Capitolo 1

Le reti wireless: evoluzioni e sviluppi

1.1 Introduzione

Il termine “telecomunicazione” prende origine dalla fusione dei concetti di comunicazione e distanza. Il concetto di *comunicazione* ha segnato fin dall'inizio lo sviluppo della nostra società, dal momento che proprio la capacità di comunicare in maniera chiara ed esplicita ha permesso l'evoluzione della nostra specie portandola ad un grado di organizzazione sociale mai raggiunto da altri esseri viventi.
L'atto del comunicare può essere definito come una *trasmissione di informazione* basata su due elementi: il *messaggio* e la *codifica*. Infatti, per trasmettere l'informazione abbiamo bisogno di un messaggio espresso attraverso un codice che sia noto a chi trasmette come a chi riceve. Tali codici possono essere di diverso tipo: ad esempio la lingua italiana, il codice Morse o il codice attraverso il quale comunicano le api, basato su dei cambiamenti di direzione durante il volo.
L'altro concetto, quello di *distanza*, identifica il metro con cui è sempre stata misurata l'evoluzione nella scienza della comunicazione, poiché i progressi fatti registrare dal genere umano in questo campo sono sempre stati evidenziati dalle maggiori distanze che di volta in volta era possibile coprire con l'ausilio delle nuove tecnologie. Infatti, le prime rudimentali tecniche di comunicazione, come i tamburi delle civiltà primitive, i fuochi di segnalazione degli antichi romani o i segnali di fumo degli indiani d'America, seguivano dei codici basati su regole ben precise che avevano però carattere locale, poiché servivano alla comunicazione cosiddetta “a vista” all'interno di territori e culture ben definiti. Con l'evolversi delle civiltà però, le maggiori distanze da coprire esigevano delle comunicazioni sempre più efficaci, mentre la necessità di far comunicare persone di culture diverse spingeva alla creazione di codici la cui comprensione potesse in qualche modo prescindere dalle rispettive culture di appartenenza.

Per soddisfare queste esigenze, si dovrà attendere tuttavia l'avvento dell'elettricità senza la quale la nascita di una vera e propria *scienza delle comunicazioni* sarebbe stata impossibile; questo evento può essere fatto coincidere con l'anno 1837, in cui Samuel Morse brevettava il suo telegrafo basato sul codice a punti e linee che nei tratti essenziali è rimasto identico fino a quando è stato messo a riposo nel 1998.

Facendo un pò di salti nel tempo tra le varie innovazioni tecnologiche, quali ad esempio la telescrivente, la pila ed il telefono, si arriva al 1901, quando Guglielmo Marconi trasmetteva la lettera "S" in alfabeto Morse, da Poldhu in Inghilterra a S.Giovanni di Terranova nel Nord America, usando un oscillatore ed una coppia di antenne. Questo segnava l'inizio delle comunicazioni via radio.

Dall'invenzione della radio alle prime reti wireless però, il passo è tutt'altro che breve; infatti per i primi sistemi cellulari analogici si dovrà aspettare la fine degli anni '70. Da allora, i sistemi wireless sono oggetto di grandi studi ed evoluzioni, che ai giorni nostri non hanno ancora avuto segnali d'arresto. Lo schema evolutivo dei sistemi wireless è riportato in Figura 1.1.

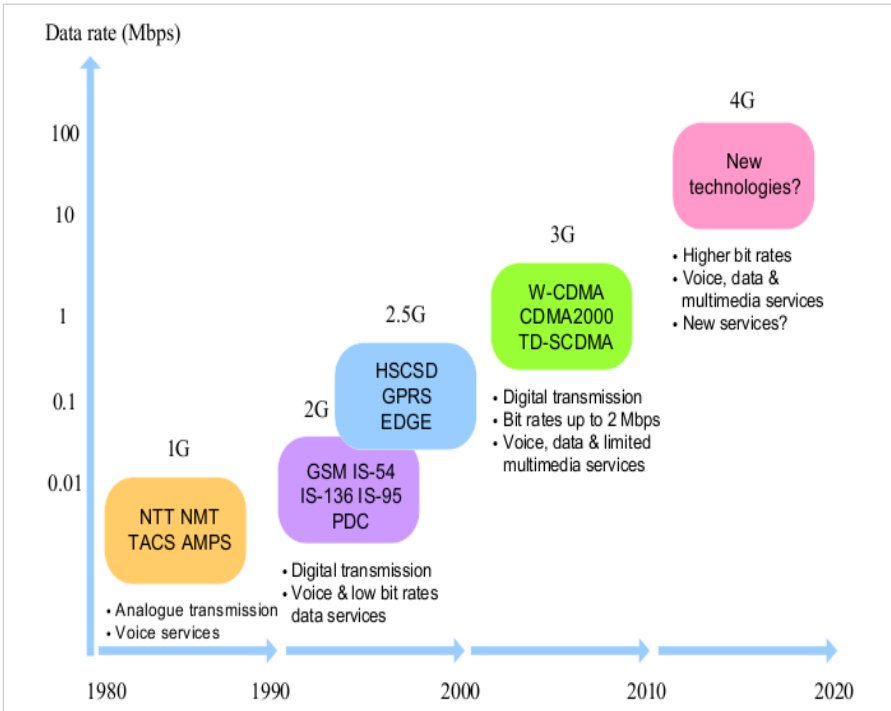

Figura 1.1. Evoluzione delle reti wireless

1.2 Sistemi di prima generazione (1G): E-TACS e AMPS

I sistemi wireless di prima generazione erano basati su trasmissioni analogiche e consentivano esclusivamente comunicazioni di tipo vocale. Il primo sistema wireless del mondo nacque nel 1979 in Giappone, creato dalla *Nippon Telephone and Telegraph* (NTT). Questo sistema utilizzava canali duplex a 600 FM nella banda degli 800 MHz, con una separazione di canali di 25 KHz. Subito dopo, nei primi anni '80, la telefonia cellulare di prima generazione fu introdotta in Europa e negli Stati Uniti. In Europa, i servizi wireless furono impiegati nel 1981 dalla *Nordic Mobile Telephone 450* (NMT-450) che operava in Norvegia, Svezia, Finlandia e Danimarca. La NMT lavorava su un range di 450 MHz ed aveva una capacità totale di banda pari a 10 MHz. Nel 1985, nel Regno Unito fu lanciato il *Total Access Communications System* (TACS) che operava sui 900 MHz. Nella stessa Europa erano presenti altri sistemi analogici minori, il *C-Netz* per la Germania, ed il *Radiocom 2000* per la Francia. Negli Stati Uniti, invece, inizialmente esisteva un solo sistema wireless, lo *Advanced Mobile Phone System* (AMPS). Il servizio AMPS entrò in funzione per la prima volta nel 1983 e la *Federal Communications Commission* (FCC) gli rese disponibili 50 MHz nella banda 824-849 MHz e 869-894 MHz. Lo spettro delle frequenze dell' AMPS era diviso in 832 canali spaziati da 30 KHz [1].

A metà degli anni '80 venne creato uno standard europeo per le comunicazioni analogiche, lo *Extends-TACS*, che derivava dal sistema americano AMPS e da quello britannico TACS. Il sistema E-TACS utilizzava una modulazione di frequenza FM, con la portante contenuta nella banda 935-960 MHz per il *downlink* e 890-915 MHz per l'*uplink.* La spaziatura dei canali fu ridotta da 30 KHz, utilizzati dallo AMPS, a 25 KHz. Per rendere possibile la comunicazione bi-direzionale, o *full-duplex*, ogni canale consisteva di fatto in due canali uni-direzionali, o *simplex,* separati da 45 MHz [2].

Tutti i sistemi cellulari di prima generazione utilizzavano la modulazione di frequenza analogica FM ed ogni chiamata individuale era trasmessa su una frequenza diversa, utilizzando l'accesso multiplo a divisione di frequenza (FDMA).

Per quanto rivoluzionari, i sistemi analogici avevano molte limitazioni, tra cui i vari

problemi legati alla interoperabilità tra sistemi diversi, la bassa capacità dei canali, la scarsa qualità del servizio, il numero dei servizi messi a disposizione, la mancanza di una cifratura delle comunicazioni e gli alti costi delle apparecchiature. A fronte di questi problemi ed anche all'incremento del numero delle utenze, si ebbe la necessità di un'evoluzione dei sistemi stessi che portarono allo sviluppo delle reti wireless di seconda generazione.

1.3 Sistemi di seconda generazione (2G): L'era digitale

La seconda generazione di telefonia cellulare si basava su tecniche di trasmissione digitali. Questo approccio portava con sé alcuni importanti vantaggi: privacy della conversazione, migliore qualità del servizio, maggiore capacità in termini di utenti serviti, servizi di accesso digitali. Inoltre le trasmissioni digitali erano più pratiche ed economiche rispetto a quelle analogiche.

La digitalizzazione permetteva di usare tecniche di accesso multiplo di tipo TDMA e *Code Division Multiple Access* (CDMA) come alternativa alla FDMA[3].

L'uso del TDMA permette di partizionare un canale radio in *timeslot* e ad ogni utente è assegnata una combinazione *frequenza/timeslot* , mentre il CDMA permette a più utenti di occupare la stessa banda senza interferire tra loro assegnando a ciascun utente un unico codice ortogonale[1]. In Figura 1.2 vengono mostrate le tecniche di accesso radio FDMA, TDMA e CDMA.

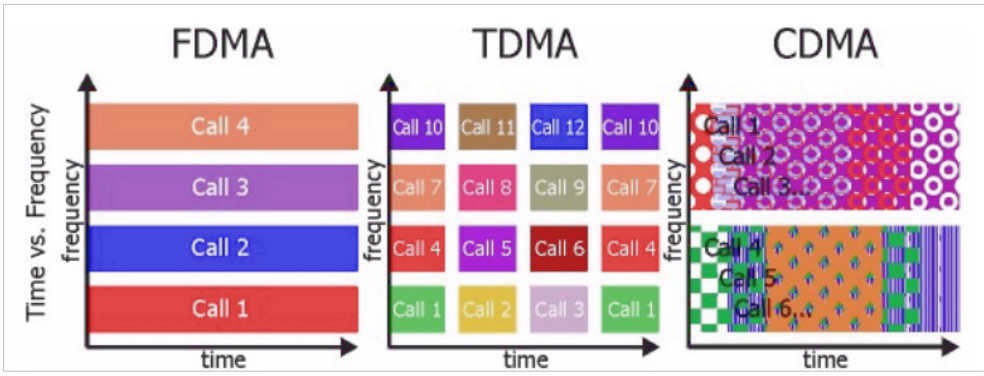

Figura 1.2. Confronto tra tecniche di accesso radio

1 Due codici pseudo-casuali si dicono tra loro ortogonali se la loro correlazione è minima.

Per usare un'analogia con le persone, con la FDMA si parla a gruppi, con la TDMA ognuno parla a turno ed infine con il CDMA tutti parlano in lingue diverse ed ognuno filtra solo quello che viene detto nella propria lingua.

Nel mercato europeo si adottava un solo standard per la telefonia di seconda generazione, mentre negli USA se ne prendevano in considerazione tre. Il primo, introdotto nel 1991, era il sistema *Interim Standard 54* (IS-54), meglio conosciuto come *Digital AMPS* (D-AMPS), e utilizzava TDMA come interfaccia radio riuscendo così a servire tre utenti in un singolo canale AMPS di 30 KHz. Nel 1996 fu introdotta una nuova versione del Digital AMPS (IS-136); questa aggiungeva alcune funzionalità rispetto all'originale, tra cui i messaggi di testo, il C*ircuit Switched Data* (CSD) e un protocollo di compressione migliore. Il terzo, adottato dal 1993, era IS-95 e fu sviluppato da *CDMA Development Group* (CDG) e da *Telecommunications Industry Association* (TIA). Questo sistema lavorava nella banda degli 800 MHz e si basava sul sistema CDMA. Con questa tecnica il canale veniva condiviso da un massimo di 35 utenti in una banda pari a 1.2 MHz.

In Europa, invece, il sistema *Global System for Mobile communications* (GSM) rappresentava lo standard digitale per la telefonia cellulare. Il lancio commerciale per il GSM era previsto per il mese di luglio del 1991, ma fu posticipato al 1992 per la mancanza di terminali mobili conformi allo standard. Alle spalle di questo standard, c'era un lavoro di organizzazione iniziato nel 1982, quando la *Conference Europeenne des Administrations des Postes et des Telecommunications* (CEPT) istituì un gruppo di lavoro per lo studio di un insieme uniforme di regole per lo sviluppo di una futura rete cellulare pan-europea: il *Group Special Mobile* da cui GSM, che successivamente divenne *Global System Mobile communications*.

Sviluppato inizialmente come standard europeo, il GSM si impose presto a livello mondiale a tal punto che la *US Federal Communications Commission* dispose un nuovo blocco di frequenze nella banda dei 1900 MHz in modo da far entrare il sistema GSM1900 nel mercato USA.

Negli anni 2000, il GSM diventava la rete cellulare più diffusa al mondo e vantava circa 80 milioni di utenti nella sola Europa e circa 200 milioni a livello mondiale (di cui 40 solo in Cina). Negli Stati Uniti, dove erano presenti circa 10 operatori, restava seconda solo alla D-AMPS.

1.3.1 Il sistema universale: GSM

Quello che inizialmente ci si aspettava dal "nuovo" sistema era principalmente una migliore qualità del servizio, la possibilità di un roaming pan-europeo e la trasmissione, oltre alle chiamate vocali, di dati per fax, e-mail, files, etc..
In principio, il GSM, doveva operare solo sulla banda di frequenza intorno ai 900 MHz con una larghezza di banda di 50MHz, ma nel 1989 la *UK Department of Trade and Industry* avviò una iniziativa finalizzata ad assegnare 150MHz nella banda dei 1800MHz per *Personal Communications Network* (PCN) in Europa e scelse il sistema GSM come standard sotto il nome di DCS1800. Lo schema delle frequenze assegnate al GSM ed al DCS 1800 è riportato in Figura 1.3.

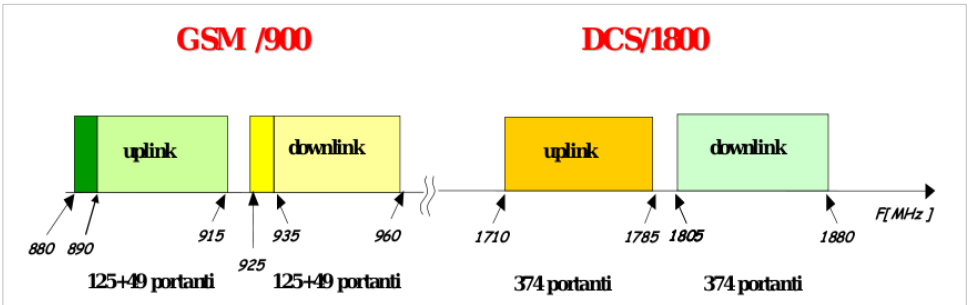

Figura 1.3. Frequenze assegnate

Il GSM utilizzava la tecnica di accesso TDMA ed ogni canale radio era diviso in otto timeslot; ogni utente trasmetteva periodicamente in ognuno degli otto slot, di durata di 0.57 ms, e riceveva nello slot corrispondente.
Le novità sostanziali per l'utente introdotte dal sistema GSM erano rappresentate principalmente dall'uso dei moduli per l'identificazione dell'utente, chiamati *Subscriber Identity Module* (SIM), dagli *Short Message Service* (SMS) e dal *Circuit Switched data service*.
Il servizio a commutazione di circuito era visto come un normale canale di traffico (TCH) capace di tarsportare dati ad un rate di 9.6 kbps e la voce codificata a 13 kbps. Il protocollo utilizzato dal GSM per l'accesso ad internet era il *Wireless Application Protocol* (WAP), che forniva connessioni a commutazione di circuito. Questi tipi di connessioni si rivelarono presto inefficaci sia dal punto di vista dell'utente, che si trovava a pagare la connessione anche in assenza di scambio di

dati (tanto da ribattezzare la tecnologia WAP come “Wait And Pay”), e sia per gli operatori di rete nella gestione delle risorse radio.
Le SIM erano, e sono tutt'oggi, delle schedine intelligenti, dotate di processore e memoria, di tipo *smart card*, su cui sono memorizzate varie informazioni, come il numero identificativo dell'utente, il tipo di rete, gli algoritmi di autorizzazione e di cifratura, i servizi a cui l'utente è abbonato e molto altro; dunque, l'utente viene associato ad una SIM piuttosto che ad un terminale.
Gli SMS sono stati introdotti nel 1994 e permettevano ad una stazione mobile di inviare e ricevere messaggi di testo di 160 caratteri in codifica ASCII a 7 bit. Gli SMS venivano instradati sul canale di controllo che era normalmente utilizzato per comunicazioni tra la rete ed il telefono; ciò significa che era possibile ricevere ed inviare SMS anche se l'utente era occupato in una chiamata vocale.

1.3.2 Generazione 2.5

Con l'incremento del numero di utenze mobili e dei nuovi tipi di servizi, il basso data rates offerto dal sistema GSM divenne obsoleto. Si svilupparono allora nuove tecnologie basate sul sistema GSM: *High Speed Circuit Switched Data* (HSCSD), *General Packet Radio Service* (GPRS) e *Enhanced Data rates for Global Evolution* (EDGE).
Lo scopo del HSCSD [4] era quello di fornire diversi servizi con differenti interfacce radio su una singola struttura di livello fisico. Questa tecnologia permetteva di trasmettere contemporaneamente diversi *timeslot*, in teoria sei, riuscendo ad ottenere un elevato *data-rate* di trasmissione, pari circa a 58 Kbps (utilizzando quattro timeslot a 14.4 Kbps).
L'HSCSD era utilizzabile nelle reti GSM senza apportare alcuna modifica a livello hardware. Anche questa soluzione non era del tutto ottimale nella gestione delle risorse radio, ad esempio la difficoltà di trovare abbastanza timeslot liberi, piuttosto che l'uso continuo delle risorse radio anche in assenza di comunicazione. Gli altri svantaggi del HSCDS erano la tariffazione a tempo, quindi costi elevati da parte dell'utente, una difficile differenzazione delle tariffe, l'elevata probabilità di blocco e di perdita delle chiamate e l'inefficienza per i servizi a bit-rate variabile. Le

caratteristiche desiderabili erano tutt'altre di quelle offerte, tra cui una connessione a commutazione di pacchetto anziché di circuito.
Il GPRS [5] approda sul mercato nel 2000 e si propone di ottimizzare le capacità di accesso alla rete Internet/Intranet ed in effetti "standardizza" l'accesso wireless alla rete con commutazione di pacchetto. La tecnologia a commutazione di pacchetto (*packet-switched*) rende ottimale la gestione delle risorse di rete permettendo di allocare le risorse radio non in maniera continua, ma bensì al bisogno; nell'interfaccia radio, gli slot TDMA sono occupati solo nel momento della trasmissione dei pacchetti. Con l'ausilio del GPRS, l'accesso alla rete IP (internet), è diretto, con la semplice aggiunta, alla parte fissa della rete, di nuovi nodi che sono di fatto *ruoter* IP. Con il GPRS si ha un *data-rates* variabile che va da pochi bits per secondo fino ad un massimo di 171.2 Kbps e offre la possibilità all'utente di rimanere collegato tutto il giorno pagando solo i dati effettivamente trasferiti. Questa tecnologia è particolarmente adatta per applicazioni "non-real time" tipo E-mail e Web surfing.
Un' altra tecnologia a supporto del GSM è l'EDGE [6]. Questa è un "add-on" ai sistemi digitali esistenti ed è capace di aumentare il rate di trasmissione. EDGE usa le stesse strutture radio e la stessa frammentazione TDMA del GSM, ma con un nuovo sistema di modulazione chiamato *eight-phase shift keying* (8PSK) capace di incrementare il *data rates* dello standard GSM di circa tre volte. Con l'EDGE è possibile utilizzare servizi wireless multimediali quali la video chiamata e la video conferenza ad un *data rates* di 384 Kbps.

1.4 Sistemi di terza generazione (3G)

L' aumento del traffico mobile e lo sviluppo di nuove applicazioni hanno reso la capacità del sistema 2G insufficiente rendendo necessario uno standard di terza generazione.
I sistemi di terza generazione sono stati definiti dall' *International Telecommunications Union* (ITU) insieme ai corpi regionali di standardizzazione; lo standard è l'*International Mobile Telecommunications 2000* (IMT-2000). IMT-2000

fornisce un framework per il “worldwide wireless access” concatenando diversi sistemi di reti sia terrestri che satellitari [7]. Il sitema di terza generazione europeo è l' *Universal Mobile Telecommunications System* (UMTS) definito dall' *European Telecommunications Standards Institute* (ETSI). Le frequenze assegnate dalla *World Administrive Radio Convention* (WARC 92) per il 3G sono 1.885-2.025 MHz e 2.110-2.200 MHz comprendenti sia i sistemi terrestri che satellitari.

I punti chiave e gli obiettivi dell'IMT-2000 includono [8]:

- l'integrazione dei sistemi di telecomunicazioni di prima e seconda generazione, sia terrestri che satellitari, nei sistemi di terza generazione;
- Bit-rates superiore a 2Mbps;
- Bit-rates variabile, larghezza di banda “on demand”;
- trasmissioni sincrone ed asincrone;
- trasmissione a commutazione di circuito e a commutazione di pacchetto;
- capacità di trasporto di Internet Protocol (IP) traffic;
- global roaming;
- alta flessibilità a supporto di servizi con diversi requisiti di *Quality of Service* (QoS).

Gli standard che rispettano i requisiti IMT-2000 sono più di uno, al contrario di quanto si pensava in origine: *Wideband CDMA* (WCDMA), *CDMA2000* e *Time Division - Synchronous CDMA* (TD-SCDMA).

Il CDMA2000 è un'evoluzione del primo standard 2G CDMA IS-95. È gestito dalla 3GPP2 che è un'organizzazione del tutto separata ed indipendente dalla 3GPP. Gli operatori del Nord America e dell'Asia Pacifica adottano questo sistema.

Il TD-SCDMA è stato sviluppato dalla *Chinese telecommunications company Datang* ed è stato approvato dall'ITU nel maggio '00 come uno degli standard 3G. In Cina, dove è presente il più grande mercato per le comunicazioni wireless, si adotta questo standard per i sistemi 3G.

Il WCDMA [9] è uno standard 3G, proposto ed elaborato dalla *Third Generation Partnership Project* (3GPP) per fornire comunicazioni multimediali e servizi integrati. I maggiori operatori di reti mobili europei hanno scelto questo standard per le loro soluzioni 3G (UMTS). Il WCDMA è un sistema a banda larga a divisione di codice e supporta due modalità operative: WCDMA/TDD (*Time Division*

Duplexing) e WCDMA/FDD (*Frequency Division Duplexing*). Un esempio delle due modalità di funzionamento è riportato in Figura 1.4.

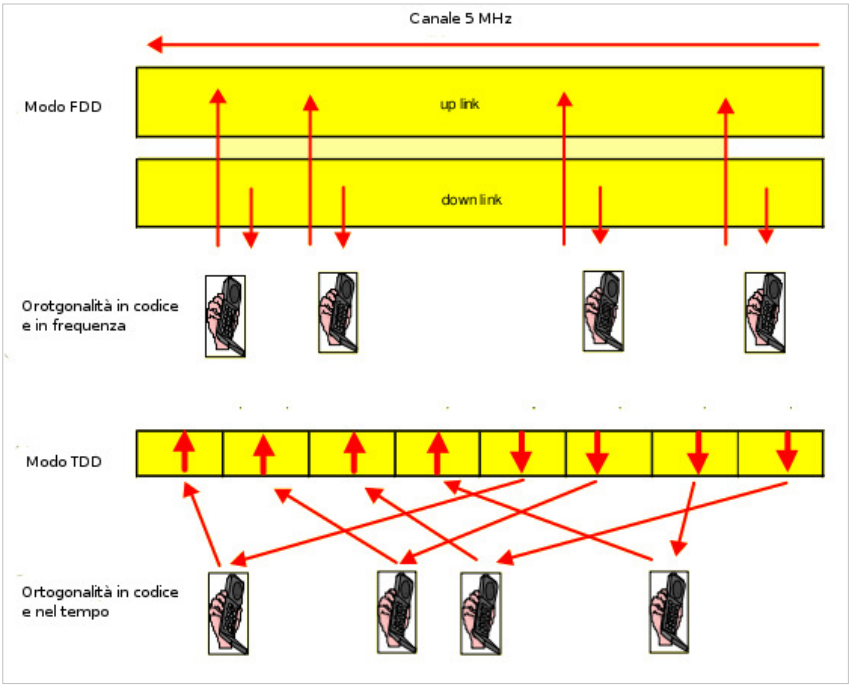

Figura 1.4. Modi FDD e TDD

La modalità FDD si basa su una tecnica CDMA a sequenza diretta (DS-CDMA). Essa comporta dei meccanismi (ad esempio dei canali piloti multiplexati nel tempo e il funzionamento asincrono) utilizzati per ottenere una capacità superiore a quella offerta dai sistemi DS-CDMA esistenti. Per questa modalità sono necessarie due portanti: una per l'uplink e l'altra per il downlink.

La modalità TDD è stata concepita ed ottimizzata per le applicazioni ad intenso traffico. In questa modalità vengono utilizzati due meccanismi differenti per mantenere l'ortogonalità tra gli utenti (vale a dire per permettere agli utenti di condividere la stessa frequenza senza avere un doppio disturbo). Si basano su un multiplexing temporale e dei codici. Questa modalità assicura una grande protezione al disturbo emesso dagli altri utenti.

Alcuni parametri radio, come la larghezza di banda (5 GHz) e la velocità di modulazione (3,48 Mbit/sec), sono comuni ad entrambe le modalità FDD e TDD; le stesse modalità di funzionamento, invece, possono differire per potenzialità dei servizi: è possibile avere delle erogazioni/potenzialità di servizio che raggiungono i

2,048 Mbit/s in modalità TDD e 384 Mbit/s in modalità FDD.

Per le sue caratteristiche, l'interfaccia radio UMTS si presenta come il dispositivo ideale per soddisfare sia le esigenza degli utenti finali che quelle degli operatori di rete. L'architettura di rete di accesso radio terrestre UMTS (UTRAN) è riportata in Figura 1.5; e tutte le interfacce terrestri si basano sulla modalità di trasferimento asincrono ATM.

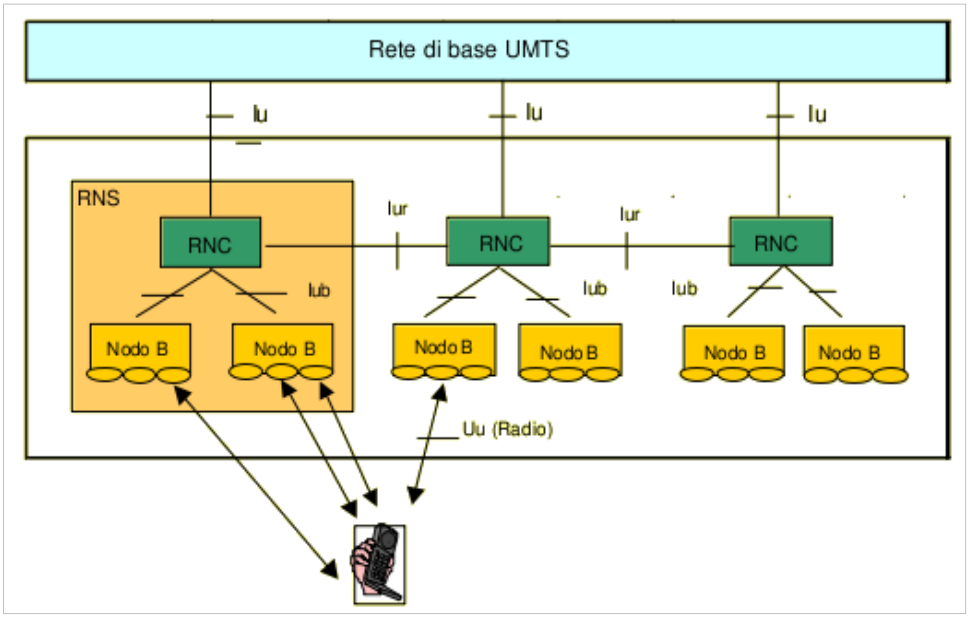

Figura 1.5. Architettura di accesso radioterrestre UMTS

Il terminale UMTS comunica con la rete d'accesso UTRAN attraverso l'interfaccia radio *Uu*. La rete di accesso è costituita da elementi quali il *nodo B*, che può essere paragonato alla stazione base di un sistema GSM, collegato ad un controllore della rete radio (RNC) attraverso l'interfaccia *Iub*, mentre il collegamento tra il sistema di accesso e la rete di base UMTS si ottiene tramite l'interfaccia *Iu*.

L'architettura della rete di base dell'UMTS è un'evoluzione delle reti GSM, in quanto si è dovuto garantire una migrazione a partire dalle reti esistenti; in particolare, si è dovuto tener conto degli investimenti fatti nel sistema GSM preservando l'architettura del sotto-sistema di rete e dare continuità all'uso di tecnologie di commutazione e di trasmissione dei pacchetti principalmente IP, per poter beneficiare di costi ridotti di gestione.

Successivamente, in aggiunta al WCDMA (3GPP release 99), è stato sviluppato, in vari Paesi, l'*High Speed Downlink Packet Access* (HSDPA). Introdotto nella release 5 della 3GPP, fornisce un canale rapido per il *downlink* ad un data rates di 3.6 Mbps. A limitare le performance di molte applicazioni rimaneva il WCDMA *uplink*, perciò, nella release 6 fu introdotto l'*High Speed Uplink Packet Access* (HSUPA) capace di raggiungere un *data rates* di 5.7 Mbps. HSDPA e HSUPA costituiscono insieme l'*High Speed Packet Access* (HSPA) [10]. Rispetto al WCDMA, oltre a fornire un maggiore *bit rates*, l'HSPA risulta essere più sensibile al ritardo, permettendo così, l'utilizzo di applicazioni *real-time*, quali il *Voice over IP* (VoIP), la video conferenza e *l'online gaming*. A seguito di HSPA è stato introdotto *HSPA+* o *Evolved* [11], in grado di fornire un significante incremento sulle performance delle reti ed erogare servizi ad una velocità di 42 Mb/s in *downlink* e 11 Mb/s in *uplink*.

1.5 La quarta generazione (4G)

Le reti wireless di quarta generazione stanno riscuotendo un enorme interesse da parte sia di ricercatori che di produttori. Con i sistemi 4G si punta al *global roaming*, cioè ad utilizzare indifferentemente ed in maniera trasparente all'utente finale, reti cellulari anziché reti satellitari o WLAN. Una grande infrastruttura di rete eterogenea che comprende sia i sistemi di accesso wireless che wireline, con una estesa area di copertura, capace di includere il *Digital Audio Broadcasting* (DAB) ed il *Digital Video Broadcasting* (DVB), che si propone di utilizzare l' IP come meccanismo di integrazione delle diverse reti, garantendo una connessione veloce ovunque ci si trovi e quando lo si vuole, sono solo alcune degli obiettivi delle reti di ultima generazione. Affinché ci sia una transazione graduale dalle reti 2G e 3G verso la nuova rete a commutazione di pacchetto, è stato creato un gruppo di lavoro, *Next Generation Mobile Networks* (NGMN), a cui compete la stesura delle norme per la costruzione della rete futura.
Un esempio di veduta delle future reti 4G è riportato in Figura 1.6.

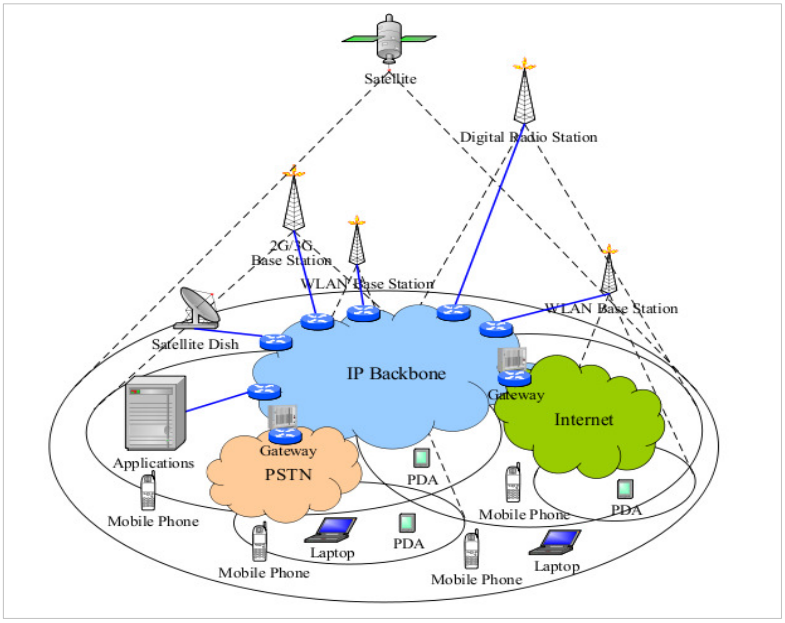

Figura 1.6. Veduta di una rete 4G

1.5.1 Long Term Evolution

Recentemente, la 3GPP ha approvato la *Long Term Evolution* (LTE) [12] definendola una interfaccia radio altamente flessibile per sistemi di quarta generazione.

La prima release della LTE prevede un picco di rates di 300 Mb/s, un ritardo radio-network inferiore a 5ms, un significante incremento in termini di efficienza di spettro rispetto ai sistemi cellulari precedenti ed una nuova architettura di rete progettata per semplificare le operazioni e ridurre i costi. LTE supporta sia la *Frequency-Division Duplex* (FDD), la *Time-Division Duplex* (TDD) e mira ad evolvere i sistemi 3GPP in uso nonché ad affermarsi come standard mondiale; non a caso include molte delle specifiche che erano state pensate per i sistemi 4G.

Il cuore della LTE per le trasmissioni radio in downlink è costituito dallo *Orthogonal Frequency-Division Multiplexing* (OFDM); questa tecnica di modulazione, infatti, utilizza frequenze radio-mobili multiple, cioè un certo numero di portanti (in genere elevato), per trasmettere dati in parallelo, con una spaziatura scelta in maniera opportuna in modo da garantire l'ortogonalità. L'incremento della

capacità, invece, è il frutto dell'utilizzo della tecnologia *Multiple Input Multiple Output* (MIMO) in grado di far ricevere ad uno stesso terminale mobile dati provenienti da più centraline.

1.6 Wireless LAN

Negli ultimi anni, quando ci si riferisce alle reti wireless si intendono principalmente le WLAN, ossia le *Wireless Local Area Network*; una reti locale, meglio conosciuta come LAN, rappresenta la forma più semplice di rete, ed è costituita da un gruppo di computer, che solitamente appartengono ad una stessa organizzazione, collegati tra loro in una piccola area geografica, che può essere una casa, un ufficio o un gruppo di edifici, ad esempio una scuola o un aeroporto.
Le reti senza fili in aree locali sono nate come alternativa infrastrutturale a queste tradizionali reti cablate, e negli ultimi tempi si stanno consolidando come tecnologia d'accesso alle reti più ampie, quale internet. Le caratteristiche principali di una WLAN sono:

- un raggio di copertura limitato, tipicamente al di sotto del chilometro;
- potenze massime non superiori a qualche centinaio di mW;
- una complessità nettamente inferiore a una rete cellulare, sia del terminale utente che dalla parte fissa di rete, che si avvale in modo essenziale dell'esistente infrastruttura di rete fissa (ad esempio una LAN tradizionale o un accesso a rete pubblica mediante un router e un rilegamento xDSL).

Tale tecnologia si affianca ad altre innovazioni wireless nel garantire il requisito della “mobilità totale”, considerato il nuovo imperativo tecnologico dalla stragrande maggioranza degli operatori di telefonia e di informatica.
La collegabilità di un sistema mobile ad internet, o comunque ad una rete aziendale, si può oggi ottenere sostanzialmente tramite due approcci: mediante l'uso di un cellulare di seconda o terza generazione, come emerge dai paragrafi precedenti, oppure tramite una rete locale wireless connessa a sua volta ad una più grande rete locale fissa e quest'ultima connessa alla rete IP in maniera tradizionale. Come si è già ampiamente descritto in precedenza, uno dei vantaggi principali offerti dalle reti cellulari è la garanzia di una connettività globale, senza limiti geografici: da

qualunque posto, purché coperto dalla rete cellulare, a qualsiasi località del globo.
Il concetto che sta alla base delle soluzioni WLAN, invece, è leggermente diverso: si realizza una rete locale basata su tecnologia radio in grado di offrire una copertura ad alta densità di traffico con estensione tipica variabile da qualche decina di metri a qualche chilometro. Tale rete può essere collegata mediante un apparato concentratore, detto *access point*, ad una tradizionale rete locale aziendale oppure, mediante una *base station*, ad un sistema di connettività geografica di un gestore di telecomunicazioni che ne garantirà il collegamento ad internet.
La tecnologia WLAN si basa sullo standard *IEEE 802.11* [13], che definisce le modalità operative ai primi due livelli del modello *ISO/OSI*[2]; tale standard viene comunemente definito Wi-Fi (*Wireless Fidelity*) e deriva dalla *Wireless Alliance*, un raggruppamento di aziende che garantiscono attraverso il marchio Wi-Fi la compatibilità degli apparati allo standard 802.11.
L'architettura dell'IEEE 802.11 è costituita da diversi componenti che interagiscono tra loro in modo da ottenere una WLAN capace di supportare la mobilità delle "stazioni", cioè l'insieme dei dispositivi che si possono connettere alla rete wireless. Tutte le stazioni devono essere equipaggiate con un'interfaccia di rete wireless, denominata *Wireless Network Interface Cards* (WNICs). La struttura base di una LAN 802.11 è il *Basic Service Set* (BSS), che rappresenta l'insieme di tutte le stazioni che possono comunicare tra loro. La Figura 1.7 mostra due BSS, ciascuna delle quali ha due stazioni che sono membri della BSS.

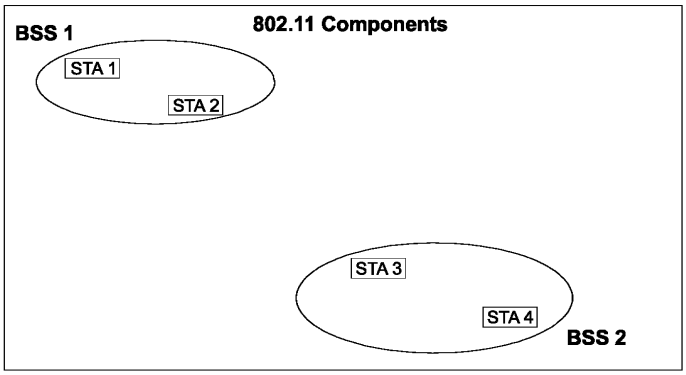

Figura 1.7. BSSs

2 *Open System Interconnection*; nasce alla fine degli anni '70 proponendosi come livello di riferimento dei sistemi aperti. I livelli in cui opera lo standard 802.11 sono il primo ed il secondo, rispettivamente Fisico e Mac.

L' "ovale" utilizzato per dipingere un BSS si può interpretare come l'area di copertura entro la quale le stazioni membri del BSS possono rimanere in comunicazione, questa zona di copertura è chiamata *Basic Service Area* (BSA). Se una stazione si muove al di fuori della sua BSA, non può più comunicare direttamente con le altre stazioni presenti nella medesima area. Esistono due tipi di BSS: *Independent BSS*, o IBSS, e *Infrastructure BSS*. L'IBSS è il tipo di WLAN più semplice, e può essere costituito anche da due soli dispositivi. Questa modalità di funzionamento può essere adottata quando le stazioni sono abilitate a comunicare in maniera diretta, quindi in assenza di strutture centrali, ed è spesso indicata come rete "ad-hoc". Un BSS ad infrastruttura, invece, prevede un organo centrale, l'access point, come unico tramite tra la rete wireless e la rete internet. Lo standard 802.11, inoltre, consente la formazione di una rete estesa, *Extended Service Set* (ESS), costituita da un gruppo di BSS, attraverso l'utilizzo di un'architettura di interconnessione chiamata *Distribution System* (DS). Con questo meccanismo, si riesce ad ottenere una rete di dimensioni arbitrarie, con la possibilità, per una stazione mobile, di passare da un BSS ad un altro, sempre all'interno della stessa ESS, mantenendo attiva la connessione.

802.11

La prima versione dello standard 802.11, sviluppato dal gruppo 11 dell'IEEE 802, venne presentata nel 1997 e viene chiamata "802.1y", specificava velocità di trasferimento comprese tra 1 e 2 Mb/s e utilizzava i raggi infrarossi o le onde radio nella frequenza di 2,4 GHz per la trasmissione del segnale. La trasmissione infrarosso venne eliminata dalle versioni successive dato lo scarso successo. La maggior parte dei costruttori infatti non aveva optato per lo standard IrDA, preferendo la trasmissione radio. Il supporto di questo standard per quanto riguarda la trasmissione via infrarossi è incluso delle evoluzioni dello standard 802.11 per ragioni di compatibilità. Poco dopo questo standard vennero realizzati da due produttori indipendenti delle evoluzioni dello standard 802.1y che una volta riunite e migliorate portarono alla definizione dello standard 802.11b.

802.11b

802.11b ha la capacità di trasmettere al massimo 11Mbit/s e utilizza il *Carrier Sense Multiple Access con Collision Avoidance* (CSMA/CA) come metodo di trasmissione delle informazioni. Una buona parte della banda disponibile viene utilizzata dal CSMA/CA. In pratica il massimo trasferimento ottenibile è di 5,9 Mbit/s in TCP e di 7,1 Mbit/s in UDP. Metallo, acqua e in generale ostacoli solidi riducono drasticamente la portata del segnale. Il protocollo utilizza le frequenze nell'intorno dei 2,4 GHz. Utilizzando antenne direzionali esterne dotate di alto guadagno si è in grado di stabilire delle connessioni punto a punto del raggio di molti chilometri. Utilizzando ricevitori con guadagno di 80 decibel si può arrivare a 8 chilometri o se le condizioni del tempo sono favorevoli anche a distanze maggiori ma sono situazioni temporanee che non consentono una copertura affidabile. Quando il segnale è troppo disturbato o debole lo standard prevede di ridurre la velocità massima a 5.5, 2 o 1 Mb/s per consentire al segnale di essere decodificato correttamente.

802.11a

Il protocollo 802.11a venne approvato nel 1999 e rettificato nel 2001. Questo standard utilizza lo spazio di frequenze nell'intorno dei 5 GHz e opera con una velocità massima di 54 Mb/s, sebbene nella realtà la velocità reale disponibile all'utente sia di circa 20 Mb/s. La velocità massima può essere ridotta a 48, 36, 24, 18, 9 o 6 se le interferenze elettromagnetiche lo impongono. Lo standard definisce 12 canali non sovrapposti, 8 dedicati alle comunicazioni interne e 4 per le comunicazioni punto a punto. Quasi ogni stato ha emanato una direttiva diversa per regolare le frequenze ma dopo la conferenza mondiale per la radiocomunicazione del 2003 l'autorità federale americana ha deciso di rendere libere secondo i criteri già visti le frequenze utilizzate dallo standard 802.11a.Questo standard non ha riscosso i favori del pubblico dato che l'802.11b si era già molto diffuso e in molti paesi l'uso delle frequenze a 5 GHz è tuttora riservato. In Europa lo standard 802.11a non è stato autorizzato all'utilizzo dato che quelle frequenze erano riservate

all'HIPERLAN[3]; solo a metà del 2002 tali frequenze vennero liberalizzate.

802.11g

Questo protocollo fu introdotto per colmare l'incompatibilità tra lo standard 802.11a e quello 802.11b, dato che essi operavano su frequenza differenti. Questo standard venne ratificato nel giugno del 2003. Utilizza le stesse frequenze dello standard 802.11b cioè la banda di 2,4 GHz e fornisce una banda teorica di 54 Mb/s che nella realtà si traduce in una banda netta di 24,7 Mb/s, simile a quella dello standard 802.11a. È totalmente compatibile con lo standard b ma quando si trova ad operare con periferiche *b* deve ovviamente ridurre la sua velocità a quella dello standard omonimo. Prima della ratifica ufficiale dello standard 802.11g avvenuta nell'estate del 2003 vi erano dei produttori indipendenti che fornivano delle apparecchiature basate su specifiche non definitive dello standard. I principali produttori comunque preferirono aderire alle specifiche ufficiali e quando queste vennero pubblicate molti dei loro prodotti furono adeguati al nuovo standard. Alcuni produttori introdussero delle ulteriori varianti chiamate *g+* o *Super G* nei loro prodotti. Queste varianti utilizzavano l'accoppiata di due canali per raddoppiare la banda disponibile anche se questo induceva interferenze con le altre reti e non era supportato da tutte le schede.

Gli standard 802.11b e 802.11g dividono lo spettro in 14 sottocanali da 22 MHz l'uno. I canali sono parzialmente sovrapposti tra loro in frequenza, quindi tra due canali consecutivi esiste una forte interferenza. I 2 gruppi di canali 1, 6, 11 e 2, 7 e 12 non si sovrappongono fra loro e vengono utilizzati negli ambienti con altre reti wireless.

802.11n

In questi ultimi mesi è stato approvato in via definitiva dalla IEEE lo standard 802.11n per la realizzazione di reti wireless metropolitane, e la sua pubblicazione è prevista per l'inizio del 2010. Progettato per essere comunque compatibile con gli standard precedente, l'802.11n permette di realizzare collegamenti con una velocità

[3] HIPERLAN (HIgh PErformance RAdio LAN) è il nome di uno standard WLAN. Descrive una serie di soluzioni europee alternative agli standard statunitensi IEEE 802.11.

nettamente superiore all'802.11g. In termini di frequenza, i dispositivi n possono lavorare sia sui 2.4 GHz dello standard b che sui 5GHz dello standard a, mentre la tecnica di modulazione adottata per questo protocollo è la OFDM. Una delle soluzioni trovate per ovviare ai problemi di portata e di velocità, è data dalla tecnologia *Multiple Input Multiple Output* (MIMO): questa tecnologia si basa sul principio della riflessione multipla (chiamata anche *multipath*) dei segnali sugli ostacoli fisici, in modo da assicurare la trasmissione contemporanea di più flussi separati nella stessa banda di frequenza. Per migliorare ulteriormente le prestazione, si aggiunge un raddoppio della larghezza di banda di canale a livello fisico da 20 a 40 MHz e una velocità di trasmissione massima maggiore rispetto allo standard 802.11g, prossima a 300 Mb/s (che nell'uso reale si riduce a 100 Mb/s), anche se nelle specifiche IEEE sono previste versioni di 802.11n capaci di raggiungere i 600 Mb/s teorici. Sul versante delle modalità di funzionamento, l'802.11n prevede diversi sistemi. Il primo è quello chiamato *legacy* che praticamente opera come i protocolli antecedenti; ciò significa che i dispositivi devono operare nella stessa gamma di frequenza degli altri standard e supportare i canali con larghezza di banda di 20 MHz. Il secondo sistema è di tipo *mixed* e permette le comunicazioni tra le schede precedenti e quelle 802.11n ed, infine, la terza modalità è quella chiamata *nativa*. Le prime due modalità, comunque, offrono dei vantaggi rispetto all'802.11g.

Uno dei problemi, invece, è quello delle interferenze con i dispositivi 802.11 b e g, a cui si è trovato rimedio spostando le frequenze operative e le modalità di funzionamento grazie a particolari algoritmi. A questo va aggiunta la tecnologia di aggregazione dei frame che è un sistema che permette di migliorare aspetti come efficienza e velocità in presenza di pacchetti di piccole dimensioni, consentendo di ridurre la latenza quando un device cerca di comunicare con l'access point. L'efficienza, inoltre, è stata migliorata anche riducendo l'impiego di larghezza banda per i dati di controllo.

Questa “famiglia” di standard appena descritta, si può considerare la base dell' 802.11, e riferisce alla trasmissione dell'informazione; altri standard, come l'802.11i, riguardano la sicurezza, mentre altri ancora sono estensioni dei servizi base e miglioramenti di servizi già disponibili.

Capitolo 2

Le reti wireless: la necessità di organizzare

2.1 Introduzione

In questo capitolo verrà esplorato ciò che si nasconde dietro ad una comunicazione di tipo wireless, si individueranno le componenti principali di una rete, come sono organizzate, quali sono le risorse da gestire, il modo in cui esse vengono gestite e quali sono i parametri che determinano la qualità in una comunicazione.
Quando ci si riferisce ad una rete wireless, si intende il modo d'accesso alla rete, che è appunto senza fili. Il mezzo trasmissivo utilizzato è l'etere, ed il segnale viene propagato sotto forma di onde radio. La propagazione via radio viene oramai utilizzata da più di cento anni e, come si appreso dal capitolo precedente, si è passati dalle semplici e dispendiose comunicazioni analogiche alle più complesse, sicure, ed economiche comunicazioni digitali. Migliorare, anche in termini di utilizzo delle frequenze di trasmissione, significa permettere sempre a più utenti di accedere alla rete, quindi di comunicare; e dato l'enorme numero di utilizzatori delle reti cellulari, anche le reti stesse, intese come strutture fisiche, hanno bisogno di una gestione ottimale nella loro interezza. In Figura 2.1 è riportata una stima[4] degli utenti radiomobili nel mondo. Dalla figura si può notare come nel corso degli anni i sottoscriventi alle reti cellulari hanno superato gli utenti di telefonia fissa; si ha quindi una necessità di organizzare al meglio le reti, evitando strutture centralizzate a favore di un approccio gerarchico e a vari livelli di "granularità". L'obiettivo è di ottenere una struttura flessibile ed adattabile ad eventuali evoluzioni territoriali, urbanistiche e sopratutto tecnologiche.

4 Dati da "www.etforecasts.com"

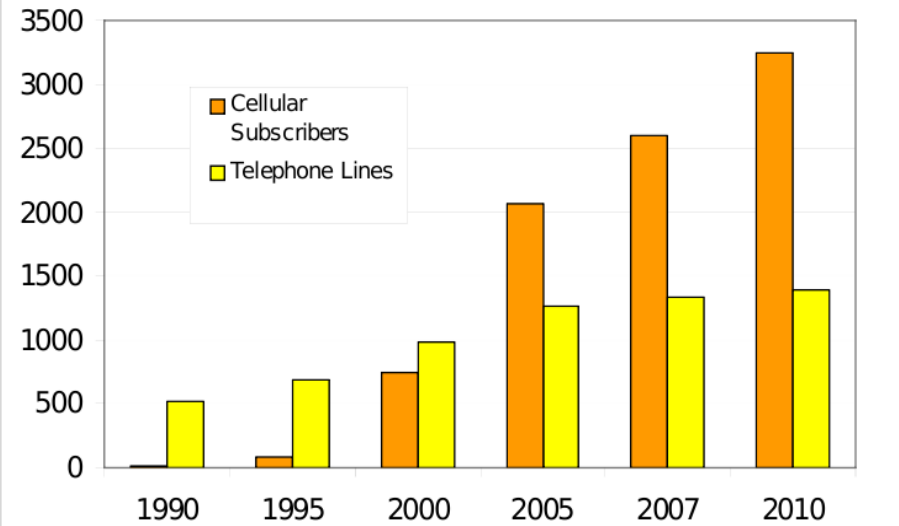

Figura 2.1. Utenti radiomobili nel mondo.

2.2 Struttura di una rete wireless cellulare

Genericamente, una rete di comunicazione senza fili è composta da una parte fissa e da un sistema di accesso wireless. La copertura geografica è ottenuta con una tassellatura di aree adiacenti e/o sovrapposte dette *celle*. I componenti principali di una rete sono tre [14]:

- Terminale mobile (MT);
- Stazione base (BS);
- Mobile switching centre (MSC).

Il *terminale mobile* rappresenta l'equipaggiamento fisico di comunicazione di un utente mobile, che può essere un semplice telefono, uno *smartphone*, un PC o un qualunque altro dispositivo capace di connettersi ad una rete wireless. Le *stazioni radio base*, invece, forniscono l'accesso radio ai terminali mobili in una data zona di copertura o di servizio, la cella. La MSC, a cui è collegato un gruppo di BS tra loro adiacenti, può essere visto come il tramite tra il sistema di accesso wireless e la rete fissa. Nella Figura 2.2 è riportato un ipotetico schema di organizzazione di una rete wireless.

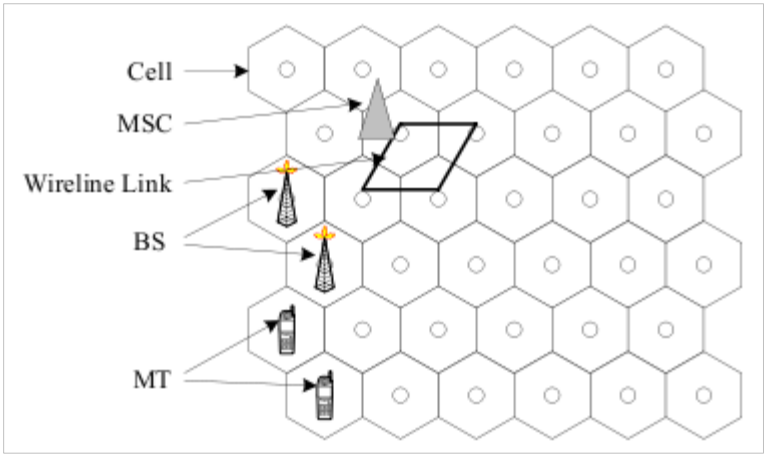

Figura 2.2. Organizzazione di una rete wireless.

Le principali funzioni di un sistema radiomobile, in linea di massima, possono essere descritte nei seguenti punti:

- Registrazione di un'utenza;
- Instaurare ed eliminare le chiamate;
- Assegnazione dell'utente mobile alla stazione base più vicina;
- Commutazione della chiamata tra due stazioni diverse in caso di movimento;
- Controllo delle potenze irradiate;
- Sicurezza e/o segretezza;
- Altre funzioni (es. tariffazione);

Inoltre, i sistemi radiomobili devono supportare una struttura di controllo efficente in grado di determinare la posizione del terminale mobile all'interno della rete.

Copertura radioelettrica del territorio

Come già accennato, la copertura geografica è ottenuta grazie all'utilizzo di celle, e da qui il nome di *rete cellulare*. In ogni cella è presente una stazione base a cui è assegnata una certa frequenza dalla quale, grazie alle tecniche di multiplazione, è possibile ottenere diversi canali radio. Dato, però, il gran numero di celle che occorrono per coprire un intero territorio nazionale, anche se si assegnasse un solo canale per ogni cella, non si riuscirebbe a coprire tutte le celle; la strategia che viene adottata per fronteggiare tale problema è il riuso delle frequenze, che consiste nell'assegnare la stessa frequenza a celle ben distanti tra loro in modo che non ci sia

interferenza radioelettrica. Le stesse celle sono solitamente organizzate in gruppi detti *cluster*; all'interno di un cluster ogni cella utilizza un sottoinsieme unico di canali radio. I sistemi analogici di prima generazione utilizzavano cluster formati da 19 o 21 celle, contro le 7 o 9 dei sistemi numerici con accesso di tipo TDMA. La grandezza del cluster, inteso come numero di celle, è una misura dell'efficienza delle reti cellulari: più sono grandi i cluster e meno efficiente è il sistema.

Da questa breve descrizione si intravede il ruolo vitale delle celle nelle radio-comunicazioni. Nello studio di algoritmi di allocazione di banda la struttura delle celle è caratterizzata da una forma regolare, che solitamente è circolare o esagonale; nella realtà la forma delle celle è ben diversa e può dipendere da diversi fattori, quale la potenza degli apparati, i ritardi di propagazione, la densità di popolazione e la struttura territoriale (in quanto si deve tener conto di eventuali ostacoli radioelettrici). Le diverse forme che una cella può assumere ha un impatto diretto sulla loro dimensione. L'utilizzo di celle piccole, o *micro-celle*, permette di ottenere una maggiore capacità di trasmissione e quindi la possibilità di allocare più utenti mobili nella cella, d'altro canto si ha un aumento del rate di handoff e cambiamenti repentini delle condizioni della rete, rendendo più difficile la gestione dello stesso handoff [15]. In alcuni casi, per favorire la gestione della mobilità degli utenti, si utilizzano delle celle organizzate in maniera gerarchica, comunemente chiamate celle ad “ombrello”. In Figura 2.3 viene riportato un esempio di copertura cellulare con celle di dimensione diversa per aree a diversa intensità di traffico.

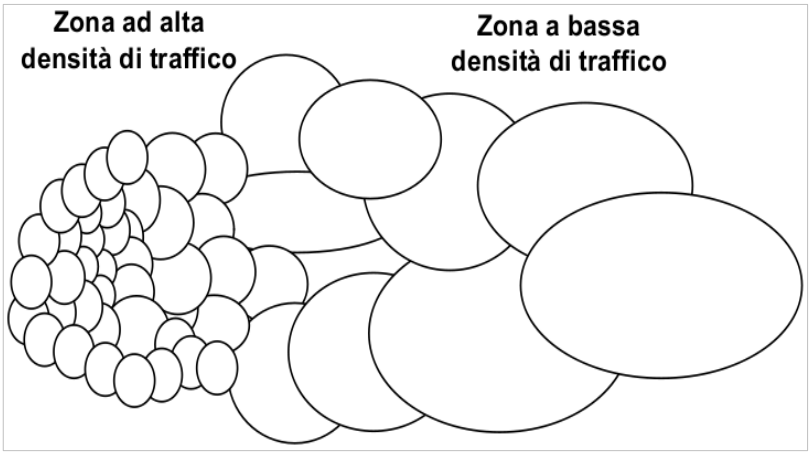

Figura 2.3. Esempio di copertura cellulare.

Tipologia di chiamate

All'interno dell'area di copertura di ogni stazione base, si possono presentare due tipologie di traffico: *nuove chiamate* e chiamate di *handoff* o *handover*.

Per nuove chiamate si intendono le connessioni attivate all'interno della cella corrente, mentre le chiamate di handoff riferiscono a connessioni iniziate in un'altra cella ed "arrivate" nella cella attuale. Si ricorda che gli utenti di una rete wireless sono considerati mobili e quando un MT si sposta da una cella all'altra, è responsabilità della stazione base e della MSC assicurare la continuità del servizio nel modo più trasparente possibile [16]. Di fatto, l'handover è la procedura che consente il trasferimento di una chiamata da una cella alla successiva, mentre il terminale mobile si sposta all'interno della rete. Si può affermare che l'handoff è l'elemento distintivo tra le reti cellulari ed ogni altro tipo di reti di telecomunicazioni. In Figura 2.4 viene mostrato un esempio di handoff.

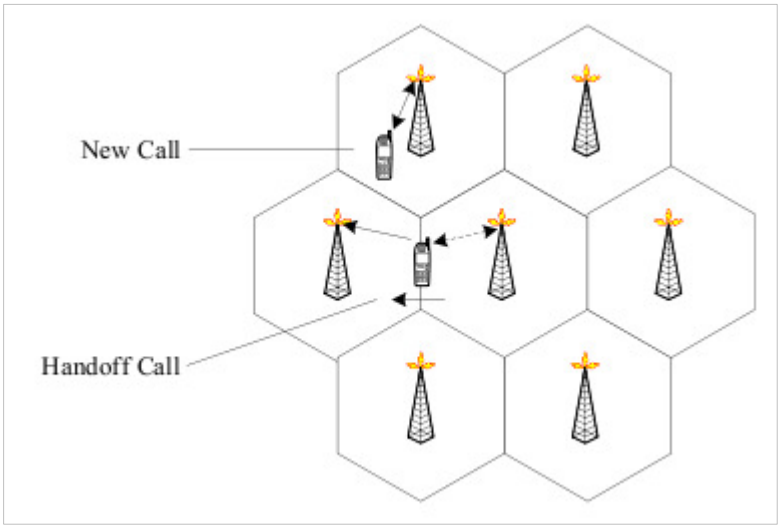

Figura 2.4. Esempio di handoff.

La procedura di handoff è sempre iniziata dalla rete e si basa sulla misura di alcuni parametri, quali la potenza del segnale ricevuto, la qualità, etc.; tali misure vengono effettuate sia dal lato rete che dal lato utente. Esistono diversi metodi per generare un evento di handoff, ad esempio, quello del segnale più forte: considerando due celle contigue *A* e *B*, e supponendo un nodo in movimento dalla cella A a quella B, quando il segnale ricevuto dalla cella B è più forte di quello corrente, allora viene effettuato l'handoff. Ovviamente, assieme alla reti, anche la procedura di handoff ha

subito della evoluzioni; infatti, nei sistemi GSM, come in tutti gli altri sistemi 2G, l'handoff, tecnicamente "Hard-Handover", veniva fatto in maniera drastica, in quanto un nodo sotto la copertura di una BS non aveva la possibilità di comunicare con un'altra BS, perciò, al verificarsi dell'handoff, il link radio corrente veniva semplicemente abbattuto e ne veniva instaurato un altro nella cella di destinazione, senza la possibilità di avere informazioni sulla sua situazione di traffico.
Nelle reti di terza generazione, invece, l'handoff viene gestito in un modo più flessibile, "Soft-Handover", dove il MS non è "agganciato" ad una singola cella, ma ad un set di celle attive (per un massimo di tre), chiamato *Active Set*. In situazioni di co-esistenza dei due sistemi (2G e 3G), attorno ad una cella UMTS devono esserci delle celle 2G, affinché, in caso di perdita del segnale UMTS o di cattive condizioni radio, si può effettuare un backup della chiamata (voce o GPRS) sulla rete GSM, evitando così di eliminare la chiamate.

2.3 Handoff tra reti eterogenee

Nella attuale evoluzione delle reti si vuole assicurare l'interoperabilità tra reti eterogenee, in previsione di un futuro popolato da molteplici standard d'accesso, come l'802.11, il Bluetooth, il WiMax, l'UMTS, etc.. Un ruolo cruciale per la co-esistenza funzionale di reti diverse, lo svolge l'handover che, come si è già accennato nei paragrafi precedenti, è l'evento che si verifica quando un terminale mobile si muove da una cella ad un'altra. L'handover può essere classificato come:

- *orizzontale* (intra-system): indica l'handover entro la stessa tecnologia wireless, offrendo una mobilità localizzata;
- *verticale* (inter-system): indica l'handover tra tecnologie wireless diverse, fornendo una mobilità globale.

Un hadover verticale è ulteriormente classificabile in handoff:

- *verso l'alto*: un handoff verso una tecnologia con la dimensione delle celle superiore (ad esempio da Wi-Fi a UMTS);
- *verso il basso*: un handoff verso una tecnologia con la dimensione delle celle inferiore(ad esempio da UMTS a Wi-Fi).

Un importante aspetto da considerare nella decisione di eseguire un handoff è ottimizzare le performance di esecuzione dello stesso. A differenza degli algoritmi di gestione degli handoff orizzontali, prendere decisioni relative agli handoff verticali è molto più complesso, dato che esso comporta la conoscenza di diversi parametri relativi a due tipologie di reti diverse, evitando, inoltre un effetto poco gradito conosciuto come effetto *ping-pong*, in cui un terminale mobile effettua continuamente l'handoff tra due stazioni base. Ovviamente, per poter effettuare un handover verticale, ad esempio tra una WLAN e una rete UMTS, il terminale mobile deve essere provvisto di doppia interfaccia radio. Lo standard IEEE per la gestione dell'handoff verticale è lo *802.21*, ed è stato sviluppato per facilitare l'interazione e l'handover tra le tecnologie 802 ed altre tecnologie wireless. Questo standard consiste in un'architettura che consente continuità di servizio, in modo trasparente, quando un terminale mobile passa da una tecnologia ad un'altra. L'802.21 definisce una struttura per poter effettuare vertical handover tra reti eterogenee ed include specifici protocolli per:

- Cellulari (GSM/GPRS/UMTS);
- Wi-Fi (802.11a/b/g);
- WiMAX (802.16e);
- Bluetooth (802.15.1).

Vi sono, all'interno dello *stack* protocollare sulla gestione della mobilità, una serie di funzioni che permettono l'handover e vanno a formare una nuova entità, che si va a collocare tra il livello due ed il livello tre: il *Media Independent Handover Function* (MIHF). In Figura 2.5 è riportata la collocazione del MIHF all'interno dello stack protocollare per la famiglia IEEE 802; L'MIHF utilizza le informazioni presenti a livello del protocollo LLC (*Link Layer Control*) e scambia informazioni con i livelli inferiori, MAC e Fisico dell'IEEE 802, per determinare se eseguire un handoff verticale.

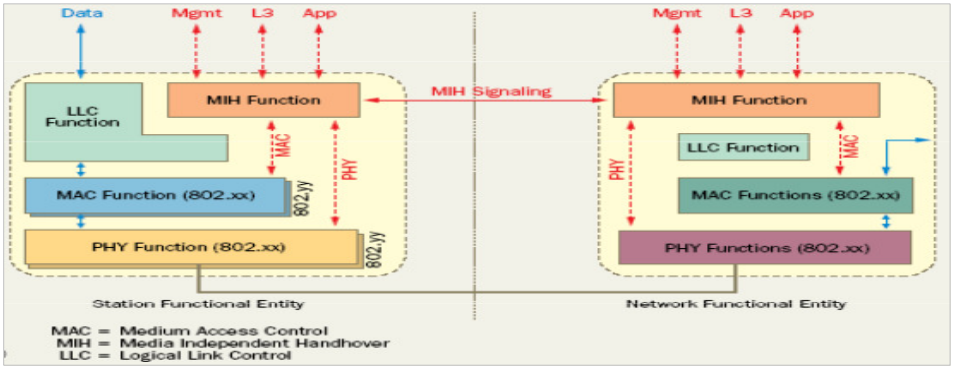

Figura 2.5. MIHF per la famiglia IEEE 802.

All'interno dello stack, l'MIH fornisce servizi agli strati superiori attraverso una singola interfaccia, "technology-independent", (*MIH-Service Access Point*), ed ottiene servizi dagli strati inferiori attraverso una molteplicità di interfacce, "technology-dependent", (*media-specific SAP*); l'obiettivo è quello di fornire la generica interfaccia tra il link layer e gli attuali media-specific link layer, come quelli specificati dal 3GPP e nella famiglia di standard IEEE 802, quindi:

1. MIH_SAP permette l'accesso al MIH dagli strati superiori;
2. MIH_NMS_SAP per la gestione della rete;
3. MIH_LINK_SAP, ogni rete di accesso s'interfaccia direttamente con il MIHF mediante il proprio SAP.

I servizi offerti dall'MIH sono: *event*, *command* e *information*; a questi tre si aggiungono il *Network Discovery*, che permette al terminale mobile di determinare tutte le reti vicine, cioè a distanza di rilevamento, a cui può accedere ed il *Network Selection*, che insieme alla network discovery fornisce informazioni sulle reti adiacenti e permette al terminale mobile di selezionare un link. Il *MI Event Service* (MIES) classifica, filtra e riporta gli eventi corrispondenti a cambiamenti dinamici delle caratteristiche di un link. Il flusso degli eventi passa attraverso il MIHF, che ne registra le notifiche e lo invia ai livelli superiori. Il servizio command è gestito, invece, dal *MI Command Service* (MICS), che permette di controllare il comportamento di un link per l'handover e la mobilità. Il *MI Information Service* (MIIS) fornisce le informazioni per gli handovers, come le mappe delle reti vicine, il link layer information, e la disponibilità dei servizi. Questo servizio condivide *elementi di informazione* (IEs), che vengono utilizzate per prendere decisioni di

handover. Tali IE sono definiti come messaggi *Type-Length-Value* (TLV) e possono essere di carattere generale (*general information*) che riguardano gli operatori, di accesso alla rete (*access network*) e riguardano il roaming, i costi, la sicurezza e la QoS, ed infine gli elementi di informazione dei punti d'accesso radio; ad esempio un elemento di informazione può essere una lista di tutte le reti disponibili in base alla posizione dell'utente, il nome di un provider di rete, la lista di altri operatori diretti di roaming, i parametri di QoS di un link.
La rete complessiva, in cui andrà ad operare lo standard, dovrà includere micro-celle (IEEE 802.11, 802.15, etc.), celle (GSM, UMTS, IEEE 802.16, etc.) ed anche macro-celle satellitari. La funzionalità MIH risiede sia nella stazione mobile che nel punto d'accesso radio (stazione base, access point, etc.), che devono pertanto essere multimodali.

2.4 Sicurezza nelle reti IEEE 802.11

Le onde radioelettriche hanno intrinsecamente una grande capacità di diffondersi in tutte le direzioni con una portata relativamente grande. È quindi molto difficile arrivare a confinare le emissioni radio in un perimetro limitato. La conseguenza principale di questa “propagazione selvaggia” delle onde radio è la facilità con cui si riesce ad “ascoltare” la rete, esponendo, la rete stesse ad innumerevoli rischi, tra cui:

- *intercettazione dei dati*, che consiste nell'ascoltare le trasmissioni dei diversi utenti nella rete senza fili;
- *furto di connessione*, con lo scopo di ottenere l'accesso alla rete locale o ad internet;
- *disturbo delle trasmissioni*, che consiste nell'emettere segnali radio in grado di produrre delle interferenze;
- *blocco di servizio*, che rende la rete inutilizzabile tramite l'invio di segnali fittizi.

Questi sono solo alcuni dei problemi che si possono riscontrare in una rete wireless, e sono sufficienti a capire la necessità di adottare dei meccanismi in grado di rendere una rete sicura ai vari tipi di “attacchi” a cui può essere soggetta. Nelle reti WLAN, i

metodi adottati per rendere sicure le trasmissioni sono basati sulla crittografia dei dati, in particolare [17]:

- Wep (*Wired Equivalent Privacy*);
- WPA (*Wi-Fi Protected Access*);
- WPA2/802.11i (*Wi-Fi Protection Access, Version 2*).

Il WEP è stato progettato per fornire una sicurezza comparabile con quella delle LAN, o almeno in principio era il risultato che si sperava di ottenere, in quanto nel corso degli anni sono emerse un numero indefinito di falle, tanto da rendere il WEP sconsigliabile. Il WEP ricorre ad una chiave segreta condivisa da tutte le stazioni mobili partecipanti alla rete e dall'access point. La chiave segreta è utilizzata per cifrare i pacchetti prima della loro trasmissione nell'etere e può essere composta da 64, 128 o 256 bit. Per codificare i pacchetti, il WEP si avvale dell'algoritmo di cifratura RC4, un noto cifratore a flusso molto veloce ed efficiente. Essenzialmente, un cifratore a flusso è un generatore di numeri pseudo-casuali (PRNG), che viene inizializzato da una specifica chiave segreta: data una chiave, il PRNG genera una sequenza virtualmente infinita di bit, chiamata *Keystream*, che viene successivamente messa in *XOR*[5] *bit-a-bit* con il testo in chiaro. Il risultato dello XOR è il testo cifrato. Il ricevente del messaggio, dato che la chiave segreta è condivisa, può generare lo stesso Keystream e decodificare il messaggio. Senza troppo argomentare i motivi "logici" che hanno portato al fallimento del WEP, in questa sede ci si limita a riportare alcune delle sue debolezze:

- Il WEP non previene la falsificazione dei pacchetti;
- Non previene gli attacchi cosiddetti "replay"; infatti i pacchetti possono essere registrati e replicati, e la rete li considererà ugualmente legittimi;
- Il WEP utilizza l'algoritmo *RC4* in maniera impropria: la chiave utilizzata è molto "debole" e può subire attacchi di tipo *brute-force* in poche ore se non minuti.
- Carenza nella gestione delle chiavi.

A fronte dell'inaffidabilità del WEP, la Wi-Fi Alliance ha elaborato una nuova

5 L'operatore **XOR** (detto anche **EX-OR**, **OR esclusivo** o **somma modulo 2**) nella sua versione a due elementi restituisce 1 (*vero*) se e solo se *un unico* dei due operandi è 1, mentre restituisce 0 (*falso*) in tutti gli altri casi.

specifica per la sicurezza delle reti Wi-Fi al fine di garantire un maggior livello di protezione. Il protocollo in questione è il *Wireless Protected Access* (WPA) e rappresenta solo alcune delle funzioni presenti nello standard 802.11i, che è lo standard relativo appunto alla sicurezza nelle reti wireless. WPA può essere utilizzato in due modalità: *Personal* ed *Enterprise*; la prima è adatta per piccoli uffici o in generale per reti di piccole dimensioni, mentre la seconda permette di gestire reti più estese. La successiva versione del WPA è il WPA2 che implementa interamente le funzioni definite nello standard 802.11i e prevede anch'esso i due modi di funzionamento del protocollo precedente. La differenza sostanziale con il WPA è che questo fornisce un migliore meccanismo di criptaggio. Rispetto al WEP, il WPA introduce una migliore crittografia dei dati utilizzando l'RC4 con una chiave a 128 bit ed un vettore di inizializzazione a 48 bit, assieme ad *Temporary Key Integrity Protocol* (TKIP) che permette di cambiare, dopo un certo numero di messaggi scambiati, le chiavi di crittografia utilizzate. Per l'autenticazione dei messaggi, è stato introdotto un codice più sicuro rispetto a quello utilizzato nel WEP, il *Message Integrity Code* (MIC, la cui implementazione in WPA prende il nome di *Michael*), che include anche un contatore di frame con l'obiettivo di prevenire i replay attack. Infine è stato aggiunto un meccanismo di mutua autenticazione, funzione quest'ultima assente nel WEP. Un elemento critico della sicurezza di una rete Wi-Fi è l'accesso degli utenti alla rete stessa, che deve essere negato agli utenti non autorizzati. Grazie all'autenticazione è possibile conoscere l'identità di un utente o di una macchina che tenta di accedere alla rete, e non appena questa viene verificata, si può prendere la decisione di permettere o meno l'accesso. È chiaro quindi che senza un attento controllo delle identità, dei possibili attaccanti potrebbero avere accesso alle risorse di rete protette. Inoltre, è necessario che anche la rete si autentichi all'utente per evitare che un utente wireless entri accidentalmente in una rete non sicura nella quale le sue credenziali potrebbero essere rubate. L'autenticazione in WPA è basata sul protocollo *Extensible Autentication Protocol* (EAP), un protocollo generico di autenticazione tra client e server che supporta numerosi metodi con password, certificati digitali o altri tipi di credenziali. Come già detto, WPA supporta due metodi di accesso: Personal ed Enterprise. Il modo personal, chiamato anche *Preshared Key Mode*, è appropriato

per reti di piccole dimensioni, ed è basata sull'immissione manuale di una password nell'AP e condivisa da tutti i client. Tale password può essere lunga da 8 a 63 caratteri, in generale 256 bit. In caso di reti estese, invece, è conveniente utilizzare il WPA Enterprise, che usa un server *Remote Authentication Dial-In User Service* (RADIUS) per l'autenticazione ed un protocollo EAP per la gestione dello scambio di dati di autenticazione tra client e server. Quando si utilizza un server per l'autenticazione, il punto di accesso senza fili impedisce l'inoltro del traffico dei dati a una rete cablata o ad un client senza fili che non dispone di una chiave valida. Per ottenerne una, il client deve seguire questo procedimento:

- Quando un client senza fili entra nel raggio di azione di un punto d'accesso senza fili, questo ultimo richiede la verifica del client;
- Il client wireless si identifica e dall'AP le informazioni vengono inviate al server RADIUS;
- Il server verifica le credenziali del client: se queste sono valide, il server invia una chiave di autenticazione crittografata all'AP;
- L'AP utilizza la chiave per trasmettere, in modo protetto, le chiavi di crittografia.

Anche il protocollo WPA, però, presenta delle debolezze, soprattutto in modalità Personal, restando comunque molto più affidabile dei sistemi precedenti.

2.5 Gestione delle risorse in una rete wireless

Lo scopo della gestione delle risorse in una rete wireless è quello di fornire garanzie sulla qualità dei servizi al traffico multimediale in accordo con i loro requisiti di banda, mantenendo il più alto possibile l'utilizzo delle risorse della rete stessa. La gestione delle risorse nelle reti wireless può essere implementata in due livelli [18]:

Macro-livello, che include il *Call Admission Control* (CAC), l'allocazione delle risorse, il riservo di banda, etc.

Micro-livello, che riguarda il controllo di potenza, media access control (MAC), la schedulazione dei pacchetti, etc.

Questo lavoro di tesi è orientato alla gestione delle risorse nelle reti multimediali a livello macro, quindi la *capacità* di una rete è interpretata in termini di larghezza di banda. Sotto queste considerazioni, l'unica risorsa da considerare nelle reti wireless multimediali, è appunto la banda.
Per evitare la complessità di un coordinamento centralizzato, la gestione delle risorse viene effettuata in maniera distribuita su ogni singola cella della rete, quindi in ogni stazione base.

Call Admission Control

Il controllo d'ammissione delle chiamate, o più semplicemente CAC, è uno dei componenti più importanti nella gestione della banda nelle reti wireless. L'obiettivo del CAC è fornire garanzie di *QoS* (Quality of Service) alle chiamate che richiedono l'accesso alla rete mantenendo alta l'utilizzazione della rete medesima [19] [20]. L''accesso o il rifiuto di una chiamata nella rete è quindi determinato dal CAC, che opera le sue scelte in base a criteri pre-stabiliti ed in accordo al carico di traffico della cella; inoltre, dato che le reti wireless sono caratterizzate dalla mobilità degli utenti, l'algoritmo di CAC è invocato non solo quando viene generata una nuova chiamata, ma anche quando una chiamata già attiva si sposta da una cella all'altra. In linea di principio, il funzionamento di un algoritmo di controllo di ammissione è molto banale: all'arrivo di una nuova chiamata o di una chiamata di handoff, l'algoritmo controlla la banda a disposizione nella cella; se la banda disponibile può soddisfare la richiesta di banda della chiamata, la chiamata stessa viene accettata, altrimenti viene reiettata. Implementare un buon algoritmo di CAC è molto importante nelle reti wireless in quanto questo ha un impatto diretto sulla QoS sia della nuove chiamate che delle chiamate di handoff.

Allocazione di banda

Nelle reti wireless la banda è una risorsa tanto scarsa quanto preziosa perciò deve essere utilizzata nel modo più efficiente possibile. Il ruolo dell'allocazione di banda è di decidere come la banda viene distribuita tra tutte le chiamate in corso nella rete in ordine di soddisfazione dei diversi requisiti di banda [21]. L'allocazione di banda può essere divisa in due categorie: *allocazione dinamica* o *adattativa* e *allocazione*

statica o *non-adattativa*.

Con l'allocazione di banda non-adattativa, alle chiamate che vengono ammesse nella rete viene allocata una certa quantità di banda, e questa rimane tale per tutto il ciclo di vita della chiamata stessa. In caso di saturazione della rete, quando una nuova chiamata o di una chiamata di handoff richiede un certo ammontare di banda, questa viene semplicemente reiettata. Questo metodo di gestione della banda è molto restrittivo e non è propriamente adatto alle reti wireless che presentano traffico multimediale e diversi requisiti di QoS.

L'allocazione di banda adattativa, invece, consente una gestione della banda più flessibile in quanto si può dinamicamente variare la banda allocata alle chiamate in corso in accordo alla situazione di traffico della rete. Ad esempio, quando arriva una nuova richiesta di connessione o una richiesta di handoff in una cella sovraccarica o in congestione, la banda allocata alle chiamate già attive all'interno della rete può essere degradata a livelli inferiori per poter accomodare la nuova richiesta. Quando una chiamata termina o si sposta in un'altra cella, la banda liberata può essere utilizzata per aumentare la banda alle chiamate che avevano subito un processo di degrado. Questo permette di aumentare l'utilizzo totale di banda della rete[22].

Bandwidth reservation

Come già detto in precedenza, a differenza delle reti *wireline*, le reti wireless sono caratterizzate dalla mobilità degli utenti.

Dopo che una chiamata si è spostata (handed-off) dalla cella di origine alla cella di destinazione, al fine di garantire il servizio di continuità, la cella di destinazione deve allocare, alla chiamata entrante, la stessa quantità di banda della cella precedente; in caso contrario, la chiamata di handoff viene reiettata.

Generalmente, in una rete wireless, le nuove chiamate e le chiamate di handoff sono in competizione per l'uso della banda, che è una risorsa limitata; il blocco delle nuove chiamate ed il dropping di quelle di handoff non possono essere ridotte simultaneamente, ma c'è bisogno di un compromesso. Dal punto di vista dell'utente finale, è ampiamente accettato che è molto meglio bloccare una nuova chiamata piuttosto che eliminare una chiamata di handoff [23]. Al fine di offrire una sorta di protezione alle chiamate di handoff, la rete può trattarle con una priorità maggiore

rispetto alle nuove chiamate riservando della banda per il loro uso esclusivo[24].
Le procedure per riservare la banda possono essere di due tipi [25]: statica o dinamica. Con un approccio statico si riserva alle chiamate di handoff una percentuale fissa di banda in ogni cella. Da un punto di vista implementativo, riservare la banda in modo statico è molto conveniente in quanto non necessita di alcuna interazione tra le celle. L'altro metodo, quello dinamico, prevede una quantità variabile di banda per gli handoff, in accordo con la situazione del traffico della cella. Se confrontato con il riservo statico di banda, questo ottiene performance migliori, riducendo le probabilità di dropping delle chiamate; il prezzo da pagare è una più elevata complessità nell'implementazione, senza tralasciare il numero di messaggi di inter-scambio tra le celle che può toccare picchi altissimi dovuto alla difficile predizione del traffico di handoff nelle reti wireless.

2.6 Quality of Service

Nelle reti cellulari, la *Quality of Service* (QoS) è definita come la capacità di fornire servizi soddisfacenti, che prevedono la qualità della voce, la forza del segnale, la bassa probabilità di blocco delle chiamate, l'alto data-rate per le applicazioni multimediali, etc. I fattori da cui dipende la QoS nelle reti wireless possono essere raggruppati nei seguenti punti:

- *Throughput*, che rappresenta il rate con cui i pacchetti viaggiano nella rete;
- *Delay*, cioè il tempo che un pacchetto impiega per andare da un dato punto ad un altro;
- *Packet Loss Rate*, il rate con cui vengono persi i pacchetti;
- *Packet Error Rate*, gli errori contenuti in un pacchetto dovuto alla presenza di bit compromessi.

Uno schema che offre garanzie di QoS tipicamente riflette una, o più, delle caratteristiche sopra elencate; ad esempio, uno schema che offre garanzie sulla perdita dei pacchetti assicura sia al mittente (*sender*) che al destinatario (*receiver*) che non più di uno specifico numero di pacchetti verrà perso durante la trasmissione.
Fornire delle garanzie sui servizi non è un operazione banale in quanto esistono

impedimenti di vario tipo, sia nelle reti wireline che wireless. In particolare, le reti wireless includono tutti gli ostacoli presenti nelle reti cablate, con l'aggiunta di altri dovuti alla diversa natura della rete; i link wireless, ad esempio, sono molto più esposti alla perdita di pacchetti, che può essere dovuto ad una semplice interferenza radio-elettrica, o ad un fading di canale, in cui al ricevitore arrivano multipli dello stesso segnale. Un altro ostacolo nelle comunicazioni wireless può essere la propagazione del ritardo, dato che alcune reti wireless coprono distanze dell'ordine del chilometro; garantire quindi delle garanzie su ritardo totale può divenire un'impresa estremamente difficile. In entrambe le tipologie di rete, la congestione rappresenta uno dei più grossi ostacoli per fornire una garanzia sul servizio; infatti, in presenza di congestione aumenta inevitabilmente il ritardo *end-to-end*, per via di un maggior tempo di attesa nelle varie code presenti in una rete, e di conseguenza può aumentare il tasso di perdita dei pacchetti, per esempio, quando una coda è sovraccarica, i pacchetti in arrivo vengono eliminati: tutto questo pesa sul throughput della rete, dato che ci saranno una quantità maggiore di pacchetti che dovranno essere ritrasmessi.

Misura della QoS

Nelle reti wireless multimediali, la QoS può essere misurata a due livelli di astrazione [26]: connection-level e application-level.

Il livello connessione rappresenta il livello base di QoS ed è legato all'instauro ed alla gestione delle connessioni; copre un ruolo molto importante, specialmente quando viene generata una richiesta di handoff da parte di un utente.

A questo livello di astrazione viene misurata la connettività e la continuità del servizio, principalmente in termini di due parametri: il *Call Blocking Probability* (CBP) ed il *Call Dropping Probability* (CDP). Il primo rappresenta il rapporto tra il numero di nuove chiamate bloccate, dovuto ad un'insufficienza di banda, con il numero totale di nuove chiamate generate all'interno di una cella; tale indice misura la connettività in presenza di richieste di nuove chiamate. Il CDP, o Handoff Dropping Probability, è il rapporto tra il numero di chiamate di handoff eliminate con il numero totale delle richieste di handoff in una data cella; il CDP misura la continuità del servizio durante un handoff.

Anche se estremamente necessario, il solo livello di connessione non è sufficiente, in particolare nel valutare la qualità delle applicazioni percepite dagli utenti finali, le quali vengono connesse e continuate appunto dal livello di connessione. Il livello di applicazione è stato introdotto come supplemento del livello precedente e si riferisce alla qualità delle applicazioni offerte dalla rete agli utenti finali in termini di parametri di QoS che includono la banda, le variazioni di ritardo, il rate di perdita dei pacchetti, il rate di errore, etc.

Nell'ultimo decennio si sta affermando a livello applicazionale di QoS l'uso della funzione di utilità. Questa, come si vedrà nel capitolo prossimo, è un concetto derivato dall'economia e rappresenta il "livello di soddisfazione" dell'utente finale. Generalmente, l'utilità di un'applicazione è in funzione della banda, del ritardo e delle perdite di pacchetti della rete; ma è ragionevole assumere che le richieste in termini di ritardi e di perdita dei pacchetti di un'applicazione siano indipendenti dalla sua banda attualmente allocata. Di conseguenza, l'utilità dipenderà solo dalla banda utilizzabile che la rete può allocare ad un'applicazione. Il perché si usi l'utilità anziché la banda come misura del livello di applicazione di QoS è che l'utente finale non è interessato all'ammontare di banda che può utilizzare l'applicazione, ma bensì alla utilità (qualità) che l'applicazione ottiene da quella quantità di banda.

Capitolo 3

Allocazione di banda: metodologie

3.1 Introduzione

Nei capitoli precedenti si è appreso cos'è una rete wireless e com'è strutturata, prestando particolare attenzione alle reti cellulari ed alle Wireless LAN che, al momento della stesura, sono le tipologie di reti wireless più utilizzate al mondo, e si candidano a giocare un ruolo da protagoniste nelle imminenti reti di quarta generazione. Qualunque sia la loro tipologia, a livello "macro", la risorsa principale da gestire è la banda; il problema di gestire tale risorsa nasce da una caratteristica che accomuna le risorse in genere, ovvero la loro limitatezza. Specie nel campo telematico, tale problema diventa molto stringente dato l'elevato ed in costante crescita, numero di utilizzatori finali ed il "sovra-affollamento" radioelettrico a cui sono soggetti i nostri cieli. Lo sviluppo di nuove applicazioni ed il miglioramento di quelle già esistenti porta ad un utilizzo sempre maggiore in termini di banda, e considerando quest'ultima come una quantità fissa, il supporto a nuove applicazione è dato dagli algoritmi di allocazione dinamica della banda. Ogni applicazione, sia essa un semplice invio di mail o una più elaborata video conferenza, richiede una certa quantità di banda, che solitamente è compresa tra un minimo ed un massimo. Se la banda venisse allocata in maniera statica ad ogni richiesta di connessione verrebbe allocato un certo ammontare di banda che rimarrebbe tale per tutto il ciclo di vita della chiamata. L'adozione di un meccanismo non adattativo nella gestione della banda limiterebbe l'utilizzo della rete stessa a poche utenze. Per migliorare l'efficienza d'utilizzo della rete, quindi, si preferisce adottare degli schemi dinamici, cioè capaci di adattarsi alle condizioni di traffico presente nella rete. Con l'allocazione dinamica della banda, una connessione può subire variazioni sulla sua quantità di banda assegnata, mantenendo però una buona qualità. Ad esempio, in una situazione di congestione di rete, una chiamata attiva a cui è assegnato un certo

ammontare di banda, può essere degradata ad un livello inferiore al fine di ammettere nella rete una nuova richiesta di connessione ed aumentare quindi l'efficienza della rete medesima, cosa che non si sarebbe potuta fare con una allocazione di banda statica. In questo capitolo verranno esposti alcuni algoritmi di allocazione dinamica della banda con particolare interesse agli schemi basati sul concetto di utilità che ha oramai preso piede nel campo telematico.

3.2 Funzione di utilità e traffico multimediale

Il concetto di *utilità*, originariamente usato in economia, è stato introdotto negli ultimi anni nel campo dei sistemi telematici[27][28][29]. L'utilità rappresenta il "livello di soddisfazione" di un utente o delle performance che un'applicazione può raggiungere.

La funzione di utilità è una curva monotòna non-decrescente che descrive l'ammontare della banda ricevuta e che, al variare di questa, l'utilità percepita dall'utente finale può anch'essa subire delle variazioni. Il vantaggio primo dello uso della funzione di utilità è che questa riesce a riflettere, in maniera intrinseca, le richieste di QoS da parte dell'utente e quantificare l'adattabilità di una applicazione.

Come generare funzioni di utilità per applicazioni multimediali in grado di tener conto della natura adattativa delle applicazioni stesse ha rappresentato fin da subito un problema notevole. Solitamente, le reti wireless sono caratterizzate da un numero elevato di applicazioni multimediali, che possono avere delle caratteristiche adattative ben diverse tra loro. Nasce cosi l'esigenza di una categorizzazione del traffico multimediale in accordo con le loro caratteristiche adattative; N. Lu e J. Bigham [30] propongono un modello in cui il traffico viene suddiviso in due diverse classi e per ogni classe viene formulata una funzione di utilità che rispecchia la loro adattabilità:

- *Classe I – traffico real-time;*
  - *Real –time adattativo;*
  - *Real-time non adattativo;*
- *Classe II – traffico non-real-time*.

Traffico Real-Time Adattativo o Hard Real-Time

Il traffico real-time adattativo riferisce ad applicazioni che hanno requisiti di banda "flessibili" e in caso di congestione, possono adattare il loro rate di trasmissione in accordo con le condizioni della rete. Tipici esempi di applicazioni che appartengono a questa classe sono i servizi multimediali adattativi e il *video on demand*. La funzione di utilità associata a questo tipo di traffico può essere modellata come segue:

$$u(b) = 1 - e^{-\frac{k_1 b^2}{k_2 + b}} \tag{3.1}$$

dove k_1 e k_2 sono due parametri positivi che determinano la forma della funzione di utilità ed assicurano che quando un'applicazione riceve la sua massima banda richiesta b^{max}, l'utilità ottenuta u^{max} è approssimativamente uguale ad 1.

Una generica forma di funzione di utilità è riportata in Figura 3.1.

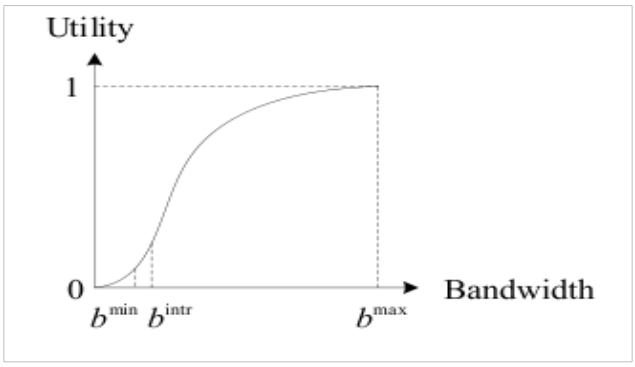

Figura 3.1. Funzione di utilità per traffico real-time adattativo

Per determinare la forma esatta di questa funzione, è necessario calcolare i parameri k_1 e k_2 , e per far ciò, si parte da una semplice considerazione: quando b, che è la banda allocata, tende a b^{max}, la corrispondente utilità tende a u^{max}, quindi la (3.1) può essere riscritta come:

$$1 - e^{-\frac{k_1 (b^{\max})^2}{k_2 + b^{\max}}} = u^{\max} \tag{3.2}$$

e dalla (3.2) si può ricavare la relazione che lega k_1 a k_2:

$$k_1 = \frac{\ln(1-u^{\max}) \cdot (k_2 + b^{\max})}{-(b^{\max})^2} \tag{3.3}$$

Come si può notare dalla Figura 3.1, prima del punto b^{intr} la curva è convessa, e dopo di esso la curva diventa concava. Tale punto può essere ricavato eguagliando a zero la derivata di ordine secondo della funzione di utilità:

$$\left(1 - e^{-\frac{k_1 (b^{intr})^2}{k_2 + b^{intr}}}\right)'' = 0 \tag{3.4}$$

e dalla (3.4), risolvendo la derivata, viene fuori la seguente equazione (3.5):

$$2k_1 \cdot \frac{e^{\frac{k_1 (b^{intr})^2}{k_2 + b^{intr}}}}{(k_2 + b^{intr})^4} \cdot \left(k_2^3 + \left(b^{intr} - 2(b^{intr})^2 k_1 \right) k_2^2 - 2(b^{intr})^3 k_1 k_2 - \frac{(b^{intr})^4 k_1}{2} \right) = 0 \tag{3.5}$$

dato che k_1 e k_2 sono entrambi positivi, l'equazione (3.5) è uguale a zero solo quando il polinomio cubico all'interno delle parentesi è zero (3.6):

$$k_2^3 + \left(b^{intr} - 2(b^{intr})^2 k_1 \right) k_2^2 - 2(b^{intr})^3 k_1 k_2 - \frac{(b^{intr})^4 k_1}{2} = 0 \tag{3.6}$$

sostituendo, infine, k_1 con la (3.3), l'equazione (3.6) diventa (3.7):

$$\begin{aligned} &\left(1 + 2\ln(1-u^{\max}) \cdot \left(\frac{b^{intr}}{b^{\max}}\right)^2\right) \cdot k_2^3 + \\ &\left(b^{intr} + 2\ln(1-u^{\max}) \cdot \frac{(b^{intr})^2}{b^{\max}} + 2\ln(1-u^{\max}) \cdot \frac{(b^{intr})^3}{(b^{\max})^2}\right) \cdot k_2^2 + \\ &\left(2\ln(1-u^{\max}) \cdot \frac{(b^{intr})^3}{b^{\max}} + \ln(1-u^{\max}) \cdot \frac{(b^{intr})^4}{2(b^{\max})^2}\right) \cdot k_2 + \\ &\ln(1-u^{\max}) \cdot \frac{(b^{intr})^4}{2b^{\max}} = 0 \end{aligned} \tag{3.7}$$

Considerando che b^{intr}, b^{max} e u^{max} sono tutte costanti che possono essere determinate dalla rete o dall'utente finale, la (3.7) si riduce ad un'equazione di terzo grado con tutti i coefficenti costanti da risolvere in funzione di k_2. Ottenuto k_2, k_1 può essere calcolato usando l'equazione (3.3).

Traffico Hard-Real-Time

Il traffico real-time non-adattativo riferisce ad applicazioni aventi requisiti stringenti di larghezza di banda, e dunque non presentano nessuna proprietà di adattamento. Una chiamata appartenente a questa classe di traffico non può essere autorizzata ad entrare nella rete se la stessa non può soddisfare la sua richiesta di banda minima. La banda allocata non può variare durante il ciclo di vita della chiamata, pena un'utilità pari a zero. Applicazioni tipiche sono telefonate audio/video e video conferenza.
Per modellare questa classe di traffico si utilizza la seguente funzione di utilità:

$$u(b) = \begin{cases} 1, \text{ when } b \geq b^{\min} \\ 0, \text{ when } b < b^{\min} \end{cases} \tag{3.8}$$

mentre in Figura 3.2 se ne riporta la forma.

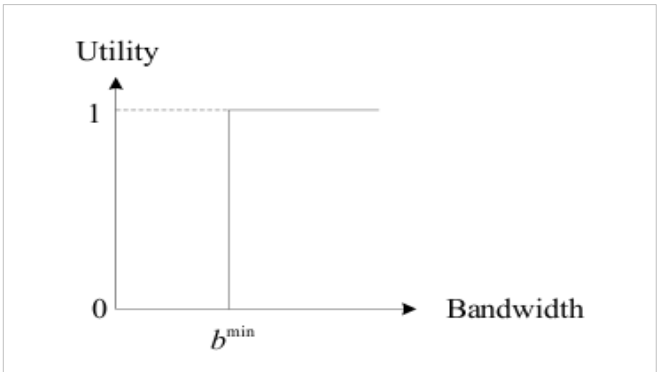

Figura 3.2. Funzione di utilità per traffico Hard-Real-Time.

Traffico Non-Real-Time

A questa classe di traffico appartengono applicazioni che sono piuttosto tolleranti ai ritardi, e di conseguenza non hanno un requisito minimo di banda. Le applicazioni più popolari non-real-time sono le e-mail, il file trasfer ed il remote login. Per modellare questa classe viene utilizzata la seguente funzione di utilità:

$$u(b) = 1 - e^{-\frac{kb}{b^{\max}}} \tag{3.9}$$

dove k è un parametro positivo che determina la forma della funzione, riportata in Figura 3.3. Anche in questo caso, quando l'applicazione riceve il suo massimo di banda richiesto, l'utilità è approssimabile ad uno.

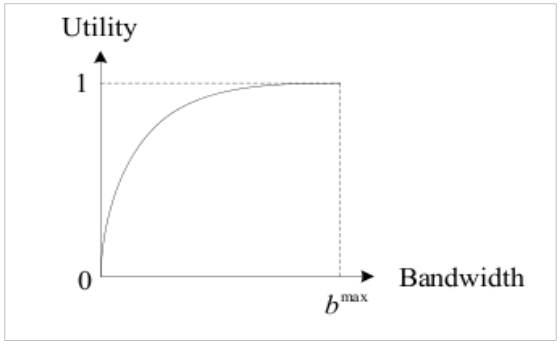

Figura 3.3 Funzioni di utilità per traffico non-real-time

Come si può notare dalla figura, la curva è concava ovunque. Per determinare l'esatta forma è opporuno calcolare il parametro k, quindi, considerando che quando la banda b tende a b^{max} la corrispondente utilità tende ad u^{max}, si può utilizzare la seguente equazione:

$$1 - e^{-\frac{kb^{\max}}{b^{\max}}} = u^{\max} \tag{3.10}$$

dalla (3.10) si può ricavare il parametro k come segue:

$$k = -\ln(1 - u^{\max}) \tag{3.11}$$

e dato che u^{max} è un parametro che può essere pre-determinato, k può essere facilmente calcolato utilizzando l'equazione (3.10).

3.2.2 Algoritmo di allocazione di banda: Connection Swapping e Bandwidth Borrowing

Il primo algoritmo preso in esame è stato presentato da T. Nyandeni et al.[31], e fa uso delle tre classi di traffico presentate da N.Lu et al.[30]. Lo scopo di questo algoritmo è quello di ridurre il *Call Blocking Probabilty* ed il *Call Dropping Probabilty* per le nuove chiamate e per quelle di handoff rispettivamente. Per la gestione delle chiamate di handoff, alle quali non è riservata alcuna quantità di banda, viene utilizzato un meccanismo chiamato *Connection Swapping*, mentre per le nuove chiamate viene utilizzato un meccanismo diverso basato sul prestito di banda, meglio conosciuto come *Bandwidth Borrowing*. Per meglio comprendere tale ultimo concetto è bene fare alcune anticipazioni: quando un nodo mobile fa una richiesta ad una data cella per una nuova connessione, esso deve specificare alcuni parametri, quali la classe di traffico di appartenenza, la quantità di banda che si aspetta di ricevere e la banda minima che è disposto a tollerare in accordo con la sua classe di traffico. La differenza tra la banda attesa e la banda minima rappresenta la quantità di banda che l'algoritmo può prendere in prestito da quella connessione in caso di *overload* della cella ospitante. Tale procedura non può essere adottata in presenza di una classe di traffico *hard-real-time*, la quale non ammette adattamenti di banda. Per quanto riguarda il meccanismo di *Connection Swapping*, invece, può essere inteso come una sorta di scambio di banda tra due connessioni appartenenti a due celle diverse, ma adiacenti.

Nuova chiamata

Una volta ricevuta una richiesta, una nuova connessione di una qualsiasi classe, è accettata nella cella solo se:

- la sua banda attesa è minore o uguale della banda libera, cioè dellla banda che la cella stessa può utilizzare;
- la sua banda minima è minore o uguale della banda utilizzabile della cella;
- la sua banda minima è minore o uguale della banda totale che si può chiedere in prestito a tutte le connessioni appartenenti alle classi II e III, e cioè le classi real-time adattative e quelle non real-time.

Se usando tutta la banda libera e tutta la banda che si può chiedere in prestito non si riesce a coprire la banda richiesta o la banda minima di una connessione, la stessa connessione viene reiettata, quindi non ammessa.

Di seguito si riporta il comportamento di tale algoritmo, sotto forma di *pseudo-codice*, per quanto riguarda le nuove chiamate.

Siano:

T_{AV} la banda libera totale della cella;

Exp_n la banda attesa di una nuova richiesta;

Min_n la banda minima di una nuova richiesta;

T_{BB} la banda totale *"borrowable"*;

e ricordando che la banda che si può chiedere in prestito è data da (3.12):

$$Borrowable\ bandwidth = Exp_n - Min_n$$

allora la banda totale *borrowable* può essere espressa come (3.13):

$$T_{BB} = \sum_{i=1}^{n} (Exp_i - Min_i) \qquad (3.13)$$

dove *n* rappresenta il numero di connessioni attive nella cella appartenenti alle classi II e III. A questo punto si riporta l'algoritmo in grado di accettare o meno una nuova richiesta di connessione:

```
If ( Exp_n <= T_AV ) Then
        Accept request
  Else If ( Min_n <= T_AV ) Then
        Accept request
  Else If ( Min_n <= T_AV + T_BB ) Then
        Accept request
    Else
        Block request
        End If
     End If
End If
```

Chiamata di handoff

Il modo in cui vengono gestite le richieste di handoff varia in base al tipo di classe di traffico di appartenenza. Si è assunto che le connessioni appartenenti alla classe III, riferenti ad applicazioni *non-real-time*, vengono lasciate attive anche se la loro banda allocata è molto piccola, dato che non hanno requisiti stringenti di qualità e sono molto tolleranti ai ritardi; quindi le connessioni di classe III non vengono *"droppate"* finché c'è banda a sufficienza nella nuova cella. Per le richieste di handoff appartenenti alle classi I e II, il discorso è chiaramente diverso; queste vengono accettate o meno dalla cella se:

- la loro banda richiesta è minore o uguale alla banda libera totale della cella;
- la loro banda minima è minore o uguale alla banda libera più la banda *borrowable* totale;
- la loro banda minima è minore o uguale alla banda libera più la banda borrowable più la banda ottenuta dopo una *"connection swapping"*.

Prima di introdurre il meccanismo di *"swap"* di una connessione, che come già anticipato si basa su uno scambio di banda tra due celle adiacenti, è necessario indagare sulla struttura stessa della cella utilzzata in questo schema, che viene mostrata in Figura 3.4. Questo tipo di struttura cellulare è stata presentata da Nkambule et al.[32], e segue un approccio di tipo *sectored-cell*, ovvero di cella settorizzata. La parte ombreggiata della cella è detta *"regione critica"* e rappresenta una zona di sovrapposizione di due o più celle. Se un nodo mobile si trova nella regione critica della cella può comunicare con le *Base Station* delle celle adiacenti.

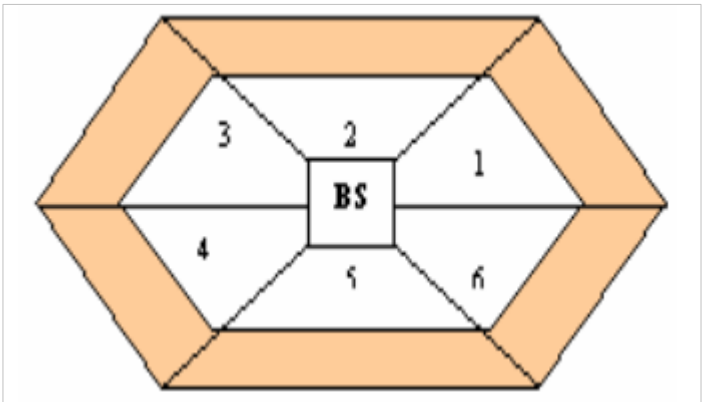

Figura 3.4. Cella Settorizzata.

Connection Swapping

Il meccanismo di *connection swapping* entra in gioco quando un nodo mobile fa una richiesta di handoff in una cella che non ha abbastanza banda. Allora, quella particolare cella dovrà scandire tutte le connessioni dei nodi mobili che si trovano nella regione critica e prenderne la classe di traffico e la direzione. Se il nodo mobile appartiene alla III classe di traffico e non è in movimento oppure si muove lentamente, allora la stazione base corrente può iniziare una negoziazione con la stazione base del nodo mobile, in particolare: la stazione base corrente, cioè quella in cui un nodo ha fatto richiesta di handoff, rilascia la banda occupata dal nodo stesso; questa banda verrà assegnata al nodo mobile scandito, in virtù del fatto che le connessioni di classe III avranno sempre una banda minore delle connessioni di classe I o II. Questo meccanismo prende il nome di "*connection swapping*" e, alla fine del processo, la richiesta di handoff sarà accettata e la connessione di classe III sarà presa dalla cella del nodo mobile che ha richiesto l'handoff. La banda ottenuta da una connection swapping è detta "*achieved bandwidth*", mentre la "*total achieved bandwidth*" è la somma della banda ottenuta dallo swapping di due o più connessioni. Al fine di una più chiara lettura del comportamento di questo algoritmo con le richieste di handoff, si riporta di seguito lo *pseudo-codice* sia per le richieste *non-real-time* che per quelle *real-time*.

Siano:

T_{AB} la banda totale ottenuta da tutti gli swapping di connessioni;

Ab la banda ottenuta da una Connection Swapping;

Exp_h la banda attesa da una nuova richiesta;

Min_h la banda minima di una nuova richiesta.

dove T_{AB} è data da (3.14):

$$T_{AB} = \sum_{j=1}^{k} (Ab_j) \tag{3.14}$$

con k pari al numero di tutte le connessioni scambiate.

Di seguito si riporta il codice per le richieste di handoff di connessioni appartenenti alla III classe di traffico:

```
If (T_AV > 0 OR T_BB > 0) Then
        Accept request
        Else
            Drop request
End If
```

Il codice per le richieste di handoff per connessioni di classe I e II è di seguito riportato:

```
If ( Exp_h <= T_AV ) Then
        Accept request
  Else If ( Min_h <= T_AV ) Then
        Accept request
  Else If ( Min_h <= (T_AV + T_BB )) Then
        Accept request
  Else If ( Min_h <= (T_AV + T_BB + T_AB ))
        Accept request
        Else
            Drop call
    End If
   End If
 End If
End If
```

Quando una connessione termina, la banda liberata è usata per colmare la banda delle connessioni che lavorano al di sotto della loro banda richiesta, a causa, per esempio, di un prestito di banda. Per questa procedura si utilizza il meccanismo di *Max-Min Fairness*[33]. Nel caso in cui, dopo questa spartizione, risulta esserci ancora banda a disposizione, questa verrà usata per incrementare il livello di banda libera della cella.

3.2.3 Algoritmo di adattamento di banda Utility-Fair

L'algoritmo di allocazione di banda che si andrà ad analizzare è stato proposto da

N.Lu e J.Bigham[34] e fa uso delle tre classi di traffico da loro stessi presentate [30]. L'obiettivo maestro che si propone di raggiungere questo algoritmo, da un punto di vista dell'utente finale, è quello di trattare tutti gli utenti allo stesso modo, e cioè la qualità ricevuta da un'applicazione deve essere il più possibile uguale a quella di tutti gli altri utilizzatori. L'*Utility-Fair Bandwidth Adaptation* si basa sulla funzione di utilità che mette in relazione la banda ricevuta con la soddisfazione dell'utente; trattare gli utenti allo stesso modo significa far loro ricevere eguale utilità.
Prima di esplorare nel dettaglio l'algoritmo proposto, è bene evidenziare alcune caratteristiche della funzione di utilità, e in che modo questa viene "preparata" per soddisfare le esigenze dell'algoritmo stesso.

Quantizzazione della funzione di utilità
Nelle reti wireless con traffico multimediale, ad ogni chiamata è associata una funzione di utilità, la cui forma dipende dalla classe di traffico di appartenenza. Quando una chiamata inoltra una richiesta di connessione ad una data cella, si presume che la stessa chiamata fornisca le seguenti informazioni:

- Classe di traffico;
- Richiesta di banda;
- Funzione di utilità;

Con il paradigma dell'allocazione di banda adattativa, se c'è abbastanza banda libera nella rete, alla chiamata viene allocata la banda massima richiesta b^{max}; d'altro canto, in accordo con il carico, o il sovraccarico, della rete, alla chiamata viene allocata una quantità di banda contenuta nell'intervallo $[b^{min},b^{max}]$. La funzione di utilità può riflettere, in teoria, le caratteristiche di un'applicazione multimediale in maniera molto accurata. Da un punto di vista pratico però, è preferibile avere una funzione di utilità il più semplice possibile, in modo da venire incontro agli schemi di adattamento di banda nelle reti wireless. Il compromesso tra accuratezza e semplicità è dato dall'uso di funzioni lineari a tratti che semplificano notevolmente le funzioni di utilità mantenendo un livello accettabile di accuratezza[35]. Una funzione di utilità può essere quantizzata in funzioni lineari a tratti dividendo il suo range di utilità o di banda in un numero di intervalli uguali. Sia *u(b)* una qualsiasi funzione di utilità, dopo la quantizzazione essa può essere rappresentata da una lista

di punti <*bandwidth,utility*> in ordine crescente di banda, ad esempio:

$$u(b) = (< b^1, u^1 >, < b^2, u^2 >, \ldots, < b^K, u^K >)$$

dove *1* è il livello più basso e *k* quello più alto della funzione lineare a tratti.

Nelle Figure 3.5 e 3.6 vengono mostrate le quantizzazioni delle funzioni di utilità relative alle tre classi di traffico ormai note; nella prima le funzioni sono state quantizzate con intervalli uguali di utilità Δu, mentre nella seconda con intervalli uguali di banda Δb. Da notare come le funzioni d'utilità relative al traffico Hard-Real-Time rimangono uguali sia prima che dopo la quantizzazione e contengono un solo punto <*bandwidth,utility*>, ad esempio :

$$u(b) = (< b^1, u^1 >).$$

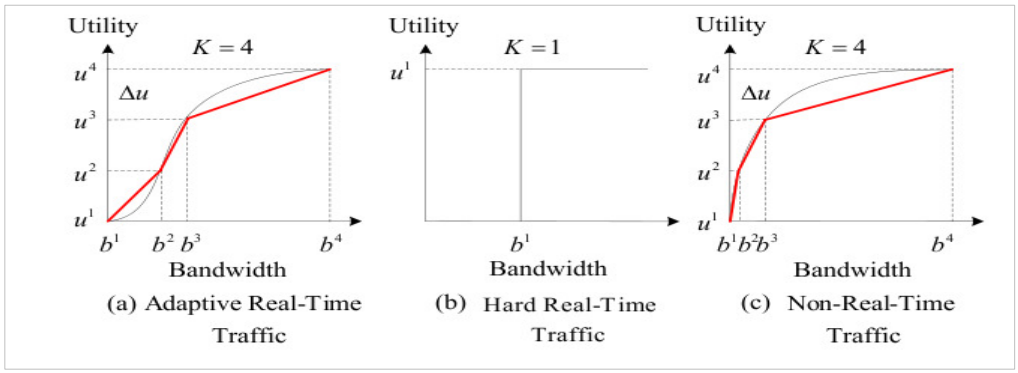

Figura 3.5. Funzione di utilità quantizzata usando intervalli uguali di utilità.

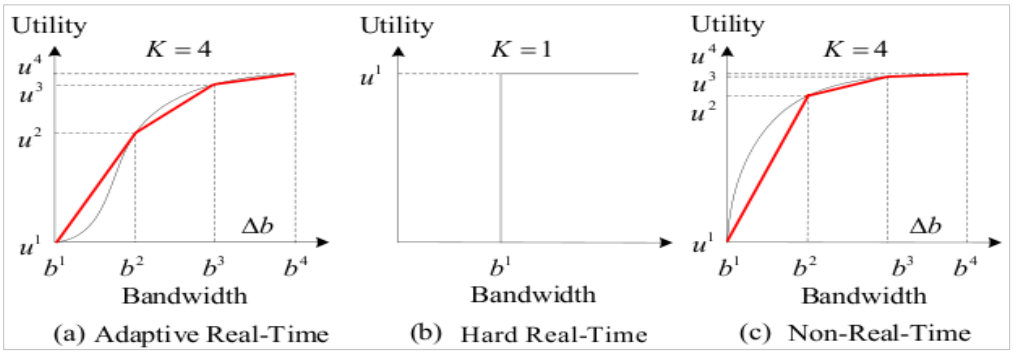

Figura 3.6. Funzione di utilità quantizzata usando intervalli uguali di banda.

Procedure di adattamento: Bandwidth Degrades e Bandwidth Upgrades

La procedura di adattamento di banda viene eseguita in ogni stazione base di ogni cella; essa consiste in due processi, bandwidth degrades e bandwidth upgrades, pilotati da eventi di arrivo di una chiamata, nuova o di handoff che sia, e da eventi di uscita, che includono la conclusione di una chiamata o il suo passaggio in un'altra cella.

Bandwidth Degrades

Il processo di degradazione viene attivato in presenza di un arrivo di una nuova chiamata od una chiamata di handoff, in una cella di una rete che non ha banda libera a sufficienza. In una situazione del genere, la banda delle chiamate in corso può essere degradata ad un valore più basso di quello attualmente ricevuto, al fine di riuscire ad ammettere la nuova richiesta. Nel processo di degradazione si deve tener conto:

- Delle chiamate in corso che possono essere degradate;
- Di quanto si deve diminuire la banda ad esse allocata;
- Di quanta banda necessita la nuova chiamata o di handoff.

Nella Figura 3.7 viene illustrato questo processo, ipotizzando una cella con n chiamate in corso di tipo adattative; b_i^{cur} e b_i^{adapt} rappresentano rispettivamente la banda correntemente allocata e la banda allocata dopo la degradazione della *i-esima* chiamata attiva.

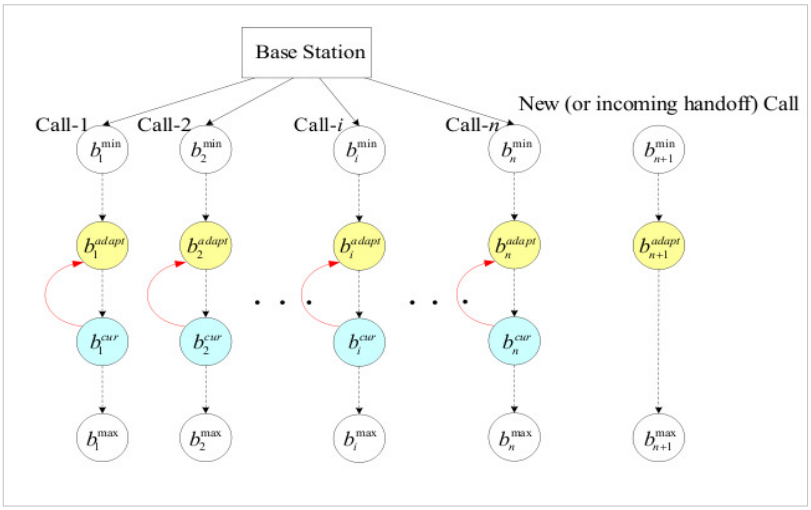

Figura 3.7. Processo di degradazione di banda.

Bandwidth Upgrades

Il processo di aggiornamento entra in gioco quando una chiamata termina il suo ciclo di vita o "passa" in un'altra cella. In ogni caso si libera della banda, e l'utilizzo di essa dipende dalla situazione in cui si trova la cella. Se tutte le chiamate adattative in corso ricevono la loro banda massima, allora la banda in eccesso è stipata per usi futuri; al contrario, la banda rilasciata può essere utilizzata per aumentare la banda delle chiamate che non stanno ricevendo il massimo di banda da loro richiesto. I fattori che si devono tener presenti durante questa procedura sono:

- Le chiamate attive che possono ricevere un upgrade;
- L'ammontare, in termini di banda, dell'*upgrade* stesso.

In Figura 3.8 è mostrato il processo di *upgrades* per chiamate adattative.

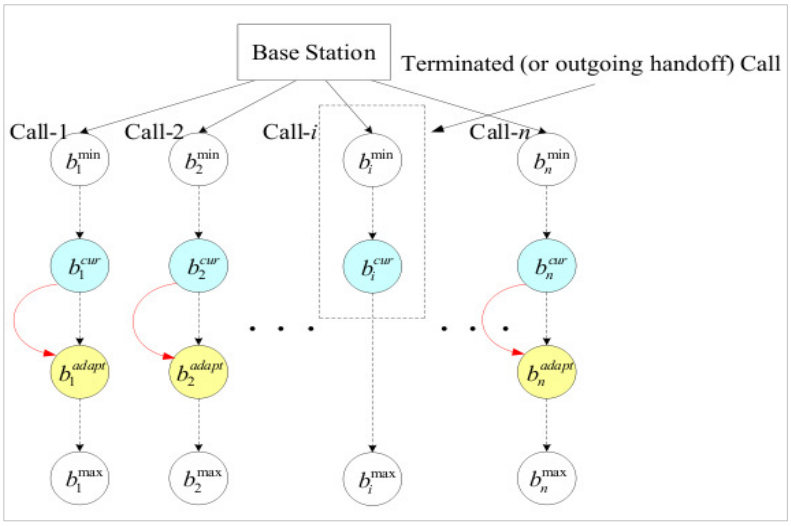

Figura 3.8. Processo di upgrades di banda.

Dove n rappresenta il numero delle chiamate adattative in corso e b_i^{adapt} è la banda allocata dopo il processo di *upgrades* della *i-esima* chiamata.

Definizione di utility-fairness

La definizione di *Utility Fairness* varia a seconda se ci si riferisce ad una rete wired oppure ad una rete wireless. In questo lavoro di tesi si prende in considerazione la definizione in riguardo alle reti wireless: Considerando una cella di una rete wireless e denotando con N il set di tutte le chiamate nella cella, il vettore dell'allocazione di banda $b=(b_1,b_2,...,b_n)$ è *utility-fair* se per ogni due chiamate $i,j \in N$, l'allocazione di

banda soddisfa la seguente equazione (3.15):

$$u_i(b_i)=u_j(b_j)$$

Degrades

Si consideri una cella satura con *n* chiamate adattative in corso; quando arriva una richiesta di nuova connessione o di handoff, le chiamate attive possono essere degradate ad un valore più piccolo per poter accettare la richiesta. Si può indicare la funzione di utilità della *i-esima* chiamata attiva come $u_i(b_i)$ con $(1 \le i \le n)$ e la sua banda allocata come β_i ; detto ciò, la funzione di utilità degradabile della *i-esima* chiamata può essere scritta come (3.16):

$$u_i^{\downarrow}(b_i^{\downarrow}) = u_i(\beta_i - b_i^{\downarrow}) \quad (0 \le b_i^{\downarrow} \le \beta_i - b_i^{\min}) \tag{3.16}$$

dove $b_i^{\downarrow}$e b^{min} sono il degrado di banda e la richiesta minima di banda della chiamata. La funzione di utilità della nuova chiamata, assunta anch'essa adattativa, la si può indicare come $u_{n+1}(b_{n+1})$. L'obiettivo del degrado di banda è quello di trovare un profilo di degrado $\{b_i^{\downarrow}\}$ per le *n* chiamate attive e l'allocazione di banda b_{n+1} per la nuova richiesta in modo che ricevano tutte un'eguale utilità (3.17):

$$u_i(\beta_i - b_i^{\downarrow}) = u_j(\beta_j - b_j^{\downarrow}) = u_{n+1}(b_{n+1}), \quad 1 \le \forall i, j \le n, \tag{3.17}$$

sotto il vincolo (3.18):

$$0 \le b_i^{\downarrow} \le \beta_i - b_i^{\min}, \tag{3.18}$$

per (3.19)

$$\left(\sum_{i=1}^{n}(\beta_i - b_i^{\downarrow})\right) + b_{n+1} \le B \tag{3.19}$$

dove *B* è il totale della banda utilizzabile.

Upgrades

Si consideri una cella sovraccarica dove al termine di una chiamata, o al suo

passaggio in un'altra cella, rimangono *n* chiamate che non ricevono la loro massima richiesta di banda. La banda rilasciata dalla chiamata conclusa, denotata con β, può essere utilizzata per fare un *upgrade* alle *n* chiamate attive, aumentando la loro qualità e l'utilizzazione di banda della rete. Denotando la funzione di utilità della *i-esima* chiamata con $u_i(b_i)$, $(1 \le i \le n)$, e la sua banda corrente con β_i , allora la funzione di utilità che può essere soggetta ad *upgrade* per la *i-esima* chiamata attiva, la si può scrivere come (3.20):

$$u_i^{\uparrow}(b_i^{\uparrow}) = u_i(\beta_i + b_i^{\uparrow}) \ (0 \le b_i^{\uparrow} \le b_i^{\max} - \beta_i) \tag{3.20}$$

dove $b_i^{\uparrow}$e b_i^{max} sono rispettivamente l'*upgrade* e la massima richiesta di di banda della chiamata. L'obiettivo di questo processo è di ottenere un profilo in cui tutte le chiamate ricevano uguale utilità (3.21):

$$u_i(\beta_i + b_i^{\uparrow}) = u_j(\beta_j + b_j^{\uparrow}), \quad 1 \le \forall i, j \le n, \tag{3.21}$$

sotto il vincolo (3.22):

$$\begin{aligned} & 0 \le b_i^{\uparrow} \le b_i^{\max} - \beta_i, \\ & \sum_{i=1}^{n} b_i^{\uparrow} \le \beta . \end{aligned} \tag{3.22}$$

Algoritmo utility-fair

Come già accennato, le funzioni di utilità vengono quantizzate in intervalli uguali di utilità Δu; dopo questa procedura, la funzione di utilità $u_i(b_i)$ divienta una funzione lineare a tratti rappresentata da un set di punti <*banda,utilità*> (3.23):

$$(< b_i^1, u_i^1 >, < b_i^2, u_i^2 >, \ldots, < b_i^{K_i}, u_i^{K_i} >) \tag{3.23}$$

dove k_i è il livello massimo di <*bandwidth,utility*>.

Per una funzione di utilità lineare a tratti $u_i(b_i)$, la sua allocazione di banda è

localizzata o su un segmento di linea l_i^{ki} tra due punti contigui di discontinuità $<b_i^{ki},u_i^{ki}>$ e $<b_i^{ki+1},u_i^{ki+1}>$, o esattamente su un punto $<b_i^{ki},u_i^{ki}>$. Questi due parametri b_i^{ki} e b_i^{ki+1}, sono utilizzati per definire la posizione approssimata dell'allocazione di banda della funzione di utilità $u_i(b_i)$. Quando la banda allocata è esattamente su un punto di discontinuità <*banda,utilità*>, è necessario solo un parametro, e cioè b_i^{ki}. Dato che le funzioni di utilità sono quantizzate con intervalli uguali di utilità, quando tutte le funzioni lineari a tratti ricevono uguale utilità, allora $k_1=k_2,...,=k_n=k$. Il primo passo dell'algoritmo di adattamento di banda è quello di individuare la posizione approssimata della allocazione di banda per ogni funzione d'utilità ricercando il suo valore di utilità tra *0* e *1* finché non si trovano i valori di b_i^{ki} e b_i^{ki+1}. La ricerca parte dal primo livello di utilità u_i^1, e ad ogni iterata il livello viene incrementato di uno. La stessa termina quando vengono trovati i valori di b_i^{ki} e b_i^{ki+1} per ogni funzione di utilità (3.24):

$$\sum_{i=1}^{n} b_i^k \le B < \sum_{i=1}^{n} b_i^{k+1} \tag{3.24}$$

dove *B* è la banda totale da assegnare. Se

$$\sum_{i=1}^{n} b_i^k = B \tag{3.25}$$

l'allocazione di banda b_i della funzione di utilità $u_i(b_i)$ è esattamente individuata su un punto di discontinuità $<b_i^k,u_i^k>$ e $b_i=b_i^k$.

Quando tutte le funzioni di utilità ricevono un uguale utilità, le loro utilità ottenute sul segmento l_i^k è uguale a tutti gli altri, cosi come descritto nella seguente equazione (3.26):

$$(b_1 - b_1^k) \times s_1^k = (b_2 - b_2^k) \times s_2^k = \cdots = (b_n - b_n^k) \times s_n^k \tag{3.26}$$

dove s_i^k rappresenta la pendenza del segmento l_i^k (3.27):

$$s_i^k = \frac{\Delta u}{b_i^{k+1} - b_i^k} \tag{3.27}$$

dalla (3.26) si possono ricavare b_2,..., b_n come segue:

$$\begin{cases} b_2 = (b_1 - b_1^k) \times \dfrac{s_1^k}{s_2^k} + b_2^k \\ \vdots \\ b_n = (b_1 - b_1^k) \times \dfrac{s_1^k}{s_n^k} + b_n^k \end{cases} \tag{3.28}$$

e dato che $b_1+b_2+...+b_n=B$, b_1 può essere calcolato come (3.29):

$$b_1 + ((b_1 - b_1^k) \times \frac{s_1^k}{s_2^k} + b_2^k) + \cdots + ((b_1 - b_1^k) \times \frac{s_1^k}{s_n^k} + b_n^k) = b^{avail} \tag{3.29}$$

ed infine, b_2,..., b_n si possono calcolare utilizzando la (3.28).

In Figura 3.9 viene mostrato un semplice esempio di ricerca dell'allocazione di banda utility-fair per due funzioni di utilità.

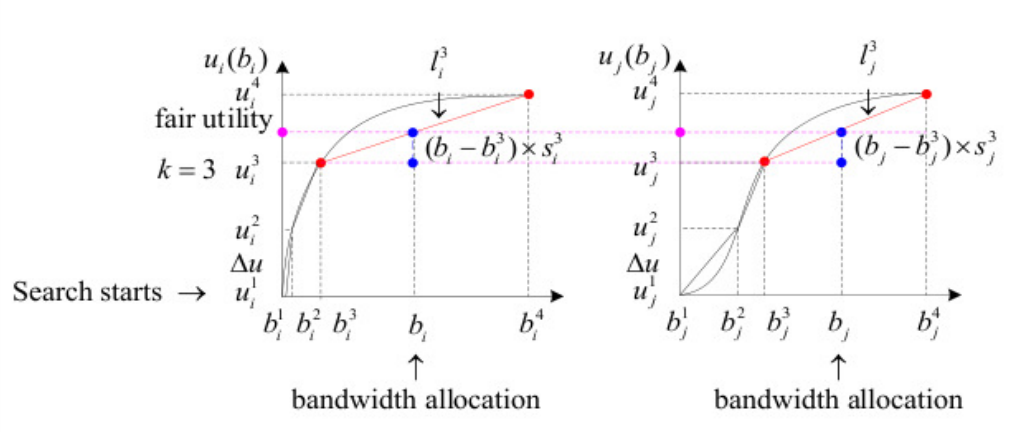

Figura 3.9. Allocazione di banda utility-fair per due funzioni di utilità.

3.3 Schemi Guard Channel

Uno dei temi più importanti nell'allocazione dinamica della banda nelle reti wireless è rappresentato dalla mobilità degli utenti, perciò molti schemi presenti in letteratura

si prestano ad affrontare tale tematica dando maggiore priorità alle chiamate di handoff. Quando si parla di schemi di priorità per le chiamate di handoff, ci si riferisce a strategie di assegnamento di canali trasmissivi, in cui vengono preferite le chiamate di handoff alle nuove chiamate. Il modo più semplice di dare priorità alle chiamate di handoff è quello di riservare loro una quantità di canali in ogni cella, e questo meccanismo prende il nome di *Guard Channel* (GC). Il primo schema GC è stato introdotto da Hong e Rappaport[36], e fin da subito si è potuto notare come l'adozione di uno schema del genere porta a miglioramenti in termini di probabilità di dropping per le chiamate di handoff, ma tende a ridurre l'utilizzazione totale della cella e ad aumentare la probabilità di blocco delle nuove chiamate, in quanto proprio queste ultime si trovano ad avere meno canali a disposizione. Nel corso degli anni sono stati proposti altri schemi basati sul concetto di GC, apportando migliorie in termini di probabilistici di blocco e dropping delle chiamate. Ad esempio, I. Candan e M. Salamah[37], propongono uno schema, *Fixed-Time-Threshold-based bandwidth allocation Schema* (FTTS), in cui viene monitorato il tempo realmente trascorso di una chiamata vocale e, a seconda se questo tempo è maggiore o minore di una data soglia temporale, la chiamata stessa viene considerata priorizzata o meno, e questo ne determina l'allocazione di banda. Successivamente è stato proposto uno schema dinamico di threshold*, Dynamic Time-Threshold-based Schema* (DTTS)[38], basato sempre sul monito delle tempo trascorso delle chiamate di tipo voce di handoff. A differenza dello FTTS, però, in questo schema le soglie di banda possono variare dinamicamente in accordo con una soglia di tempo t_e, ma diminuire questa soglia può avere un impatto diretto sulla QoS di una chiamata di handoff.

Nel successivo paragrafo viene esplorato nel dettaglio uno schema GC adattativo, proposto da J.Cai [39], per chiamate multimediali, ossia sia voce che dati.

3.3.1 Adaptive Time-Threshold-Based Schema per chiamate multimediali

L'algoritmo considerato fa anch'esso uso del monito del tempo realmente trascorso di una chiamata e facendo un confronto con delle soglie temporali, stabilisce se una data chiamata è priorizzata oppure no. I parametri delle soglie di banda possono

essere “aggiustati” adattativamente in accordo con la situazione corrente del traffico multimediale in una data cella di una rete wireless. Nella stesura di questo algoritmo si è presa in considerazione una singola cella di una rete multimediale come riferimento. Si è inoltre assunto che la cella abbia una capacità B unità di banda (Bandwidth Units) ed ogni chiamata voce richieda una BU, mentre una chiamata dati richieda due BUs. Un modello statistico di “Poisson” è stato adottato per descrivere l'andamento delle chiamate in arrivo, in modo da ottenere un largo numero di arrivi indipendenti. Come la maggior parte, se non tutti, degli schemi che utilizzano soglie temporali per differenziare le chiamate, anche questo schema si basa su un principio fondamentale, ossia: “risulta essere molto meglio bloccare una nuova chiamata che lasciar cadere una chiamata in corso”. In particolare, in questo contesto, viene fatta una considerazione più estesa al riguardo, data la presenza delle chiamate multimediali, e viene utilizzata la parola *noia*, *seccante*, appunto per sottolineare una sorta di insoddisfazione da parte di un utente al verificarsi di un tale evento: “per una chiamata attiva, cioè in corso, di tipo dati, risulta essere molto seccante se questa viene eliminata quando sta per volgere al termine, o meglio si avvicina al suo completamento, mentre è meglio eliminarla quando essa è appena iniziata; al contrario, per le chiamate vocali, risulta essere meglio se queste vengono eliminate qundo volgono al termine del loro ciclo di vita”. Come stabilire quando una chiamata, sia essa voce che dati, sta per completarsi, è una procedura basata su soglie di tempo, stabilite a priori, e su un ipotetico tempo minimo di conversazione che un utente può tollerare. In Figura 3.10 è riportata l'allocazione di banda dello ATTSM per le chiamate multimediali sia nuove che di handoff.

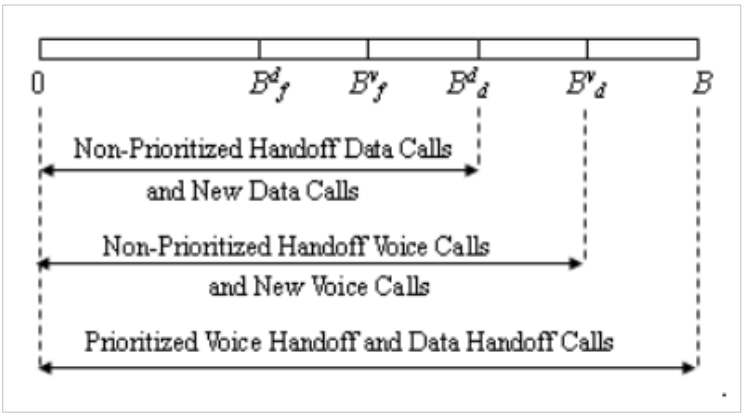

Figura 3.10 Allocazione di banda dello ATTSM

Le soglie di banda B_d^v e B_f^v sono rispettivamente la soglia dinamica e quella fissa per le chiamate di tipo voce, mentre B_d^d e B_f^d sono rispettivamente la soglia dinamica e fisse per le chiamate di tipo dati, con B_f^v maggiore di B_f^d . Le soglie di banda B_d^v e B_d^d possono cambiare dinamicamente in base al rate di arrivo delle chiamate di handoff (λ_h) ed al rate di arrivo delle nuove chiamate (λ_n) della cella corrente.

t_e^v e t_e^d rappresentano, invece, due soglie di tempo in base alle quali verranno categorizzate le chiamate. La prima riferisce alle chiamate vocali, ed è assunta essere maggiore della *durata minima di conversazione* "accettata" dall'utente, mentre la seconda riferisce alle chiamate dati, ed è assunta essere minore della *durata massima di dropping* accettata dall'utente mobile.

Come già accennato, le chiamate possono essere *priorizzate* e *non-priorizzate*; affinchè una chiamata di tipo voce si possa considerare priorizzata, il suo tempo trascorso deve essere più piccolo della soglia di tempo t_e^v, mentre, per una chiamata di tipo dati, il suo tempo trascorso deve essere maggiore della soglia t_e^d affinché essa sia ritenuta priorizzata. Tali chiamate sono ammesse nella cella finché l'ammontare della banda occupata nella cella è minore della capacità totale B della cella stessa. Per le chiamate di handoff e le nuove chiamate di tipo voce non-priorizzate, l'ammissione nella cella avviene finché l'ammontare della banda occupata nella cella è minore della soglia dinamica di banda B_d^v, mentre le chiamate di handoff e le nuove chiamate di tipo dati sono ammesse finché la banda occupata è minore della soglia dinamica di banda B_d^d . In definitiva, solo le chiamate non-priorizzate sono soggette ad adattamenti, o meglio, le soglie di banda adattative riferiscono alle chiamate non-priorizzate.

L'algoritmo adattativo per B_d^v e B_d^d può essere definito come:

$$\Delta B = \begin{cases} B - \left[\dfrac{|\lambda_h - \lambda_n|}{\lambda_n} \right] - B_{\max}^M & , \text{ if } \lambda_h \geqslant 2\lambda_n \\ B - B_{\max}^M & , \text{ if } \lambda_h < 2\lambda_n \end{cases} \qquad (3.30)$$

con B_d^v pari a (3.31):

$$B_d^v = max\{\ \Delta B\ ,\ B_f^v\ \} \tag{3.31}$$

e B_d^d pari a (3.32):

$$B_d^d = max\{\ \Delta B - (B_f^v - B_f^d),\ B_f^d\ \} \tag{3.32}$$

dove *[x]* denota il più grande intero di *x* e B^M_{max} denota la massima richiesta di unità di banda delle chiamate multimediali.

Capitolo 4

Algoritmi proposti: Analisi dei risultati

4.1 Introduzione

La simulazione, in generale, è la fase che segue alla stesura di un qualsiasi algoritmo, e permette di trarre informazioni sul suo comportamento, astraendo, in qualche modo, la realtà in cui si andrà a contestualizzare. L'elemento principale di questa fase è appunto il simulatore, che da la possibilità di implementare un algoritmo ed eseguirlo tenendo conto degli eventi che possono verificarsi in una situazione reale. In linea di massima, quindi, un simulatore deve mettere a disposizione degli sviluppatori un ambiente il più generico possibile, dando l'opportunità di creare un contesto *ad hoc*, quale la possibilità di scegliere la struttura della cella (circolare, esagonale,etc.), il suo diametro (in genere dell'ordine del Km), l'utilizzo di una singola cella oppure di un *cluster* di celle. Oltre al contesto, si deve avere la possibilità di generare chiamate di vario tipo (nuove chiamate, chiamate di handoff) con diversi requisiti di banda e la possibilità di gestire il modo d'arrivo, cioè il modo in cui viene modellato il rate d'arrivo, con cui queste chiamate giungono alla cella, che solitamente segue una distribuzione di Poisson. Al termine del ciclo di simulazione, ovviamente, si devono produrre dei risultati numerici consistenti, al fine di poter esprimere in maniera grafica il comportamento, appunto, di un algoritmo sottoposto a testing.

Seguendo l'ordine di esposizione del capitolo precedente, di seguito verranno illustrati, analizzati e, ove possibile, confrontati i risultati simulativi degli algoritmi considerati. Nella maggior parte delle analisi numeriche, il risultato grafico verrà espresso in termini di *Call Blocking Probaility* per le nuove chiamate, di *Call Dropping Probability* per le chiamate di handoff, nonché di utilizzazione totale della banda. Ogni autore si riserva la possibilità di utilizzare strumenti diversi, quale l'ambiente simulativo ed il numero di celle utilizzato per l'analisi, di generare numeri

arbitrari di chiamate multimediali e di sceglierne la loro distribuzione.

4.2 Connection Swapping e Bandwidth Borrowing: risultati simulativi

L'analisi delle performance dell'algoritmo presentato in [31] è stata eseguita su di un simulatore sviluppato in *VB.net* utilizzando un modello di simulazione descritto in [1].

Entrando subito nel dettaglio, in Figura 4.1 viene mostrato il numero di chiamate generate dal simulatore in 86 secondi.

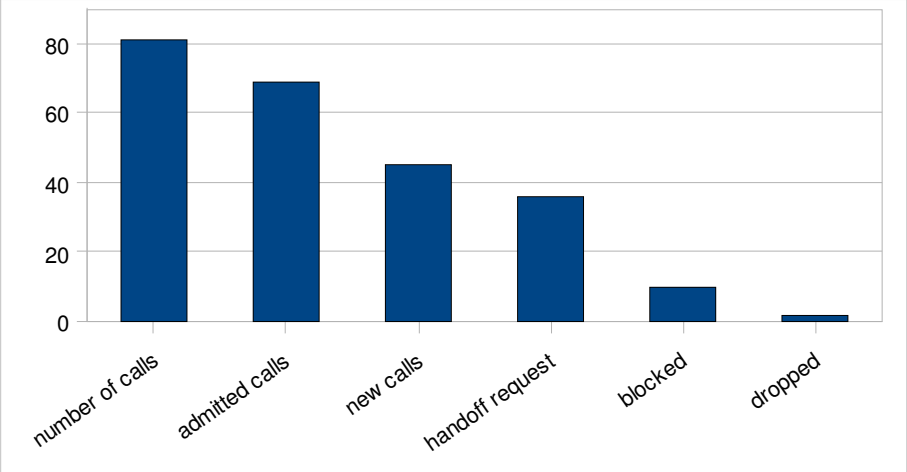

Figura 4.1. Numero delle chiamate.

Dalla Figura 4.1 si può notare che vengono generate 81 chiamate, di cui 45 sono nuove chiamate (originate all'interno della cella) e 36 sono chiamate di handoff (provenienti da cella adiacenti), ed ognuna di queste ha il proprio requisito di banda. Di queste 81 chiamate, 69 sono ammesse nella cella, mentre le restanti 12 non sono accettate. Tra le chiamate escluse, 10 sono bloccate, mentre solo 2 vengono droppate. Con questi dati numerici, si ha la possibilità di calcolare il *CBP* secondo l' equazione (4.1) riportata di seguito:

$$CBP = C_b / C_n \tag{4.1}$$

Dove C_b rappresenta la somma delle chiamate bloccate e C_n è il totale delle nuove

chiamate, quindi, con riferimento ai risultati sopracitati si ottiene:

$$CBP = 10 / 45 = 0,22 \tag{4.2}$$

Mentre per il calcolo del *CDP* si utilizza l'equazione (4.3):

$$CDP = C_d / C_h \tag{4.3}$$

Dove C_d è la somma delle chiamate droppate e C_h è il totale delle chiamate di handoff; sostituendo i valori numerici si ottiene:

$$CDP = 2 / 362 = 0,05 \tag{4.4}$$

Un valore di CDP pari a quello ottenuto nella (4.4) è un buon risultato, cioè la probabilità che una chiamata di handoff venga droppata è bassa, e questo è un punto a favore nella analisi della "bontà" di un algoritmo, ma il contesto di simulazione si presenta troppo generico, infatti non vi è riferimento alcuno a come le chiamate siano distribuite o a quanto ammonta il diametro della cella presa in considerazione (si rammenta che tale algoritmo fa uso di una cella esagonale settorizzata [32]).

Si ricorda che per ridurre il CDP, questo algoritmo fa uso del meccanismo di Connection Swapping (*vedi* cap. 3.2.2); nelle Figure 4.2 e 4.3 viene mostrato il risultato di un'analoga simulazione senza e con Connection Swapping, rispettivamente, considerando un intervallo temporale di dieci secondi. Questo da una più chiara visione, in termini di numero di chiamate droppate, dell'effetto che provoca tale meccanismo.

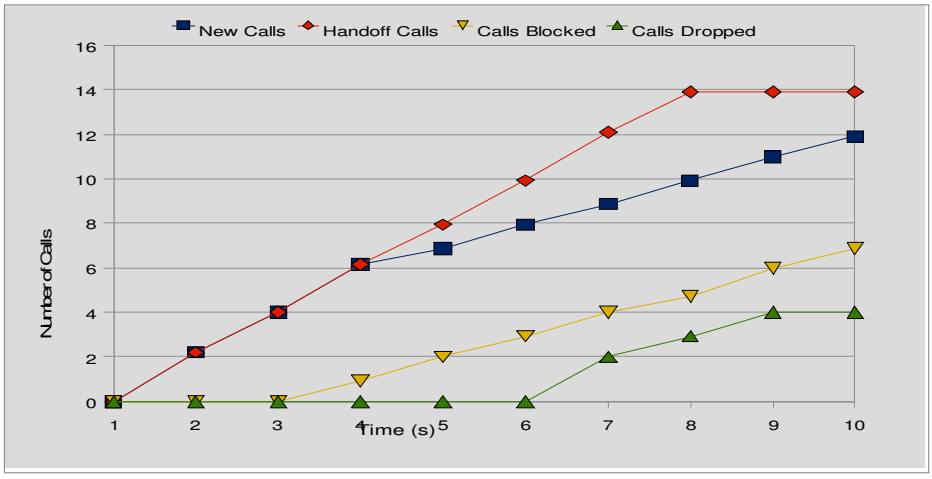

Figura 4.2. CDP e CBP senza Connection Swapping.

Dalla figura si può notare che il numero delle chiamate bloccate, riferito quindi alle nuove chiamate, è maggiore del numero delle chiamate di handoff droppate; questo, genericamente, è dovuto ad una preferenza da parte dell'utente finale, infatti risulta essere di più alto gradimento il blocco di una nuova connessione piuttosto che di una chiamata già avviata. Con riferimento alla linea gialla, che indica le chiamate bloccate, si può notare come questa cresca in maniera costante, ed è dovuto al fatto che l'algoritmo usa la somma della banda totale *borrowable* con la banda totale utilizzabile come condizione per accettare le nuove chiamate nella cella; alla fine della simulazione, comunque, risultano poche chiamate bloccate.
La linea verde, invece, indica le chiamate droppate, e si può notare come questa cresca rapidamente dopo i 6 secondi; questo significa che il prestito di banda da solo non apporta miglioramenti in termini di CDP.

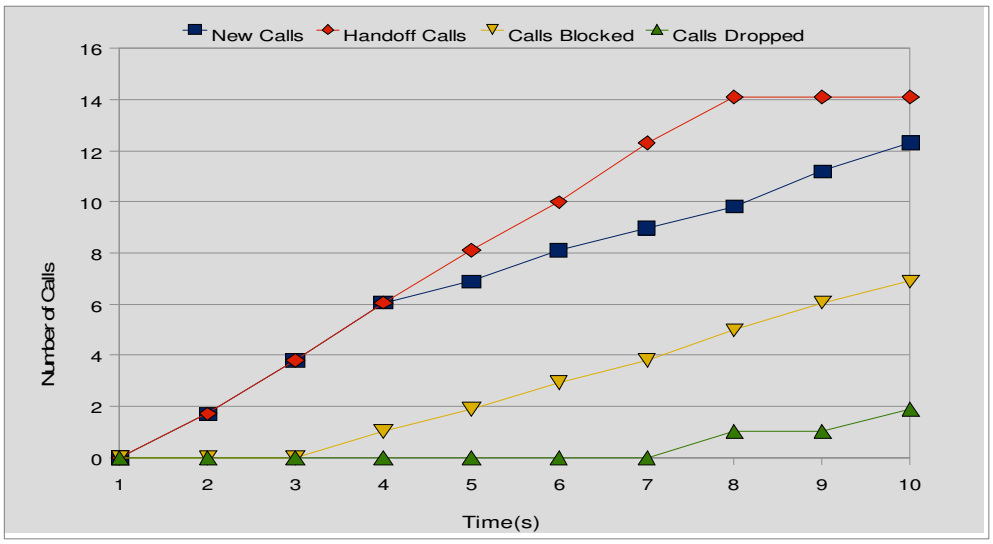

Figura 4.3. CDP e CBP con Connection Swapping.

In Figura 4.3 è stata utilizzata la Connection Swapping e si può subito notare che il numero delle chiamate droppate è molto basso, all'incirca pari a 2, e che da 0s a 7s il CDP è zero, mentre dopo i 7s cresce lentamente. L'uso del meccanismo di swapping ha, in effetti, migliorato la probabilità di dropping delle chiamate di handoff.
Nonostante il buon risultato ottenuto, a causa della mancanza di un confronto diretto con altri tipi di algoritmi di allocazione, ad esempio un algoritmo non adattativo, fa sì che il risultato simulativo medesimo resti fine a se stesso. Inoltre, il meccanismo

di Connection Swapping è strettamente legato ad un determinato tipo di cella, e questo può limitare, in termini di espandibilità, lo stesso algoritmo, e le sue ipotetiche applicazioni in ambienti già consolidati.

4.3 Analisi dei risultati dell'algoritmo Utility-Fair

Per valutare le performances dell'algoritmo utility-fair [34], i risultati simulativi del medesimo algoritmo,vengono confrontati con altri due schemi di adattamento di banda: uno non-adattativo ed il *Rate-Based Borrowing Schema* (RBBS) [41], che fa uso del meccanismo del prestito di banda. Il modello di rete adottato nella simulazione è composto da 36 (6X6) celle esagonali; ogni cella ha il diametro di 1 Km ed una capacità di 30 Mbps. La struttura della rete è riportata in Figura 4.4.

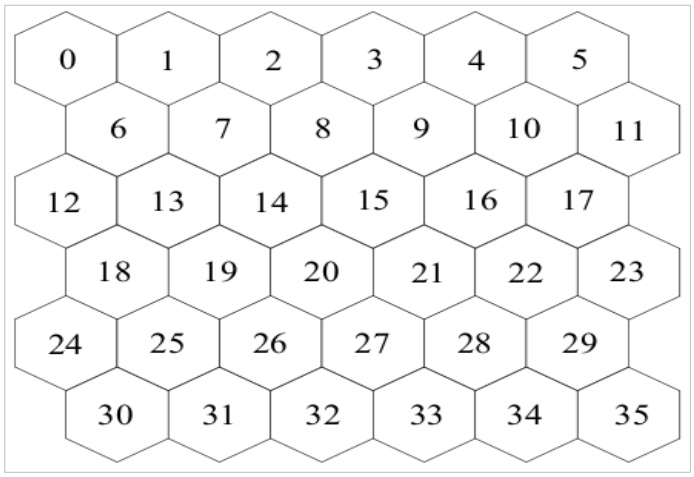

Figura 4.4. Modello di rete wireless.

Il modello di traffico, invece, è costituito da sei gruppi multimediali, in accordo con le tre classi di traffico utilizzate dall'Utility-Fair (*vedi* cap. 3.2.1). Ad ogni gruppo di traffico è associata una funzione di utilità, e tutte le chiamate appartenenti allo stesso gruppo di traffico presentano gli stessi requisiti di banda e la stessa funzione di utilità. Le esatte caratteristiche del traffico usato nella simulazione sono riportate nella tabella 4.1.

Tabella 4.1. Caratteristiche di traffico per la simulazione.

App. Group	Traffic Class	Bandwidth Requirement (Mbps)	Average Connection Duration	Example	Utility Function (b is Mbps)
0	I (Hard Real-Time)	$b^{min} = 0.03$ $b^{desired} = 0.03$	3 minutes	Voice Service & Audio Phone	$\begin{cases} 1, b \geq 0.03 \\ 0, b < 0.03 \end{cases}$ $u^{max} = 1$
1	I (Hard Real-Time)	$b^{min} = 0.25$ $b^{desired} = 0.25$	5 minutes	Video Phone & Video Conference	$\begin{cases} 1, b \geq 0.25 \\ 0, b < 0.25 \end{cases}$ $u^{max} = 1$
2	I (Adaptive Real-Time)	$b^{min} = 1$ $b^{intr} = 1.5$ $b^{desired} = 2$ $b^{max} = 6$	10 minutes	Interact. Multimedia & Video on Demand	$1 - e^{-\frac{1.8b^2}{8.3+b}}$ $u^{max} = 0.99$
3	II (Non-Real-Time)	$b^{min} = 0$ $b^{desired} = 0.003$ $b^{max} = 0.02$	30 seconds	E-mail, Paging & Fax	$1 - e^{-\frac{4.6b}{0.02}}$ $u^{max} = 0.99$
4	II (Non-Real-Time)	$b^{min} = 0$ $b^{desired} = 0.1$ $b^{max} = 0.5$	3 minutes	Remote Login & Data on Demand	$1 - e^{-\frac{4.6b}{0.5}}$ $u^{max} = 0.99$
5	II (Non-Real-Time)	$b^{min} = 0$ $b^{desired} = 1.5$ $b^{max} = 10$	2 minutes	File Transfer & Retrieval Service	$1 - e^{-\frac{4.6b}{10}}$ $u^{max} = 0.99$

Gli arrivi delle nuove chiamate seguono una distribuzione di Poisson; il rate di arrivo delle chiamate di handoff è proporzionale a quello delle nuove chiamate. La durata delle chiamate, detta *Call Holding Time* (CHT) segue una distribuzione esponenziale $1/\mu$; un'eccezione sulla durata è concessa alla classe di traffico II (gruppi 2,3,4), dato che risulta essere molto difficile pre-definirla, in quanto non dipende solo dal servizio, ma anche dall'adattamento dinamico di banda a cui è soggetta. Per questi motivi, alle chiamate di Classe II è assegnata una durata fissa. Tutti i sei gruppi di traffico sono stati generati con eguale probabilità. In ogni cella, il 5% della capacità totale viene riservato alle chiamate di handoff appartenenti alla Classe I (Hard-Real Time e Adaptive-Real Time), dato che esse presentano un requisito minimo di banda. Al fine di valutare sotto diversi aspetti le prestazioni dell'algoritmo in esame, oltre alle tradizionali CBP e CDP, gli autori propongono una nuova metrica a livello applicazionale, detta *Utility Fairness Deviation*, che misura in maniera quantitativa l'equità,o l'imparzialità, di utilità di tutte le chiamate adattative in ogni cella della rete. La *Utility Fairness Deviation* è definita come la

deviazione standard dell'attuale utilità ricevuta da ogni chiamata adattativa in corso dalla sua utilità aspettata, in relazione alla situazione di carico attuale della cella, e può essere calcolata secondo l'equazione (4.5):

$$d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(u_i - u_i^{fair})^2} \qquad (4.5)$$

Dove u_i e u_i^{fair} sono rispettivamente l'utilità attualmente ricevuta e l'utilità attesa della *i-esima* chiamata. Dalla definizione è noto che maggiore è l'utility fairness deviation, più "ingiusta" è la distribuzione dell'utilità tra tutte le chiamate adattative.

In Figura 4.5 è riportata la deviazione di utilità, per i tre schemi presi in considerazione, in funzione del rate di arrivo delle chiamate.

Lo schema RBBS è quello che presenta la più alta deviazione di utilità, mentre lo schema utility-fair e quello non adattativo hanno una utility fairness deviation pari a zero; questo risultato è dovuto al fatto che, per ogni chiamata in corso, l'utilità ricevuta dall'utility-fair e dallo schema non adattativo è uguale alla loro utilità attesa ($u_i = u_i^{fair}$ per l'utility-fair e $u_i = u_i^{fair} = u_i^{max}$ per il non adattativo).

In termini di distribuzione "imparziale" dell'utilità tra le chiamate in corso, si ha un primo risultato a favore dello Utility-Fair rispetto al RBBS.

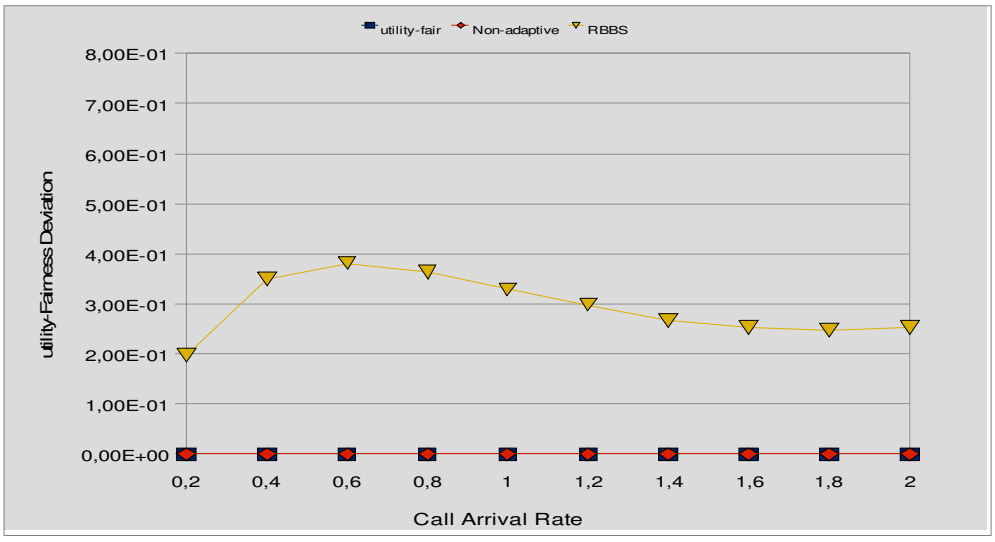

Figura 4.5. Utility Fairness Deviation.

Ritornando alle metriche classiche di valutazione, in Figura 4.6 vengono mostrate le probabilità di blocco delle nuove chiamate per gli schemi considerati.

L'algoritmo Utility-Fair presenta il minor CBP rispetto agli altri due fino ad un rate di 1.2, e si avvicina a quello del RBBS da 1.4 a 2.0. Ad un rate d'arrivo pari a 2.0, l'Utility-Fair accomoda circa il 6% di nuove chiamate in più rispetto allo schema non-adattativo e circa l' 1% rispetto al RBBS.

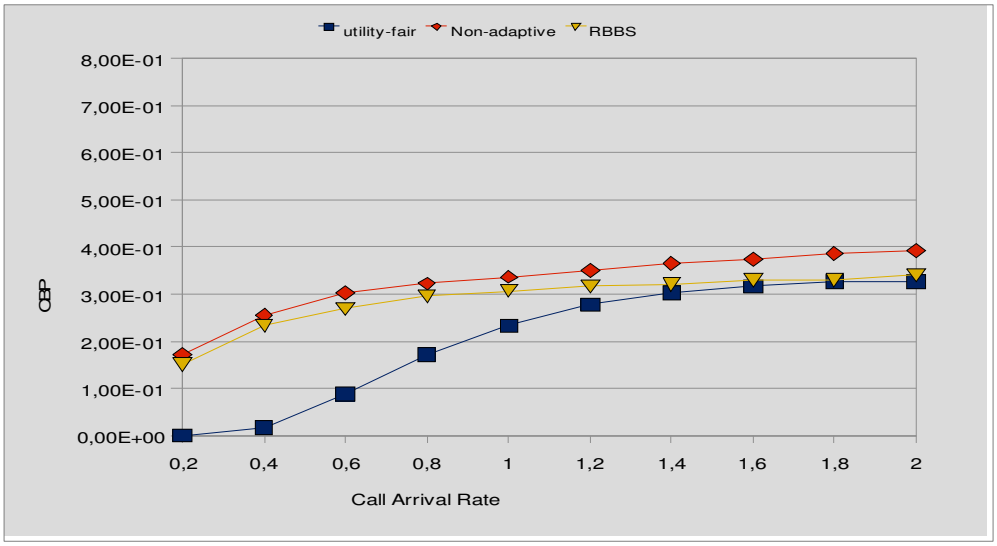

Figura 4.6. Call Blocking Probability con traffico multimediale.

In Figura 4.7 è riportata, invece, la probabilità di dropping delle chiamate di handoff, dove è possibile notare come l'algoritmo Utility-Fair mantiene una probabilità di dropping prossima allo zero, che rispetto agli altri schemi è la più bassa. Un risultato cosi ottimale è raggiunto grazie al meccanismo del bandwidth degrades, che mantenendo imparziale l'utilità ricevuta dalle chiamate in corso, può degradare, per quanto possibile, la banda alle medesime chiamate attive al fine di ammettere altre chiamate di handoff, le quali possono essere accettate con i loro requisiti minimi di banda. Lo schema non-adattativo è quello che ottiene i risultati più scarsi, dovuto al fatto che non utilizza alcun meccanismo di adattamento per liberare della banda; da qui l'impossibilità di accomodare nuove chiamate o chiamate di handoff in una situazione di congestione della cella. In termini di CBP e CDP, l'algoritmo Utility-Fair supera parzialmente lo schema RBBS, in quanto con il RBBS, ogni volta che avviene l'adattamento di banda, solo una quota della banda adattativa può essere presa in prestito (la quota di banda è un parametro pre-definito), mentre con l'Utility-Fair, la banda allocata alle chiamate in corso può essere degradata al livello minimo.

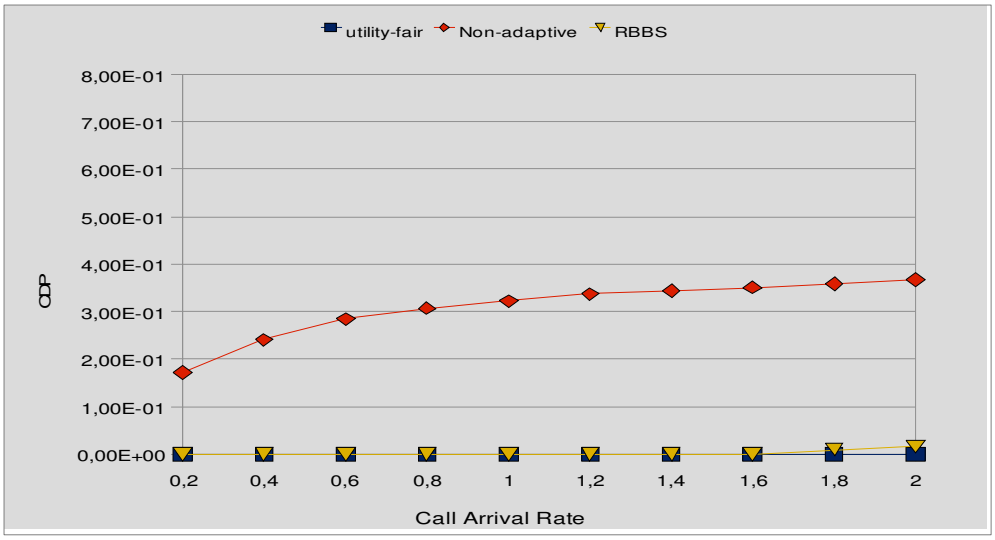

Figura 4.7. Call Dropping Probability con traffico multimediale.

In Figura 4.8 è riportato l'utilizzo di banda dei tre schemi.

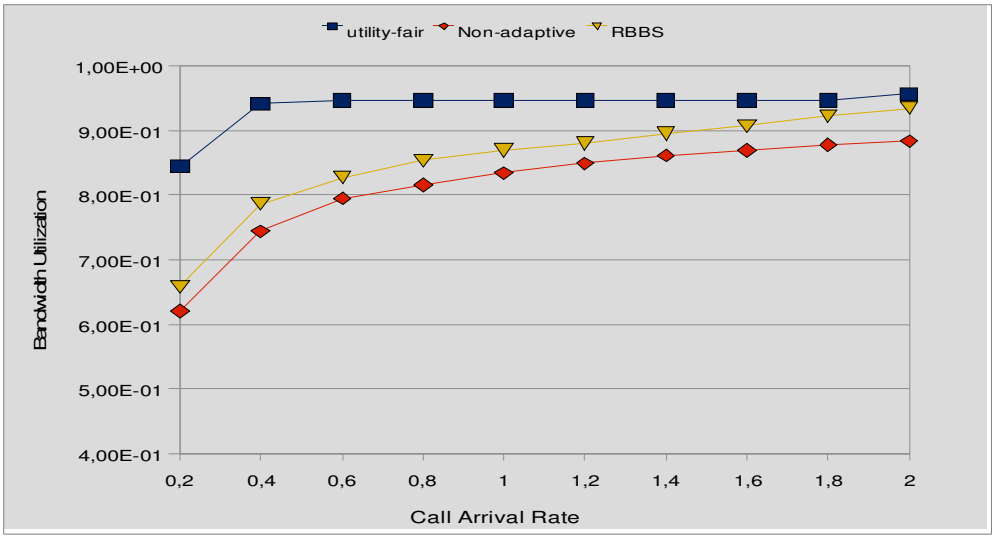

Figura 4.8. Utilizzo di banda.

L'algoritmo Utility-Fair sfrutta meglio la banda rispetto allo schema non adattativo ed a quello RBBS. Si noti come dopo un certo rate d'arrivo, 0.4 precisamente, l'utilizzo della banda da parte dell'Utility-Fair è pressoché costante; questo si può ricondurre alla possibilità di variazioni dell'utilità percepita da un utente al variare della sua banda allocata (*vedi* cap. 3.2.1). Ricevendo, ogni utente pari utilità, l'occupazione di banda nel complesso, si stabilizza intorno ad un valore (che dipende dalle condizioni di carico della rete).

4.4 Schema threshold-based per chiamate adattative: risultati e analisi

Questo algoritmo, presentato in [39], a differenza degli altri citati in questo lavoro di tesi, non fa uso della funzione di utilità, ma si basa su soglie di banda e soglie temporali (*vedi* cap. 3.3.1). Per la simulazione è stato utilizzato il simulatore di eventi discreti *OPtimized Network Engineering Tool* (OPNET) [42] e si è considerata una singola cella. Si ricorda che l'algoritmo in questione differenzia le chiamate di handoff, in base al tempo trascorso, come *priorizzate* e *non-priorizzate*; di conseguenza, i risultati simulativi relativi al Call Dropping Probability, sono espressi in termini di handoff priorizzati e non-priorizzati. Il traffico multimediale si è assunto essere di due tipi, voce e dati, dove il primo occupa una unità di banda, mentre il secondo due. Del traffico generato, facenti parte nuove chiamate e chiamate di handoff, il 70% è rappresentato da chiamate voce ed il restante 30% da applicazioni dati. Il rate d'arrivo delle chiamate multimediali, λ_h per le chiamate di handoff e λ_n per le nuove chiamate, segue una distribuzione di Poisson, mentre il *Call Holding Time* segue una distribuzione esponenziale pari a 180s. Gli altri parametri utilizzati nella simulazione, quali la capacità totale della cella *B*, la soglia di banda fissa per le chiamate di tipo voce B_f^v e per quelle di tipo dati B_f^d sono pari riapettivamente a 100, 90 e 85 unità di banda. La media, uniformemente distribuita, del tempo trascorso delle chiamate multimediali di handoff è di 90s, mentre le soglie di tempo per le chiamate voce t_e^v e per le chiamate dati t_e^d sono rispettivamente 120s e 30s.

I risultati simulativi, che verranno di seguito esposti, sono stati messi a confronto con un altro schema di threshold, il *Fixed Time-Threshold-Based Schema* (FTTSM) [43], che si basa, allo stesso modo dello ATTSM, sul monito del tempo trascorso relativo alle chiamate multimediali, ma fa uso di soglie di banda fisse, quindi non adattabili alla situazione di carico della rete.

Le performance dell'algoritmo, in termini di CBP, vengono valutate tenendo conto del rate d'arrivo delle chiamate di handoff e di quello delle nuove chiamate, presi separatamente. Le probabilità di blocco delle nuove chiamate sono riportate rispettivamente in Figura 4.9 e 4.10.

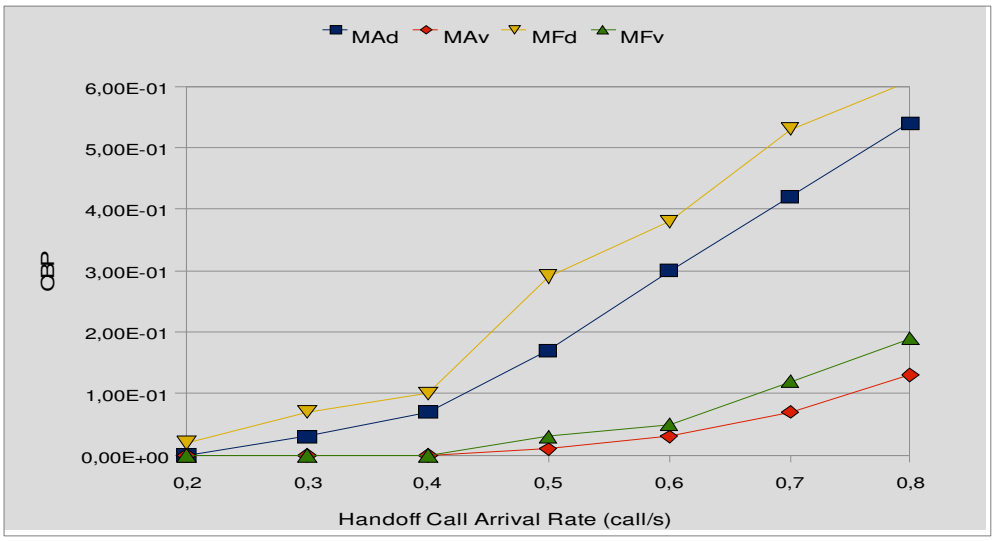

Figura 4.9. CBP vs Handoff arrival rate con λ_n=0.2call/s.

Dove *MAd* e *MAv* denotano il traffico dati e quello voce dello ATTSM, mentre *MFd* e *MFv* sono il traffico dati e voce dello FTTSM.

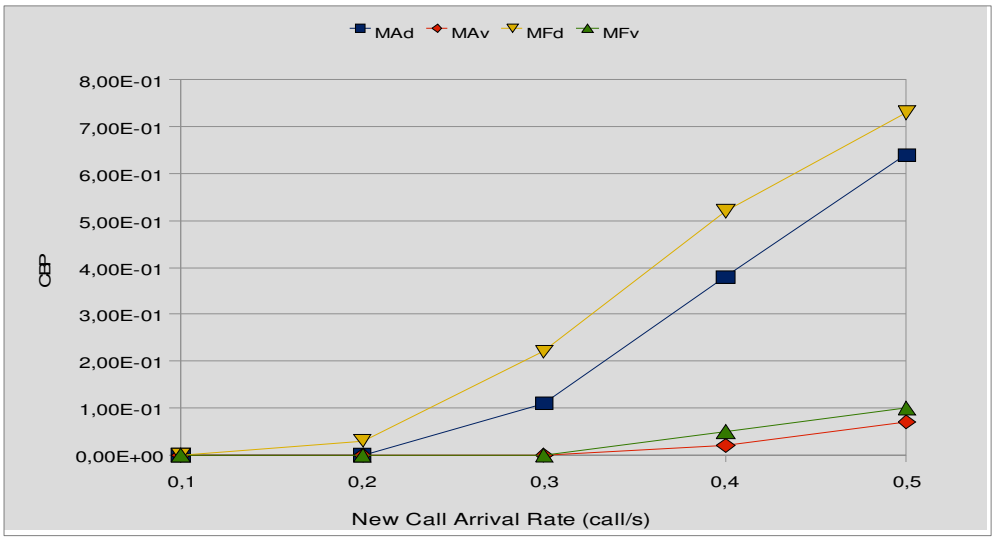

Figura 4.10. CBP vs New Call arrival rate con λ_h=0.2 call/s.

Dalle figure si può notare come il CBP dell'algoritmo in esame, rispetto allo FTTSM, scende notevolmente con il crescere di λ_h o λ_n; questo perché più unità di banda possono essere utilizzate ad ammettere nuove chiamate multimediali. Più interessante è il discorso sulla probabilità di dropping delle chiamate di handoff. In Figura 4.11 è riportato il CDP per le chiamate di handoff non-priorizzate.

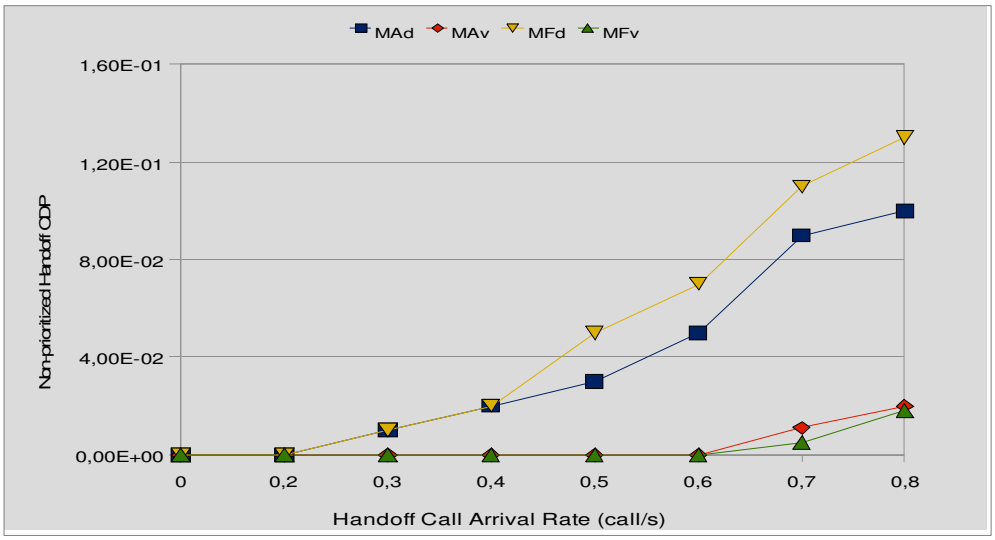

Figura 4.11 Non-prioritized CDP vs Handoff arrival rate con λ_n=0.2 call/s

Con riferimento alla Figura 4.11, per λ_h che tende a 0.8 chiamate/s, il CDP dello ATTSM per le chiamate dati è 1.07E-1, mentre quello dello FTTSM è 1.3E-1 che, in termini percentuale si traduce in un miglioramento del 17.69% a favore del threshold adattativo. Un risultato analogo può essere osservato in Figura 4.12, che mostra il CDP in relazione al rate d'arrivo delle nuove chiamate multimediali.

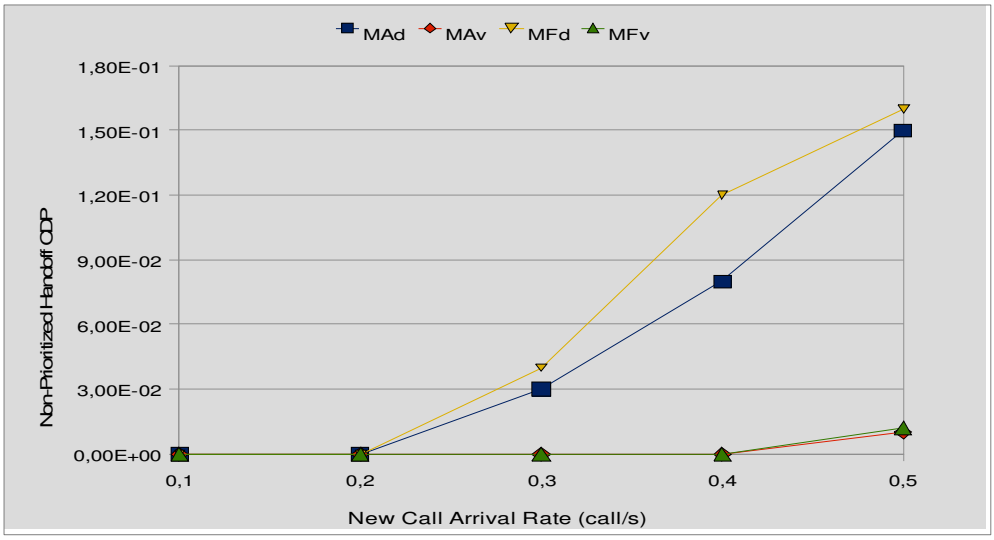

Figura 4.12. Non-prioritized CDP vs New Call arrival rate con λ_h=0.2 call/s.

Il CDP delle chiamate di handoff di tipo dati per lo schema ATTSM e quello FTTSM sono rispettivamente 7.86E-2 e 1.23E-1 con λ_n pari a 0.4 chiamate/s. In termini

percentuali si ha un miglioramento del 36.10%. Per quanto riguarda la probabilità di dropping delle chiamate di handoff priorizzate, seguendo lo stesso criterio di confronto, lo schema FTTSM si presenta con risultati migliori rispetto allo ATTSM, dovuto al maggior numero di unità di banda che lo schema a soglie fisse riserva alle chiamate multimediali priorizzate. Lo scarto di CDP tra i due schemi è comunque piccolo. Nelle Figure 4.13 e 4.14 vengono mostrati i risultati simulativi per gli handoff priorizzati in termini di CDP.

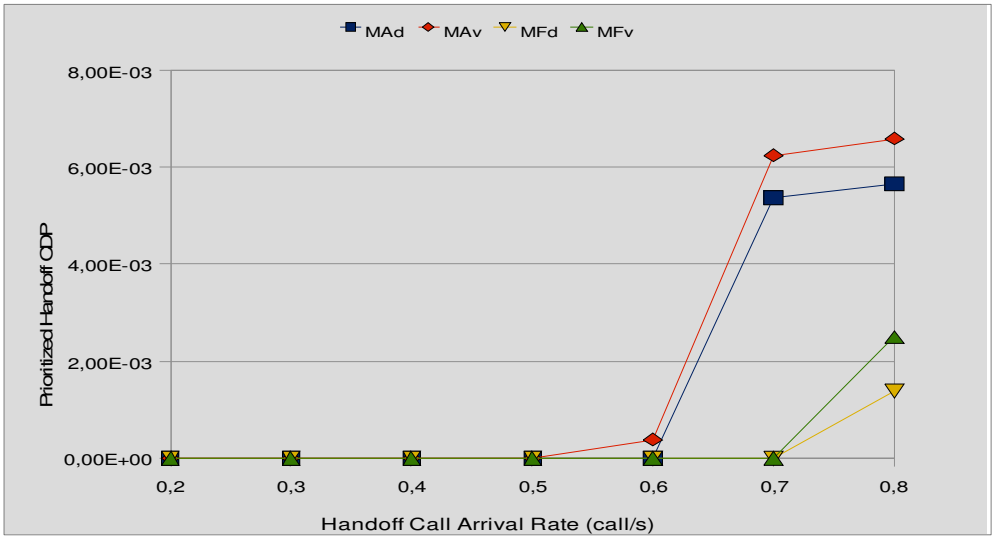

Figura 4.13. Prioritized CDP vs Handoff arrival rate con λ_n=0.2 call/s.

Con riferimento alla Figura 4.13, il valore di CDP per le chiamate voce e dati dello schema ATTSM sono rispettivamente 6.57E-3 e 5.58E-3 contro i 2.50E-3 e 1.39E-3 dello schema FTTSM.

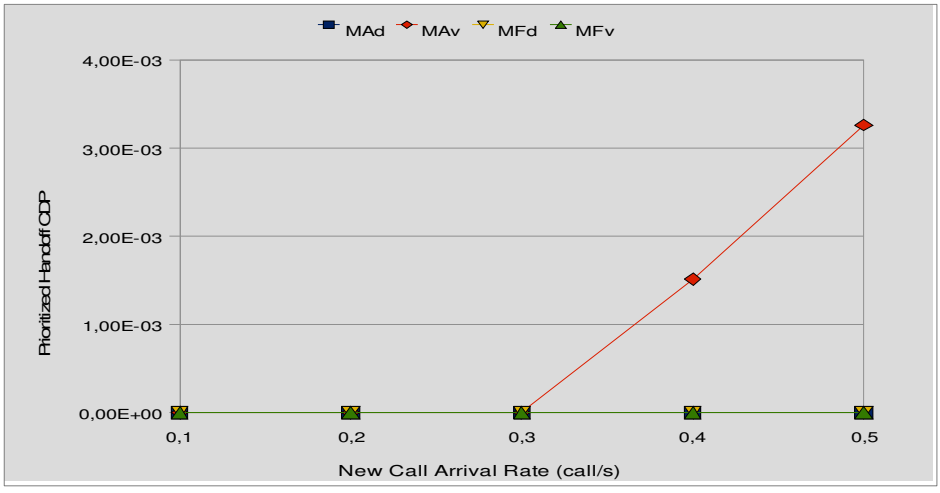

Figura 4.14. Prioritized CDP vs New Call arrival rate con λ_h=0.2 call/s.

Lo schema ATTSM, rispetto allo schema FTTSM, apporta dei miglioramenti significativi in termini di CBP e CDP delle chiamate non-priorizzate, al costo di un leggero aumento della probabilità di dropping delle chiamate di handoff priorizzate.
Nelle Figure 4.15 e 4.16 viene mostrato l'utilizzo di banda dello schema in esame, e vengono introdotte due nuove variabili: MA, che indica l'utilizzo totale di banda dello ATTSM, e MF, che ne indica quello dello FTTSM.

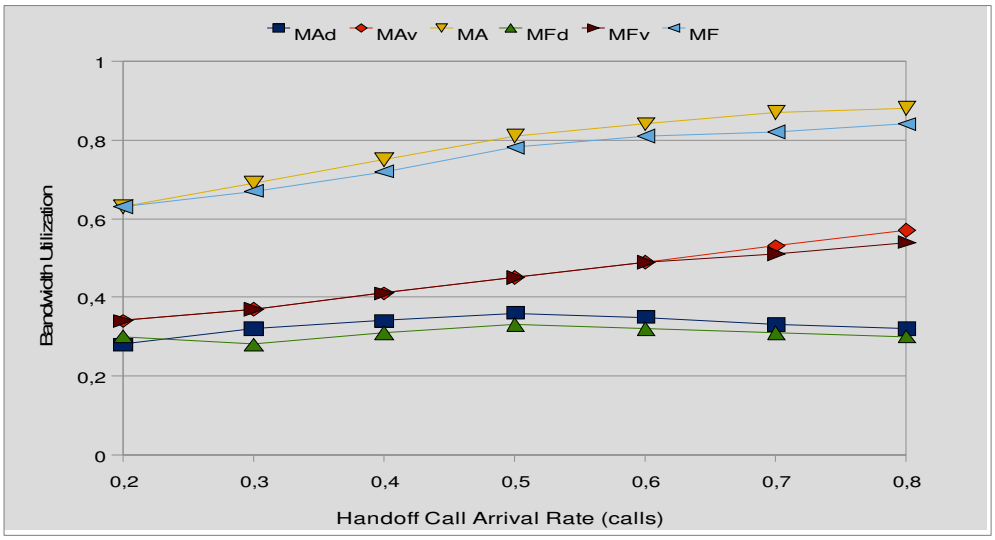

Figura 4.15 Utilizzo di banda vs Handoff arrival rate con λ_n=0.2 call/s

Con riferimento alla Figura 4.15, considerando la banda totale e quella del traffico dati dello schema ATTSM si ha un miglioramento, in termini di percentuale, rispettivamente di 4.88% e 9.68% rispetto allo schema FTTSM, mentre, considerando la Figura 4.16, che mette in relazione l'utilizzo di banda con il rate di arrivo delle nuove chiamate, si ha un miglioramento sulla banda totale del 8.24% e del 19.05% per il traffico dati.

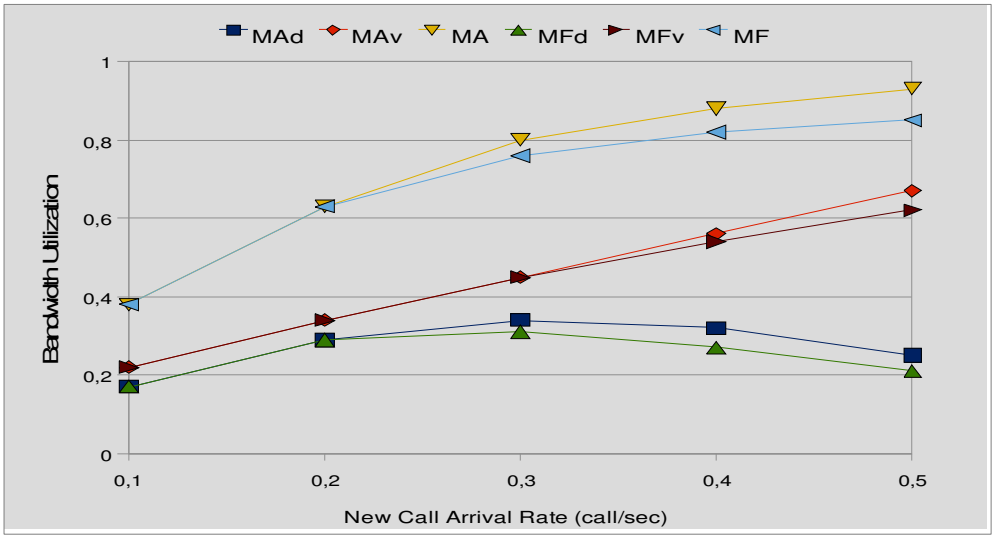

Figura 4.16. Utilizzo di banda vs New Call arrival rate con λ_h=0.2 call/s.

4.5 Conclusioni

In questo lavoro di tesi, sono stati esaminati alcuni algoritmi di allocazione dinamica della banda presenti nella letteratura più recente. La scelta non è dipesa soltanto dalla data di pubblicazione, ma ci si è soffermati anche sulla natura degli stessi, sulla strada che ogni autore ha deciso di intraprendere per risolvere l'annoso problema della gestione di banda nelle reti wireless. Il primo algoritmo esposto si basa su due meccanismi: il prestito di banda e lo scambio di connessione. Le chiamate sono raggruppate in tre diverse classi di traffico, e ad ogni classe è associata una funzione di utilità. Il secondo fa sempre uso delle tre classi di traffico sopracitate, ma opera su di esse una quantizzazione, in modo da ottenere una funzione continua a tratti. A fronte di ciò, questo algoritmo distribuisce la banda tra le chiamate multimediali in modo che ognuna di esse riceva pari utilità. Il terzo è uno schema di threshold dinamico, ovvero la quantità di banda allocata, descritta da una soglia variabile, ad un particolare tipo di chiamata dipende dal carico della cella che ospita la chiamata medesima. Il suo funzionamento è basato principalmente sul monito del tempo trascorso di ogni chiamata. L'obiettivo comune è quello di garantire la qualità del servizio agli utenti mobili, che nelle reti wireless può essere quantificata in termini

di parametri probabilistici di QoS, come il Call Blocking Probability, il Call Dropping Probability ed anche la quantità totale di banda che viene utilizzata, o meglio sfruttata, da un algoritmo. Come si può notare dai risultati, tutti gli algoritmi presentano valori di CBP superiori a quelli di CDP; di fatto, è preferibile non fare iniziare una connessione, quindi bloccare una nuova chiamata, piuttosto che eliminarne una proveniente da una cella limitrofa, quindi che ha già un suo trascorso alle spalle. Purtroppo, la sostanziale differenza nella struttura dei risultati simulativi, non permette un confronto approfondito tra gli algoritmi, specie tra gli ultimi due che sono, da un punto di vista tecnico, più rilevanti. Il modo migliore per confrontare i due algoritmi, sarebbe quello di creare una nuova simulazione sotto lo stesso contesto e le stesse condizioni; tenendo conto dei dati forniti dagli autori, si è comunque in grado di ottenere qualche risultato. In Figura 4.17 viene mostrato l'utilizzo della banda relativo agli algoritmi ATTSM, RBBS e Utility-Fair. Per lo schema ATTSM si è scelto di utilizzare il rate di arrivo delle chiamate di handoff, mantenendo costante quello delle nuove chiamate, pari a 0.2 call/s.

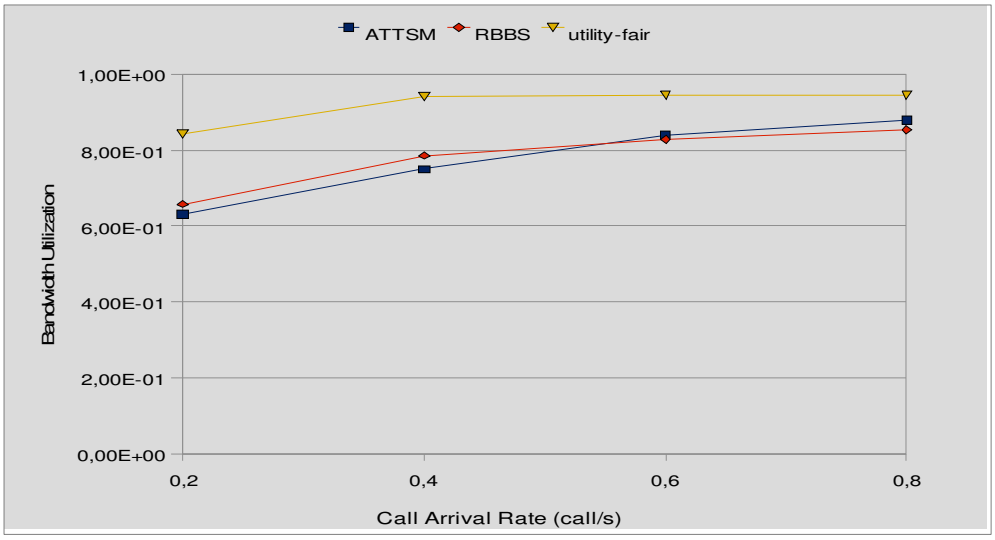

Figura 4.17. Confronto tra gli utilizzi di banda.

Sotto tali considerazioni, dalla figura è possibile notare che l'algoritmo Utility-Fair, descritto dalla linea gialla, è quello che sfrutta meglio la banda, mentre lo schema ATTSM supera lo RBBS dopo un certo rate. Tutti e tre gli schemi hanno comunque un utilizzo ottimale della banda, e lo scarto tra loro è relativamente piccolo.
Un altro confronto si può fare in termini di CBP, come mostrato in Figura 4.18.

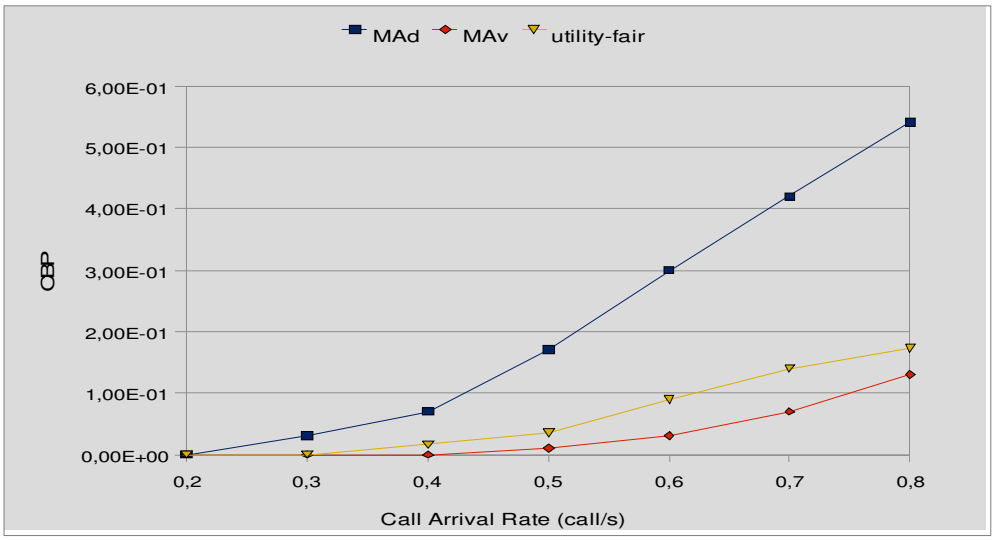

Figura 4.18. Confronto tra CBP.

Nel confronto tra le probabilità di blocco è necessario fare alcune considerazioni; lo schema ATTSM tiene conto, in maniera separata, delle chiamate voce e dati, mentre l'Utility-Fair fa riferimento a chiamate multimediali, a cui fanno parte chiamate voce e dati non distinte tra loro. Dalla Figura 4.18 si può notare che lo schema orientato all'utilità presenta un CDP minore se confrontato con il traffico dati dello ATTSM, ma è maggiore se confrontato con quello voce; è da tener comunque presente che nella simulazione dello ATTSM, le chiamate generate sono per il 70% voce e 30% dati. Un problema diverso si riscontra nel confronto tra le probabilità di dropping, in quanto, nonostante si adoperino le stesse metriche per la valutazione, lo schema ATTSM divide le chiamate di handoff in due categorie: chiamate priorizzate e non-priorizzate; nello schema Utility-Fair, invece, non è prevista nessuna distinzione tra gli handoff, in quanto si basa su principi diversi. Da un ipotetico confronto diretto potrebbero quindi affiorare risultati incongruenti. Osservando, però, i singoli grafici relativi ai CDP dei singoli algoritmi, è facile verificare che entrambi mantengono la probabilità di dropping nell'ordine dello n=10^-3, quindi relativamente prossimi allo zero. Lasciando da parte le metriche classiche di valutazione, un altro criterio di confronto può essere espresso in termini di estendibilità e di adattabilità, tenendo anche conto delle oramai prossime reti di quarta generazione e delle nuove applicazioni che queste potranno supportare. Grazie all'uso della funzione di utilità ed alla sua quantizzazione, lo schema Utility-Fair è quello che sembra essere meno

rigido, cioè più aperto ad operare su nuovi generi di applicazioni; in più, il modo in cui viene ricercato il livello di utilità tra le chiamate attive, segue un processo iterativo, che probabilmente può essere migliorato in termini di efficienza computazionale.

BIBLIOGRAFIA

[1] J.E. Padgett, C.G. Gunther, and T. Hattori, "Overview of Wireless Personal Communications," IEEE Communications Magazine, vol. 33, no. 1, pp. 28-41, Jan. 1995.

[2] L.W.Couch, "Digital and Analog Communication System, 6th ed.", Prentice Hall, Inc.

[3] A.-L. I. Semeia, "Wireless Network Performance Analysis for Adaptive Bandwidth Resource Allocations". PhD Thesis, Stevens Institute of Technology, USA, 2003.

[4] J. Hamalainen, "Design of GSM High Speed Data Service", Doctor of Technology Thesis, Tampere University of Technology, Oct.1996.

[5] N.R.Prasad, "An Overview of General Packet Radio Services (GPRS)", Proc. of WPMC '98, Yokosuma, Japan, Nov. 4-6, 1998.

[6] N.R. Prasad, "GSM Evolution Towards Third Generation UMTS/IMT2000," Proceedings of the IEEE International Conference on Personal Wireless Communication, pp.50 - 54, Feb. 1999.

[7] ITU, "International Mobile Telecommunications-2000 (IMT-2000)," http://www.itu.int/home/imt.html, 2000 (Internet).

[8] N. Padovan, M. Ryan, and L. Godara, "An Overview of Third Generation Mobile Communications Systems: IMT-2000," the IEEE Region 10 International Conference on Global Connectivity in Energy, Computer, Communication and Control (TENCON '98), vol. 2, pp. 360-364, Dec. 1998.

[9] T. Ojanpera and R. Prasad, "An Overview of Third-Generation Wireless Personal Communications: A European Perspective," IEEE Personal Communications, vol. 5, no. 6, pp. 59-65, Dec. 1998.

[10] Prokkola, P.H.J Perala, M. Hanski, E. Piri, "3G/HSPA Performance in live Networks from the End User Perspective",IEEE

[11] "About HSPA", http://hspa.gsmworld.com/abouthspa (Internet).

[12] D. Astely, E. Dahaliman, A. Furuscar, Y. Jading, M. Lindstram, S. Parvall, "LTE: the evolution of mobile broadband-[LTE part II: 3GPP release 8]", Communications Magazine, IEEE,vol.47, issue 4, pp. 44-51, April 2009.

[13] IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999), "PART 11: Wireless LAN Medium Access Control (MAC) and Phisical Layer (PHY) Specifications", 12 Juna 2007.

[14] A. Malla, M. El-Kadi, S. Olariu, P. Todorova, "A Fair Resource Allocation Protocol for Multimedial Wireless Networks", IEEE Transaction on Parallel and Distribuited Systems, vol. 14, no. 1, pp. 63-72, Jan. 2003.

[15] J.E. Padgett, C.G. Gunther, and T. Hattori, "Overview of Wireless Personal Communications," IEEE Communications Magazine, vol. 33, no. 1, pp. 28-41, Jan. 1995.

[16] Y-B. Lin, I. Chilamtac, "Wireless and Mobile Network Architectures", Wiley, 2000.

[17] A.B.Lashkari, M.M.M.Danesh, B.Samadi, "A Survey on wireless security Protocols (WEP,

WPA and WPA2/802.11i)", IEEE 2009.

[18] N. Banerjec, K. Basu, S.K. Das, "Adaptive Resource Management for Multimedia Applications in Wireless Networks", Proceedings of the 6th IEEEInternational Symposium on a W.o.W.M.o.M '05, pp. 250-257, June 2005.

[19] M.H.Ahmed, "Call Admission Control in Wireless Networks: A Comprehnsive Survey", IEEE Communications Surveys, first quarter 2005, vol. 7, no.1.

[20] S.Jagannathan, A. Tohmaz, A. Chronopoulus, H.G. Cheung, "Adaptive Admission Control of Multimedia Traffic in High-Speed networks", Proceedings of the IEEE International Symposium on Intelligent Control, pp. 728- 733, Oct. 2002.

[21] A. M. Rashwan, A-EM: Taha, H. S. Hassanein, "Considerations for bandwidth Adapatation Mechanisms in Wireless Networks", IEEE 2008.

[22] J. Elias, F. Martignon, A. Capone, G. Pujolle, "A New Approach to Dynamic Bandwidth Allocation in QoS Netwoks: Performance and Bounds", IEEE Computer Networks, vol. 51, pp. 2833-2835, 2007.

[23] J. Cai, "An Adaptive Time-Threshold Based Schema for Voice Handoff Calls in Cellular Wireless Networks", in 2007, Proc. WICOM Conference, pp. 1749-1752.

[24] S.S. Rappaport, "The Multiple-Call Hand-off Problem in High Capacity Cellular Communications System", IEEE Transaction on Vehicular Technology, vol. 40, no.3, pp. 546-557, Aug.1991.

[25] S. Chai, K.G. Shin, "Adaptive Bandwidth reservation and Admission Control in QoS-Sensitive Cellular Networks", IEEE Transaction on Parallel and Distribuited Systems, vol. 13. no. 9, pp. 882-897, Sept. 2002.

[26] L. Huang, S. Kumar, C.-C.S. Kou, "Adaptive Resource Allocation for Multimedia QoS Management in Wireless Networks", IEEE Transaction on Vehicular Technology, vol. 53, no. 2, pp. 547-558, March 2004.

[27] R.R-F Liao, A.T. Compbell, "A Utility-Based Approach for Quantitative Adaptation in Wireless Packet Networks",. Wireless Networks, vol. 7, no. 5, pp 541-557, Sept. 2001.

[28] M.Xiao, N.B. Shrott, Z.K.P.Chong, "A Utility-Based Power Control Schema in Wireless Cellular Systems" IEEE Trans. New. Vol. 11, no.2, 2002.

[29] W.H. Kou, W. Liao, "Utility-Based Radio Resource Allocation for QoS Traffic in Wireless Networks", IEEE Trans. On Wireless Communications, vol. 7, no. 7, July 2008.

[30] N.Lu, J.Bigham, "QoS Provisiong in Wireless Multimedia: Utility-Maximization Bandwidth Adaptation for Multiclass Traffic QoS Provisionig in Wireless Networks", Prooc.of 1st ACM International Workshop in QoS and Security in Wireless and Mobile Networks, 2005.

[31] T.Nyandeni, M.Adigun, J. Emuoyibofarhe, S.Sibiya, "A Bandwidth Allocation Schema for Mobile Wireless IP Networks", IEEE 2007.

[32] M.Nkambule, G.Ojong, M.Adigun, "Predicting Mobility on the Wireless Internet using Sectored-Cell Approach",Prooc. SATNAC 2004.

[33] A.Malla, M.El-Kadi, S.Olariu,P. Todorova, "A Fair Resource Allocation Protocol for Multimedia Wireless Networks", IEEE Transaction on Parallel and Distribuited Systems, vol.14, no.1, pp. 63-71, Jan 2003.

[34] N.Lu, J.Bigham, "On Utility-Fair Bandwidth Adaptation for Multimedia Wireless Networks", IEEE 2006.

[35] C.Curescu, S.Nadjin-Tehrani, "Time-Aware Utility-Based Resource Allocation in Wireless Networks", IEEE Transactions on Parallel and Distribuited Systems, vol.16, no.7, pp.624-636, July 2005.

[36] D.Hong, S.Rappaport, "Traffic Model and Performance Analisys for Cellular Mobile Radio Telephone Systems with Prioritizede and Non prioritized Handoff Procedures", IEEE Trans. Vehicular Tech. vol VT 52, 1986, pp 77-92.

[37] I.Candan, M. Salamah, "A Time-Threshold Based Multi-Guard Bandwidth Allocation Schema for Cellular Networks", Proc. Of the 7th IEEE International Conference on Computer Networks, ISCN 2006, Turkey.

[38] I. Candan, M. Salamah, "Dinamic Time-Threshold Based Schema for Voice Calls in Cellular Networks", Prooc. New 2 AN Conference, pp. 189-199, 2006.

[39] J. Cai "An Adaptive Time-Threshold-Based Schema for Multimedial Calls in Wireless Networks",IEEE 2008.

[40] M.El-Kadi, S.Olariu, H. Abdel-Wahab, "A Rate-Based Borrowing Schema for QoS Provisioning in Multimedia Wireless Networks", IEEE Trans. on Parallel and Distr. Systems, vol. 13, pp.156-167, Feb. 2002.

[41] I.Karzela, Modeling and simulating communication networks: a hands-on approach using OPNET, Prentice Hall, New Jersey, August 1998.

[42] M.salamah, I. Candan, "A Fair Bandwidth Allocation Schema for Multimedia Handoff Calls in Cellular Networks", New Trends in Computer Networks, pp. 40-49, ISBN. 1-86094-611-9, Imperial College Press, 2005.

ALTRE FONTI:

- B.Fagarazzi, R.Zane, "Telecomunicazione e Telematica vol. 1-2", Edizioni Calderini, Gennaio 1998.
- Bates Regis J.,"Manuale di Telecomunicazioni a banda larga", McGraw-Hill Companies, 2000.
- http://www.gsmworld.com (Internet)
- http://www.comlab.uniroma3.it/ (Internet)
- http://home.dei.polimi.it/capone (Internet)
- http://www2.polito.it/didattica (Internet)
- http://www.pcpedia.it
- http://www.wardrive.net