

Introduzione

Il lavoro di tesi, qui presentato, si occupa fundamentalmente della gestione delle risorse condivise in ambiente wireless, mediante un nuovo algoritmo che tiene conto sia della mobilità degli utenti, sia della qualità dei links, tramite i quali viaggiano i segnali trasmessi; si presenta anche una tecnica di predittività, che cerca di gestire in maniera più intelligente la prenotazione delle risorse passive.

La rapida e sempre crescente evoluzione delle attività umane ha reso l'informazione uno dei beni più importanti in termini economici e sociali. L'aumento della richiesta di telecomunicazioni ha portato da una parte allo sviluppo vertiginoso di *Internet* e dall'altra alla nascita di nuove applicazioni, come ad esempio le videoconferenze, la telemedicina e il telelavoro.

E' ben nota, ormai, la costante crescita del mercato dei dispositivi per comunicazioni *wireless* (senza filo). Fino a qualche tempo fa la gran parte di essi era dedicata a servizi vocali, mentre ora tutti i terminali si stanno orientando verso il trasporto di dati, immagini e persino video, al fine di permettere una sempre più piena connettività ad *Internet*.

I dispositivi wireless consentono una completa mobilità da parte degli utenti all'interno di determinate aree di copertura, che possono essere anche molto ampie. Tale possibilità, insieme alla comodità di non essere collegati fisicamente ad una rete di telecomunicazione, sta portando questi dispositivi ad una crescita costante, molto insidiosa per il primato dei dispositivi wired (secondo molti esperti, fra non molti anni, esisteranno esclusivamente dispositivi senza filo).

Una rete di telecomunicazioni può essere vista come una struttura in grado di instaurare, su richiesta di servizio da parte dell'utente, un collegamento per la comunicazione con un altro utente o gruppo di utenti. Le principali funzioni di una rete di telecomunicazioni sono: scegliere le opportune risorse per fare comunicare i due utenti, trasmettere effettivamente l'informazione, trasferire all'interno della rete le

segnalazioni opportune per il funzionamento e amministrare le risorse di rete per permetterne l'effettivo utilizzo e la manutenzione.

La fornitura di un servizio da parte di una rete di telecomunicazioni avviene attraverso il dialogo e la cooperazione fra varie entità, residenti su macchine differenti e collegate l'una all'altra attraverso i mezzi fisici di comunicazione. In qualsiasi attività di dialogo, ogni entità si coordina con le altre mediante *protocolli di comunicazione*, generalmente basati sul concetto di *strato*: l'insieme delle funzioni viene suddiviso in una pila di strati sovrapposti, ognuno dei quali è preposto a fornire un sottoinsieme delle funzionalità necessarie alla comunicazione con un altro sistema; il più noto, fra tutti i modelli di riferimento, è il modello ISO-OSI (International Standard Organization - Open System Interconnection) che si basa su una pila a sette strati.

Esistono due differenti tipologie di reti LAN (Local Area Network): quelle di tipo cablato (wired-LAN) e quelle che usano le onde radio come mezzo trasmissivo (wireless-LAN).

Le wireless-LAN (WLAN) trasmettono e ricevono dati via etere, minimizzando il bisogno di collegamenti cablati, combinando, così, la connettività e la mobilità degli utenti, i quali possono accedere ad informazioni condivise, senza cercare un luogo per collegarsi in rete; gli amministratori del sistema, inoltre, possono implementare o ingrandire una WLAN senza installare o trasportare fili (tali opportunità di mobilità e di servizio non sono possibili con reti cablate). Altri vantaggi possono essere la velocità di installazione, l'elevata scalabilità e i costi di gestione ridotti. Uno dei principali svantaggi di questo tipo di reti è l'utilizzo di onde radio, come mezzo trasmissivo delle informazioni, che risente dei disturbi elettromagnetici (interferenze varie) e della presenza di ostacoli composti da materia non perfettamente penetrabile dalle onde radio. Un altro svantaggio, dovuto all'uso delle onde radio, riguarda tutte le problematiche relative alla limitatezza di banda e al riuso delle frequenze.

Per il traffico real-time su Internet viene utilizzato, come protocollo di prenotazione di risorse, il ReSerVation Protocol, noto come RSVP, la cui funzione principale è quella di trasportare le informazioni necessarie all'instaurarsi nei singoli nodi, lungo il percorso, del *reservation-state*, ossia alla prenotazione di risorse da parte del singolo flusso.

La mobilità degli utenti ha un impatto notevole sui parametri di QoS (Quality of Service) delle applicazioni real-time. L'architettura esistente per i servizi *real-time* su una rete con hosts fissi risulta, quindi, inadeguata per supportare utenti mobili. Infatti, per effettuare la prenotazione delle risorse e fornire servizi real-time ad un utente mobile, sono necessarie alcune caratteristiche addizionali, che non sono presenti nel protocollo RSVP. Ad esempio, per ridurre al minimo gli effetti della mobilità, è necessario richiedere una prenotazione delle risorse su tutte le locazioni (celle) che l'utente mobile andrà a visitare durante la sua connessione (si parla di politica delle *prenotazioni passive*). Quando un host si muove tra le celle con una connessione aperta, il ritardo di propagazione dei pacchetti può cambiare (la congestione nei routers lungo il nuovo percorso potrebbe essere diversa rispetto al precedente); se il nuovo nodo verso il quale l'host si è spostato risulta sovraccarico, la banda disponibile nella nuova locazione potrebbe non essere sufficiente per soddisfare la richiesta, quindi l'utente mobile, che si muove con una connessione attiva, potrebbe essere costretto ad "adattare" le sue richieste di QoS. Inoltre, durante l'hand-off, l'utente mobile potrebbe avere dei disturbi temporanei sul servizio (addirittura in alcuni casi la connessione potrebbe essere interrotta). Per evitare questi inconvenienti, nell'ammettere un nuovo flusso, è necessario eseguire il controllo di ammissione su tutti i routers e gli access-points (nodi di copertura) lungo il percorso, che va dalla sorgente alle locazioni, che l'utente mobile potrà visitare durante la durata della connessione. Un flusso è ammesso nella rete solo se supera tutti i controlli di ammissione dove il flusso stesso ha fatto richiesta di prenotazione di risorse.

Tali operazioni sono rese possibili tramite il protocollo MRSVP (Mobile ReSerVation Protocol), il quale nasce come estensione dell'RSVP per il supporto della mobilità degli hosts (soprattutto nelle reti ISPN – Integrated Services Packet Networks): si utilizzano due tipi di prenotazioni, quella *attiva* (effettuata tra il sender e la cella corrente dell'host mobile) e quella *passiva* (effettuata tra il sender e le celle future presenti nelle specifiche di mobilità dell'host mobile). In tal modo si riesce a riservare in anticipo la giusta quantità di banda necessaria al mantenimento del flusso in seguito ad un evento di hand-off.

Un'altra estensione del protocollo RSVP è il DRSVP (Dynamic ReSerVation Protocol), in grado di fornire QoS dinamica in una rete a banda variabile, in cui

esistono link wireless (le cui prestazioni in termini di probabilità di errore variano nel tempo) e nodi intermedi mobili (per cui la topologia della rete varia, modificando la qualità del cammino esistente tra due generici punti della rete); di volta in volta, il protocollo limita la quantità di banda assegnata ai flussi, eseguendo un'analisi accurata di tutti i *bottlenecks* presenti nella rete: non ha senso riservare molta banda per un flusso in alcuni nodi, se in altri punti della rete, a causa della presenza di “strozzature”, non potrà essere possibile utilizzarla; in tal modo si evitano gli sprechi di risorse.

Come accennato, ai numerosi vantaggi offerti dalle comunicazioni wireless (portabilità, alta connettività e bassi costi) si affiancano delle problematiche, che non bisogna per niente trascurare, legate soprattutto alla propagazione del segnale (in termini di degradazione e interferenze). Uno dei più importanti fenomeni nella trasmissione di segnali in sistemi radiomobili digitali è costituito dall'attenuazione ambientale, derivante sia dalla distanza tra ricevitore e trasmettitore, sia dalla presenza di ostacoli lungo il link che, bloccando il segnale, creano delle distorsioni e delle vere e proprie zone d'ombra per l'utente (si pensi agli edifici, alle colline e agli altri veicoli che s'incontrano durante una comunicazione mobile). Se consideriamo la propagazione esclusivamente nello spazio aperto, l'attenuazione del segnale è data principalmente dal *path-loss*, mentre nelle zone urbane e rurali si verificano ulteriori fenomeni, dovuti all'atmosfera e alla presenza di ostacoli lungo il cammino: dispersione, trasmissione, rifrazione, shadowing e assorbimento atmosferico. A questi si deve aggiungere il fenomeno della *riflessione* del fascio elettromagnetico, che genera percorsi aggiuntivi lungo i quali il segnale si propaga (multipath fading): il fascio è soggetto ad una diffusione ritardata, ovvero le singole componenti riflesse giungono a destinazione ad istanti diversi, creando una vera e propria frantumazione del segnale. Mentre nelle comunicazioni satellitari e nelle trasmissioni a microonde punto-a-punto tale fenomeno è trascurabile, nella telefonia mobile vi è una grande quantità di ostacoli (quindi di riflessioni), per cui il ricevitore rileverà più copie del segnale, che si presentano a ritardi variabili. Il termine *evanescenza* (o fading) fa riferimento alla variazione temporale della potenza del segnale ricevuto provocata da cambiamenti nel mezzo di trasmissione o nei percorsi. In un ambiente fisso il fading è influenzato dalle variazioni nelle condizioni atmosferiche (per esempio dalle piogge); in un ambiente mobile, dove un'antenna è in costante movimento rispetto all'altra, la posizione

relativa dei vari ostacoli cambia nel corso del tempo, creando effetti di trasmissione complessi. Provando a trasmettere un segnale digitale in due istanti temporali diversi e ben separati, i due segnali ricevuti saranno differenti, anche se si usa lo stesso canale di comunicazione: ciò è dovuto ai continui cambiamenti delle caratteristiche fisiche del mezzo, che avvengono in maniera del tutto casuale (si dice perciò che il link è dotato di una risposta all'impulso random e tempo-variante). Infatti, se effettuiamo l'esperimento con un segnale di durata molto piccola (idealmente un impulso) attraverso un canale soggetto a multipath fading, il segnale ricevuto appare come un treno d'impulsi, la cui forma varia ad ogni tentativo (numero, ampiezza e ritardo relativo degli impulsi ricevuti non sono mai identici nelle diverse prove).

L'unico modo di descrivere il link, allora, è quello di ricorrere ad una trattazione statistica, che consideri la natura tempo-variante del canale stesso; in particolare il nostro studio si è basato su un particolare processo stocastico, quale quello di *Markov*, che si presta bene a descrivere l'andamento del canale nel tempo (in termini di errori di trasmissione e degradazione). I modelli di fading più utilizzati sono due: quello di *Rayleigh* (spesso usato quando sono presenti più percorsi indiretti tra il trasmettitore e il ricevitore e non esiste un percorso dominante, come per esempio il percorso ottico diretto) e quello di *Rice* (che caratterizza meglio una situazione in cui vi è un percorso diretto più una serie di segnali indiretti multipath). Il modello di Rice viene frequentemente applicato per gli ambienti interni, mentre il modello di Rayleigh caratterizza principalmente gli ambienti esterni.

Per legare il processo stocastico al modello di canale occorre partizionare i possibili valori del rapporto segnale-rumore ricevuto in un numero finito d'intervalli, associandone uno a ciascuno stato della catena e ottenendone le grandezze caratteristiche (in questo modo si tiene conto delle caratteristiche fisiche e del comportamento tempo-variante del canale, per il quale assumiamo che l'involuppo della funzione di trasferimento sia distribuito secondo Rayleigh).

Nello specifico, il modello adottato nel nostro lavoro è stato particolarizzato per lo standard IEEE 802.11b, basato su una frequenza di lavoro di 2.4GHz, un data rate di 5.5Mbps (o 11Mbps) e una modulazione di tipo CCK (*Complementary Code Keying*); la catena di Markov relativa al link prevede 4 stati, ognuno dei quali rappresenta un intervallo di rapporto S/N (Signal to Noise ratio) e ne riassume le caratteristiche di

BER; l'evanescenza descritta dal processo markoviano è di tipo *slow*, ovvero l'evoluzione della catena può essere considerata "lenta" rispetto alla trasmissione dei pacchetti.

Nella trasmissione dei dati in ambiente wireless, quindi, conviene mettere da parte il concetto di "canale ideale" e tener conto delle perdite subite anche in fase di assegnazione di banda: come avviene nei sistemi ISPN, quando un utente richiede un servizio in cui si specifica il livello minimo di banda voluto, la rete deve garantire tale soglia e, nel caso di links soggetti a degradazioni, il sistema deve allocare una quantità maggiore di risorse, per rispettare la quantità minima *effettiva* di banda ricevuta dal flusso. Questo ragionamento dev'essere valido per tutti gli utenti mobili presenti all'interno di una cella e si mette in atto mediante le politiche di C.A.C. (Call Admission Control) e Rate-Adaptation (riadattamento di banda); senza queste tecniche di gestione delle richieste, il sovraccarico delle celle potrebbe divenire incontrollabile e, come risultato, si potrebbe avere una degradazione della QoS offerta, che porterebbe ad una violazione dei requisiti minimi richiesti.

L'obiettivo principale delle politiche di C.A.C. è quello di gestire il numero di utenti presenti in una data cella, garantendo allo stesso tempo un'alta utilizzazione del sistema e un basso livello di congestione; per poter decidere se accettare o meno un nuovo utente, la rete valuta le caratteristiche statistiche del traffico che l'utente intende trattare (velocità, bit-rate, ecc.): un flusso viene ammesso nella rete solo se supera il controllo di ammissione su ognuno degli *switch* o *base station* sui quali aveva fatto richiesta di prenotazione di risorse. Quindi l'utente si presenta alla rete comunicando le proprie caratteristiche statistiche e i propri parametri di qualità di servizio, dopodiché il sistema calcola la QoS che può offrire (sempre rispettando i vincoli accordati con gli altri utenti sotto servizio) e la confronta con quella richiesta: se è sufficiente decide di accettare l'utente.

Le politiche di riallocazione di banda, invece, sono rivolte a cambiare dinamicamente il rate offerto ai singoli flussi, rispettando i bounds richiesti dagli utenti (come il *packet-delay* e il *delay-jitter*) e mettendo in pratica i principi di *fairness* (imparzialità tra i flussi) e di minimo *overhead* (minimo sovraccarico protocollare). L'architettura della rete considerata nel nostro lavoro è di tipo cellulare, per cui un'area geografica viene divisa in un certo numero di regioni, dette celle. Ognuna di esse è coperta da una *base-station*

(detta anche access-point), che si preoccupa di garantire agli utenti mobili l'accesso alla rete in maniera wireless. Inoltre, ipotizziamo che la banda disponibile sui links wired (necessari per collegare fra loro i diversi access-points) sia molto più elevata della banda disponibile sui links wireless. Così facendo quando un elevato numero di utenti si concentra in una zona geografica, solo la relativa base-station viene considerata sovraccarica; il link wired ha quindi la capacità di soddisfare un numero arbitrario di richieste.

In questo scenario, i riadattamenti di banda si rendono necessari quando si verificano gli eventi di *hand-in* (un flusso entra in una cella a seguito di un hand-off o dopo aver superato il controllo di ammissione) o *hand-out* (un flusso lascia la cella corrente); nel primo caso, se la cella risulta altamente utilizzata, è necessaria un'operazione di *degrade*, con cui si abbassa il livello di rate offerto agli altri flussi presenti; nel secondo caso la banda rilasciata è ridistribuita mediante un'operazione di *upgrade*, con cui si aumenta il livello di rate offerto agli altri flussi presenti. Nelle nostre applicazioni è prevista la possibilità di far lavorare i flussi su diversi livelli di banda (4 nello specifico) e le operazioni di upgrade/degrade consistono nell'incrementare o diminuire il rate assegnato a passi costanti.

Diversi sono gli algoritmi proposti in letteratura per risolvere queste problematiche: alcuni si basano su stime statistiche di alcuni parametri di QoS, come il *DR* (Degradation Ratio, ovvero il rapporto fra il tempo durante il quale una connessione riceve servizio degradato e il tempo di occupazione totale di una cella) e la *UDF* (Upgrade/Degrade Frequency, ovvero la frequenza di scambio tra servizio pieno e degradato di una connessione ammessa), derivati da attente analisi dell'evoluzione del sistema mediante catene di Markov; altri mirano ad aumentare l'utilizzazione del sistema definendo e fissando delle *Thresholds* (soglie di accesso), ovvero offrendo differenti gradi di ammissione delle richieste, basati sulla situazione corrente di occupazione delle risorse disponibili e sulla priorità dei flussi in arrivo.

Altre tipologie di algoritmi sono basati sulle stime dinamiche di alcuni parametri (utilizzazione, numero di utenti presenti) e seguono la filosofia secondo la quale conviene non fissare a priori i valori su cui ci si basa per operare l'ammissione di chiamata, bensì analizzare dinamicamente il comportamento del traffico, evitando così di sovrastimare le richieste di banda e ottenere una scarsa utilizzazione della rete. Per

effettuare le valutazioni circa lo stato del sistema in maniera dinamica si usano le *time-windows* (finestre temporali), ovvero dei periodi di campionamento statici o dinamici, all'interno dei quali si effettua una valutazione dei parametri d'interesse.

L'algoritmo proposto nella tesi riguarda le problematiche appena descritte e gestisce sia la "non idealità" del link fisico, sia la mobilità degli utenti all'interno delle WLAN: ha come obiettivo fondamentale quello di garantire agli utenti la QoS richiesta anche in presenza di canali tempo-varianti e aggiunge alle tecniche di correzione degli errori di basso livello alcuni meccanismi che, già ad alto livello, mirano a rendere minimi gli effetti del cambiamento di qualità del link durante la connessione. Inoltre tale schema, oltre a fornire garanzie sulla QoS, assicura una certa imparzialità a lungo termine e un altissimo utilizzo del sistema. Per la gestione della mobilità degli hosts, l'algoritmo è stato integrato con il protocollo MRSVP.

L'algoritmo è detto "*orientato all'utilità*" poiché è definita una funzione di utilità strettamente legata alla banda ricevuta, che dà una misura del livello di soddisfacimento ottenuto dall'utente per il servizio offertogli (uno degli obiettivi dell'algoritmo è quello di massimizzare tale funzione, adattandosi dinamicamente all'evoluzione del canale fisico). Il vantaggio principale della funzione d'utilità è che riflette in maniera implicita le richieste di QoS dell'utente e quantifica l'adattabilità di un utente o di un'applicazione.

Per gestire la variabilità del link fisico, si usano generalmente meccanismi di basso livello, come i codici di correzione d'errore e le tecniche di swap nello scheduling dei pacchetti; ma tali meccanismi appaiono inadeguati per variazioni lente del link (*slow fading*): nella codifica adattativa, un utente con una qualità di link davvero pessima potrebbe sprecare un grande quantitativo di banda per l'overhead di codifica durante tutto il tempo di permanenza del link nello stato degradato. In presenza di *slow-fading*, si preferisce allora, oltre che cambiare la lunghezza dei codici, variare anche l'ammontare di banda riservata ogni qualvolta il link cambia stato, permettendo all'utente di ricevere ancora il servizio quando il canale subisce una degradazione; così facendo, l'intera banda del sistema sarà meglio utilizzata, poiché la si alloca a quegli utenti che godono di una degradazione più bassa. Quindi si rende necessario un meccanismo che, ad alto livello, gestisca la banda assegnata ad ogni utente in modo da minimizzare gli sprechi al variare della qualità dei links.

Ogni link wireless viene modellato da una catena di Markov e assumiamo che tutti gli utenti si muovono liberamente nella stessa regione, per cui i links risultano identici e indipendenti.

Per catturare le caratteristiche di ogni stato della catena, associamo ad ogni stato m un parametro detto *rapporto di degradazione di banda* D_m (Bandwidth Degradation Ratio), che rappresenta lo spreco totale di banda causato da trasmissioni non andate a buon fine, overhead di codifica e altri fattori; inoltre $0 \leq D_m < 1$. In pratica, se la banda allocata all'utente è r e il suo link si trova nello stato m , allora $D_m * r$ rappresenta lo spreco di banda e $(1 - D_m) * r$ è la quantità di banda effettiva che l'utente riceve.

Mediante l'algoritmo proposto, la rete prova, dinamicamente, ad allocare una quantità di banda tale che l'utilità istantanea di ogni utente sia mantenuta al di sopra di un *lower-bound* $u_{i,\min}$ e, a lungo termine, sia garantita una certa imparzialità ed un'utilizzazione efficiente delle risorse del sistema. Quindi, gli obiettivi che l'algoritmo si prefigge sono:

- a) *requisiti di QoS*: si deve tenere conto dell'evento di *utility-outage*, cioè dello sconfinamento del lower-bound, che si verifica quando l'utilità istantanea dell'utente i -esimo scende al di sotto di $u_{i,\min}$ (si pensi, per esempio, a quando un utente, che gode di ottima qualità di link e che sta ricevendo un'allocazione di banda pari al minimo, subisce una degradazione del canale, per cui la banda effettiva ricevuta risulterà inferiore ad $r_{i,\min}$ e l'utilità istantanea si annulla, per come è definita la generica funzione ad essa associata); l'algoritmo deve garantire che la probabilità di *outage* per ogni utente si mantenga sotto una certa soglia p_{outage} ;
- b) *imparzialità*: per garantire una certa *fairness* tra i flussi ci si basa ancora una volta sul concetto di utilità; per tutti gli utenti presenti si può definire una grandezza detta *gap-normalizzato*, che tiene conto, per ciascun flusso, della differenza fra l'utilità media ricevuta e il livello minimo previsto; l'algoritmo tende ad ottenere, a lungo termine, l'uguaglianza fra i gap di tutti gli utenti per raggiungere l'obiettivo di *fairness*;
- c) *alta utilizzazione del sistema*: l'utilità totale fornita è il criterio per misurare l'utilizzazione della banda disponibile.

L'idea fondamentale dello schema di rate-adaptation proposto è questa: quando un utente subisce una degradazione del link, potrebbe cedere una parte della propria

banda ad un altro utente con un gap-normalizzato più basso; quando un link “migliora”, un utente può ricevere più banda da qualcuno dei presenti, che abbia un gap-normalizzato più grande. In questo modo si ottiene un guadagno di utilità istantanea. Se assumiamo che il sistema conosca con molta accuratezza lo stato del link di ogni singolo utente, allora il meccanismo di riallocazione viene invocato in due casi: ogni qualvolta si verifica un cambiamento di stato di un link e/o quando un access-point osserva un’entrata o un’uscita di qualche utente (un flusso entra in una cella se supera il CAC, mentre esce se opera un hand-off, se subisce un dropping o se termina la connessione).

Risulta evidente che, per garantire il livello minimo di utilità a tutti gli utenti, l’algoritmo deve prevedere un meccanismo di controllo di ammissione per limitare il numero dei presenti nel sistema. Se la banda totale richiesta per mantenere tutti i flussi sopra il livello minimo di utilità eccede la banda totale disponibile allora ci sarà almeno un utente in outage e la probabilità p_0 che tale fenomeno si verifichi ad un qualsiasi istante è facilmente calcolabile, una volta stabilite le caratteristiche della catena di Markov di riferimento. E’ ovvio che risulta sempre $p_{i,0} \leq p_0$, se con $p_{i,0}$ indichiamo la probabilità di outage dell’utente i . Se riusciamo a garantire che $p_0 \leq p_{\text{outage}}$, dove p_{outage} è la soglia desiderata, sarà sempre vero che $p_{i,0} \leq p_{\text{outage}}$; quindi, al momento dell’arrivo di una nuova richiesta, il sistema calcola p_0 e se risulta $p_0 \leq p_{\text{outage}}$, allora il nuovo utente è ammesso nella rete, altrimenti la richiesta è rifiutata.

Dopo aver definito il comportamento generale dell’algoritmo, lo abbiamo particolarizzato per le reti ISPN, in cui sono previste diverse classi di servizio, fra cui quelle *predictive*; a tale classe appartengono gli utenti *MIP* (che richiedono un servizio indipendente dalla mobilità) e gli utenti *MDP* (il cui servizio può essere soggetto a dropping, derivanti dalla mobilità).

In particolare abbiamo previsto due vie per gestire le due classi di servizio:

- gestione *congiunta*: si stabilisce un’unica soglia per la probabilità di outage che sarà valida per entrambe le classi;
- gestione *disgiunta*: si stabilisce una soglia per gli utenti MIP e una soglia per gli utenti MDP, con la possibilità di scegliere due diversi gradi di ammissione.

Il protocollo MRSVP garantisce la continuità del servizio per i flussi MIP effettuando le prenotazioni passive su tutti gli access-points (AP) presenti nella rete; questo tipo di

gestione, però, è fonte d'inefficienza, in quanto si mantiene inutilizzata una certa quantità di banda finché l'utente non effettua un hand-in, con una conseguente commutazione di banda da passiva ad attiva. Il sotto-utilizzo del sistema si rende ancora più evidente nel caso in cui ai flussi MDP non sia consentito l'utilizzo della banda passiva. Di certo tale fenomeno non può essere eliminato (se si vuole garantire l'indipendenza dalla mobilità per i flussi MIP), ma abbiamo cercato di ridurlo effettuando le seguenti riflessioni:

- ha poco senso prenotare risorse passive in tutta la rete, in quanto esistono alcuni AP ai quali l'utente non richiederà mai servizio in quanto la propria velocità media potrebbe essere così moderata che non arriverà mai fisicamente nelle zone coperte da quegli AP; basandoci sul tempo medio di permanenza dell'utente nella cella (C.S.T. - *Cell Stay Time*), abbiamo constatato che i flussi MIP possono ottenere lo stesso trattamento (in termini di continuità del servizio), anche se il protocollo non occupa risorse passive in tutti i nodi della rete;
- per ottenere un'ulteriore riduzione di banda non utilizzata, si tiene conto della qualità prevista del link al momento di ogni hand-off futuro, evitando di effettuare prenotazioni passive che impieghino il livello massimo previsto di banda (in altri termini si effettuano delle stime che, a meno di una certa percentuale d'errore, indicano in anticipo quale sarà il livello minimo di banda richiesto per la continuità del servizio in fase di hand-off, sicché il protocollo riserverà tale quantità).

Al fine di stabilire in che modo le bande si distribuiscono nel tempo e valutare la degradazione del link agli istanti di hand-off, abbiamo fatto uso di un modello markoviano, estrapolato da alcune simulazioni effettuate; mediante tale modello, che consiste, come vedremo, in due catene di Markov incastrate, è possibile analizzare l'evoluzione dei livelli di rate assegnabili (quindi della catena relativa alla qualità del canale), in termini di probabilità di transizione da un livello all'altro e di tempi medi di soggiorno nei livelli.

Per valutare e confrontare le performance dell'algoritmo seguendo le diverse politiche di prenotazione e gestione di banda sono state eseguite diverse simulazioni, utilizzando un simulatore di rete da noi implementato in C++. Il modello di rete simulato prevede la presenza di 5 celle wireless, ognuna coperta da una base station, e un sender con il quale ognuna delle celle è connessa da un link di banda infinita.

Siamo partiti da una gestione elementare della banda, fino ad arrivare, per passi successivi, ad un miglioramento dell'utilizzazione del sistema, mediante l'introduzione delle tecniche di predittività sopra citate. Inizialmente abbiamo confrontato le due politiche di gestione delle classi (congiunta e disgiunta), osservando cosa accade se si permette o meno ai flussi MDP di riutilizzare la banda passiva; poi abbiamo valutato gli eventuali miglioramenti ottenuti mediante le stime sul C.S.T. e l'efficienza della tecnica di predittività basata sulla stima esatta dello stato del canale previsto agli istanti di hand-off dei flussi MIP.

In particolare, in assenza del riutilizzo della banda passiva, abbiamo notato che, per qualunque percentuale di traffico MIP/MDP, conviene una gestione disgiunta per i flussi MIP: nonostante siano pochi quelli che riescono ad entrare in rete, vengono privilegiati in termini di degradazioni, visto che, nella politica disgiunta, le numerose riallocazioni degli MDP non influenzano la banda attiva dei MIP. Per qualunque percentuale di traffico, la politica congiunta offre una minore possibilità ai MIP di utilizzare il sistema, contrariamente a quanto accade per gli MDP. Il sistema, comunque, viene maggiormente utilizzato tramite la politica congiunta (per qualunque percentuale di traffico), visto che il contributo dato dai MIP (anche se minore) risulta molto basso. La gestione congiunta limita il numero di MIP presenti in rete, ma ne aumenta il dropping, visto che i numerosi MDP presenti nel sistema possono influenzarne la banda, aumentando la possibilità di avere eventi di outage non colmabili. Quindi, la gestione disgiunta, proprio perché mantiene le due classi separate, per valori di soglia meno permissivi, tende a privilegiare i flussi MIP (in termini di banda allocata, di utilizzazione e di taglio), mentre mantiene bassa l'utilizzazione dei flussi MDP, quindi del sistema.

Anche quando si dà la possibilità ai flussi MDP di riutilizzare la banda, la politica congiunta, nei confronti dei flussi MIP, si comporta bene solo per bassi valori di soglia (nessun flusso tagliato per outage non colmato). Per gli MDP, la banda ricevuta risulta circa identica per alte percentuali di traffico, mentre migliora nel caso di politica congiunta per basse percentuali di traffico. Il sistema risulta egualmente utilizzato tramite la politica congiunta o disgiunta (per qualunque percentuale di traffico).

Effettuando le prenotazioni secondo la prima tecnica di predittività, ossia quella basata sulla stima del C.S.T., con o senza il riutilizzo della banda passiva, abbiamo

osservato dei miglioramenti per i flussi MIP: riescono ad entrare più facilmente nel sistema e ad ottenere un trattamento migliore in termini di rate. Per percentuali alte di traffico MDP, l'incremento dell'utilizzazione da parte dei MIP è maggiore nel caso di politica congiunta (si arriva anche al 70% circa per alti valori di soglia, contro il 37% della politica disgiunta), per il fatto che le risorse occupate dagli MDP sono contese anche dai MIP; per percentuali basse di traffico MDP, il miglioramento è simile per entrambe le politiche. Da questo punto di vista, la politica congiunta privilegia i flussi MIP, ma, per ciò che riguarda la percentuale di dropping degli stessi, conviene di certo una politica disgiunta, che garantisce l'assenza di tagli per qualunque valore di soglia imposto (anche la politica congiunta si dimostra accettabile, ma solo per valori di soglia bassi).

A prescindere dal tipo di politica adottato, dalle percentuali di traffico considerate e dalla possibilità o meno di utilizzare la banda passiva, abbiamo potuto constatare che l'introduzione di una certa predittività, basata sulla distribuzione del tempo medio di soggiorno nella cella, porta dei discreti miglioramenti, a conferma delle nostre aspettative: i flussi MIP possono partecipare adeguatamente all'occupazione della rete (in assenza di predittività, la loro utilizzazione si manteneva, nella maggior parte dei casi, inferiore al 10%), risultando privilegiati anche in termini di rate ricevute, che diventa confrontabile, se non superiore, a quello ricevuto dagli MDP; diminuiscono le risorse passive riservate che, in alcuni casi, potrebbero rimanere inutilizzate.

Infine, abbiamo ottenuto alcune curve riguardanti l'introduzione della seconda tecnica predittiva (basata sul modello di Markov ottenuto), riscontrando ulteriori miglioramenti: l'analisi della qualità del link di un flusso serve ad eliminare gli sprechi residui di risorse passive, soprattutto nei casi in cui sia vietato il riutilizzo della banda passiva da parte degli MDP. La stima del C.S.T. può portare degli incrementi che, come abbiamo avuto modo di apprezzare, arrivano anche al 60% (rispetto al caso in cui si prenotano risorse passive su tutti gli A.P. della rete), mentre la stima della degradazione del link, rispetto alla precedente tecnica, porta ulteriori miglioramenti, anche del 7%-8%. Combinando le due tecniche, cioè evitando le prenotazioni passive su tutti i nodi della rete e riservando una certa quantità di banda su ogni A.P. interessato (che non sia necessariamente il massimo), si ottiene un incremento nell'utilizzazione del sistema da parte dei flussi MIP che può arrivare anche al 70%

circa, in base al tipo di politica di gestione utilizzata, al possibile riutilizzo della banda passiva e alle percentuali di traffico scelte. In ogni caso, si dà la possibilità al traffico MIP di essere sottoposto ad un controllo di ammissione molto meno selettivo, potendo meglio competere con i flussi MDP all'occupazione del sistema.