



UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Tesi di Laurea Triennale in
INGEGNERIA DELL'INFORMAZIONE

Comunicazione multimediale su reti LTE
Multimedial Communication over LTE
networks

Relatore
Leonardo Badia

Candidato
Davide Rigon

Anno Accademico 2011/2012

Ringraziamenti

A seguito di questo breve lavoro di tesi, ritengo doveroso ringraziare una serie di persone.

Il ringraziamento maggiore va innanzitutto alla mia famiglia, per la presenza costante e per il sostegno morale ed economico in questo periodo della mia vita.

Un grazie di cuore a Eleonora, per l'ascolto, la comprensione e il sostegno nelle difficoltà, e a tutti gli amici che mi hanno concesso momenti di svago in innumerevoli occasioni.

Un grazie sincero inoltre ai compagni di corso con cui ho condiviso momenti di ascolto, attenzione e studio ma anche risate e gioia in questi tre anni di Università.

Desidero infine ringraziare il relatore di questa tesi, il prof. Leonardo Badia, per la gentilezza e i commenti costruttivi.

Sommario

La Long Term Evolution della trasmissione UMTS rappresenta un notevole passo avanti nella serie dei sistemi di telecomunicazione mobile e costituisce una soluzione alla crescente richiesta di servizi multimediali.

Scopo di questa tesi è quello di descrivere le performance e la struttura delle reti LTE, in rapporto alla trasmissione di contenuti aventi diverse richieste di qualità di servizio e in particolare alla trasmissione di video. Verrà trattata l'allocazione dinamica delle risorse, ponendo l'accento sulle moderne ed efficienti tecniche di modulazione multiportante che consentono la gestione di un grande numero di utenti e su alcuni algoritmi di scheduling di cui vengono valutate le prestazioni nei vari contesti e vincoli più o meno stringenti di ritardo.

Indice

1	Introduzione	1
2	Stato dell'arte	7
2.1	Architettura di rete LTE	8
2.1.1	Access Network	9
2.1.2	Core Network	10
2.2	Data-Link Layer	12
2.3	Tecniche di Modulazione	15
2.3.1	OFDM	15
2.3.2	LTE downlink: OFDMA	21
2.3.3	LTE uplink: SC-FDMA	23
3	LTE downlink e multimedia	25
3.1	Algoritmo Rate Utility-Function Maximization	27
3.2	Algoritmo Two-Steps	28
3.3	Algoritmo Delay-Buffer Based	29
3.4	Algoritmo Weighted Round-Robin	31
4	Valutazione delle prestazioni	35
5	Conclusioni	43

Capitolo 1

Introduzione

Negli ultimi anni, l'evoluzione tecnologica nelle comunicazioni mobili è stata motivata da diversi fattori [6], tra cui la crescita esponenziale della domanda di servizi di telecomunicazione avanzati.

Le applicazioni che dovranno essere supportate dai sistemi cellulari sono infatti molteplici e presentano ciascuna diverse richieste di Quality of Service: alcune di esse sono applicazioni real-time sensibili al ritardo (ad esempio i servizi voce e videoconferenza), altre richiedono integrità, alti data-rate e vincoli stringenti sulla latenza (si pensi al video streaming, o al gaming online) [6]. In particolare, l'esplosione di dispositivi mobili nel mercato capaci di gestire contenuti video di alta qualità fa prevedere che le reti future dovranno essere ottimizzate per l'invio e la ricezione di traffico video e per il supporto di applicazioni video-based, da ogni luogo, in ogni momento e con ogni dispositivo [15].

Il supporto simultaneo di una tale varietà di servizi è una delle maggiori sfide che i sistemi cellulari stanno affrontando, unitamente ad una limitata disponibilità di potenza e di banda, che richiede che tali sistemi siano sviluppati con elevate efficienze spettrali, e ad una non-idealità del canale che invece si presenta selettivo in frequenza [1].

Ecco che dunque la Long Term Evolution del 3G rappresenta uno dei passi avanti più grandi nella comunicazione radio mobile per permettere all'utente di usufruire al meglio di questi servizi. LTE prevede un cambiamento della tecnologia di accesso al mezzo, passando da *Wideband-Code Division Multiple Access*, *W-CDMA*, a *Orthogonal Frequency Division Multiple Access*, *OFDMA* (come era avvenuto nel passaggio tra GSM e UMTS, in cui

era stato necessario abbandonare la metodologia di accesso al mezzo di tipo *Time Division Multiple Access*, *TDMA*, divenuta obsoleta, in favore di una metodologia di tipo W-CDMA); inoltre mira all'ottimizzazione della rete d'accesso *Universal Terrestrial Radio Access Network (UTRAN)* di UMTS. Una visione dell'evoluzione degli standard definiti da 3GPP è illustrata in figura 1.1.

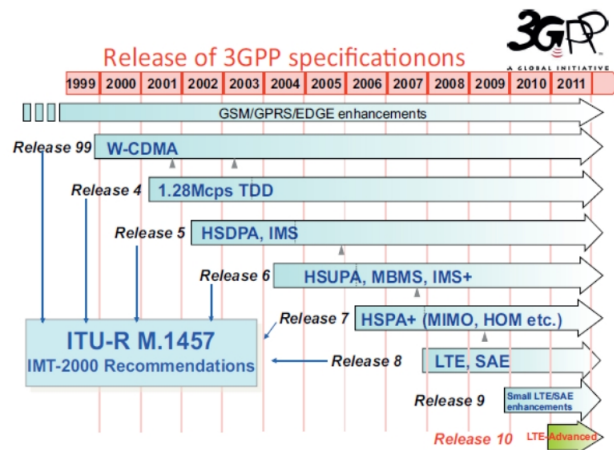


Fig. 1.1: Evoluzione degli standard 3GPP

Il passaggio dalla terza alla quarta generazione, secondo gli standard di 3GPP, prevede due fasi, come illustrato in [6]: la prima riguarda il completamento del primo standard LTE (Release 8), le cui specifiche sono state approvate nel dicembre 2008; la seconda intende adattare LTE ai requisiti di 4G attraverso una nuova tecnologia chiamata LTE-Advanced (Release 9 e 10), il cui piano è stato definito nell'ottobre 2009.

LTE si presenta come una rete packet-oriented all-IP, cioè è stata fin dalle origini progettata come rete a commutazione di pacchetto, con il target di supportare efficientemente l'utilizzo di massa di ogni servizio IP-based [11]. Gli obiettivi di LTE nella release 8 sono i seguenti, enunciati in confronto alla release 6 HSDPA (come illustrato in figura 1.2):

- Data rates più elevati, con un picco garantito di 100 Mbps in downlink e di 50 Mbps in uplink: il data rate di picco è definito come il massimo throughput per utente assumendo che l'intera banda a disposizione sia

allocata ad un singolo utente, che venga usato lo schema di modulazione più efficiente e che sia utilizzato il massimo numero di antenne supportate, cioè 2x2 in downlink e 1x2 in uplink [13]. Si arriva a 340 Mbps con canali a 20 MHz [15].

- Costo per bit ridotto, cioè aumento dell'efficienza spettrale di 3-4 volte in downlink e di 2-3 volte in uplink in media: con un'ampiezza di banda di 20 MHz, si ha un picco di efficienza spettrale rispettivamente di 5 e 2.5 bps/Hz in downlink e uplink [13].
- Aumento del bit rate a bordo cella, per la garanzia di un servizio il più uniforme possibile tra tutti gli utenti [13].
- Flessibilità nell'utilizzo dello spettro a disposizione, con una larghezza di banda variabile tra i 1.4 e i 20 MHz [13].
- Riduzione dei ritardi sia nello stabilire la connessione sia in termini di latenza di trasmissione: questa deve essere inferiore a 5 ms nel percorso tra terminale utente e rete fissa [13].
- Supporto di comunicazione in mobilità fino a 350 km/h (500 km/h per alcune frequenze), per garantire un handover trasparente e senza interruzione per l'utente [13].
- Supporto di trasmissione contemporanea di flussi downlink e uplink sia di tipo *Frequency Division Duplexing (FDD)* sia di tipo *Time Division Duplexing (TDD)* [13]. FDD prevede la suddivisione dei flussi downlink e uplink su due diverse bande di frequenza separate, risultando molto efficace in caso di traffico simmetrico e più semplice da implementare. TDD prevede invece di allocare intervalli temporali diversi per downlink e uplink pur impiegando la stessa banda di frequenza, dimostrandosi più efficiente nel caso di traffico sbilanciato e permettendo un'allocazione dinamica, pena una maggiore complessità realizzativa.
- Coesistenza con le tecnologie precedenti, semplificazione dell'architettura di rete e ottimizzazione del consumo di potenza del terminale utente [13].

Con il completamento della Release 8, 3GPP ha iniziato a guardare a come evolvere LTE per prestazioni future migliori. La risposta si è avuta a partire

		Absolute requirement	Release 6 (for comparison)	Comments
Downlink	Peak transmission rate	> 100 Mbps	14.4 Mbps	LTE in 20 MHz FDD, 2×2 spatial multiplexing. Reference: HSDPA in 5 MHz FDD, single antenna transmission
	Peak spectral efficiency	> 5 bps/Hz	3 bps/Hz	
	Average cell spectral efficiency	> 1.6–2.1 bps/Hz/cell	0.53 bps/Hz/cell	LTE: 2×2 spatial multiplexing, Interference Rejection Combining (IRC) receiver [3]. Reference: HSDPA, Rake receiver [4], 2 receive antennas
	Cell edge spectral efficiency	> 0.04–0.06 bps/Hz/user	0.02 bps/Hz/user	As above, 10 users assumed per cell
	Broadcast spectral efficiency	> 1 bps/Hz	N/A	Dedicated carrier for broadcast mode
Uplink	Peak transmission rate	> 50 Mbps	11 Mbps	LTE in 20 MHz FDD, single antenna transmission. Reference: HSUPA in 5 MHz FDD, single antenna transmission
	Peak spectral efficiency	> 2.5 bps/Hz	2 bps/Hz	
	Average cell spectral efficiency	> 0.66–1.0 bps/Hz/cell	0.33 bps/Hz/cell	LTE: single antenna transmission, IRC receiver [3]. Reference: HSUPA, Rake receiver [4], 2 receive antennas
	Cell edge spectral efficiency	> 0.02–0.03 bps/Hz/user	0.01 bps/Hz/user	As above, 10 users assumed per cell
System	User plane latency (two way radio delay)	< 10 ms		LTE target approximately one fifth of Reference.
	Connection set-up latency	< 100 ms		Idle state \rightarrow active state
	Operating bandwidth	1.4–20 MHz	5 MHz	(initial requirement started at 1.25 MHz)
	VoIP capacity	NGMN preferred target expressed in [2] is > 60 sessions/MHz/cell		

Fig. 1.2: Confronto tra lo standard 3GPP Release 6 e lo standard LTE Release 8 (tratto da [13])

dalla Release 10, che costituisce la prima versione di LTE-Advanced. Oltre ovviamente a garanzie di compatibilità con la prima versione di LTE, nonché con le tecnologie 3GPP precedenti, essa promette [13]:

- crescita significativa dei data-rate di picco istantanei, di 1 Gbps in downlink e 500 Mbps in uplink;
- latenza massima di 10 ms per la sincronizzazione di un utente;
- picco di efficienza spettrale di 30 bps/Hz in downlink e di 15 bps/Hz in uplink con una configurazione di antenne 8x8 in DL e 4x4 in UL;
- efficienza spettrale ulteriormente migliorata in media e a bordo cella;
- larghezza di banda scalabile fino a 100 MHz.

La più importante tecnologia utilizzata in LTE per perseguire i requisiti richiesti è la modulazione **Orthogonal Frequency Division Multiplexing**,

OFDM, una tecnica di trasmissione multiportante che consiste nell'assegnazione di piccole porzioni non sovrapposte dello spettro disponibile ai diversi utenti: essa verrà descritta nel dettaglio nella sezione 2.3.1.

L'assegnazione viene regolata dinamicamente nel tempo e prende il nome di *scheduling*. In un sistema a singolo utente, l'utente può sfruttare la potenza totale per trasmettere su tutte le sottoportanti, mentre in un sistema multiutente, come di fatto è una rete cellulare, c'è la necessità di uno schema di accesso multiplo per allocare le sottoportanti e la potenza [1]. L'allocazione può essere *statica*, se ad ogni utente vengono assegnate risorse indipendentemente dalle condizioni di canale che esso percepisce, o *dinamica* se invece essa ne tiene conto. Anche se più difficile da implementare, chiaramente si preferisce l'allocazione dinamica poichè porta ad uno sfruttamento migliore delle risorse in termini di potenza e bit rate, quando invece l'allocazione statica comporterebbe degli sprechi: lo scheduler ha dunque gli obiettivi di massimizzare l'efficienza del sistema, allocando le risorse agli utenti più appropriati, e garantire equità (o *fairness*) tra di essi [2].

Lo scopo della tesi è quindi duplice: descrivere quali tecniche consentano a LTE di raggiungere elevate efficienze spettrali e brevi latenze e soprattutto analizzare, attraverso il confronto tra alcuni algoritmi di scheduling, quali siano i principali parametri da tenere in considerazione nella allocazione dinamica delle risorse e quali algoritmi raggiungano le prestazioni migliori per determinate richieste di QoS.

Nella parte rimanente, la tesi è organizzata come segue. Il capitolo 2 descrive l'architettura di LTE, le tecniche di modulazione e di accesso al mezzo in relazione alle problematiche di trasmissione di contenuti video. Il capitolo 3 tratta dello scheduling più nel dettaglio, confrontando quattro algoritmi di allocazione delle risorse e rimandando al capitolo 4 la valutazione delle loro prestazioni. Infine nel capitolo 5 seguono alcune conclusioni.

Capitolo 2

Stato dell'arte

Le future reti wireless dovranno essere ottimizzate per l'invio di grandi moli di dati, soprattutto contenuti video. È stato predetto (in accordo con [15]) che il traffico mobile sarebbe cresciuto di un fattore di 39 volte tra il 2009 e il 2014, e che il 66 per cento di tale traffico sarebbe stato di tipo video. Le applicazioni di video streaming real-time (RT) sono sempre più diffuse, e i pacchetti di un utente video RT devono essere ricevuti dall'utente entro una certa soglia di ritardo, altrimenti essi vengono scartati e considerati pacchetti persi, degradando la qualità percepita. Per soddisfare i requisiti di QoS degli utenti RT dunque deve essere minimizzato il **Packet Loss Rate (PLR)** [5].

Secondo [15], standard come WiMAX ed LTE sono in grado di fornire servizi video in mobilità usando tecnologie sia di tipo unicast sia di tipo broadcast, dal momento che una parte considerevole e crescente dei servizi è indirizzata a gruppi che condividono interessi simili (film, eventi sportivi ecc.). La modalità *unicast*, adatta a piccoli gruppi di utenza, presenta i vantaggi che le risorse di rete sono consumate solo quando un utente è effettivamente connesso e che la rete può ottimizzare la trasmissione per ogni utente singolarmente: ciò richiede un feedback da parte dell'utente in termini di *Channel Quality Indicator (CQI)* e di messaggi di ACK/NACK durante i protocolli di HARQ. Per la trasmissione a gruppi numerosi di utenti, sono state sviluppate tecnologie di broadcast quali *MBMS* in LTE, che distribuiscono i contenuti multimediali simultaneamente a tutti gli utenti in ricezione, in modo da risparmiare sulla capacità di rete: in questo contesto non sono però facilmente implementabili meccanismi di feedback dall'utente, e per miglio-

rare l'affidabilità si utilizzano protocolli di *Forward Error Correction (FEC)*. In entrambi i casi, la valutazione della capacità video fornita da [15] mostra come LTE sia molto promettente in termini di supporto alla trasmissione video. Nella simulazione viene calcolato il picco massimo di *video users* all'interno di un settore, assumendo che il 50 per cento delle risorse siano riservate al traffico video; si prendono in considerazione data-rate di servizio video di 384 kbps, 784 kbps e 1.536 Mbps. I risultati in figura 2.1 evidenziano un aumento generale del numero di utenti video supportabili all'interno della stessa cella, in confronto alla tecnologia 3GPP release 6; la crescita è ulteriore in LTE Advanced.

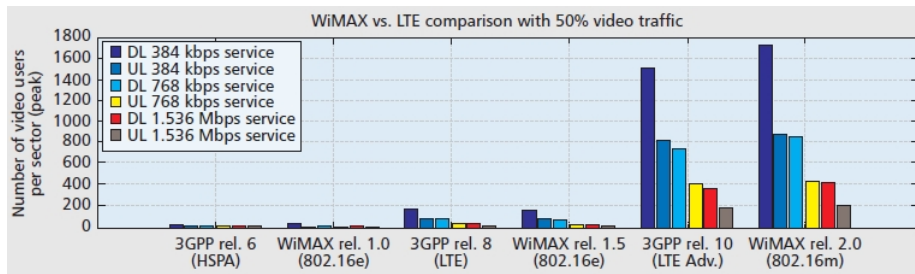


Fig. 2.1: Utenti video supportabili in diverse versioni di 3GPP e WiMAX [15]

La fornitura di servizi multimediali ad elevate capacità ed efficienza spettrale e brevi latenze dipende strettamente dall'adozione di un'architettura di rete semplificata, di meccanismi di ritrasmissione che siano efficienti e non portino ad un eccessivo overhead di segnalazione, nonché di tecniche di modulazione avanzate.

2.1 Architettura di rete LTE

Al fine di supportare servizi esclusivamente a commutazione di pacchetto, gli standard LTE prevedono una evoluzione sia dell'interfaccia di accesso radio che degli aspetti non-radio del sistema, indicati con il termine *System Architecture Evolution (SAE)*, che includono la *Evolved Packet Core (EPC)* network. Insieme, LTE (che indica l'evoluzione dell'interfaccia radio) e SAE danno origine all'*Evolved Packet System (EPS)*. EPS utilizza il concetto di *bearers*, cioè di flussi di pacchetti IP con un preciso QoS, per direzionare il

traffico da un gateway nella *Packet Data Network (PDN)* a un *User Equipment (UE)*; bearers multipli possono essere associati allo stesso utente per fornire flussi di QoS diverse o connessioni a PDN differenti, in accordo con [13].

Un confronto tra l'architettura di rete dei sistemi UMTS Release 6 e LTE Release 8 è fornito in figura 2.2.

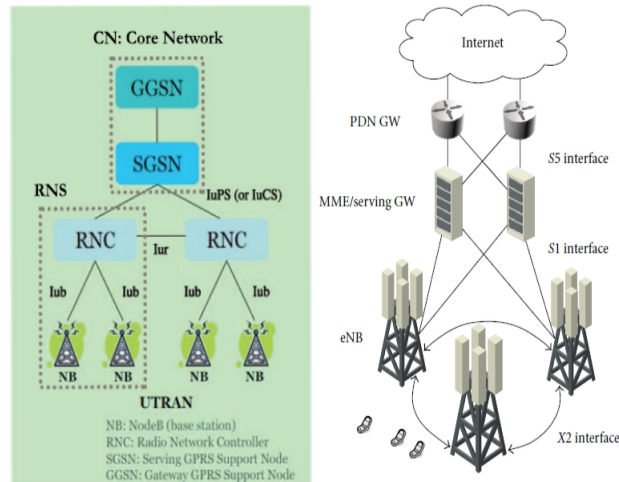


Fig. 2.2: Architetture delle reti UMTS e LTE a confronto (tratto da [6])

2.1.1 Access Network

La filosofia dominante dietro LTE è quella di minimizzare il numero di nodi: dunque gli sviluppatori hanno optato per un'architettura single-node. La rete di accesso radio di LTE consiste in stazioni radio base denominate *Enhanced NodeB (eNB)*, connesse tra di loro per mezzo dell'interfaccia X2 e alla core network tramite l'interfaccia S1. L'architettura E-UTRAN è detta *flat*, intendendo che nel normale traffico utente (opposto al broadcast) non si ha un controllo centralizzato, ma ciascuna stazione si coordina indipendentemente dalle altre [13].

Attraverso l'utilizzo di protocolli denominati *Access Stratum (AS)* protocols, la E-UTRAN è responsabile delle funzioni di connessione radio tra le specifiche eNB e gli UE, che possono essere così riassunte:

- **Gestione delle risorse radio:** controllo dei portatori radio, controllo dell'ammissione e della mobilità radio, nonché lo scheduling e l'allocazione dinamica delle risorse agli utenti sia in uplink che in downlink.
- **Header Compression:** tecnica che mira ad un utilizzo più efficiente dell'interfaccia radio comprimendo gli header dei pacchetti IP che potrebbero rappresentare altrimenti un carico eccessivo, soprattutto per piccoli pacchetti come quelli VoIP.
- **Sicurezza:** crittazione dei dati inviati sull'interfaccia radio.
- **Posizione:** fornitura delle misure necessarie e di altri dati per trovare la posizione dello UE.
- **Connessione allo EPC:** segnalazione per l'accesso alla core network.

Tutte queste funzioni hanno sede negli eNB, ciascuno dei quali gestisce celle multiple. LTE integra le funzioni di controllo radio nell'eNB stesso, fattore che permette una stretta interazione tra i diversi protocol layers, riducendo la latenza e aumentando l'efficienza; inoltre, attraverso l'eliminazione della necessità di un controllore centrale, si sono ottenuti vantaggi in termini di riduzione dei costi e robustezza della rete, eliminando i cosiddetti *single points of failure*. Nelle reti precedenti, l'UTRAN prevede un controllo non integrato nelle NB ma esterno: ciascuna base station gestisce un gruppo di 3-6 celle ed è controllata da un controllore di rete *Radio Network Controller (RNC)* con cui comunica attraverso un'apposita interfaccia (Iu). Di conseguenza, funzioni come la cifratura e la header compression, prima svolte da RNC in UTRAN, ora sono affidate direttamente a eNB; analogamente, gli handover tra stazioni base sono gestiti dall'interfaccia X2 laddove prima esisteva un'entità centrale di ARQ in RNC [9].

2.1.2 Core Network

La core network, chiamata EPC in SAE, è responsabile del controllo globale dello User Equipment e dell'instaurazione dei bearers. EPC consiste in [13]:

- **Mobile Management Entity (MME)**: nodo di controllo, che si occupa della segnalazione tra UE e core network attraverso protocolli *Non-Access Stratum (NAS)*. Le sue funzioni principali prevedono la stabilizzazione, il mantenimento ed il rilascio dei bearers; la gestione della connessione e della sicurezza tra la rete e lo UE; la gestione di connessioni con altre reti.
- **Packet-data network gateway (P-GW)**: nodo che gestisce l'allocazione degli indirizzi IP per gli utenti e l'invio dei pacchetti IP in downlink nei bearers aventi diversi QoS. Inoltre rappresenta l'ancora di mobilità per il collegamento a reti di tipo non-3GPP, come CDMA2000 e WiMAX.
- **Serving gateway (S-GW)**: nodo che costituisce l'ancora di mobilità locale per i data bearers quando l'utente si muove tra le eNB e che conserva le informazioni riguardo i bearers quando l'utente è in stato idle. Inoltre effettua alcune funzioni amministrative nella rete visitata, raccogliendo informazioni come i volumi di dati scambiati dall'utente. Infine rappresenta l'ancora di mobilità per il collegamento ad altre tecnologie 3GPP (come GPRS e UMTS).
- **Evolved Serving Mobile Location Centre (E-SMLC)**: nodo che gestisce le risorse richieste per trovare la locazione di un utente collegato a una eNB e stimarne altri dati quali la velocità e l'accuratezza raggiunta.

La funzionalità di MME è tenuta separata dai gateway per facilitare lo sfruttamento della rete e l'evoluzione indipendente della tecnologia. L'integrazione dei sistemi GSM e WCDMA/HSPA avviene mediante interfacce tra il *SGSN (Serving GPRS Support Node)* e l'EPC network, mentre il nodo di gateway GW funziona anche da nodo *GGSN (Gateway GPRS Support Node)* per i terminali GSM e WCDMA/HSPA [11]. Altri nodi logici di supporto presenti sono il **Policy Control and Charging Rules Function (PCRF)** che controlla la tariffazione e la configurazione IP Multimedia Subsystem (IMS) di ciascun utente, il **Gateway Mobile Location Centre (GMLC)** che contiene funzionalità per la localizzazione, il **Home Subscriber Server (HSS)** che contiene dinamicamente traccia dei dati riguardanti la mobilità di ciascun utente. In figura 2.3 uno schema globale più dettagliato degli

elementi di rete dell'EPS.

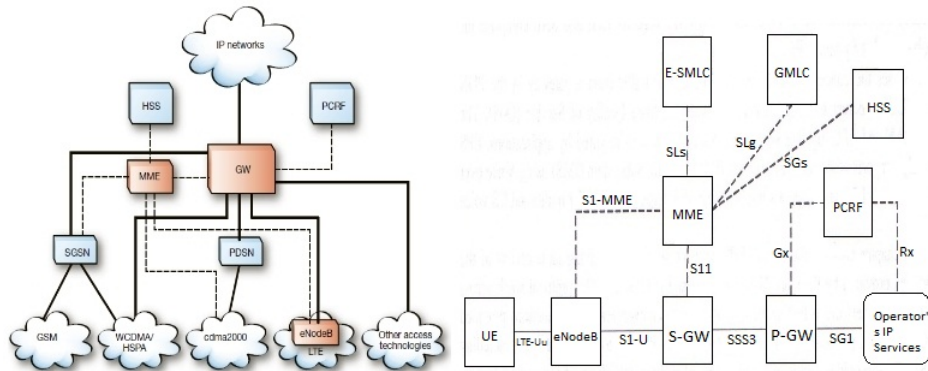


Fig. 2.3: Elementi di rete dell'Evolved Packet System, in relazione agli elementi di rete di UMTS Release 6 (a destra, tratto da [11]) e con tanto di interfacce (a sinistra, tratto da [13])

È evidente dunque una semplificazione nell'architettura LTE-SAE rispetto ai sistemi GSM e WCDMA/HSPA sullo user-plane, con soli due nodi (eNB e S/P-GW, quest'ultimo implementato come nodo unico diviso in due entità logiche) contro quattro (NodeB, RNC, SGSN, GGSN), riducendo in questo modo la latenza [9].

2.2 Data-Link Layer

I protocolli di link-layer hanno le funzioni di arricchire il servizio agli strati sovrastanti in termini di affidabilità, sicurezza e integrità, oltre che di effettuare l'accesso multiplo e lo scheduling. Una delle maggiori sfide è quella di garantire la riduzione dei ritardi nei flussi di pacchetti IP, riducendo l'overhead [9]. Lo strato di livello 2 (data-link layer) della tecnologia LTE è composto da tre sottostrati, ciascuno dei quali riceve un Service Data Unit (SDU) dallo strato sovrastante, effettua un servizio e rilascia un Protocol Data Unit (PDU) allo strato sottostrante, come spiegato in [13] e illustrato in figura 2.4.

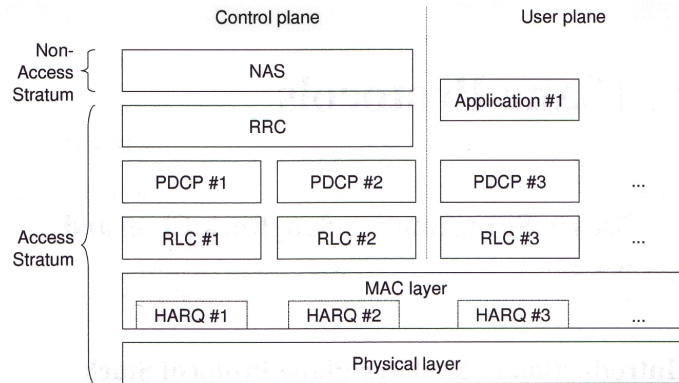


Fig. 2.4: Strutture del data-link layer in LTE (tratto da [13])

- Packet Data Convergence Protocol (PDCP) layer:** elabora i messaggi Radio Resource Control nel piano di controllo e i pacchetti IP nel piano utente, svolgendo funzioni di header compression dei dati, di sicurezza (cifatura, protezione integrità), di supporto all'handover; inoltre si occupa di scartare i pacchetti utente esaurito il loro tempo di timeout.
- Radio Link Control (RLC) layer:** ha in compito di segmentare e riassemblare i pacchetti provenienti dagli strati superiori per adattarli alle dimensioni con cui possono essere trasmessi sull'interfaccia radio; in più, per i radio bearers che necessitano una trasmissione error-free, questo strato effettua la ritrasmissione dei pacchetti. Infine si occupa del riordino per compensare la ricezione disordinata dovuta ai meccanismi di Hybrid Automatic Repeat reQuest (HARQ) dello strato sottostante. HARQ è una tecnica di ritrasmissione dei pacchetti dati affetti da errori: essa combina la *Forward Error Correction (FEC)*, una tecnica di codifica di canale atta a rilevare e correggere al ricevitore un certo numero di bit errati, con la ritrasmissione vera e propria dei dati ancora affetti da errori attraverso lo scambio di messaggi di *acknowledgement (ACK/NACK)* tra trasmettitore e ricevitore riguardo lo stato del PDU appena inviato sul canale. HARQ dunque si presenta come un compromesso tra il FEC puro, che potrebbe introdurre eccessiva ridondanza, e l'ARQ puro che, richiedendo un canale di controllo, rischierebbe di aumentare troppo i ritardi [16].

Le funzioni del sottostrato RLC sono affidate a delle 'entità RLC', che operano in modalità *Transparent Mode (TM)* se non viene effettuata nessuna elaborazione ai PDUs che passano attraverso esse; *Unacknowledged Mode (UM)* per i servizi di trasferimento dati unidirezionali, soprattutto per applicazioni real-time sensibili al ritardo che non richiedono ritrasmissioni; *Acknowledged Mode (AM)* per i servizi di trasferimento dati bidirezionali, per favorire i meccanismi di ritrasmissione ARQ in applicazioni non real-time sensibili agli errori, dove non ci sono vincoli stringenti di delay.

- **Medium Access Control (MAC) layer:** effettua il multiplexing dei dati dai diversi radio bearers, prendendo in considerazione il QoS per ciascun radio bearer per decidere la quantità di dati da trasmettere per ciascun bearer e comunicare allo strato RLC la dimensione dei pacchetti da trasmettere. C'è una sola entità MAC per ogni UE. Lo strato MAC si connette allo strato fisico sottostante tramite *canali di trasporto* e allo strato RLC tramite *canali logici*, questi ultimi divisi in canali logici di *controllo* e di *traffico*.

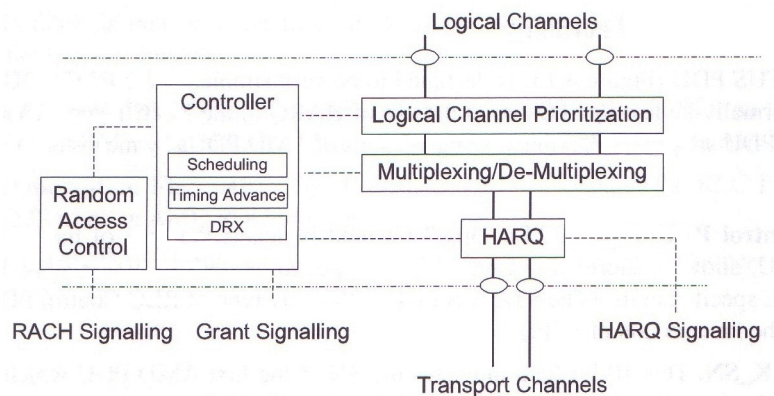


Fig. 2.5: Entità presenti nello strato MAC (tratto da [13])

Una scelta di design per LTE è stata quella non di lasciar propagare errori sui bit verso gli strati più alti in ricezione, ma di ritrasmettere in toto le data unit contenenti errori, grazie a un codice a ridondanza ciclica (*CRC checksum*) attaccato dal physical layer alle data unit. L'entità di HARQ presente nel

MAC effettua la ritrasmissione, con protocolli che usano processi di stop-and-wait HARQ multipli. Tuttavia, invece di un messaggio contenente un numero in sequenza, per dare l'informazione di risposta (ACK/NACK) viene inviato un solo bit con una fissa relazione temporale con il tentativo di trasmissione. Questo porta ad un guadagno in termini di ritardo, semplicità e overhead di controllo: velocità del protocollo HARQ e suo basso consumo di risorse radio sono requisiti importanti [17].

2.3 Tecniche di Modulazione

La scelta di tecniche di modulazione e di accesso multiplo adeguate è un fattore critico per raggiungere buone performance di sistema: i canali radio infatti sono caratterizzati tipicamente dalla tendenza ad essere tempo-varianti e selettivi in frequenza, a causa della mobilità dei terminali e del fading causato dalla multipropagazione dovuta agli ostacoli [7]. Di conseguenza, in LTE come in altri sistemi di comunicazione attuali, si è guardato con molto interesse ai sistemi multiportante, in particolare all'**Orthogonal Frequency Division Multiplex (OFDM)**.

2.3.1 OFDM

OFDM è una tecnica di modulazione che prevede di suddividere il segnale di informazione ad alta velocità trasmissiva in flussi paralleli a bassa velocità multiplati a divisione di frequenza su portanti ortogonali e quindi non interferenti tra loro. Tale tecnica si è dimostrata la scelta migliore per far fronte al problema dell'elevatezza del *multipath* nei canali radio: questa interferenza multi-cammino, dovuta alle riflessioni del segnale da parte degli ostacoli che incontra nella trasmissione, genera repliche del segnale trasmesso che arrivano al ricevitore attenuate, sfasate e ritardate, causando un degrado del segnale stesso [12]. Il principio dello OFDM consiste nel dividere la banda disponibile B in M parti, in modo tale da ottenere canali a banda stretta di larghezza B/M (molto minore della larghezza di banda di coerenza del canale), aventi un guadagno approssimativamente costante e quindi idealmente non distorcenti e non selettivi in frequenza.

In ciascuna sottobanda sono trasmessi dati a ritmo binario ridotto, che vengono modulati su portanti *ortogonali* tra loro: la distanza tra esse viene

scelta in modo che la risposta in frequenza di ciascuna sottoportante abbia un massimo laddove la risposta in frequenza dei canali adiacenti ha invece un valore nullo. Ciò rende idealmente nulla l'interferenza tra i canali in parallelo (ICI) e permette di evitare l'utilizzo di bande di guardia, garantendo elevata efficienza spettrale [8]. Un esempio del confronto tra trasmissione sequenziale su singola portante e trasmissione multiportante su canali in parallelo è riportato in figura 2.6.

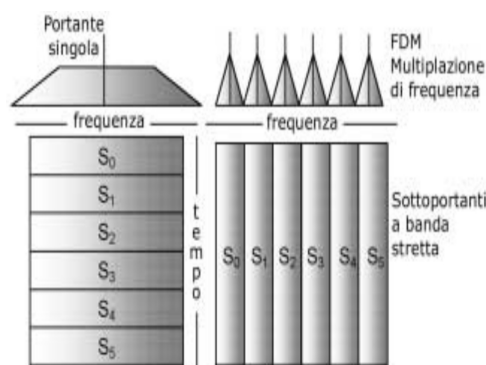


Fig. 2.6: Confronto tra trasmissione sequenziale su portante singola e trasmissione multiportante su canali in parallelo (tratto da [8])

Lo spettro del segnale trasmesso da ogni singola portante ha la forma di una funzione $\text{sinc}(f)/f$, dunque lo spettro del segnale integrale è una successione di tali funzioni spaziate di $1/T$ (si veda la figura 2.7), con T periodo di modulazione. La spaziatura tra i sottocanali in OFDM è tale da permettere una perfetta separazione di essi al ricevitore: ciò si traduce in una implementazione a bassa complessità [14].

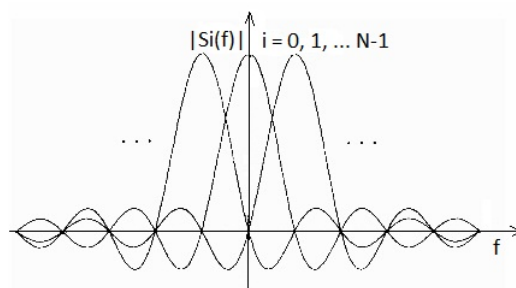


Fig. 2.7: Spettro del segnale OFDM (tratto da [14])

L'aggiunta di intervalli di guardia (a scapito comunque della capacità trasmissiva), o **prefissi ciclici** permette di eliminare l'interferenza intersimbolo (ISI) dovuta, per ogni simbolo utile, ai simboli che lo precedono: un prefisso ciclico tipicamente consiste nella ripetizione dell'ultima porzione di segnale che viene collocata all'inizio del successivo blocco di dati [12]. Dal ricevitore questo intervallo di guardia di durata T_g viene scartato, cosicchè gli echi che raggiungono il ricevitore con un ritardo inferiore a T_g non generano ISI, come mostrato in figura 2.8.

Il multiplexing ortogonale comprende diversi blocchi funzionali in trasmissione, come descritto in [13]:

- **Convertitore S/P**: Il segnale d'ingresso ad alto data-rate, definito nel dominio della frequenza, viene distribuito sugli M canali tramite una conversione serie/parallelo (S/P). Questa prima conversione consente di far fronte al problema degli echi: nella trasmissione seriale il periodo di simbolo T_s è molto minore del channel delay spread T_d e ciò genera ISI poichè l'eco ha effetto sull'intero simbolo. Invece, trasmettendo in parallelo, la durata di simbolo in ciascuna sottoportante è aumentata di un fattore M e diventa molto maggiore del channel delay spread; dunque l'eco ha effetto solo su una piccola porzione del simbolo [12], come illustrato in figura 2.9.

All'uscita di questo convertitore dunque si hanno M flussi a basso symbol rate: se si pensa a questi flussi come a array, il k -esimo simbolo OFDM è un vettore di dimensione M costituito dagli elementi di indice k di ciascun array; in altre parole, un simbolo OFDM è una combinazione lineare dei simboli presenti in un dato istante su ciascuna sottoportante.

- **Mapper**: Gli M flussi dati paralleli sono modulati ciascuno in modo indipendente. Entro ciascun canale i segnali possono essere modulati per mezzo di una qualsiasi delle leggi di modulazione convenzionali, scelta in base alle condizioni del canale:
 - **64-QAM**: Presenta una costellazione di 64 punti, ciascuno con ampiezza e fase diverse dagli altri. Permette di raggiungere efficienze spettrali elevate, di 6 bit/s/Hz, ma la sua sensibilità al rumore (essendo i punti vicini) lo rende utilizzabile solo in condizioni

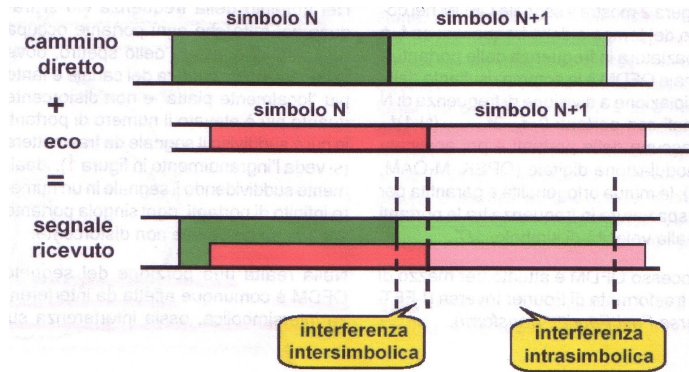


Fig. 2.8: Interferenza intersimbolo (tratto da [12])

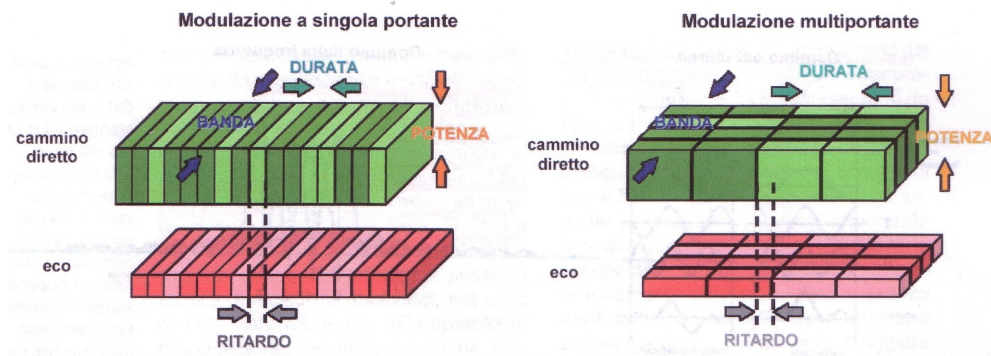


Fig. 2.9: Confronto tra modulazione a singola e a multi portante (tratto da [12])

di canale molto buone, con un elevato rapporto segnale-rumore (SNR).

- **16-QAM**: Presenta una costellazione di 16 punti, ciascuno con ampiezza e fase diverse dagli altri. Permette di raggiungere buone efficienze spettrali, di 4 bit/s/Hz, richiedendo un SNR meno elevato della modulazione 64-QAM, a parità di BER (Bit Error Ratio).
- **QPSK**: Presenta una costellazione di 4 punti, di uguale ampiezza e fase diversa. Adeguato a condizioni di canale peggiori, permette di trasmettere con basso tasso di errore anche con un SNR basso, al prezzo di un'efficienza spettrale più limitata, di 2 bit/s/Hz.

Si parla dunque di *modulazione adattativa*.

- **banco IFFT**: Siccome l'utilizzo di oscillatori analogici costerebbe troppo e comporterebbe imperfezioni e malfunzionamenti, si preferisce implementare la modulazione via software in maniera digitale [14], servendosi di un banco che effettua la Inverse Fast Fourier Transform. Viene dunque calcolata di volta in volta la IFFT su un insieme di simboli trasmessi durante il periodo di modulazione, e il segnale ottenuto da questa trasformazione, dopo la conversione digitale-analogico, costituisce il simbolo OFDM vero e proprio.
- **Convertitore P/S**: I flussi vengono riportati tutti sulla medesima linea fisica, e nel frattempo viene creato un periodo di guardia all'inizio di ciascun simbolo OFDM. Il prefisso ciclico viene generato duplicando gli ultimi G campioni dell'uscita del banco IFFT e, per evitare l'ISI completamente, il prefisso deve avere lunghezza G maggiore della lunghezza della risposta impulsiva del canale.
- **Convertitore Digitale/Analogico, filtro passa banda, convertitore RF**: Il segnale, dopo la conversione D/A che lo riporta nel dominio del tempo e il filtraggio, subisce una conversione in frequenza allo scopo di trasportare il segnale completo nella banda a radiofrequenza ad esso allocata. Infine, viene trasmesso sul canale.

In ricezione si operano le funzioni duali, come illustrato in figura 2.10.

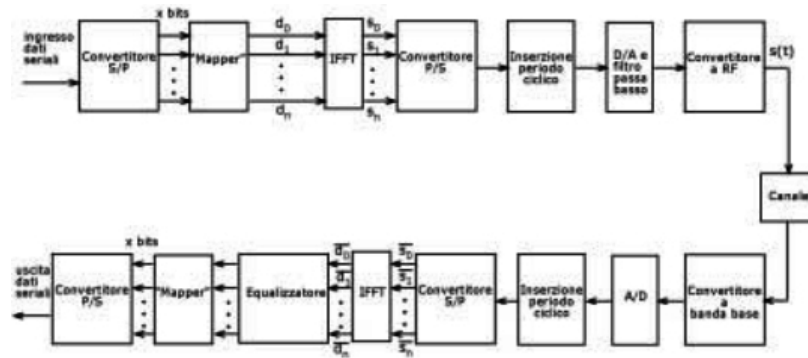


Fig. 2.10: Generazione e ricezione di un segnale OFDM (tratto da [8])

E' dunque evidente come la tecnologia OFDM presenti molti vantaggi per quanto riguarda la trasmissione multipla su un canale tempo-variante e selettivo in frequenza. I miglioramenti in termini di efficienza spettrale di LTE rispetto a HSPA sono in gran parte attribuibili proprio alle proprietà garantite da OFDM: ortogonalità tra gli utenti grazie all'equalizzazione nel dominio della frequenza, scheduling che permette di trasmettere utilizzando le sottoportanti con le migliori condizioni di canale [10]. Esistono tuttavia dei limiti nell'utilizzo di questa tecnologia [13]:

- elevato *Peak-to-Average Power Ratio (PAPR)*: Questo problema nasce dal fatto che le variazioni in ampiezza del segnale OFDM modulato possono essere molto accentuate, dal momento che i simboli OFDM modulati possono essere approssimati come forme d'onda Gaussiane (in virtù del teorema limite centrale), mentre gli amplificatori di potenza dei trasmettitori a radiofrequenza mostrano un comportamento lineare solo entro un intervallo limitato: ciò comporta una distorsione del segnale OFDM, per evitare la quale è necessaria una grande potenza di amplificazione, cioè trasmettitori molto costosi. Il PAPR è appunto una misura del range dinamico dell'ampiezza dell'input, e dunque della sua degradazione. È proprio a causa dell'elevato PAPR di OFDM, che richiederebbe di costruire amplificatori di potenza troppo costosi sui terminali mobili, che in uplink si preferisce l'utilizzo di una tecnica di accesso multiplo al mezzo di tipo Single-Carrier Frequency Division Multiple Access (SC-FDMA), variante di OFDM [10].

- sensibilità al *Carrier Frequency Offset*: a causa del non perfetto allineamento tra i riferimenti di frequenza degli oscillatori del trasmettitore e del ricevitore, dovuto al cambiamento di parametri come la temperatura e il voltaggio, è possibile perdere l'ortogonalità delle sottoportanti, causando Inter-Carrier Interference (ICI).

Molte tecniche sono state studiate per ridurre l'impatto di questi problemi. L'utilizzo di tale tecnologia, che permette di implementare ricevitori a bassa complessità, consente di utilizzare tecniche di tipo **Multiple In Multiple Out (MIMO)**, per raggiungere data-rate elevati [6]. Si tratta di un modello di trasmissione multicanale che effettua la diversità di spazio utilizzando configurazioni di più antenne - tipicamente in LTE downlink la configurazione base è 2 antenne in trasmissione e 2 in ricezione - per le trasmissioni multiple di flussi di dati in parallelo, al fine di migliorare la capacità di dati trasmessi per singolo utente [7]. La modulazione OFDM è alla base della tecnologia di accesso al mezzo Orthogonal Frequency Division Multiple Access (OFDMA), utilizzato in downlink.

2.3.2 LTE downlink: OFDMA

Orthogonal Frequency Division Multiple Access è un'estensione di OFDM per l'implementazione di un sistema di comunicazione multiutente e consiste nell'assegnazione a ciascun utente di un sottoinsieme delle sottoportanti in cui viene suddivisa la banda disponibile in un determinato intervallo di tempo. Di solito, le sottoportanti sono allocate in gruppi contigui per semplicità e per ridurre il sovraccarico di indicare quali sottoportanti sono state allocate per ogni utente [13].

OFDMA può essere considerata in linea di massima una tecnica che combina la modulazione OFDM e l'accesso TDMA, come illustrato nella figura 2.11.

Alla dimensione spaziale si ha accesso attraverso *antenna ports* presenti in ciascuna eNB. Siccome l'allocazione dinamica delle risorse avviene sulla base delle condizioni del canale radio, ogni porta fornisce un *Reference Signal (RS)* per abilitare l'UE alla stima di tali condizioni. Le risorse radio possono essere considerate come elementi di una griglia tempo-frequenza [13], come illustrato in figura 2.12.

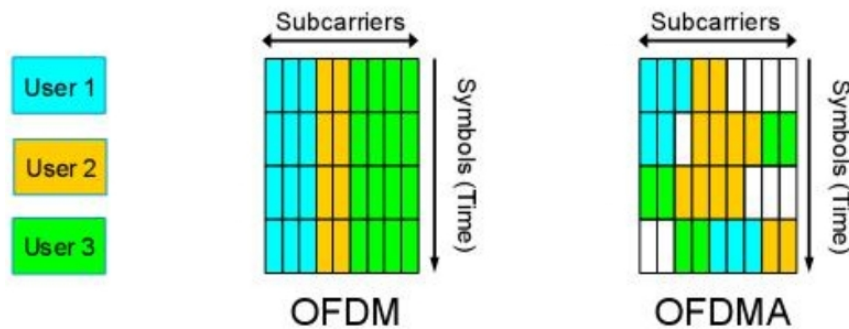


Fig. 2.11: OFDM vs OFDMA (tratto da <http://www.eefocus.com/article/08-04>)

- Nel dominio del tempo, l'unità più grande è il *frame* della durata di 10 ms, suddiviso a sua volta in 10 *subframes* di 1 ms, ciascuno dei quali è costituito da due *slot* da 0.5 ms. Ciascuno slot comprende di norma 7 simboli OFDM (6 se viene utilizzata la configurazione con prefisso ciclico esteso).
- Nel dominio della frequenza, le risorse sono raggruppate in unità di 12 sottoportanti spaziate di 15 kHz tra di loro, per un'occupazione totale di 180 kHz.

Una unità di 12 sottoportanti per la durata di uno slot è chiamata **Resource Block (RB)**, mentre l'unità di risorse più piccola viene chiamata **Resource Element (RE)** e consiste in una sottoportante per la durata di un simbolo OFDM: dunque di norma un RB comprende 84 REs (72 in configurazione con prefisso ciclico esteso).

In realtà lo schema appena descritto, in cui tutti i REs sono utilizzati per la trasmissione di dati in downlink (di solito a coppie di RBs), è valido in un contesto di Frequency Division Duplexing. Se viene invece utilizzata la tecnica di separazione Time Division Duplexing chiaramente non tutte le sottoportanti possono essere utilizzate per la trasmissione in downlink, ma una parte deve essere usata per la trasmissione di dati in uplink e di speciali *Guard Periods* che permettono lo switch tra uplink e downlink [13].

Per comunicare con una eNB che gestisce più celle, lo User Equipment deve identificare il downlink di trasmissione da una di quelle celle e sincronizzarsi con esso, tramite opportuni segnali: il **Primary Synchronization Signal (PSS)** e il **Secondary Synchronization Signal (SSS)**, trasmessi periodi-

camente in posizioni riservate, nel dominio del tempo, due volte ogni frame di 10 ms. Essi permettono di fornire all'UE l'identificazione della cella e l'informazione se essa utilizza Frequency Division Duplex (FDD) o Time Division Duplex (TDD).

2.3.3 LTE uplink: SC-FDMA

Lo schema di accesso multiplo utilizzato da LTE per l'uplink è il SC-FDMA, noto anche come modulazione OFDM *DFT-spread*, ed è una variante di OFDM avente l'obiettivo di ridurre il PAPR e quindi diminuire il consumo di potenza del terminale utente.

SC-FDMA ha in comune con OFDMA la divisione della banda in sottoportanti parallele, con l'utilizzo di prefissi ciclici per prevenire l'ISI; tuttavia, mentre in OFDM i dati vengono modulati indipendentemente su ciascuna sottoportante, in SC-FDMA il segnale modulato su una data sottoportante è una combinazione lineare di tutti i simboli trasmessi nello stesso istante di tempo: dunque in ciascun periodo di simbolo tutte le sottoportanti trasmesse portano una componente di ogni simbolo modulato, dando origine alla proprietà di *single-carrier*, con un conseguente abbassamento del PAPR [13].

La differenza tra OFDMA e SC-FDMA è illustrata in figura 2.13.

La struttura del *frame* di trasmissione uplink è identica a quella del frame in downlink e sono presenti anche in questo caso alcuni Reference Signals: **Demodulation RS (DM-RS)**, associato alla trasmissione di dati in uplink, alla segnalazione di controllo e alla stima di canale per la demodulazione coerente; **Sounding RS (SRS)**, usato per la determinazione della qualità di canale per favorire lo scheduling.

Esistono due varianti di SC-FDMA, che differiscono nell'utilizzo dello spettro. SC-FDMA *localizzato* utilizza bande di frequenza contigue per un singolo utente; SC-FDMA *distribuito* alloca ad ogni utente intervalli di frequenze non continue, sempre conservando il vantaggio di un basso PAPR in confronto a OFDMA [10].

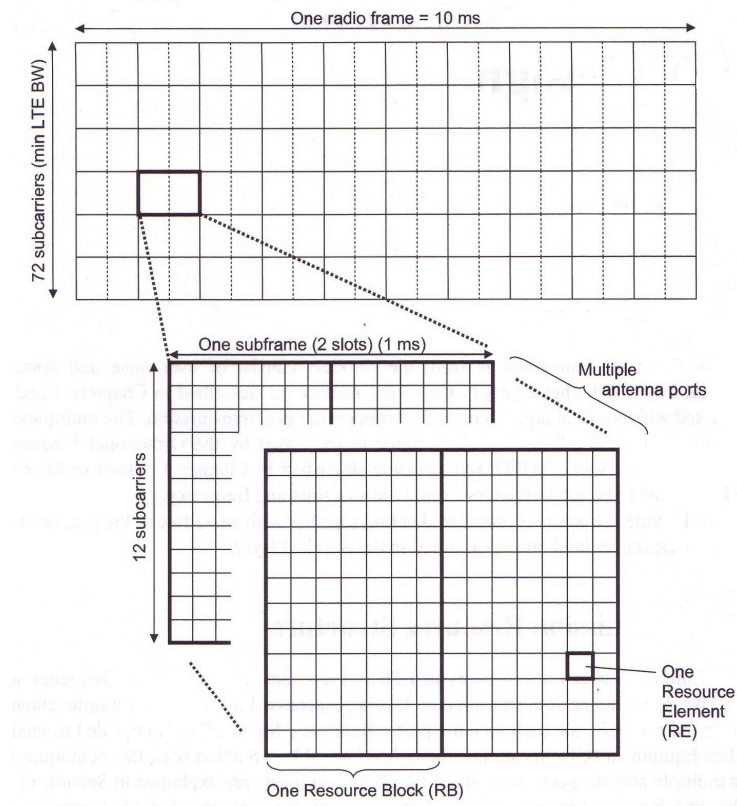


Fig. 2.12: Struttura del frame di trasmissione LTE downlink (tratto da [13])

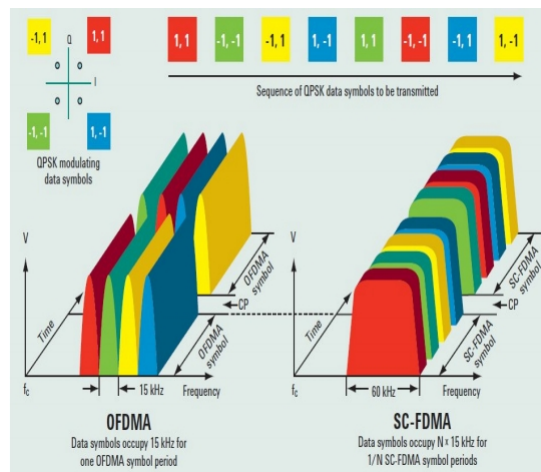


Fig. 2.13: OFDMA vs SC-FDMA (tratto da <http://www.eeweb.com/project/atitpatel>)

Capitolo 3

LTE downlink e multimedia

L'eNodeB in un sistema LTE è responsabile di gestire lo scheduling delle risorse di canale. Si definiscono **chunks** (in italiano ‘spezzoni’) le più piccole unità di risorse trasmissibili, che corrispondono a 12 sottoportanti OFDM in un millisecondo, cioè a due Resource Blocks. L’allocazione può essere modificata dinamicamente una volta ogni subframe, cioè ogni millisecondo, e la decisione di scheduling riguarda non solo l’assegnazione dei Resource Blocks, ma anche quale *Modulation and Coding Scheme (MCS)* utilizzare. L’obiettivo da raggiungere è quello di fornire il QoS desiderato in un canale condiviso: sistemi di comunicazioni tradizionali come UMTS e GSM lo facevano pre-allocando le risorse radio in canali dedicati; in LTE, dove esistono solo due canali condivisi (uno in uplink e uno in downlink), è compito dello scheduler garantire la soddisfazione dei QoS e aumentare la capacità del sistema, basando la sua decisione di scheduling su fattori come la classe di QoS, il ritardo tollerabile, le condizioni istantanee di canale o gli indicatori di *fairness* [9].

Gli algoritmi di scheduling fanno uso [13] di due tipi di informazioni: il *Channel State Information (CSI)* e altre *misure di traffico*, ottenuti da misure dirette all’eNodeB e/o via canali di feedback. In particolare è molto importante il **Channel Quality Indicator (CQI)**, trasmesso dall’UE come indicazione del data-rate che può essere supportato dal canale tenendo conto del *Signal-to-Interference-plus-Noise Ratio (SINR)* e delle caratteristiche del ricevitore.

Per valutare le condizioni di canale, si assume che gli utenti siano in grado di stimare e inviare perfettamente le informazioni alla stazione base, che tali

informazioni siano disponibili sempre alla eNB all'inizio di ciascun blocco di trasmissione e che il fading rate del canale sia lento abbastanza da poter considerare le condizioni di canale quasi statiche in ciascun blocco OFDM [1].

In generale [13], possono essere individuate due filosofie opposte di scheduling:

- *opportunistic scheduling*: mira a massimizzare la somma dei data rates trasmessi a tutti gli utenti, sfruttando il fatto che diversi utenti hanno diversi guadagni di canale e quindi sperimenteranno buone condizioni di canale a tempi e frequenze differenti. Si allocano dunque le risorse agli utenti aventi le condizioni di canale migliori per trasmettere, dal momento che è stato dimostrato [1] che così facendo il throughput totale del sistema viene massimizzato. Il limite principale risiede nel fatto che, in tal modo, risulta difficile garantire a ciascun utente un adeguato fairness e il QoS richiesto: i dati degli utenti non possono sempre attendere che le condizioni di canale siano sufficientemente favorevoli.
- *fair scheduling*: mira a garantire un data-rate minimo a ciascun utente piuttosto che a massimizzare il data-rate totale del sistema, ponendo attenzione alla latenza (fattore molto importante in applicazioni real-time). Una metrica per valutare il grado di fairness fornito da un algoritmo è la *Cumulative Distribution Function (CDF)* dei throughput di tutti gli utenti, come indicato in figura 3.1: un sistema è sufficientemente equo se la CDF dei throughput giace a destra di una particolare linea.

Nella pratica la maggior parte degli algoritmi di scheduling ricade tra le due tecniche, con l'obiettivo di evitare la penalizzazione degli utenti a bordo cella e di accrescere il throughput degli utenti in buone condizioni di canale [13].

Di seguito verranno presentati quattro schedulers proposti da diversi autori: Algoritmo *Rate Utility-Function Maximization* e Algoritmo *Two-Steps*, di Assaad et al. [2], Algoritmo *Delay-Buffer Based* ad opera di Ramli et al. [4], Algoritmo *Weighted Round-Robin* di Luo et al. [5].

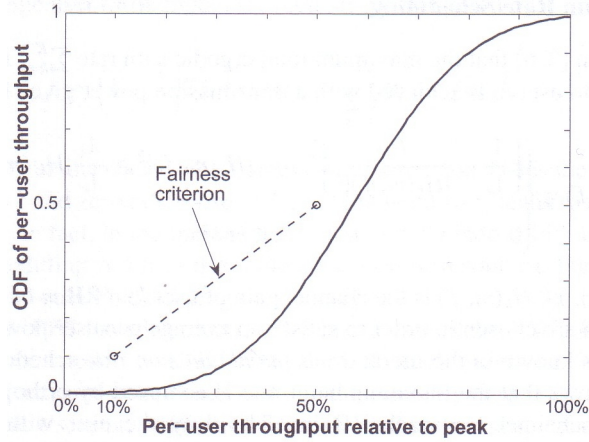


Fig. 3.1: CDF dei throughput

3.1 Algoritmo Rate Utility-Function Maximization

L'allocazione dinamica delle risorse è considerata un problema di ottimizzazione per risolvere il quale sono state usate preliminarmente due strategie: *Margin Adaptive*, che vuole minimizzare la potenza totale nel rispetto dei vincoli di rate, QoS e Bit Error Ratio (BER) dei singoli utenti; *Rate Adaptive*, che mira a massimizzare il rate totale trasmesso tenendo conto dei vincoli sulla potenza totale trasmessa [1].

A fronte di una eccessiva complessità nella soluzione di questo problema di ottimizzazione, che le rende non implementabili, [2] propone due algoritmi. Il primo, Algoritmo Rate Utility-Function Maximization (**RUFM**), si propone di allocare i chunks in ogni istante in modo tale da massimizzare la seguente funzione utilità:

$$\operatorname{argmax} \sum_{k=1}^K \frac{\frac{R_{k,min}}{R_k}}{\sum_{j=1}^K \frac{R_{j,min}}{R_j}} r_k \quad (3.1)$$

dove r_k e R_k sono rispettivamente il rate istantaneo e medio raggiunto dall'utente k -esimo all'istante t , mentre $R_{k,min}$ è il rate minimo richiesto dall'utente k -esimo.

Lo scheduler classifica i chunks nel seguente modo:

$$\frac{CQI_k^l}{\frac{1}{L} \sum_{i=1}^L CQI_k^i} \frac{CQI_k^l}{\frac{1}{K} \sum_{j=1}^K CQI_j^l} \quad (3.2)$$

in modo da valutare la qualità di canale per ciascun utente k su ciascun chunk l . Successivamente assegna il chunk avente il valore di (3.2) più elevato all'utente che massimizza la funzione utilità. Poi passa al chunk successivo, valuta il CQI *effettivo* di ciascun utente e vi attribuisce un MCS basandosi sul chunk allocato al passo precedente e su quello attuale; il CQI effettivo per l'utente k -esimo è uno scalare che viene calcolato a partire dal vettore dei CQI sperimentati dall'utente k su ciascun chunk. Sulla base del nuovo valore del CQI effettivo e del rate raggiunto, calcola la funzione utilità per ciascun utente e attribuisce il chunk all'utente che la massimizza, scegliendo di volta in volta uno tra i K utenti.

Una volta assegnati tutti i chunks, viene valutato il rate istantaneo r_k di ciascun utente, e viene aggiornato il rate medio raggiunto a lungo termine dall'utente k -esimo con la formula

$$R_k[t+1] = \left(1 - \frac{1}{t_c}\right) R_k[t] + \frac{1}{t_c} r_k[t] \quad (3.3)$$

con t_c finestra temporale su cui si impone il fairness. La scelta di una finestra temporale t_c grande tende a massimizzare il throughput medio totale; una finestra piccola fa tendere verso uno scheduling di tipo round-robin, con fairness maggiore [13]. Nell'Algoritmo RUFM, t_c è impostato al valore intermedio 1000 per ottenere un buon compromesso tra i due.

3.2 Algoritmo Two-Steps

Il secondo algoritmo proposto in [2] (Algoritmo Two-Steps, **TS**) ha l'obiettivo di semplificare il problema suddividendolo in due sotto-problemi:

- **allocazione dei chunks:** determina il numero di chunks da allocare a ciascun utente, su un totale di L . La decisione è basata sulle condizioni istantanee del canale e sul rate medio raggiunto dall'utente.

Considerando il vettore dei CQIs relativi all'utente k su ogni chunk, sia CQI_k^l l'indicatore del k -esimo utente sull' l -esimo chunk, e $ECQI_k$ il CQI effettivo. Attribuito un opportuno MCS all'utente, viene valutato il rate effettivo, $Eff.rate_k$, che rappresenta il bit rate dell'utente k -esimo come se tutti i chunks fossero attribuiti a lui. Se R_k rappresenta il rate medio raggiunto dall'utente k -esimo ad un dato istante e $R_{k,min}$ il minimo rate richiesto dall'utente k -esimo, viene valutato il

numero di chunks per ciascun utente come:

$$L_k = \left\lfloor \frac{Eff.rate_k}{\sum_{i=1}^K Eff.rate_i} \frac{\frac{R_{k,min}}{R_k}}{\sum_{i=1}^K \frac{R_{i,min}}{R_i}} L \right\rfloor \quad (3.4)$$

Se al termine del calcolo rimangono chunks, essi vengono allocati agli utenti che ne hanno di meno, per accrescere il fairness.

- **assegnazione dei chunks:** assegna agli utenti la quantità di chunks calcolata al passo 1, scegliendo i chunks dopo averli classificati in ordine decrescente del rapporto (3.2) definito nell'algoritmo precedente. Lo scheduler costruisce una matrice di L righe e $K + 2$ colonne delle quali la prima contiene il massimo valore del rapporto (3.2), con massimo preso tra tutti gli utenti, la seconda gli indici di chunk, le altre K gli indici degli utenti classificati in ordine decrescente del loro rapporto (3.2).

Lo scheduler ordina la matrice in ordine decrescente dei valori della prima colonna, e assegna il chunk avente il massimo elemento nella prima colonna all'utente che raggiunge quel massimo, cioè quello il cui indice compare nella terza colonna. Se viene raggiunto il numero massimo di chunk per un dato utente, lo scheduler assegna il chunk all'utente che raggiunge il massimo valore di (3.2) per quel chunk dopo il primo, cioè quello nella quarta colonna, e così via finché tutti i chunks sono stati allocati.

Infine, dopo le assegnazioni, viene calcolato l'*ECQI* di ogni utente e scelto un MCS con massimo rate istantaneo r_k . Il valore medio del rate raggiunto viene aggiornato con (3.3).

3.3 Algoritmo Delay-Buffer Based

L'algoritmo proposto in [4] (Algoritmo Delay-Buffer Based, **DBB**) costituisce un'estensione di un algoritmo di tipo opportunistico e si basa sul valore istantaneo dell'SNR in downlink, sul throughput medio, sul *Head Of Line (HOL)* packet delay (differenza tra il tempo corrente e il tempo di arrivo di un pacchetto) e sull'informazione riguardo il buffer di ciascun utente. Anch'esso divide lo scheduling in due parti: allocazione e assegnazione dei chunks.

Il numero dei chunks da dedicare a ciascun utente viene determinato in 3 passi:

- Si determina il numero di chunks per l'utente i -esimo *attivo* (avente dati da ricevere) all'istante t con la seguente formula:

$$n_i(t) = \left(\frac{r.avg_i(t)}{\frac{1}{N_u(t)} \sum_{j=1}^{N_u(t)} r.avg_j(t)} \frac{\frac{1}{R_i(t)}}{\frac{1}{N_u(t)} \sum_{j=1}^{N_u(t)} \frac{1}{R_j(t)}} \right) \quad (3.5)$$

dove $i, j \in \text{utenti attivi}$ e $r.avg_i(t)$, SNR medio dell'utente i -esimo all'istante t , si definisce come:

$$r.avg_i(t) = \frac{1}{N_{RB}} \sum_{k=1}^{N_{RB}} r_{i,k}(t) \quad (3.6)$$

con $N_u(t)$ numero totale di utenti attivi all'istante t , $R_i(t)$ throughput medio dell'utente i -esimo all'istante t definito come in (3.3), $r_{i,k}(t)$ SNR istantaneo dell'utente i sul RB k .

- Se dopo la determinazione di $n_i(t)$ per ogni utente rimangono ancora RBs disponibili, tutti gli utenti attivi sono ordinati in ordine crescente di priorità sulla base di

$$d_i(t) = T_i - W_i(t) \quad (3.7)$$

che rappresenta il tempo rimanente all'utente i -esimo per raggiungere la sua soglia di ritardo tollerata (con T_i soglia di ritardo e $W_i(t)$ HOL packet delay). Il passo 2 assegna un RB a ciascun utente, iniziando da quello con priorità maggiore, fino a esaurimento dei blocchi.

- Se dopo il passo 1 si ha che il numero di RBs allocati è maggiore di quelli effettivamente disponibili (N_{RB}) e quindi non ci sono abbastanza RBs per tutti gli utenti attivi, è necessario ridurre il numero di RBs calcolato per alcuni utenti sulla base della valutazione del $d_i(t)$: in questo modo, utenti aventi pacchetti vicini alla soglia di ritardo possono tenere tutti i loro RBs precedentemente

allocati, mentre può essere ridotto il numero di RBs attribuito agli utenti i cui HOL packet delay sono lontani dalla scadenza.

L'assegnazione dei RBs avviene ordinando gli utenti in ordine decrescente di priorità sulla base del numero di pacchetti scartati e attribuendo i migliori $n_i(t)$ RBs all'utente con priorità massima, passando via via agli utenti con priorità sempre minore. L'algoritmo termina quando ciascun utente ha ricevuto il numero di RBs associatogli in fase di allocazione.

3.4 Algoritmo Weighted Round-Robin

In [5] viene proposto un algoritmo di allocazione dinamica delle risorse basato su una tipologia di Weighted Round Robin, **WRR**, che considera tre fattori: vincoli sul packet delay (conosciuto a priori) variabile a seconda dell'applicazione che si intende utilizzare, qualità di canale (variabile), data-rate storico medio dell'utente n su un dato RB (si considera un sistema di N utenti e K sottocanali). Il sistema seleziona poi un MCS per l'utente n sul blocco k selezionato, per determinare il data-rate istantaneo e quello storico medio, di cui tiene conto allo scopo di evitare che alcuni utenti usufruiscano troppo a lungo delle risorse.

L'Algoritmo WRR considera tre parametri di design per ogni utente all'interno di un frame OFDM, ciascuno dei quali è ottenuto pesando l'utente su tutti i blocchi di risorse disponibili nel frame OFDM corrente: dunque ogni parametro di design si presenta come un vettore di K elementi.

- *rate di canale disponibile*: ogni RB presenta un diverso SNR, il cui valore viene restituito dall'utente che aveva precedentemente utilizzato quel blocco; il rate disponibile per un utente su un dato blocco viene valutato in accordo con il teorema di Shannon:

$$R_{max} = \frac{1}{2} \log_2 (1 + SNR) \quad (3.8)$$

Dunque si ottiene, per ogni utente, un vettore w_a i cui elementi vengono poi normalizzati dividendoli per il massimo rate di cana-

le disponibile tra tutti gli utenti su tutti i RBs. Per massimizzare il throughput totale, più alto è il rate di canale disponibile a un dato utente su un dato RB, più elevata è la necessità di assegnare quel blocco all'utente: a questo parametro sarà dato peso *positivo*.

- *vincolo sul packet delay*: ad ogni utente è attribuito un vincolo sul delay, T_n^{max} , ad ogni periodo di allocazione. Si ottiene per ogni utente un vettore w_b i cui elementi sono tutti uguali, dal momento che questo vincolo di delay non dipende dalle condizioni di canale dei RBs. Un vincolo di delay minore significa urgenza maggiore nell'attribuire risorse all'utente corrente: a questo parametro sarà dato peso *negativo*.
- *data-rate storico medio*: per garantire fairness, si tiene conto di questo terzo parametro di design. Si definisce per ogni utente il vettore w_c dei data-rate medi di tale utente su un dato RB in un intervallo temporale t_w . Gli elementi di w_c sono poi normalizzati rispetto al massimo data-rate storico medio tra tutti gli utenti su tutti i RBs: un valore maggiore di un elemento significa minore priorità necessaria nell'allocazione del RB corrente all'utente dato. A questo parametro sarà dunque dato peso *negativo*.

I tre parametri vengono sommati e combinati in un vettore che, per l'utente n , vale

$$w_n = \alpha_{a,n}w_{a,n} - \alpha_{b,n}w_{b,n} - \alpha_{c,n}w_{c,n} \quad (3.9)$$

con $\alpha_{a,n}$, $\alpha_{b,n}$, $\alpha_{c,n}$ coefficienti usati per decidere il peso dei tre parametri di design.

Successivamente gli elementi dei vettori di tutti gli N utenti vengono inseriti in un unico vettore e ordinati in senso decrescente. Partendo dal primo elemento del vettore, ciascun RB viene assegnato all'utente che su quel blocco ha peso massimo tra tutti gli utenti; se due o più utenti hanno lo stesso peso (massimo) per un blocco, l'allocazione segue la modalità *round-robin*, cioè un ordine ciclico. Dopo un'assegnazione, tutti gli elementi riguardanti il RB appena assegnato vengono eliminati dal vettore totale per evitare allocazioni multiple della stessa

risorsa. In figura 3.2 un esempio con due utenti U_1 e U_2 :



Fig. 3.2: Esempio di allocazione delle risorse con Algoritmo 4: a) vettori risultanti per ciascun utente; b) vettore ordinato dei valori per ogni utente su ogni RB; c) risultato dell'allocazione (tratto da [5])

Capitolo 4

Valutazione delle prestazioni

Nella valutazione delle prestazioni degli algoritmi di scheduling entrano in gioco vari fattori.

Il primo obiettivo da raggiungere nell'allocazione dinamica delle risorse è il giusto compromesso tra *throughput* e *fairness*: si cerca di massimizzare il primo mantenendo una certa equità tra gli utenti, in modo da favorire non solo gli utenti che godono delle migliori condizioni di canale ma anche quelli che si trovano a bordo cella o comunque sperimentano condizioni non troppo favorevoli.

Algoritmi come quelli di tipo Max-Rate [3] sfruttano efficientemente il canale dal momento che vengono sempre scelti gli utenti in condizioni di canale migliori, ma in questo modo utenti che risentono di condizioni più sfavorevoli non possono accedervi efficacemente. Viceversa, in algoritmi che allocano ciclicamente un uguale intervallo temporale agli utenti per la trasmissione (Round Robin, [3]) viene assicurata equità massima, ma al prezzo di una grande degradazione del throughput.

L'Algoritmo RUFM in questo senso (simile a quello che in [3] viene chiamato algoritmo Proportional Fair) garantisce un compromesso accettabile tra throughput e fairness; tuttavia, come illustrato in figura 4.1, la Cumulative Distribution Function del bit rate medio raggiunto per utente ottenuta per l'Algoritmo TS (in rosso) dimostra che il numero di utenti che hanno probabilità di raggiungere rates più alti di 1

Mbps è maggiore rispetto alla CDF ottenuta per l'Algoritmo RUFM (in verde), anche se i due algoritmi danno entrambi un throughput di cella di circa 22 Mbps [2]. Inoltre la probabilità che un utente abbia throughput inferiore a 500 kbps è pari a 0.2 nell'Algoritmo RUFM, mentre tende a zero per l'Algoritmo TS: questo risultato è dovuto al maggior controllo che l'Algoritmo TS ottiene suddividendo in due passi la gestione delle risorse. Tutte queste considerazioni portano a preferire l'Algoritmo TS per ottenere un buon compromesso tra throughput e fairness, anche per la sua maggiore semplicità implementativa [2].

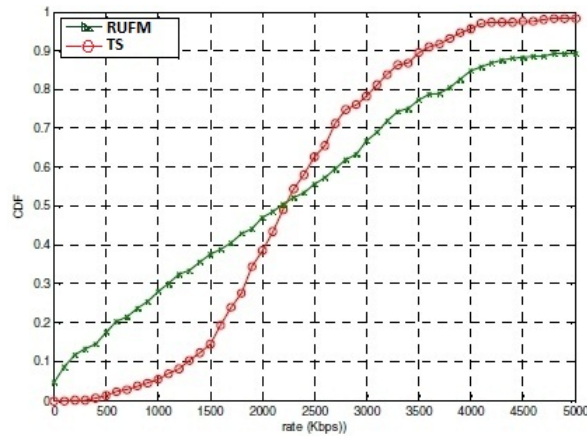


Fig. 4.1: CDF del rate utente raggiunto. Risultato ottenuto con una trasmissione su frequenza portante 2 GHz, larghezza di banda 10 MHz, 12 sottoportanti per RB, 14 simboli OFDM per TTI di lunghezza 1 ms (tratto da [2]).

Le prestazioni dell'Algoritmo DBB vengono valutate dagli autori di [4] in confronto a quelle dell'algoritmo opportunistico di cui è una variante. Si considera una cella esagonale con frequenza portante di 2 GHz e larghezza di banda di 5 MHz costituita da 25 RBs; la eNodeB si trova al centro della cella e gli utenti (da 20 a 100) sono collocati uniformemente all'interno della cella e si muovono in direzioni casuali con velocità variabile tra 1 e 100 km/h. Si assume che gli utenti richiedano traffico video real-time con soglia di ritardo di 20 ms.

I parametri di confronto sono il throughput di sistema, il Packet Loss

Rate e la utilizzazione dei RBs, definita come:

$$RB_{util} = \frac{1}{T} \sum_{t=1}^T \frac{total_{RB_{used}}(t)}{N_{RB}} \quad (4.1)$$

con $total_{RB_{used}}(t)$ numero di RBs allocati che stanno venendo usati all'istante t , T tempo totale di simulazione e N_{RB} numero totale di RBs disponibili.

È evidente dalla figura 4.2 che l'utilizzo dell'Algoritmo DBB comporta un aumento del throughput di sistema rispetto all'algoritmo opportunistico; la figura 4.3 pure dimostra un'utilizzazione più efficiente delle risorse disponibili, dovuta al fatto che l'Algoritmo DBB evita la sovrallocatione delle risorse che quindi non rimangono sotto-utilizzate dagli utenti che non ne hanno bisogno: ciò determina un aumento del throughput [4]. Si nota infine dalla figura 4.4 una forte diminuzione del PLR, che in accordo con gli autori di [4] dovrebbe essere mantenuto sotto lo 0.01 per avere un video streaming accettabile: l'Algoritmo DBB lo mantiene al di sotto di tale valore fino a un limite di 80 utenti presenti nel sistema, contro i 50 dell'algoritmo opportunistico.

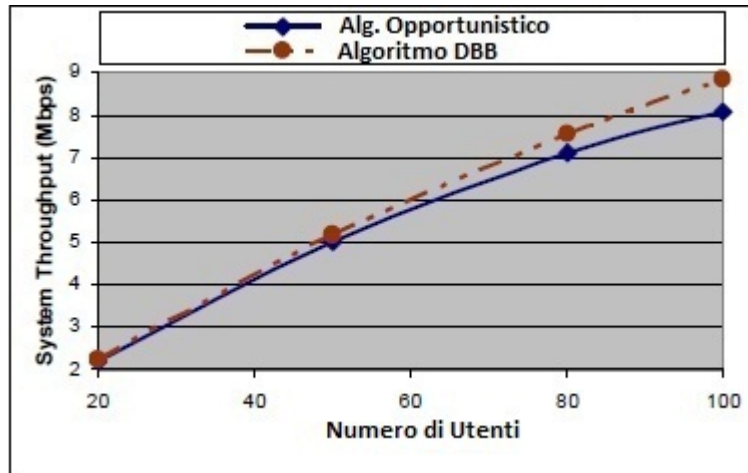


Fig. 4.2: Throughput di sistema totale in Mbps vs. numero di utenti (tratto da [4])

In [5] gli autori valutano invece le performances dell'Algoritmo WRR in confronto a un generico algoritmo di scheduling avente il solo obiet-

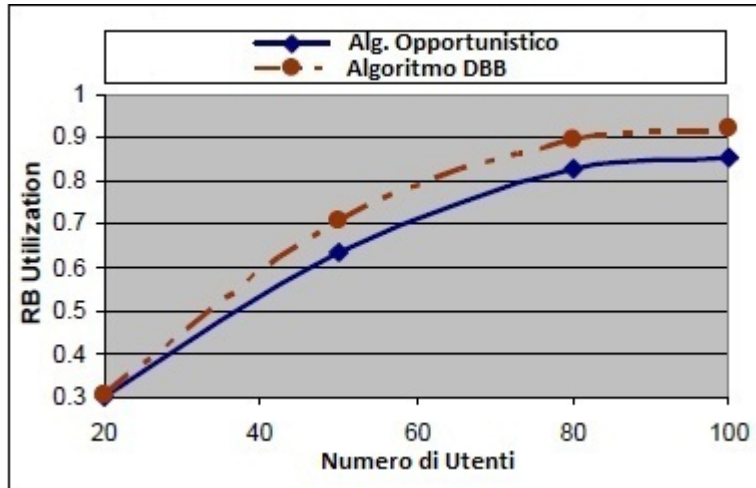


Fig. 4.3: Utilizzazione RB vs. numero di utenti (tratto da [4])

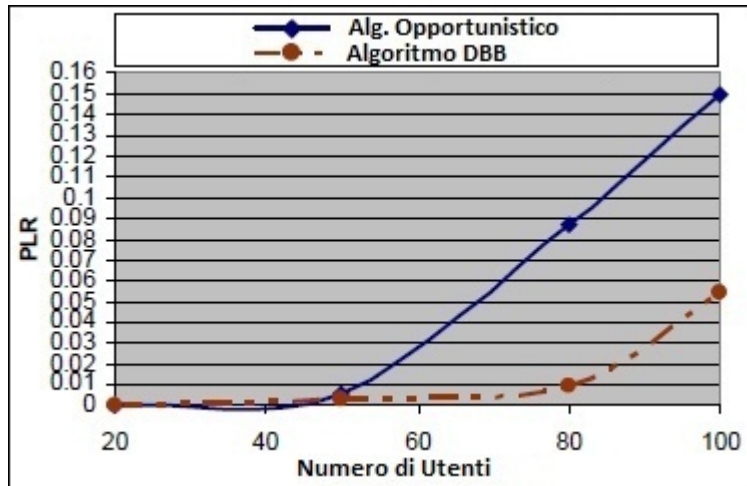


Fig. 4.4: PLR vs. numero di utenti (tratto da [4])

tivo di massimizzare il throughput. La valutazione avviene in termini di aumento di *Peak SNR (PSNR)* (rapporto tra la massima potenza possibile del segnale video e la potenza del rumore da cui è affetto), rispetto all’algoritmo opportunistico, per tre utenti con vincoli di delay di 20, 30 e 40 ms rispettivamente.

La figura 4.5 mostra come l’utilizzo dell’Algoritmo WRR porti ad un aumento del PSNR di oltre 30 dB rispetto all’algoritmo puramente opportunistico (oltre 37 dB se si utilizza anche l’ottimizzazione cross-layer trattata in [5] che mira a minimizzare la distorsione video attesa). In particolare, anche in presenza di una qualità di canale piuttosto bassa (15 dB) e di un vincolo di delay alquanto stretto (20 ms) si hanno prestazioni più elevate: merito del fatto che l’allocazione dinamica dell’Algoritmo WRR tiene conto proprio di questo vincolo di delay come parametro *negativo*; dunque più il vincolo è stretto, più il peso dato all’utente risulterà maggiore, e più risorse gli verranno assegnate.

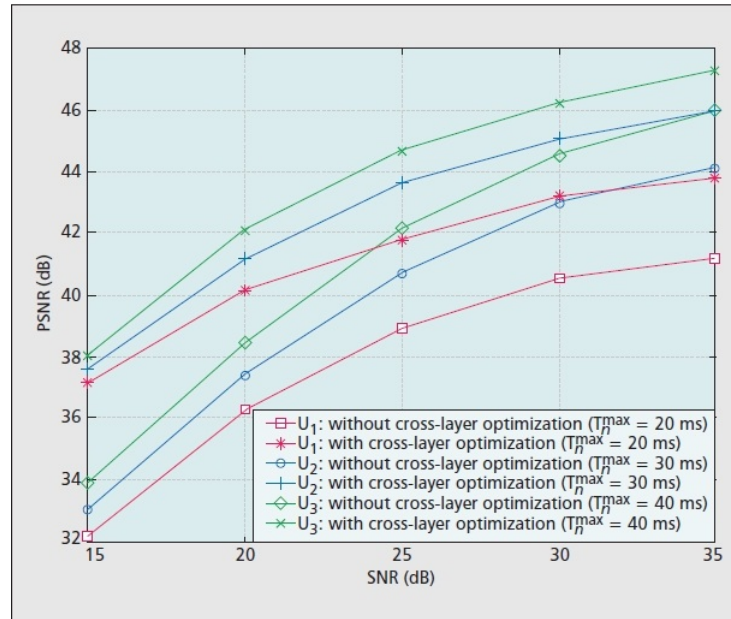


Fig. 4.5: Confronto tra PSNR di Algoritmo WRR e algoritmo opportunistico (tratto da [5])

Un confronto tra le prestazioni degli Algoritmi proposti nel capitolo 3 si può estrapolare da [3], con le dovute cautele. L’articolo [3] con-

fronta, in termini di throughput, PLR e fairness alcuni algoritmi. In particolare, tra gli altri, prende in considerazione un algoritmo *Proportional Fair (PF)* che seleziona gli utenti massimizzando, per ogni utente, un rapporto tra data-rate raggiungibile e data-rate medio, e un algoritmo *Maximum-Largest Weighted Delay First (M-LWDF)*, che tiene conto, oltre che di quei data-rate, anche dell'HOL packet delay e della soglia di ritardo di ciascun utente.

La simulazione avviene nelle stesse identiche condizioni in cui avviene la simulazione dell'Algoritmo DBB in [4], con un numero di utenti variabile tra 80 e 120. I risultati sono illustrati nelle figure 4.6, 4.7, 4.8.

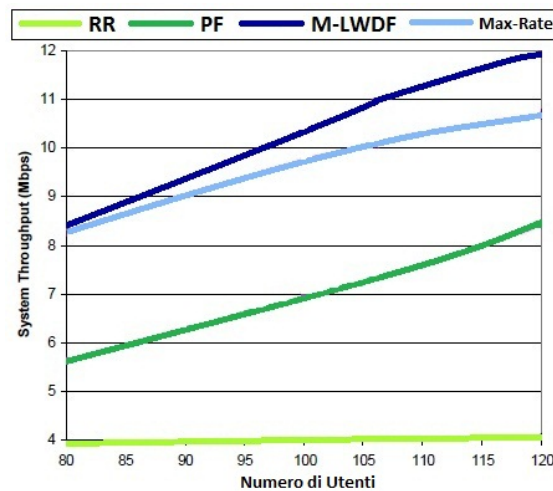


Fig. 4.6: Throughput di sistema totale in Mbps vs. numero di utenti (tratto da [3])

Dunque, si potrebbero fare rientrare gli Algoritmi RUFM e TS (simili tra loro in prestazioni salvo un compromesso peggiore tra throughput e fairness [2]) nella categoria simil-PF e gli Algoritmi DBB e WRR nella categoria simil-M-LWDF: quest'ultima considerazione non è irragionevole dal momento che, riportando il grafico di figura 4.2 in quello di figura 4.6 si ottiene che l'andamento sarebbe di circa una unità al di sotto della linea di M-LWDF; anche per quanto concerne il PLR si ha che la linea rappresentante il PLR dell'Algoritmo DBB (figura 4.4) starebbe di pochissimo al di sopra di quella relativa a M-LWDF di figura 4.7.

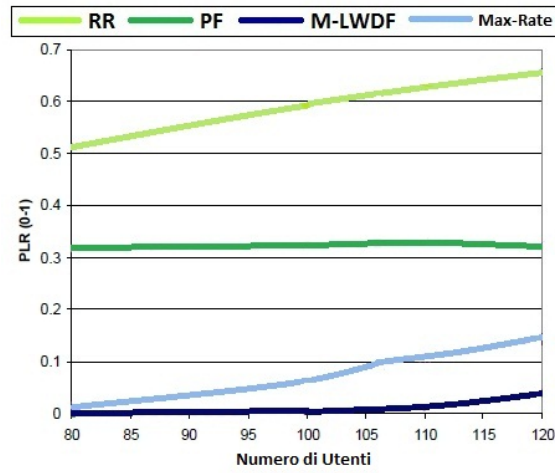


Fig. 4.7: PLR vs. numero di utenti (tratto da [3])

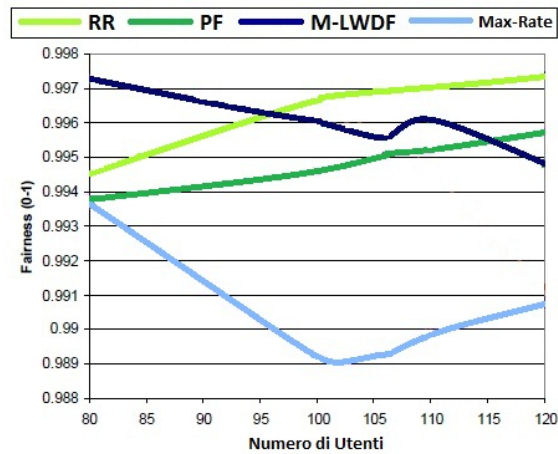


Fig. 4.8: fairness vs. numero di utenti (tratto da [3])

I risultati quindi mostrano che algoritmi di tipo M-LWDF presentano un throughput nettamente superiore a quello di algoritmi di tipo PF, e che questo throughput cresce al crescere del numero di utenti. Algoritmi di tipo M-LWDF riescono inoltre a supportare fino a 100-110 utenti al di sotto della soglia massima di PLR accettabile, fissata dagli autori di [3] e [4] al valore 0.01; lo stesso non si può dire degli algoritmi di tipo PF, che presentano una grossa perdita di pacchetti, di un fattore circa 0.3. Infine si nota come del resto gli algoritmi di tipo PF gestiscano meglio il fairness in presenza di un elevato numero di utenti, seppure il fairness degli algoritmi M-LWDF sia comunque accettabile.

In definitiva, dall'analisi degli algoritmi proposti e delle loro prestazioni emerge che l'allocazione dinamica può essere perfezionata e adattata a esigenze di QoS più elevate tenendo conto di un numero maggiore di parametri. La categoria di algoritmi PF, che basa le decisioni di scheduling unicamente sul rate utente raggiunto e sulle condizioni di canale, si presenta come una scelta adatta ad applicazioni e servizi di comunicazione *non real-time* generici, come il trasferimento file e il video sharing. La categoria di algoritmi M-LWDF invece, tenendo conto di parametri ulteriori come i vincoli di ritardo di ciascun utente e altri dati riguardanti la sua utilizzazione della rete, si presenta come una buona scelta per la gestione di applicazioni *real-time* sensibili al ritardo e alla perdita di pacchetti, come il video streaming e la telefonia IP.

Capitolo 5

Conclusioni

Questa tesi ha presentato tramite considerazioni qualitative come l'architettura semplificata di LTE, la riduzione dell'overhead di segnalazione e l'utilizzo di tecniche di modulazione multiportante OFDM garantiscano una riduzione delle latenze e un aumento dell'efficienza spettrale, per fornire agli utenti servizi di comunicazione in grado di stare al passo con la crescente richiesta di QoS.

Sono stati poi illustrati quattro algoritmi proposti da diversi autori e i risultati delle simulazioni a essi relativi. Dal confronto tra essi sono state individuate due categorie: la prima comprende algoritmi che tengono conto dei rate utente e delle qualità di canale istantanee sperimentate da ciascun utente su ogni sottoportante; la seconda comprende algoritmi che prendono in considerazione anche vincoli di ritardo dettati dalla QoS richiesta da ciascun utente e dati storici sul rate raggiunto e sull'utilizzazione effettiva delle risorse.

È dunque stato possibile verificare che gli algoritmi appartenenti alla prima categoria sono adatti a servizi generici non real-time: essi garantiscono un buon compromesso tra throughput e fairness e per essi ci si aspetta una riduzione della complessità computazionale rispetto agli algoritmi del secondo gruppo, rispecchiata da un corrispondente calo del tempo di elaborazione (la cui valutazione esula dagli scopi di questa tesi); tuttavia sono più scadenti in termini di throughput di sistema e soprattutto di perdita di pacchetti.

Gli algoritmi appartenenti al secondo gruppo invece compensano una maggiore complessità computazionale derivata dall'utilizzo di un nu-

mero maggiore di parametri con prestazioni migliori in termini sia di throughput sia di riduzione del PLR, fattore che li rende la scelta migliore per la gestione di servizi real-time.

Bibliografia

- [1] S. Sadr, A. Anpalagan, K. Raahemifar, *Radio Resource Allocation Algorithms for the Downlink of Multiuser OFDM Communication Systems*, IEEE Communication Surveys Tutorials, Vol. 11, no. 3, pp. 92–106, 2009
- [2] M. Assad, A. Mourad, *New Frequency-Time Scheduling Algorithms for 3GPP/LTE-like OFDMA Air Interface in the Downlink*, IEEE Vehicular Technology Conference, pp. 1964–1968, 2008
- [3] H. A. M. Ramli, R. Basukala, K. Sandrasegaran, R. Patachianand, *Performance of Well Known Packet Scheduling Algorithms in the Downlink 3GPP LTE System*, Proceedings of the 2009 IEEE 9th Malaysia International Conference on Communications, pp. 815–820, 2009
- [4] H. A. M. Ramli, R. Basukala, K. Sandrasegaran, R. Patachianand, M. Xue, C. C. Lin, *Resource Allocation Technique for Video Streaming Applications in the LTE System*, IEEE 19th Annual Wireless and Optical Communications Conference (WOCC), May 2010
- [5] H. Luo, S. Ci, D. Wu, J. Wu, H. Tang, *Quality-Driven Cross-Layer Optimized Video Delivery over LTE*, IEEE Communications Magazine, pp. 102–109, Febr. 2010
- [6] D. Martin-Sacrestà, J. F. Monserrat, J. Cabregas-Penuelas, D. Calabruig, S. Garrigas, N. Cardona, *On the Way towards Fourth-Generation Mobile: 3GPP LTE and LTE-Advanced*, EURASIP Journal on Wireless Communications and Networking, 2009

- [7] F. Mazzenga, C. Monti, F. Vatalaro, *Nuovi Servizi a Larga Banda e Tecnologie per la Mobilità*, Mondo Digitale, no. 1, Mar. 2007
- [8] I Quaderni, pagg. 215-220, Centro Nazionale per l'Informatica nella Pubblica Amministrazione, no. 36, pp. 215-220, Mar. 2008
- [9] A. Larmo, M. Lindström, M. Meyer, G. Pelletier, J. Torsner, H. Wiemann, *The LTE Link-Layer Design*, IEEE Communications Magazine, pp. 52-59, Apr. 2009
- [10] A. Toskala, H. Holma, K. Pajukoski, E. Tirola, *Utran Long Term Evolution in 3GPP*, IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2006
- [11] P. Beming, L. Frid, G. Hall, P. Malm, T. Noren, M. Olsson, G. Rune, *LTE-SAE architecture and performance*, Ericsson Review, no. 3, pp. 98-104, 2007
- [12] V. Mignone Rai, *Che cosa è, come funziona: La modulazione COFDM*, Elettronica e Telecomunicazioni, no. 3, pp. 43-48, Dec. 2005
- [13] S. Sesia, I. Toufik, M. Baker, *LTE - The UMTS Long Term Evolution, From Theory To Practise, Second Edition*, Wiley, Chichester, 2011
- [14] C. Regazzoni, *Trasmissione-Dati con Modulazione OFDM*, Slides del Corso di Sistemi Telecomunicazione Mobili
- [15] O. Oyman, J. Foerster, Y. Tcha, S. Lee, *Toward Enhanced Mobile Video Services over WiMAX and LTE*, IEEE Communications Magazine, pp. 68-76, Aug. 2010
- [16] L. Badia, N. Baldo, M. Levorato, M. Zorzi, *A Markov framework for error control techniques based on selective retransmission in video transmission over wireless channels*, IEEE Journal on Selected Areas in Communications, Apr. 2010
- [17] L. Badia, M. Levorato, M. Zorzi, *Markov analysis of selective repeat type II hybrid ARQ using block codes*, IEEE Transactions on Communications, vol. 56, no. 9, pp. 1434-1441, Sept. 2008.