Downlink Packet Scheduling in LTE Cellular Network-Key Design Issues and a Survey

710183124 施旻錞

F. Capozzi, G. Piro, Student Member, IEEE,
L.A. Grieco, Member, IEEE,
G. Boggia, Senior Member, IEEE, and P. Camarda

Outline

- Abstract
- Introduction
- Overview on LTE Network
- Scheduling in LTE System
- Scheduling Strategies for LTE Downlink
- New Direction and Future Challenges
- Conclusions and Lesson Learned

Abstract

- LTE systems represent an important milestone towards the so called 4G cellular networks.
- This paper provides an overview on the key issues that arise in the design of a resource allocation algorithm for LTE networks.

Introduction

- The Radio Resource Management (RRM) block exploits a mix of advanced MAC and Physical functions.
- The design of effective resource allocation strategies becomes crucial：
  - 1. meet the system performance targets.
  - 2. satisfy user needs according to specific Quality of Service (QoS) requirements.
- The downlink shared channel is the key facets of LTE scheduling.

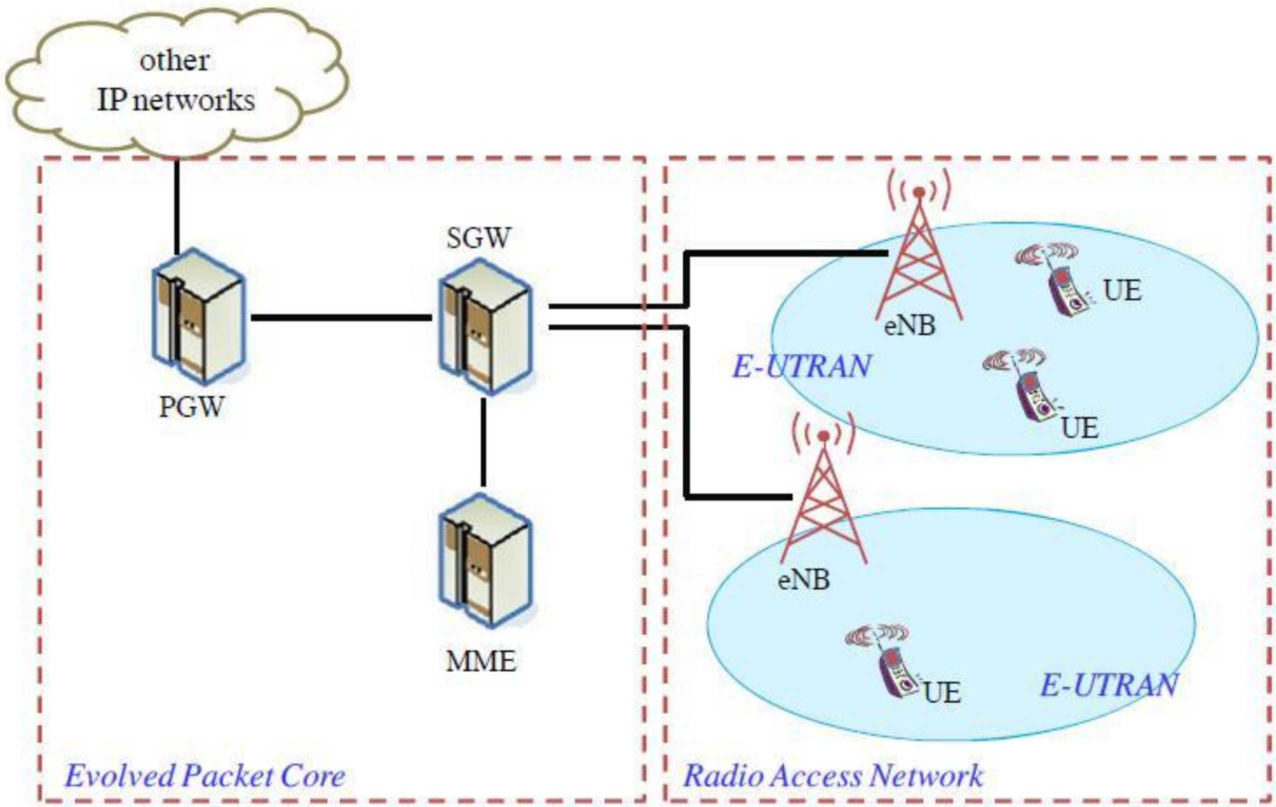
Overview on LTE Network

MAIN LTE PERFORMANCE TARGETS

Peak Data Rate	- Downlink: 100 Mbps - Uplink: 50 Mbps
Spectral Efficiency	2 - 4 times better than 3G systems
Cell-Edge Bit-Rate	Increased whilst maintaining same site locations as deployed today
User Plane Latency	Below 5 ms for 5 MHz bandwidth or higher
Mobility	- Optimized for low mobility up to 15 km/h - High performance for speed up to 120 km/h - Maintaining connection up to 350 km/h
Scalable Bandwidth	From 1.4 to 20 MHz
RRM	- Enhanced support for end-to-end QoS - Efficient transmission and operation of higher layer protocols
Service Support	- Efficient support of several services (e.g., web-browsing, FTP, video-streaming, VoIP) - VoIP should be supported with at least a good quality as voice traffic over the UMTS network

A. System Architecture and Radio Access Network

- The LTE system is based on a flat architecture.



Legend

PGW: Packet Gateway
MME: Mobility Management Entity
SGW: Service gateway
E-UTRAN: Evolved Universal Terrestrial Radio Access Network

B. Radio Bearer Management

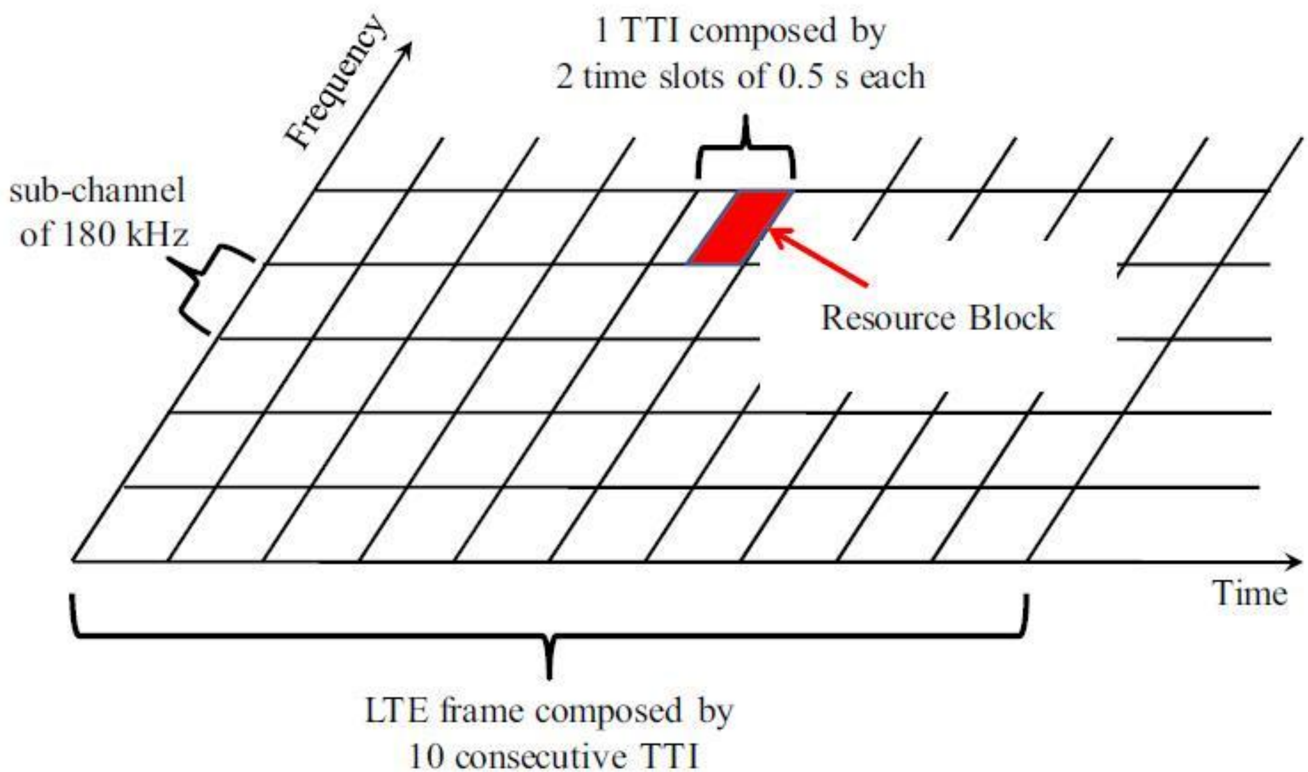
- A radio bearer is in charge of managing QoS provision on the E-UTRAN Interface.
- When an UE joins the network, a default bearer is created for basic connectivity and exchange of control messages.
- Dedicated bearers are set up every time a new specific service is issued.

STANDARDIZED QoS CLASS IDENTIFIERS FOR LTE

QCI	Resource Type	Priority	Packet Delay Budget [ms]	Packet Loss Rate	Example services
1	GBR	2	100	10^{-2}	Conversational voice
2	GBR	4	150	10^{-3}	Conversational video (live streaming)
3	GBR	5	300	10^{-6}	Non-Conversational video (buffered streaming)
4	GBR	3	50	10^{-3}	Real time gaming
5	non-GBR	1	100	10^{-6}	IMS signaling
6	non-GBR	7	100	10^{-3}	Voice, video (live streaming), interactive gaming
7	non-GBR	6	300	10^{-6}	Video (buffered streaming)
8	non-GBR	8	300	10^{-6}	TCP based (e.g., WWW, e-mail), chat, FTP, P2P file sharing
9	non-GBR	9	300	10^{-6}	

C. Physical layer

- Radio spectrum access is based on the Orthogonal Freq. Division Multiplexing (OFDM) scheme.
  - Single Carrier Freq.Multiple Access (SC-FDMA) and OFDMA are used in uplink and downlink directions.



Cont.

- LTE radio interface supports two types of frame structure
  - Frequency Division Duplex(FDD):allowing simultaneous downlink and uplink data transmissions
  - Time Division Duplex (TDD):divided into two consecutive half-frames
- The selection of the TDD frame configuration is performed by the RRM module.

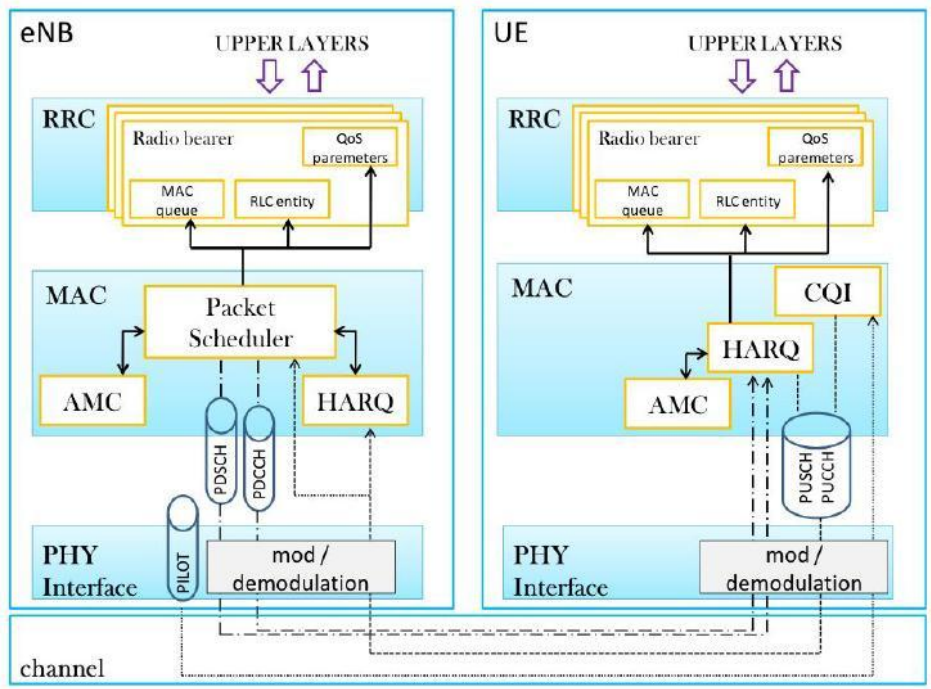
TDD FRAME CONFIGURATIONS

configuration number	sub-frame number: 1st half frame					2nd half frame				
	0	1	2	3	4	5	6	7	8	9
0	D	S	U	U	U	D	S	U	U	U
1	D	S	U	U	D	D	S	U	U	D
2	D	S	U	D	D	D	S	U	D	D
3	D	S	U	U	U	D	D	D	D	D
4	D	S	U	U	D	D	D	D	D	D
5	D	S	U	D	D	D	D	D	D	D
6	D	S	U	U	U	D	S	U	U	D

D = downlink sub-frame; U = uplink sub-frame; S = Special sub-frame.

D. Radio Resource Management

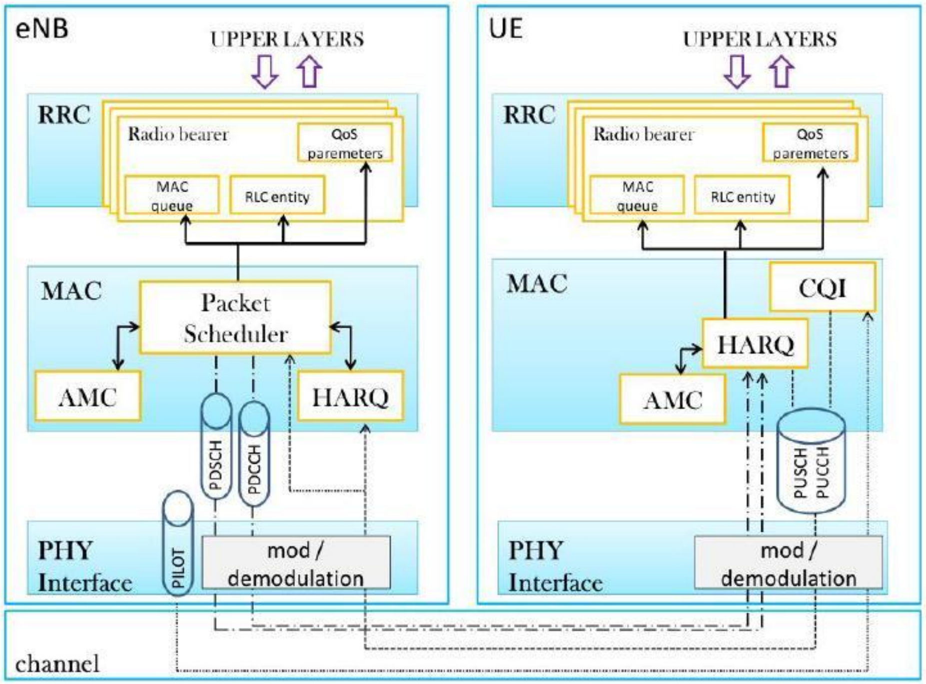
- 1. CQI reporting
  - CQI enables the estimation of the quality of the downlink channel at the eNB.
  - Each CQI is calculated as a quantized and scaled measure of the experienced Signal to Interference plus Noise Ratio (SINR).



Legend
AMC: Adaptive Modulation and Coding
RLC: Radio Link Control
PDCCH: Physical Downlink Control Channel
PUSCH: Physical Uplink Schared Channel
CQI: Channel Quality Indicator
HARQ: Hybrid Automatic Repeat Request
RRC: Radio Resource Control
PDSCH: Physical Downlink Shared Channel
PUCCH: Physical Uplink Control Channel

Cont.

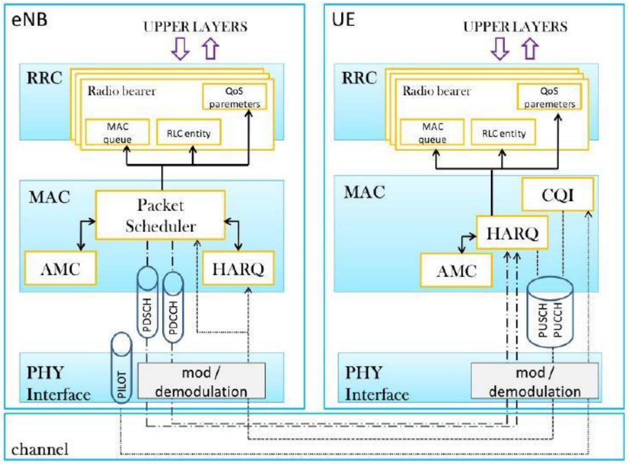
- 2. AMC and Power Control:
  - AMC module selects the proper Modulation and Coding Scheme (MCS) trying to maximize the supported throughput with CQI and a given target Block Error Rate (BLER).



Legend
AMC: Adaptive Modulation and Coding
RLC: Radio Link Control
PDCCH: Physical Downlink Control Channel
PUSCH: Physical Uplink Schared Channel
CQI: Channel Quality Indicator
HARQ: Hybrid Automatic Repeat Request
RRC: Radio Resource Control
PDSCH: Physical Downlink Shared Channel
PUCCH: Physical Uplink Control Channel

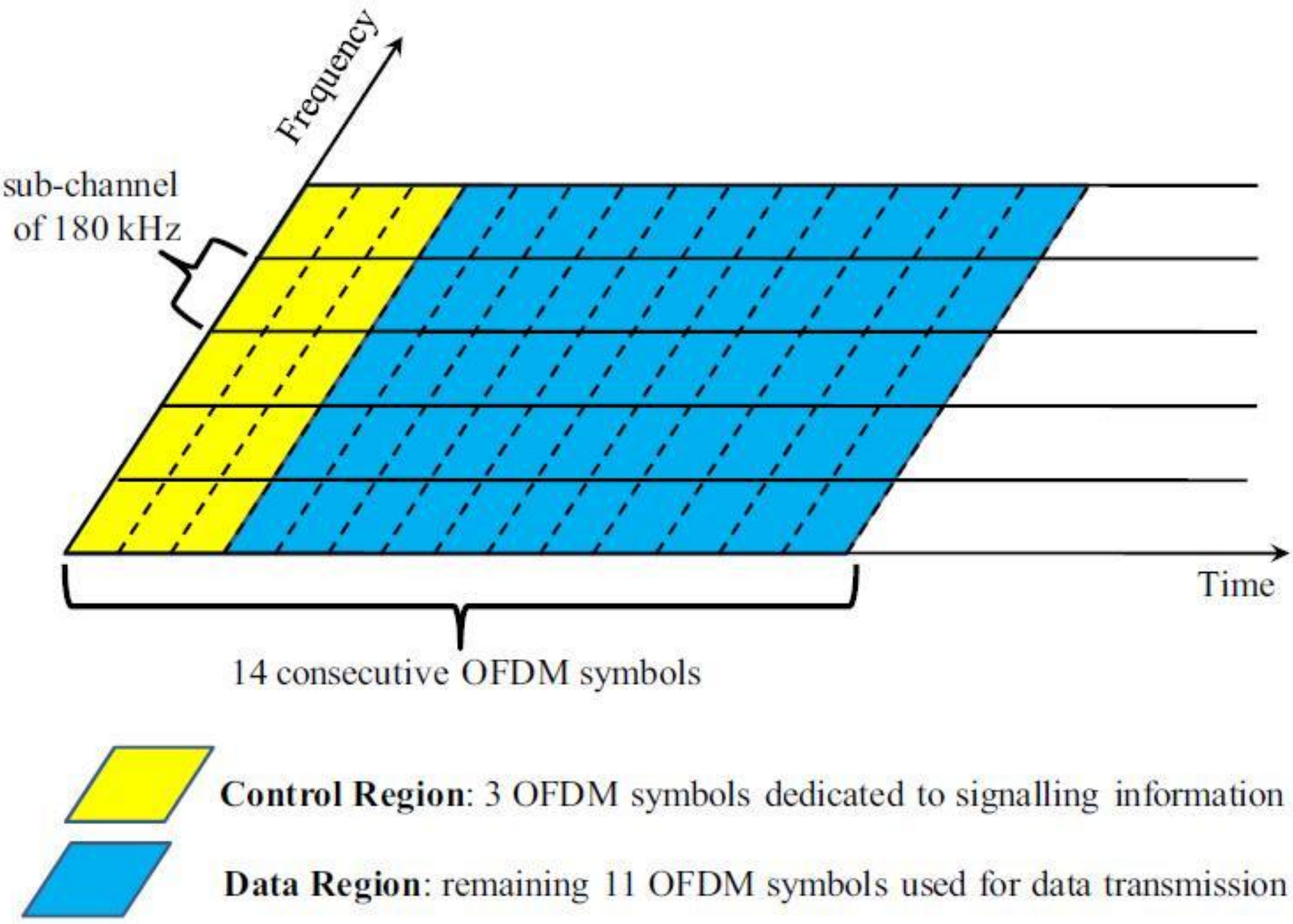
Cont.

- 3. Physical Channels:
  - Downlink data are transmitted by the eNB over the Physical Downlink Shared Channel(PDSCH).
  - Physical Downlink Control Channel (PDCCH) carries assignments for downlink resources and uplink grants, including the used MCS.
  - The Physical Uplink Control Channel (PUCCH) and the Physical Uplink Shared Channel (PUSCH) are defined in the uplink direction.
  - Due to single carrier limitations, simultaneous transmission on both channels is not allowed.



Legend
AMC: Adaptive Modulation and Coding
RLC: Radio Link Control
PDCCH: Physical Downlink Control Channel
PUSCH: Physical Uplink Schared Channel
CQI: Channel Quality Indicator
HARQ: Hybrid Automatic Repeat Request
RRC: Radio Resource Control
PDSCH: Physical Downlink Shared Channel
PUCCH: Physical Uplink Control Channel

Cont.


Frequency
sub-channel
of 180 kHz
Time
14 consecutive OFDM symbols
Control Region: 3 OFDM symbols dedicated to signalling information
Data Region: remaining 11 OFDM symbols used for data transmission

Cont.

- 4. HARQ(Hybrid Automatic Repeat request)
  - It is the retransmission procedure at MAC layer, based on stop-and-wait algorithm.
  - A NACK is sent over the PUCCH when a packet transmitted by the eNB is unsuccessfully decoded at the UE.
  - the eNB will perform a retransmission, sending the same copy of the lost packet.
  - the UE will try to decode the packet combining the retransmission with the original version, and will send an ACK message to the eNB upon a successfully decoding.

Scheduling in LTE System

- Multi-user scheduling is one of the main feature in LTE systems.
- Resource allocation for each UE is usually based on the comparison of per-RB metrics : transmission priority of each user on a specific RB

$$m_{j,k} = \max_{i}\{m_{i,k}\}$$

Cont.

- Status of transmission queues
- Channel Quality
- Resource Allocation History
- Buffer State
- Quality of Service Requirements

A. Key design aspects

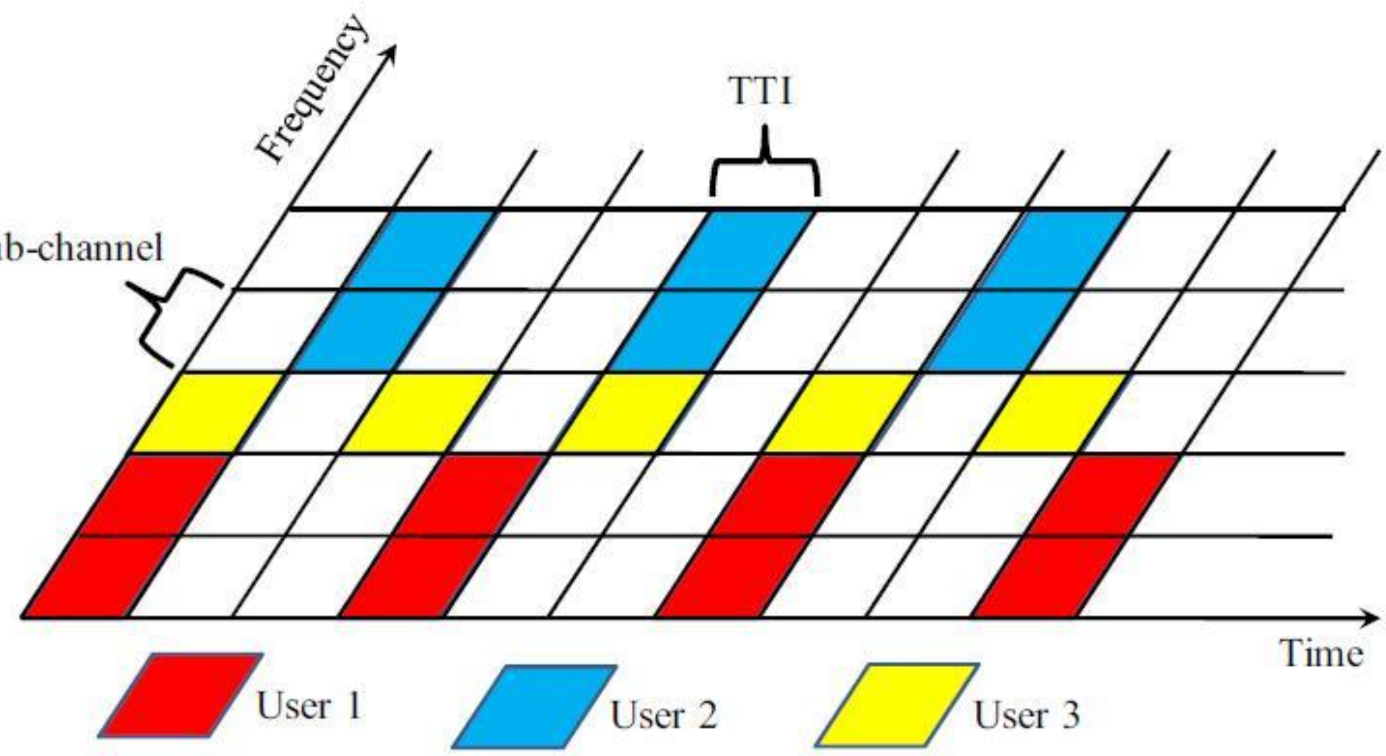
- Complexity and Scalability
- Spectral efficiency
- Fairness
- QoS Provisioning

B. Practical limitations in real LTE systems

- Uplink Limitations
- Control overhead
- Limitation on multi-user diversity gain
- Energy Consumption: Discontinuous Reception (DRX) methods

C. Persistent and semi-persistent scheduling

- Dynamic frequency domain strategies have the main benefit of exploiting multi-user diversity gain, but this comes at the cost of increased control overhead.



Scheduling Strategies for LTE Downlink

- A. Channel-unaware Strategies
- 1. First In First Out
  - t : current time
  - T_i : the time instant when the request was issued by the i-th user

$$m_{i,k}^{FIFO} = t - T_i$$

  - very simple, but both inefficient and unfair.

A. Channel-unaware Strategies

- 2. Round Robin
  - t : current time
  - T_i : last time when the user was served

$$m_{i,k}^{FIFO} = t - T_i$$

  - The concept of fairness is related to the amount of time in which the channel is occupied by users but not in terms of user throughput.

A. Channel-unaware Strategies

- 3. Blind equal throughput

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

with

$$\overline{R^i}(t) = \beta\overline{R^i}(t-1) + (1-\beta)r^i(t)$$

where $0 \leq \beta \leq 1$.

NOTATIONS USED FOR SCHEDULING METRICS

Expression	Meaning
$m_{i,k}$	Generic metric of the i-th user on the k-th RB
$r^i(t)$	Data-rate achieved by the i-th user at time t
$\overline{R^i}(t)$	Past average throughput achieved by the i-th user until time t
$\overline{R^i_{sch}}(t)$	Average throughput achieved by data flow of the i-th user when scheduled
$D_{HOL,i}$	Head of Line Delay, i.e., delay of the first packet to be transmitted by the i-th user
τ_i	Delay Threshold for the i-th user
δ_i	Acceptable packet loss rate for the i-th user
$d^i(t)$	Wideband Expected data-rate for the i-th user at time t
$d^i_k(t)$	Expected data-rate for the i-th user at time t on he k-th RB
Γ^i_k	Spectral efficiency for the i-th user over the k-th RB

A. Channel-unaware Strategies

- 4. Resource preemption
  - This approach can be fruitfully exploited to handle the differentiation among QoS (high priority) and non-QoS flows

A. Channel-unaware Strategies

- 5. Weighted Fair Queuing
  - $m_{i,k}^{RR}$ is the RR specific metric for the i-th user

$$m_{i,k}^{WFQ} = w_i \cdot m_{i,k}^{RR}$$

  - No starvation is possible because the RR metric would control that the waiting time of a given user does not indefinitely grow.

A. Channel-unaware Strategies

- 6. Guaranteed Delay
  - require that each packet has to be received within a certain deadline to avoid packet drops.

$$m_{i,k}^{EDF} = \frac{1}{(\tau_i - D_{HOL,i})} .$$

NOTATIONS USED FOR SCHEDULING METRICS

Expression	Meaning
$m_{i,k}$	Generic metric of the i-th user on the k-th RB
$r^i(t)$	Data-rate achieved by the i-th user at time t
$\overline{R^i}(t)$	Past average throughput achieved by the i-th user until time t
$\overline{R^i_{sch}}(t)$	Average throughput achieved by data flow of the i-th user when scheduled
$D_{HOL,i}$	Head of Line Delay, i.e., delay of the first packet to be transmitted by the i-th user
τ_i	Delay Threshold for the i-th user
δ_i	Acceptable packet loss rate for the i-th user
$d^i(t)$	Wideband Expected data-rate for the i-th user at time t
$d^i_k(t)$	Expected data-rate for the i-th user at time t on he k-th RB
Γ^i_k	Spectral efficiency for the i-th user over the k-th RB

A. Channel-unaware Strategies

- 7. General considerations on channel-unaware strategies
  - We referred to them as channel-unaware schedulers , because their working rationales do not account for channel quality variations, making them unsuitable in cellular networks.

B. Channel-aware/QoS-unaware Strategies

- 1. Maximum Throughput
  - maximizing the overall throughput by assigning each RB to the user that can achieve the maximum throughput (indeed) in the current TTI.
  - $d_t^i(t)$: the achievable throughput expected for the i-th user at the t-th TTI over the k-th RB

$$m_{i,k}^{MT} = d_k^i(t).$$

  - MT is obviously able to maximize cell throughput, but it performs unfair resource sharing since users with poor channel conditions.

B. Channel-aware/QoS-unaware Strategies

- 2. Proportional Fair Scheduler
  - MT : Maximum Throughput
  - BET : Blind equal throughput

$$m_{i,k}^{PF} = m_{i,k}^{MT} \cdot m_{i,k}^{BET} = d_k^i(t)/\overline{R^i}(t-1) \ .$$

  - It find a trade-off between requirements on fairness and spectral efficiency.

B. Channel-aware/QoS-unaware Strategies

- 3. Throughput to Average
  - $d_t^i(t)$: the achievable throughput expected for the i-th user at the t-th TTI over the k-th RB
  - $d^i(t)$: the achievable throughput expected for the i-th user at the t-th TTI over all the bandwidth.

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

  - It quantifies the advantage of allocating a specific RB, guaranteeing that the best RBs are allocated to each user.
  - It exploits channel awareness for guaranteeing a minimum level of service to every user.

B. Channel-aware/QoS-unaware Strategies

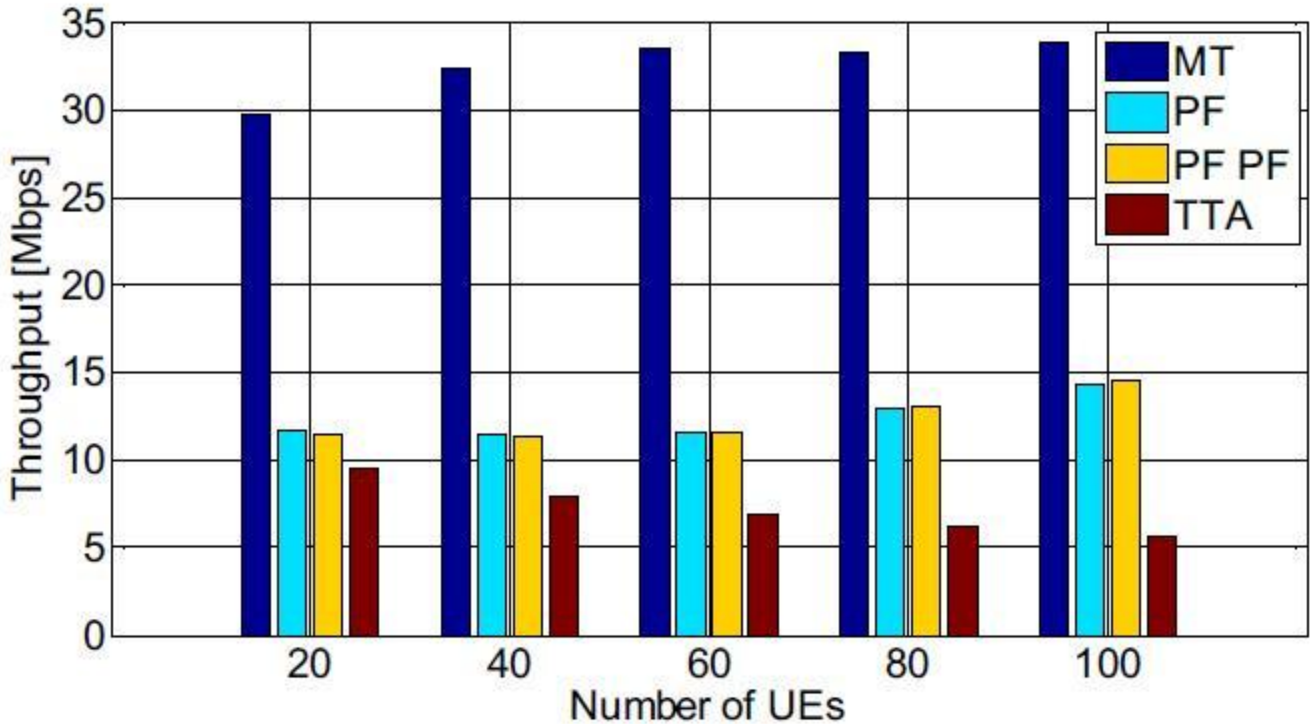
- 4. Joint Time and Frequency domain schedulers
  - i. at first, a Time Domain Packet Scheduler (TDPS) selects a subset of active users in the current TTI among those connected to the eNB.
  - ii. Then, RBs are physically allocated to each user by a FDPS.
  - The computational complexity at the FDPS is reduced, due to the number of candidate users for resource allocation decreases.
  - For each phase a different policy can be selected.

B. Channel-aware/QoS-unaware Strategies

- 5. Buffer-aware Schedulers
  - Buffer-Aware Traffic-Dependent (BATD) deal with the packet dropping probability due to a receiver buffer overflow.
  - BATD 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.

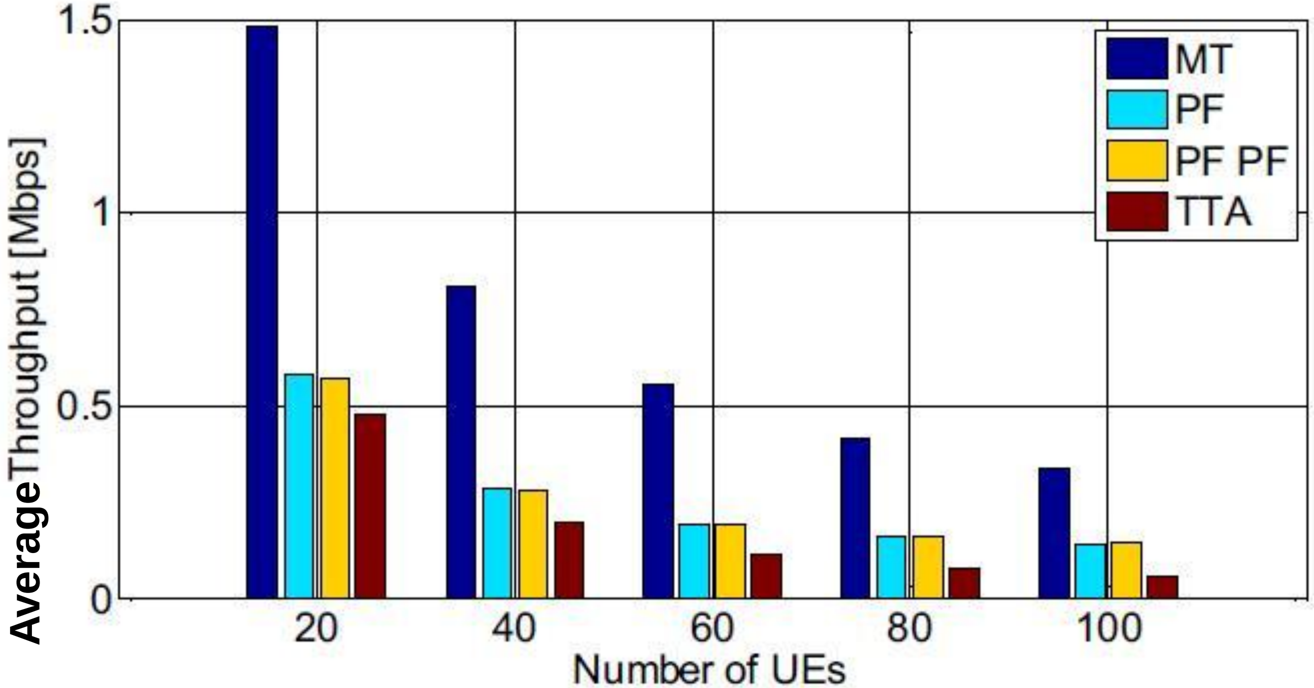
B. Channel-aware/QoS-unaware Strategies

- 6. Performance evaluation of most relevant channel-aware/QoS-unaware strategies



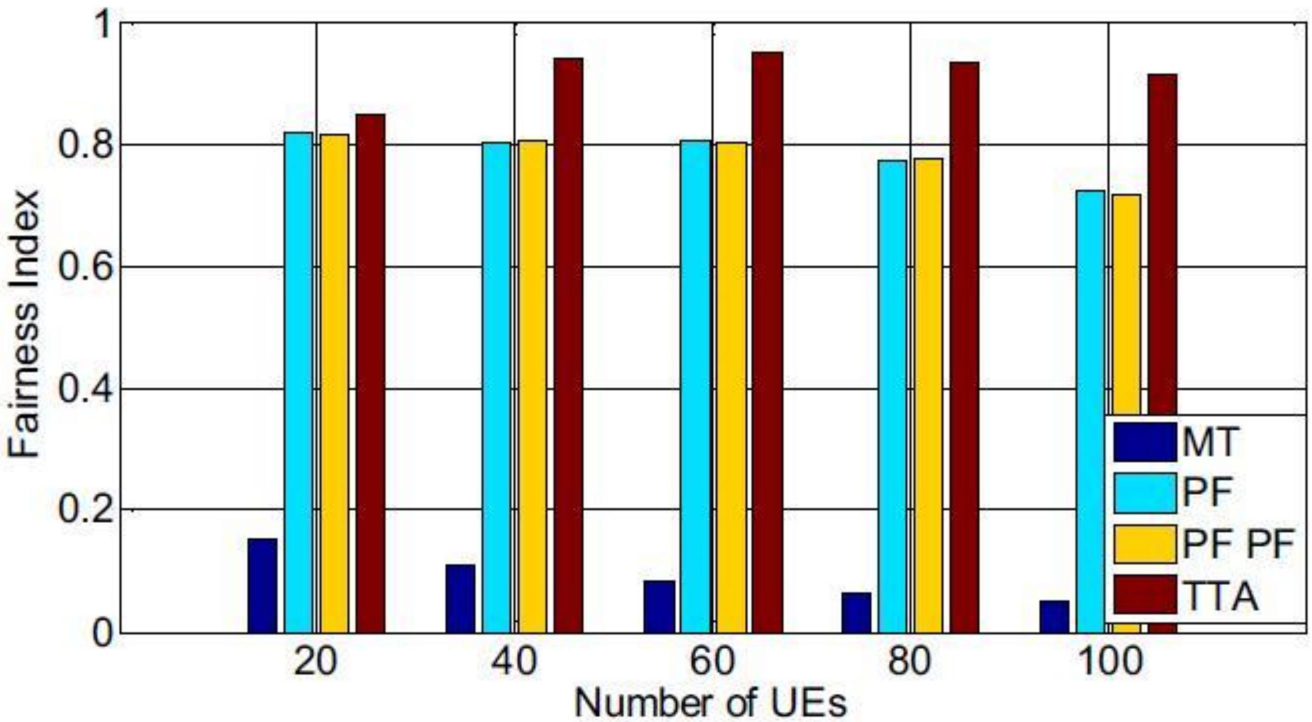
Aggregate cell throughput for different number of users

B. Channel-aware/QoS-unaware Strategies



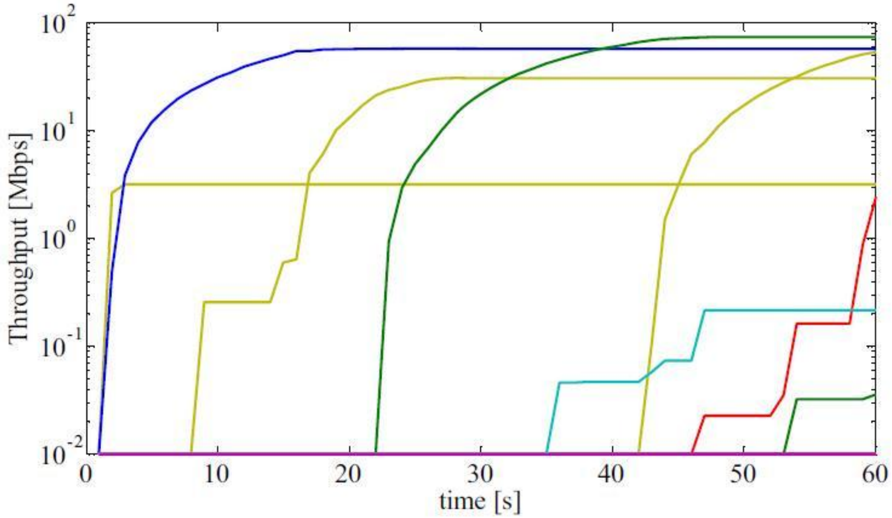
Average user throughput for different number of users

B. Channel-aware/QoS-unaware Strategies

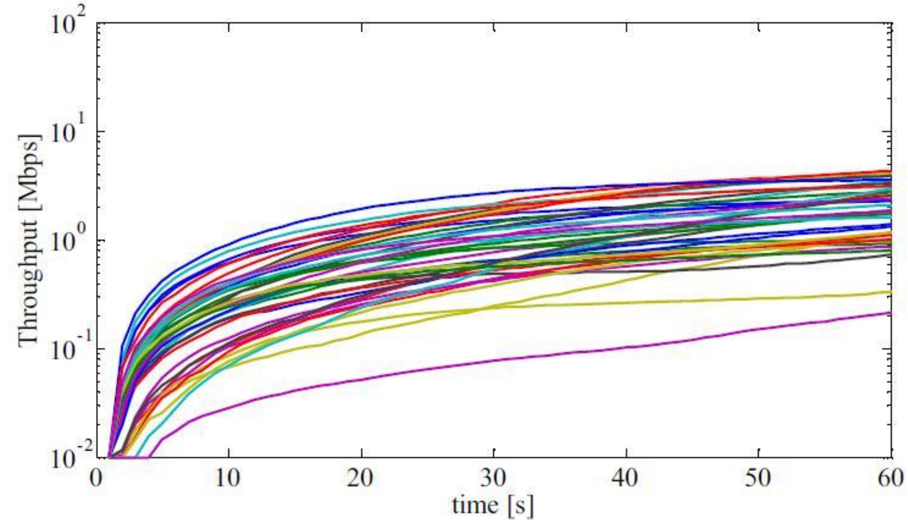


Fairness index for different number of users

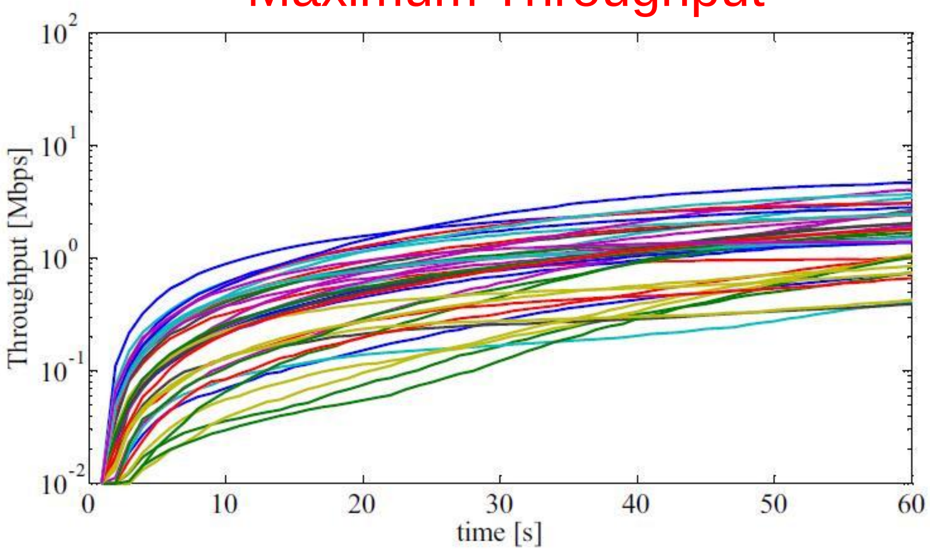
B. Channel-aware/QoS-unaware Strategies



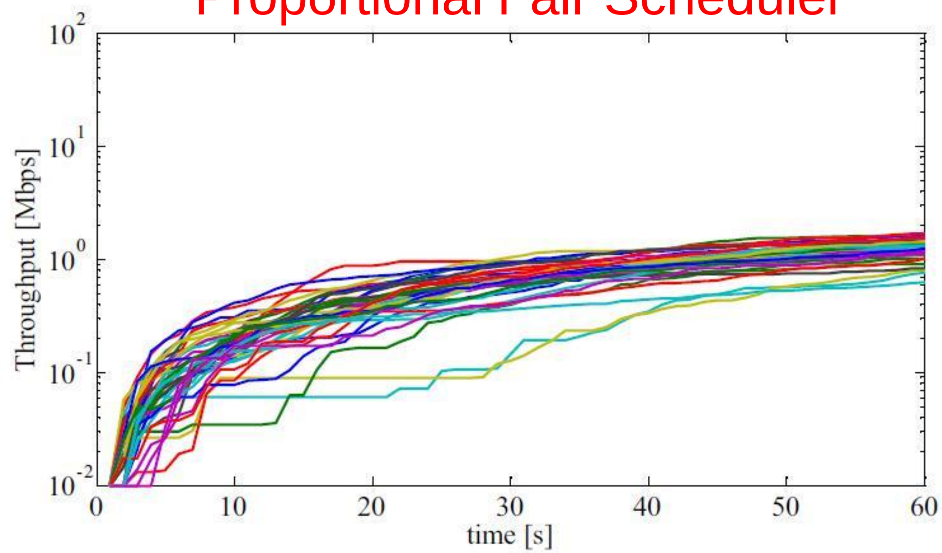
Maximum Throughput



Proportional Fair Scheduler



PF-PF



Throughput to average

B. Channel-aware/QoS-unaware Strategies

- 7. General considerations on channel-aware/QoS-unaware strategies
  - Channel-awareness is a fundamental concept for achieving high performance in a wireless environment.

C. Channel-aware/QoS-aware Strategies

- 1. Schedulers for Guaranteed Data-Rate
  - It works in both time and frequency domains
  - For the time domain, users with flows below their target bitrate form a high priority set, the rest of the users forms a lower priority set.
  - Users belonging to first and second sets are managed by using Blind equal throughput(BET) and Proportional Fair Schedule (PF) algorithms
  - The approach use ordered lists to prioritize the most delayed flows and to meet the FDPS is QoS-unaware.

C. Channel-aware/QoS-aware Strategies

- 2. Schedulers for Guaranteed Delay Requirements
- non real-time : Modified Largest Weight Delay First(M-LWDF)
  - $D_{HOL,i}$: Head of Line Delay

$$m_{i,k}^{M-LDWF} = \alpha_i D_{HOL,i} \cdot m_{i,k}^{PF} = \alpha_i D_{HOL,i} \cdot \frac{d_k^i(t)}{\overline{R^i}(t-1)}$$

  - M-LWDF uses information about the accumulated delay for shaping the behavior of PF, assuring a good balance among spectral efficiency, fairness, and QoS provisioning.

- Real-time : Exponential/PF (EXP/PF)
  - N_{rt} : the number of active downlink real-time flows
  - EXP/PF distinguishes between real-time and best effort flows

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

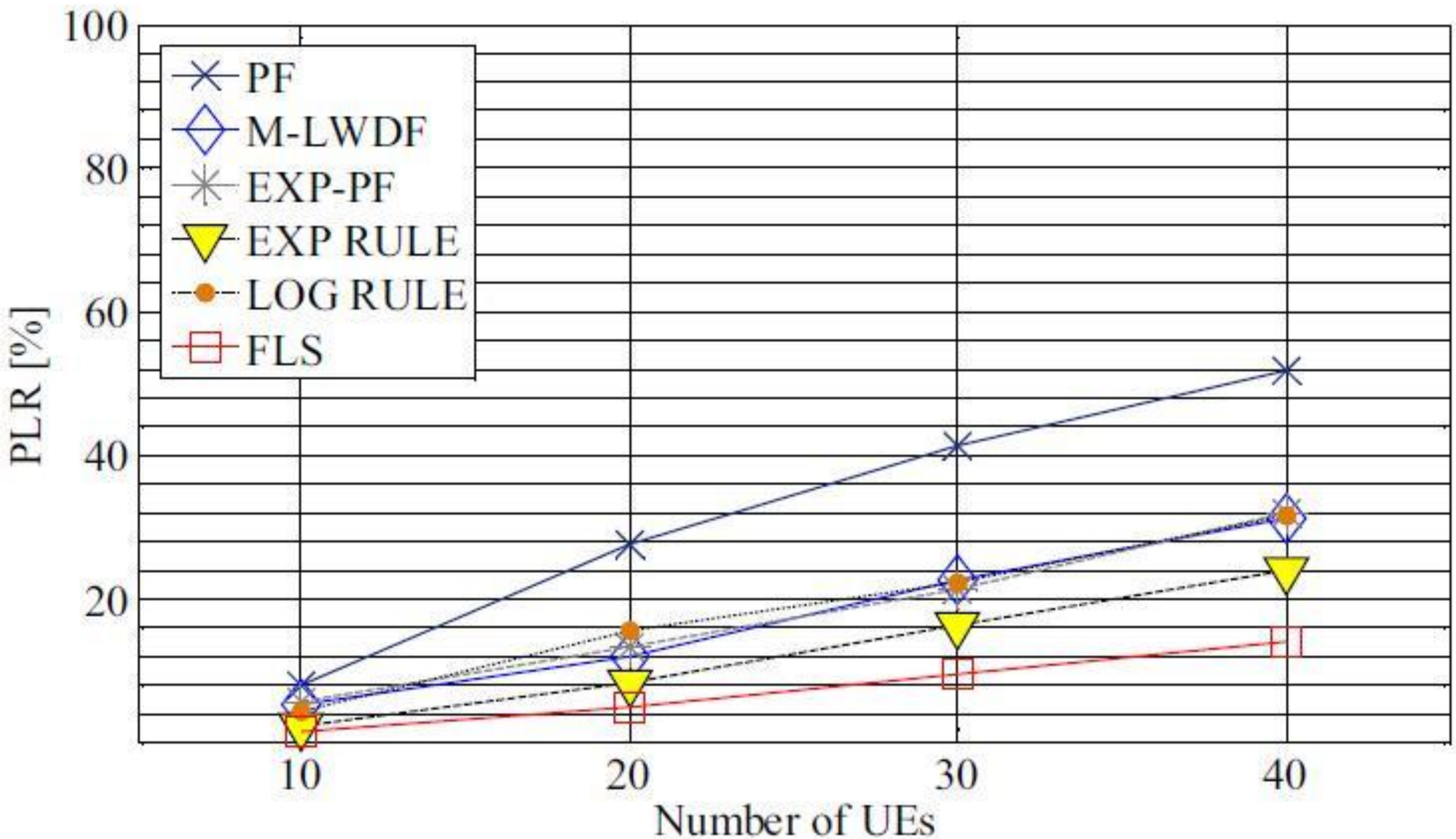
where

$$\chi = \frac{1}{N_{rt}} \sum_{i=1}^{N_{rt}} \alpha_i D_{HOL,i}$$

  - They both try to guarantee good throughput and an acceptable level of fairness.

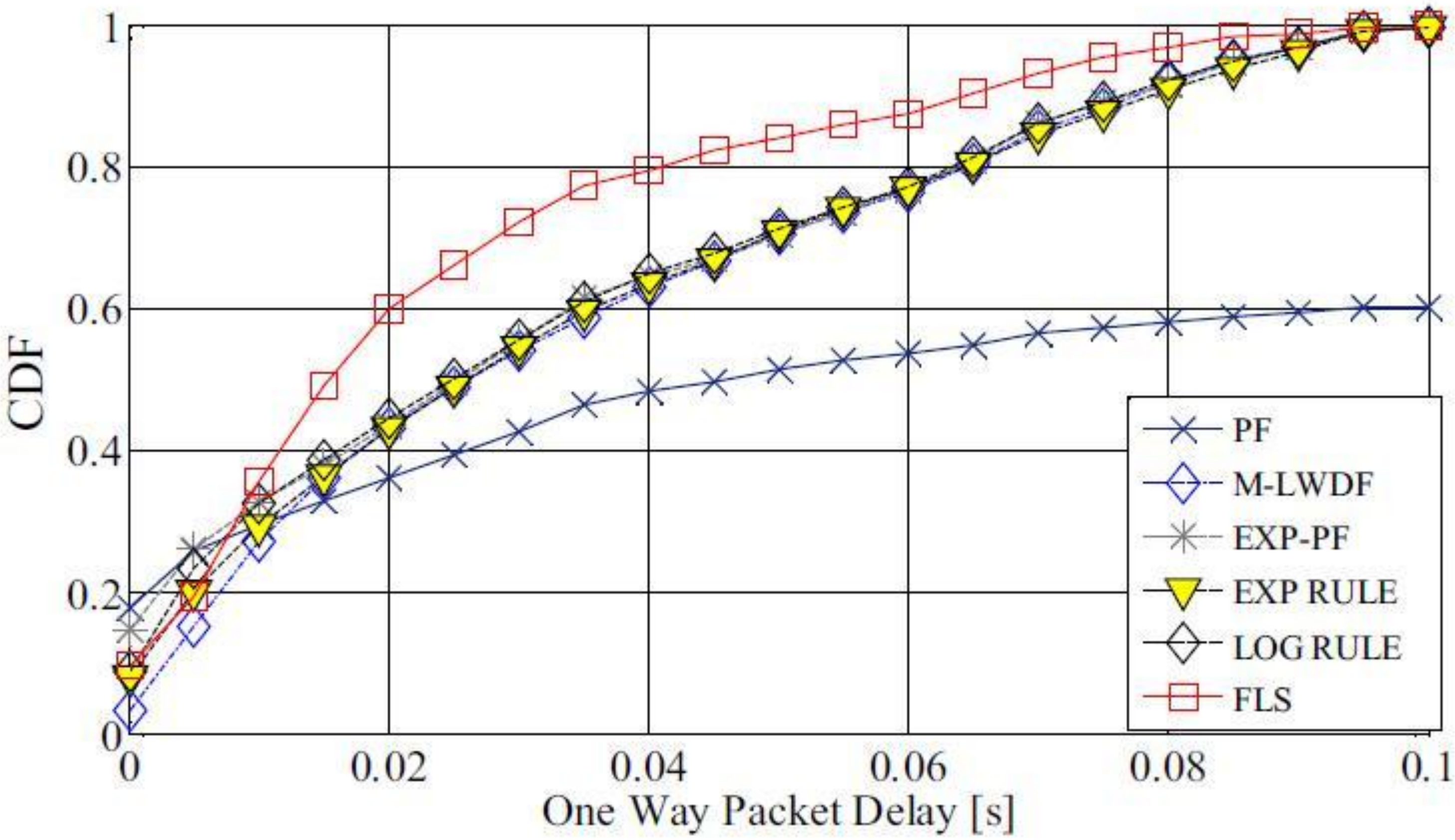
C. Channel-aware/QoS-aware Strategies

- 3. Performance evaluation of most relevant channel-aware/Qos-aware strategies



Packet Loss Rate of video flows with QoS-aware strategies

C. Channel-aware/QoS-aware Strategies



C. Channel-aware/QoS-aware Strategies

- 4. General considerations on channel-aware/Qos-aware strategies
  - support of multimedia flows is very important in LTE.
  - In fact, the working rationales of many of them are based on parameter settings whose optimality might depend on the specific scenario.
  - Further, more evolved approaches risk to be too complex and to waste resources.

D. Semi-persistent scheduling for VoIP support

- D. Semi-persistent scheduling for VoIP support
  - the radio resources are divided in several groups of RBs, and each block is pre-configurated and associated to VoIP.
  - It might be convenient to interleave persistent approaches, specific for VoIP, with dynamic ones that should perform allocation decisions for other flows.

E. Energy-aware strategies

- E. Energy-aware strategies
  - The modification of resource allocation policies on a TTI basis may not have strong impact on energy performance of a cellular network, unless under very low traffic load.
  - In this case, the best choice should be the maximization of the spectral efficiency.

New Direction and Future Challenge

- Its natural evolution, the Long Term Evolution-Advanced (LTE-A) solution, has been introduced by the 3GPP with the Release 10 for fulfilling and even surpassing IMT-A targets.
- Innovative technological solutions are standardized or foreseen for LTE-A, such as carrier aggregation, enhanced multi-antenna support, Coordinated Multi-Point (CoMP) transmission techniques, relaying, multi-userMultiple input multiple output (MIMO) communications, and Heterogeneous Networks (HetNets) deployment

Conclusion and Lesson Learned

- In this paper we provided an extensive survey on downlink packet allocation strategies recently proposed for LTE networks
- A main issue we found while studying literature on this topic, is the lack of a common reference scenario that can be used to compare different solutions.
- For this reason, we tried to identify a reference scenario and we used it to compare performance of the most representative solutions.