# Mohammad Mahmoodian, Paper Reading #2

## III. SCHEDULING IN LTE SYSTEMS

Resource allocation for each UE is usually based on the comparison of per-RB metrics: the *k*-th RB is allocated to the *j*-th user if its metric *mj,k* is the biggest one, i.e., if it satisfies the equation:

$$m_{j,k} = \max_i \{m_{i,k}\} \ .$$

This metric depends on:

- *Status of transmission queues:* The longer the queue, the higher the metric.
- *Channel Quality:* The higher the expected throughput, the higher the metric.
- *Resource Allocation History*: The lower the past achieved throughput, the higher the metric.
- *Buffer State*: The higher the available space in the receiving buffer, the higher the metric.
- *Quality of Service Requirements.*

Every TTI the scheduler performs the allocation decision valid for the next TTI and sends such information to UEs and informs UEs about RBs allocated for data transmission on the PDSCH in the downlink direction.

The whole process of downlink packet scheduler can be divided in a sequence of operations that are repeated, in general, every TTI:
1) each UE decodes the reference signals, computes the CQI, and sends it to the eNB.
2) The eNB uses the CQI information for the allocation decisions and fills up a RB “allocation mask”.
3) The AMC module selects the best MCS that should be used for the data transmission by scheduled users.
4) The information about these users, the allocated RBs, and the selected MCS are sent to the UEs on the PDCCH.
5) Each UE reads the PDCCH payload and, in case it has been scheduled, accesses to the proper PDSCH payload.

***A. Key Design Aspects***: We present a list of the main design factors that always should be taken into account before defining an allocation policy for LTE.

1) *Complexity and Scalability*: Low complexity and scalability are therefore fundamental requirements for limiting processing time and memory usage. For this reason, FDPS decisions are usually based on the computation of *per-RB* metrics for each user.
2) *Spectral Efficiency*: Effective utilization of radio resources is one of the main goals to be achieved. One of the most used efficiency indicators is the *user goodput* that is a measure of the actual transmission data rate without including layer two overheads and packet retransmissions due to physical errors.
3) *Fairness*: Fairness is therefore a major requirement that should be taken into account to guarantee minimum performance also to the cell-edge users.

4) *QoS Provisioning*: QoS constraints may vary depending on the application and they are usually mapped into some parameters: minimum guaranteed bitrate, maximum delivering delay, and packet loss rate.

***B. Practical limitations in real LTE systems***

1) *Uplink Limitations*: the scheduler for the uplink has limited degrees of freedom, it has to allocate contiguous RBs to each user without the possibility of choice among the best available ones.
2) *Control Overhead*: As a consequence, the amount of resources dedicated to the PDCCH is limited, thus decreasing the degrees of freedom for the downlink scheduler. PDCCH overhead can be reduced using specific RRM procedures.
3) *Limitations on the Multi-User Diversity Gain*: The chosen CQI reporting scheme has great impact on the multi-user diversity gain, as it defines the time and frequency resolution of the channel quality information available at the scheduler.
4) *Energy Consumption*: Energy saving is a required feature for mobile terminals, and in LTE it is achieved through Discontinuous Reception (DRX) methods.

C. ***Persistent and semi-persistent scheduling***: As already pointed out, dynamic frequency domain strategies have the main benefit of exploiting multi-user diversity gain, but this comes at the cost of increased control overhead, due to the need of forwarding DCI messages to scheduled users every TTI. For this reason, especially in scenarios with high traffic load, the limited amount of radio resources dedicated to control information transmission can become a bottleneck, with consequent degradation of QoS provisioning capabilities.

## IV. SCHEDULING STRATEGIES FOR LTE DOWNLINK

In this section, we will illustrate different allocation strategies introduced for LTE systems, they differ in terms of input parameters, objectives, and service targets. We have classified them in five groups of strategies: (i) channel-unaware; (ii) channel-aware/QoS-unaware; (iii) channel-aware/QoS-aware; (iv) semi-persistent for VoIP support; and (v) energy-aware.

We will first illustrate channel unaware approaches, which have been historically adopted to face fairness, flow priorities, and deadline expiration in all packet switching networks.

Channel aware schedulers (with and without QoS support) are then introduced and deeper analyzed, because more suitable for wireless environments.

A. ***Channel-unaware Strategies***: Firstly introduced in wired networks, channel unaware strategies are based on the assumption of time-invariant and error-free transmission media. While their direct application in LTE is not realistic, they are typically used jointly with channel-aware approaches to improve system performance.

1) *First In First Out*: The <u>simplest</u> case of channel unaware allocation policy serves users according to the order of resource requests, exactly like a First In First Out (FIFO) queue. This technique is very simple, but both <u>inefficient</u> and <u>unfair</u>.
2) *Round Robin*: It performs <u>fair</u> sharing of time resources among users Of course, this approach is not fair in terms of user throughput.
3) *Blind Equal Throughput*: <u>Throughput Fairness</u> can be achieved with Blind Equal Throughput (BET) which stores the past average throughput achieved by each user and uses it as metric.

4) *Resource Preemption*: The idea is that transmission queues are grouped in several priority classes, and a queue belonging to a given class cannot be served until all queues having higher priorities are empty.
5) *Weighted Fair Queuing*: An alternative way to introduce priorities avoiding the possibility of starvation is through the usage of an approximation of the well-known Weighted Fair Queuing (WFQ) approach.
6) *Guaranteed Delay*: Guaranteed delay services, in particular, require that each packet has to be received within a certain deadline to avoid packet drops. Earliest Deadline First (EDF) and Largest Weighted Delay First (LWDF) are two policies.

B. ***Channel-aware/QoS-unaware Strategies***: Thanks to CQI feedbacks, which are periodically sent (from UEs to the eNB) using ad hoc control messages, the scheduler can estimate the channel quality perceived by each UE; hence, it can predict the maximum achievable throughput.

1) *Maximum Throughput*: The strategy known as Maximum Throughput (MT) aims at maximizing the overall throughput by assigning each RB to the user that can achieve the maximum throughput (indeed) in the current TTI. MT is obviously able to maximize cell throughput, but, on the other hand, it performs unfair resource sharing since users with poor channel conditions will only get a low percentage of the available resources.
2) *Proportional Fair Scheduler*: A typical way to find a trade-off between requirements on fairness and spectral efficiency is the use of Proportional Fair (PF) scheme. The idea is that the past average throughput can act as a weighting factor of the expected data rate, so that users in bad conditions will be surely served within a certain amount of time.
3) *Throughput to Average*: The scheme Throughput To Average (TTA) can be considered as intermediate between MT and PF.
4) *Joint Time and Frequency domain schedulers*:
   a two-step technique for distributing radio resources is presented:
   1) At first, a Time Domain Packet Scheduler (TDPS) selects a subset of active users in the current TTI among those connected to the eNB;
   2) then, RBs are physically allocated to each user by a FDPS.

   The final allocation decision is the outcome of the decisions of two schedulers (one in the time domain and the other one in the frequency domain) that work in series.
5) *Delay sensitivity*: The idea is that, even if we neglect the problem of packet deadline expiration (typical of real-time flows), the average data delivering delay can be taken as the overall key performance indicator.
6) *Buffer-aware schedulers*: makes use of buffer status information reported by the user to the eNB and of traffic statistics for setting dynamic priorities associated to each MAC queue.
7) *Performance evaluation of most relevant channel aware/QoS-unaware strategies*: In order to enrich the analysis of the relevant MT, PF, TTA, and PF-PF schemes, an accurate quantitative performance assessment has been carried out using a system level simulator.
8) *General considerations on channel-aware/QoS-unaware*: Channel-awareness is a fundamental concept for achieving high performance in a wireless environment, and it can be used by exploiting RRM features such as CQI reporting and link adaptation.