor of the achievable throughput on the considered k-th RB. It is evident from its metric that the higher the overall expected throughput of a user is, the lower will be its metric on a single Resource Block.

C. Channel Dependent/QOS aware Scheduling Strategies

In LTE, QoS differentiation is managed by associating a set of QoS parameters to each flow. Minimum required performance can be guaranteed by the scheduler if it knows the values of QOS parameters, either in terms of guaranteed data rates or of delivery delays.

In this subsection, a comprehensive overview on QoS-aware solutions presented in literature for LTE systems is presented.

1) Schedulers for Guaranteed Data Rate: G. Monghal *et. al.* in [14] proposed QOS oriented Time and frequency domain schedulers that focus on GBR considerations. The proposed Time Domain Priority Set Scheduler (TDPSS) has been devised to select users with the highest priority. Users are separated into two sets. Set 1's users with bit rate below target bit rate are managed by using BET and prioritised over all the other users which form Set 2. Furthermore, within each set; prioritization is according to priority metrics.While TD-PSS will tend to maintain the throughput of low signal quality users to Target Bit Rate, Frequency Domain- PF i.e PF scheduled (PFsch) will tend to reduce their allocation share in the frequency domain with priority metric given as:

$$m_{i,k}^{PFsch} = \frac{d_k^i(t)}{\overline{R_{sch}^i}(t-1)} \quad (13)$$

where $\overline{R_{sch}^i}(t-1)$ is similar to the past average throughput defined in eq. (4), with the difference that it is updated only when the *i*-th user is actually served.

In [15] a Dynamic Hybrid Scheduler (DHS) composed by two basic components, corresponding to a guaranteed and a dynamic delay based rate allocation policy respectively is presented. Used priorities are calculated, for the *i*-th user, as:

$$P_i = \frac{D_{HOL,i}}{\tau_i} \quad (14)$$

It is important to note that the transmission of the head of line packet becomes more urgent, when the value of *Pi* is increased. To attain the guaranteed bit-rate, the resources are allocated to the user with the highest priority. The user with second highest priority is considered thereafter for allocation in case the RBs are left free and so on.

A similar approach is followed in [16] by Y. Zaki *et. al.*. In order to simplify the LTE MAC scheduling, two stages have been defined: Time Domain (TD) and Frequency Domain (FD) schedulers. The TDPS differentiates the users according to their QoS characteristic whereas FDPS assigns the RBs between the priority users. Based on the QoS Class Identifier (QCI), the incoming packets are categorized upon their priority order. The priority sets are classified as GBR and non-GBR set. After this step, the FDPS orderly assigns the best RB to each user in the GBR set, updating the achieved bitrate. When all users in the list have reached their target bit-rate, if RBs are still available, the scheduler assigns them to users in the non-GBR list using PF metric.

Thus, all these approaches use ordered lists to prioritize the most delayed flows and to achieve their target bit-rate.

2) Schedulers for Guaranteed Delay Requirements: Real-Time flows have more strict delay restraint than Non-Real-Time flows resulting in the reduction of influence of error correction. Scheduling strategies that aim to guarantee bounded delay fall in the category of the QoS-aware schemes.

Herein, QOS aware algorithms present in the literature that makes use of per-RB metrics are described.

The Modified Largest Weighted Delay First (M-LWDF) [17] combines both channel conditions and the state of the queue with respect to delay in making scheduling decisions. It ensures that the probability of delay packets does not exceed the discarded bound below the maximum allowable packet loss ratio i.e.

$$\Pr(D_{HOL,i} > \tau_i) \le \delta_i$$

The scheduler allocates resources to the user with the maximum priority index which is made up of the product of the HOL packet delay of the user, the channel capacity with respect to flow and the QoS differentiating factor:

$$m_{i,k}^{MLWDF} = \alpha_i \cdot D_{HOL,i} \cdot \frac{d_k^i(t)}{\overline{R^i}(t-1)} \quad (15)$$

Where $D_{HOL,i}$ is waiting time of the packet at the head of the line and $\alpha_i = -\log \delta_i / \tau_i$;δ_i represents acceptable packet loss rate (i.e. the maximum probability for HOL packet delay of user *i* to exceed the delay threshold of user *i*.) and τ_i defines Delay threshold for the *i*-th user.

EXP/PF is a QOS aware extension of PF that can support both Non-Real Time and Real Time flows at the same time [18].For real-time flows the metric is calculated as:

$$m_{i,k}^{EXP/PF} = \exp\left(\frac{\alpha_i \cdot D_{HOL,i} - \chi}{1+\sqrt{\chi}}\right) \cdot m_{i,k}^{PF} \quad (16)$$

Where $\chi = \frac{1}{N_{RT}} \cdot \sum_{i=1}^{N_{RT}} D_{HOL,i}$ and N_{RT} is the number of active downlink real-time flows.

The metric when calculated for Non-real-time flows is given as:

$$m_{i,k}^{EXP/PF} = \frac{w(t)}{M(t)} \cdot \frac{d_k^i(t)}{\bar{R}^i(t-1)} \quad (17)$$

Where w(t)={ w(t-1) - ε...........$(D_{HOL})_{max} > \tau_{max}$

{w (t-1) + ε /p.......$(D_{HOL})_{max} < \tau_{max}$

M(t) is the average number of RT packets waiting at e-Node B buffer at time *t*, *ε* and *p* are constants, $(D_{HOL})_{max}$ is the maximum HOL packet delay of all RT service users and τ_{max} is the maximum delay constraint of RT service users. Here, RT users are prioritized over NRT users when their HOL packet delays are approaching the delay deadline. The exponential term is closer to 1 if HOL delays of all users are about the same. Thus above formula becomes Proportional Fair. If one of the user's delays becomes large, the exponential term in will override the left term in (16) and dominate the selection of a user.

EXP rule [19] can be considered as modified form of the above mentioned EXP/PF and its priority metric is calculated as:

$$m_{i,k}^{EXP-Rule} = b_i \cdot \exp\left(\frac{\alpha_i \cdot D_{HOL,i}}{c + \sqrt{\left(\frac{1}{N_{RT}}\right) \sum_j D_{HOL,j}}} \right) \cdot \Gamma_k^i \quad (18)$$

Where Γ_k^i represents Spectral efficiency for the user i over the *k*-th RB and the optimal parameter set according to [19] is:

{ $a_i \in [(5/0.99\, \tau_i\ ,10/0.99\, \tau_i)]$,

b_i=1/E [Γ^i] and

c=1

In [20] performance of Exponential Rule is evaluated in comparison to PF Scheduler and MLWDF .A variant of the Exponential rule i.e. EXPQW is also proposed which assigns weights to the subscriber stations based on their queue length and waiting time. Three hierarchical schedulers which use a combination of the exponential rule for waiting time and queue-length and other scheduling rules have also been presented. The results indicate that EXPQW and the hierarchical schedulers have comparable throughput and fairness values with algorithms like PF and MLWDF in moderately loaded and heavily loaded scenarios.

In [21] M. Iturralde *et. al.* proposed a two level resource allocation scheme is to enhance the QoS for multimedia services. It corresponds to a procedure that combines cooperative game theory, a virtual token mechanism, and the EXP-RULE algorithm. It works in two phases: in the first one the game is run to partition available RBs among different groups of flows, populated depending on the type of application they carry. The second phase uses of EXP rule modified by using a virtual token mechanism in order to meet bounded delay requirements and to guarantee, at the same time, a minimum throughput to all flows. In this way, a significant performance gain over the EXP rule is achieved in terms of both packet loss rate and fairness.

LOG Rule algorithm has been described in [22]. For the LOG rule, the preference function is calculated as:

$$m_{i,k}^{LOG-Rule} = b_i \cdot \log\left(c + \alpha_i D_{HOL,i}\right) \Gamma_k^i \quad (19)$$

where b_i, c, and a_i are tuneable parameters; Γ_k^i represents the spectral efficiency for the i-th user on the k-th RB. Optimal parameters are given as :

b_i = 1/E [Γ^i] ,c = 1.1, and a_i = 5/0.99 τ_i

Prio *et al.* proposed a two-level downlink scheduling for real-time flows in LTE networks [23]. At the highest level, a discrete time linear control law is applied every LTE frame. The total amount of data that real-time flows should transmit in 10 ms is thus pre-calculated while considering their delay constraints. When FLS completes its task, the lowest layer scheduler works every TTI. The lower PF algorithm allocates radio resources by considering bandwidth requirements of FLS to flows hosted by UEs experiencing the best channel quality. In particular, the lowest layer scheduler decides the number of TTIs/RBs (and their position in the time/frequency domains) in which each real-time source will actually transmit its packets. The resources left free by real time flows are assigned to NRT flows.

D. Dynamic and Semi-persistent Scheduling for VoIP support

Dynamic packet scheduling for VoIP traffic in the LTE Downlink is presented in [24][25]. In [24] the aim is to optimize the performance of dynamic scheduling when mix of VoIP traffic and best effort flows are available. The proposed algorithm is divided into time domain and frequency domain packet schedulers. At every TTI ,scheduler called as Required Activity Detection with Delay Sensitivity (RAD-DS) prioritizes each schedulable user according to the time domain metric $\mathbf{M}^{TD}[n, t]$,which is combination of 3 metrics given as:

$$\mathbf{M}^{TD}[n, t] = m[n, t] \cdot \mathbf{RA}^{traf}[n, t] \cdot \mathbf{DS}^{traf}[n, t] \quad (20)$$

$\mathbf{RA}^{traf}[n, t]$ (i.e. the required activity) implies the time share required by user *n* where a user should be scheduled. $m[n, t]$ is a counter incremented every TTI that guarantees some fairness in resource scheduling. Finally, $\mathbf{DS}^{traf}[n, t]$ (i.e. the delay sensitivity) function imposes time constraints to users with a delay bound that increases with HOL packet delay.

The frequency domain scheduler allocates Resource Blocks to different users using the Proportional Fair scheduled (PFsch) metric.

To support high number of VoIP flows, semi-persistent allocation solutions (generally considered as channel-independent schemes) aim at increasing the VoIP capacity of the network by maximising the number of supported VoIP calls. One such scheme presented in [26], improves the VoIP capacity of the network with the use of semi-persistent scheme. Here, the radio resources are divided in several groups of RBs. Each pre-configured block is associated only to certain users. Furthermore, RB groups are associated to each user in contiguous TTIs. Resource allocation of each RB group to the associated UEs is performed dynamically. The proposed scheme reduces the control overhead with respect to the dynamic scheduling. Semi-persistent schemes for VOIP have also been proposed in [27] [28].

E. Energy Aware Solutions

Energy consumption is heavy in LTE due to tremendous processing load on UE. Energy conserving solutions curb energy waste and hence extend the battery life of UE among which Discontinuous reception (DRX) is useful. In DRX, when there are no data transmissions; UE turn off its radio equipment UE to save energy. In [29], the light sleeping mode is introduced to further improve the performance of DRX for QOS –aware traffic. The key idea is to turn off the power amplifier .Other components in transceiver cut down their power consumption while allowing fast wakeup. Proposed scheme reduces energy consumption while satisfying the delay constraints.

In [30], the impact of different scheduling schemes from an energy efficiency point of view is analysed. It is demonstrated that the MaxT scheme is more energy efficient than both PF and RR. In scenarios with low traffic load, Bandwidth Expansion Mode algorithm is used for achieving energy savings for the eNB [31]. The eNB transmission power is reduced by assigning a coding scheme with lower rate to each user. Consequently their spectrum occupation is expanded.

All aspects and targets of scheduling strategies discussed in this subsection, as well as parameters they use for computing scheduling metrics, have been summarized in Table I.

V. CONCLUSIONS

This paper provides a broad survey on downlink packet allocation strategies in LTE networks. The various key issues that should be considered when designing a new scheduling scheme are extensively studied. Starting from channel independent strategies, most recently introduced QOS aware as well as energy aware solutions have also been studied.

REFERENCES

[1] 3GPP, Technical Specifications Group Radio Access Network - Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN), 3GPP TS 25.913. [Online]. Available: http:// www.3gpp.org

[2] D. McQueen, “The momentum behind LTE adoption,” *IEEE Communication Magazine*, vol. 47, no. 2, pp. 44–45, Feb. 2009.

[3] 3GPP, Technical Specifications Group Radio Access Network - Physical channel and Modulation (release 8), 3GPP TS 36.211.

[4] F Capozzi, G Piro, La Grieco, G boggia, P Camarda, “Downlink packet scheduling in lte cellular networks: Key Design Issues and A survey,” *IEEE Commun. Surv. Tutorials*, 2012.

[5] A. S. Tanenbaum, *Modern Operating Systems*, 3rd Ed. Upper Saddle River, Nj, Usa: Prentice Hall Press, 2007.

[6] S. Hussain, “Dynamic Radio Resource Management in 3GPP LTE”, Blekinge Institute Of Technology 2009.

[7] Demers, A., Keshav, S., And Shenker,”Analysis and Simulation Of A Fair Queuing Algorithm.” Symposium *Proceedings on Comm. Architectures and Protocols*" ACM, New York, USA, 1989.

[8] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, “Dynamic Packet Scheduling Performance in UTRA Long Term Evolution Downlink,” In *Proc. of International Symposium On Wireless Pervasive Comput.,*, Santorini, Greece, May 2008, pp. 308 – 313.

[9] A. Stolyar and K. Ramanan, “Largest Weighted Delay First Scheduling: Large Deviations and Optimality,” Annals of Applied Probability, Vol. 11, pp. 1–48, 2001.

[10] K. W. Choi, W. S. Jeon, and D. G. Jeong, “Resource Allocation in OFDMA Wireless Communications Systems Supporting Multimedia Services,“ *IEEE/ACM Trans. On Networking*, Vol. 17, pp. 926-935, June 2009.

[11] R. Almatarneh, M. H. Ahmed and O. A. Dobre, “'Frequency-Time Scheduling Algorithm For OFDMA Systems,” in *Proc. IEEE Canadian Conference on Electrical And Computer Engineering*, 2009, Pp. 766 – 771.

[12] Rabie K. Almatarneh, Mohamed H. Ahmed, Octavia A. Dobre,” Performance Analysis of Proportional Fair Sceduling in OFDMA Wireless Systems,”*in Proc. of 72nd IEEE Vehicular Tech. Conference*, 6-9 Sept., 2010.

[13] R. Kwan, C. Leung, And J. Zhang, “Proportional Fair Multiuser Scheduling In Lte,” *Proc. Of Ieee Signal Processing Letters*, Vol. 16, No. 6, Pp. 461–464, Jun. 2009

[14] G. Monghal, K. I. Pedersen, I. Z. Kovacs, and P. E. Mogensen, “QOS oriented Time And Frequency Domain Packet Schedulers for the Utran Long Term Evolution,” In *Proc. of IEEE Veh. Tech. Conf., VTC-Spring*, Marina Bay, Singapore, May 2008.

[15] D. Skoutas And A. Rouskas, “Scheduling with Qos Provisioning in Mobile Broadband Wireless Systems,” In *Wireless Conference (EW), 2010 European*, April 2010, Pp. 422 –428.

[16] Y. Zaki, T. Weerawardane, C. Gorg, And A. Timm-Giel, “Multi-Qos-Aware Fair Scheduling for LTE,” in *Veh. Technology Conference (Vtc Spring), 2011 IEEE 73rd*, May 2011, Pp. 1 –5.

[17] H. Ramli, R. Basukala, K. Sandrasegaran, And R. Patachaianand, “Performance Of Well Known Packet Scheduling Algorithms In The Downlink 3gpp Lte System,” In *Proc. Of Ieee Malaysia International Conf. On Comm., Micc*, Kuala Lumpur, Malaysia, 2009, Pp. 815 –820.

[18] R. Basukala, H. Mohd Ramli, and K. Sandrasegaran,“Performance Analysis of EXP/PF And M-LWDF in Downlink 3GPP LTE System,” In *Proc. of First Asian Himalayas International Conf. on Internet, Ah-ICI*, Kathmundu, Nepal, Nov. 2009, Pp. 1 –5.

[19] B. Sadiq, R. Madan, And A. Sampath, “Downlink Scheduling for Multiclass Traffic In LTE,” *Eurasip J. Wirel. Commun. Netw.*, Vol. 2009, pp. 9–9, 2009.

[20] Preethi Chandur, Karthik R and Krishna M. Sivalingam,”Performance Evaluation Of Scheduling Algorithms for Mobile Wimax Networks,” *Eighth IEEE Percom Workshop On Pervasive Wireless Networking* 2012.

[21] M. Iturralde, A. Wei, and A. Beylot, “Resource Allocation for Real Time Services using Cooperative Game Theory and A Virtual Token Mechanism in LTE Networks,” in *IEEE Personal Indoor Mobile Radio Communications*, Jan. 2012.

[22] Bilal Sadiq, Seung Jun Baek,”Delay-Optimal Opportunistic Scheduling and Approximations: The Log Rule”, IEEE/Acm Transactions on Networking, Vol. 19, No. 2, April 2011

[23] G. Piro, L. Grieco, G. Boggia, R. Fortuna, And P. Camarda, “Two-Level Downlink Scheduling For Real-Time Multimedia Services in LTE Networks,” in *IEEE Trans. Multimedia*, Vol. 13, No. 5, Oct. 2011.

[24] G. Monghal, D. Laselva, P. Michaelsen, and J. Wigard, “Dynamic Packet Scheduling for Traffic Mixes of Best Effort and VoIP Users in E-UTRAN Downlink,” in *Proc. of IEEE Veh. Tech. Conf., VTC-Spring*, Marina Bay, Singapore, May 2010, pp. 1 –5.

[25] S. Choi, K. Jun, Y. Shin, S. Kang, and B. Choi, “MAC Scheduling Scheme for VoIP Traffic Service in 3G LTE,” in *Proc. of IEEE Veh. Tech. Conf., VTC-Fall*, Baltimore, MD, USA, Oct. 2007.

[26] Y. Fan, P. Lunden, M. Kuusela, and M. Valkama, “Efficient Semi-Persistent Scheduling for VoIP on EUTRA Downlink,” in *Proc. of IEEE Veh. Tech. Conf., VTC-Fall*, Calgary, Alberta, Sep. 2008.

[27] Y.-S. Kim, “An Efficient Scheduling Scheme to Enhance the Capacity of VoIP Services in Evolved UTRA Uplink,”*EURASIP J. Wirel. Commun. Netw.*, vol. 2008, 2008.

[28] S. Saha and R. Quazi, “Priority-coupling-a semi-persistent MAC scheduling scheme for VoIP traffic on 3G LTE,” in *Proc.of Conf. on Telecommunications, ConTEL*, Zagreb, Croatia, 2009, pp. 325 –329.

[29] Kuo chang,Hwang Cheng, Chih-Cheng Tseng,” Energy-Efficient DRX scheduling for QOS traffic in LTE networks ,”*IEEE 9th International Symposium on parallel and distributed processing with applications* ,pp 213-218 ,2011.

[30] D. Sabella, M. Caretti, and R. Fantini, “Energy efficiency evaluation of state of the art packet scheduling algorithms for lte,” *in Proc. of IEEE European Wireless Conf., EW*, pp. 1–4, Apr. 2011.

[31] S. Videv and H. Haas, "Energy-efficient scheduling and bandwidth-energy efficiency trade-off with low load," in *Proc. IEEE International Conf. on Comm.*, Jun. 2011, pp. 1 –5

TABLE II
PARAMETERS USED BY EACH SCHEDULER

Author/Year	**Scheduling Strategy**	**Requested Bit rate**	**SINR**	**Average Data Rate or Throughput**	**HOL Packet Delay**	**Target Delay**	**Max PLR**	**Queue Size**
P. Kela *et. al.* (2008)	MaxT [8]		X					
Rabie K. Almatarneh(2010)	PF [12]		X	X				
R. Kwan *et. al.* (2009)	PFMultiuser [13]		X	X				
G. Monghal *et. al.* (2008)	PSS/PFsch [14]	X	X	X				
D. Skoutas *et. al.* (2010)	[15]	X	X					
Y. Zaki *et. al.* (2011)	[16]	X	X					
H. Ramli *et. al.* (2009)	MLWDF [17]		X	X	X	X	X	
R. Basukala *et. al.* (2009)	EXP/PF [18]		X	X	X	X	X	
M. Iturralde *et. al.* (2012)	[21]		X	X	X	X		
B. Sadiq *et. al.* (2009)	EXP Rule [19]		X	X	X	X		
Seung Jun Baek *et. al.* (2011)	LOG Rule [19][22]		X	X	X	X		
G. Piro *et. al.* (2011)	FLS [23]					X		X