

UNIVERSITÀ DEGLI STUDI DI PISA

Facoltà di Ingegneria

Corso di Laurea Specialistica in Ingegneria delle Telecomunicazioni



TESI DI LAUREA

Criteria distribuiti di allocazione delle risorse nelle reti wireless ad hoc

Relatori

Candidato

Prof. Marco Luise

Riccardo Nuti

Prof. Filippo Giannetti

Ing. Antonio D. Fittipaldi

Anno Accademico 2004 – 2005

Alla mia famiglia

INDICE

INTRODUZIONE	- 1 -
CAPITOLO 1 - WIRELESS AD HOC NETWORK.....	- 5 -
1.1 Caratteristiche generali.....	- 5 -
1.2 Applicazioni.....	- 7 -
1.2.1 Legami con reti <i>infrastructure-based</i>	- 8 -
1.2.2 Scenari d'uso.....	- 9 -
1.2.3 Vantaggi.....	- 11 -
1.2.4 Svantaggi e limiti	- 13 -
1.3 Sfide per la ricerca.....	- 15 -
1.3.1 Medium Access Control.....	- 15 -
1.3.2 Routing.....	- 15 -
1.3.3 Security	- 16 -
1.3.4 Valutazione delle prestazioni	- 18 -
1.3.5 Scalabilità.....	- 19 -
1.3.6 Power conservation.....	- 19 -
1.3.7 Hardware-Software	- 23 -
CAPITOLO 2 - MAC LAYER IN AD HOC NETWORK.....	- 25 -
2.1 Livello MAC nelle reti ad hoc.....	- 25 -
2.2 Tecniche di accesso multiplo in reti ad hoc.....	- 28 -
2.2.1 Tecniche Contention-Based	- 28 -
2.2.1.1 CSMA/CA dell'IEEE 802.11	- 28 -
2.2.1.2 Procedura di backoff.....	- 31 -
2.2.1.3 Svantaggi e limiti	- 32 -

2.2.1.4 Estensioni ed alternative.....	- 34 -
2.2.2 Tecniche Contention-Free.....	- 36 -
2.2.2.1 TDMA.....	- 37 -
2.2.2.2 CDMA.....	- 43 -
2.2.2.2.1 CDMA in ad hoc network.....	- 44 -
2.2.2.2.2 Near-far effect.....	- 47 -
2.2.2.2.3 Impatto dell'interferenza da accesso multiplo.....	- 48 -
2.2.2.2.4 Caratteristiche necessarie per l'applicazione di un protocollo MAC basato su CDMA.....	- 49 -
2.3 Power Control a livello fisico.....	- 50 -
2.3.1 Open Loop Power Control.....	- 50 -
2.3.2 Closed Loop Power Control.....	- 52 -
CAPITOLO 3 – GAME THEORY E POWER CONTROL.....	- 55 -
3.1 Radio Resource Management.....	- 56 -
3.1.1 Classificazione schemi RRM.....	- 57 -
3.2 Introduzione alla <i>game theory</i>.....	- 58 -
3.2.1 Concetti fondamentali di <i>game theory</i>	- 59 -
3.2.2 Classificazione dei giochi.....	- 60 -
3.2.3 Concetto di soluzione.....	- 60 -
3.2.4 Equilibrio di Nash.....	- 61 -
3.2.5 Efficienza Paretiana.....	- 62 -
3.3 Applicazione della teoria dei giochi allo studio di una rete wireless.....	- 63 -
3.3.1 Esempi di applicazioni.....	- 64 -
3.4 Applicazione della teoria dei giochi a rete wireless ad hoc.....	- 66 -
3.4.1 Algoritmo di aggiornamento delle potenze.....	- 67 -
3.4.2 Criteri-guida e obiettivi del power control.....	- 69 -
3.4.3 Scelta della funzione utilità.....	- 72 -
3.4.4 Segnale dati vs segnale vocale.....	- 74 -
3.4.5 Costruzione della funzione utilità.....	- 75 -
3.4.6 Equilibrio di Nash.....	- 80 -
3.4.7 Formalizzazione del gioco.....	- 81 -
3.4.8 Scelta dei parametri della funzione utilità.....	- 85 -
3.4.9 Algoritmo di power control.....	- 88 -
3.4.10 Funzione utilità basata sulla capacità di Shannon.....	- 88 -

CAPITOLO 4 – SIMULAZIONI E RISULTATI	- 91 -
4.1 Parametri fisici del sistema simulato.....	- 91 -
4.2 Simulazione	- 94 -
4.3 Coefficiente di prezzo	- 95 -
4.4 Criterio <i>Power Saving</i>	- 97 -
4.5 Criterio Max-Capacity	- 112 -
4.6 Coefficiente di prezzo “adattativo”	- 126 -
CONCLUSIONI	- 129 -
BIBLIOGRAFIA	- 131 -

INTRODUZIONE

Nel corso dell'ultimo decennio, le comunicazioni e le reti "senza fili", comunemente dette *wireless*, hanno conosciuto una crescita senza precedenti e hanno raggiunto un elevato grado di pervasività molto prima di quanto chiunque avrebbe potuto immaginare. Non a caso, infatti, ci si aspetta che esse divengano il principale e onnipresente mezzo di telecomunicazione entro i prossimi decenni. La possibilità di lavorare ad un computer in maniera completamente svincolata da una particolare postazione fissa, quale può essere, ad esempio, quella di un ufficio, riscuote oggi favori sempre maggiori. A conferma di ciò, l'andamento crescente delle vendite dei laptop e, ultimamente, anche dei terminali hand-held, quali i PDA (Personal Digital Assistant): la mobilità offre oggi ai già affermati servizi informatici un valore aggiunto che sta diventando irrinunciabile. Lo scenario applicativo di riferimento in quest'ambito è l'accesso da computer mobile a quei servizi di rete che sono tradizionalmente sviluppati sulle reti fisse (Web, e-mail, etc.). Tuttavia, il successo diffuso dei sistemi wireless (cellulari e WLAN) spingono verso lo sviluppo di sistemi wireless avanzati, in grado di fornire altri tipi di servizi oltre la voce, come video-conferenza, media interattivi, giochi real-time attraverso Internet, etc. È, quindi, fondamentale lo studio di architetture di rete che consentano di soddisfare la crescente domanda di applicazioni così eterogenee. Al momento, i problemi che emergono considerando tale integrazione sono ancora piuttosto lontani dall'essere risolti. I dispositivi mobili sono computer molto diversi dalle *workstation* a cui siamo abituati: i primi sono di solito caratterizzati da capacità di calcolo e di memorizzazione molto inferiori, sono dotati di apparati di visualizzazione piuttosto piccoli e l'alimentazione è fornita da batterie di durata fortemente limitata. Da questa forte diversità discende il fatto che le soluzioni (applicative, di rete,

...) pensate per le reti fisse non si possono portare così come sono in un ambiente mobile, ma vanno riconsiderate: il completo fallimento di WAP ne è un esempio lampante.

Un aspetto particolarmente importante in questo senso è la gestione delle risorse energetiche. I dispositivi portatili che formano reti wireless, come appunto notebook, PDA, telefoni cellulari e micro-sensori, sono, come si è detto, inevitabilmente alimentati mediante batterie, la cui durata costituisce un fattore cruciale per questi tipi di rete. La tecnologia delle batterie risulta, tuttavia, notevolmente in ritardo in confronto al progresso che ha conosciuto nell'ultimo decennio quella delle telecomunicazioni e dell'informatica: esse riescono a fornire solo una breve autonomia energetica null'affatto paragonabile a quella praticamente illimitata dei terminali fissi, che invece sono collegati alla rete elettrica pubblica. Sfortunatamente, quindi, non è possibile affidarsi alle tecnologie costruttive delle batterie, la cui capacità, almeno nel breve termine, non può essere significativamente migliorata, visto che i progressi in questo campo sono lenti e modesti; gli sforzi della ricerca si devono, pertanto, concentrare nel progetto di sistemi che siano *energy-efficient*, tali da ottimizzare l'uso delle risorse dei terminali mobili, allo scopo di prelevare dalla batteria una quantità di energia che sia soltanto essenziale: d'altra parte, quando non si riesca ad aumentare la produzione di energia, non resta che risparmiare quella a disposizione!

Ovviamente, rimane comunque molto importante il bisogno che l'adozione di questi sistemi di *Power Saving* non comprometta la qualità del servizio reso all'utente: poco importa a questi il risparmio della batteria, se è costretto ad attese estenuanti davanti al monitor del portatile per completare il suo lavoro.

In questo elaborato, l'attenzione è stata rivolta verso una tecnologia di comunicazione wireless particolare e ancora piuttosto bisognosa di ricerca, quella delle *reti wireless ad hoc*. Molte delle caratteristiche e dei requisiti che si trovano in queste reti, infatti, sono assenti in tutti gli altri sistemi wireless e, pertanto, molti protocolli e algoritmi necessari per il loro funzionamento e gestione devono essere riprogettati *ex novo*. Come emergerà dall'elaborato, la natura estremamente instabile di questi sistemi, legata all'assenza di un'infrastruttura di rete, rende notevolmente più complesso il progetto di un qualsiasi tipo di protocollo.

In questo lavoro di tesi, in particolare, come era prevedibile in base a quanto detto in precedenza, è stato analizzato un algoritmo per la gestione efficiente delle potenze usate per la trasmissione dei segnali in queste reti ad hoc.

Il primo capitolo introduce le caratteristiche principali di questa tecnologia, presentando anche i principali ambiti applicativi e gli aspetti tecnici tuttora bisognosi di ricerca. Il secondo capitolo si concentra sull'analisi del livello protocollare MAC, il cui progetto, in questi sistemi, risulta estremamente critico; vengono, in particolare, esaminate le varie tecniche di accesso al canale su cui

il livello MAC può basarsi, e sono menzionate anche le principali soluzioni finora proposte dalla ricerca. Tale capitolo principalmente vuole costituire un preambolo per il successivo, in quanto in esso viene anche analizzata l'applicazione del CDMA alle reti ad hoc, il quale è stato implicitamente assunto come tecnica di accesso usata dal sistema ad hoc cui viene applicato l'algoritmo di *power control*. Il terzo capitolo è il più importante dell'elaborato: in esso, viene anzitutto presentato il problema della gestione delle risorse in una ad hoc network; successivamente si introducono i fondamenti della *teoria dei giochi*, che ha fornito la base concettuale matematica per la costruzione dell'algoritmo descritto dettagliatamente nel resto del capitolo. Il quarto ed ultimo capitolo illustra l'architettura del sistema, le principali ipotesi alla base delle simulazioni ed i risultati ottenuti. Per le simulazioni è stato utilizzato un simulatore sviluppato in C++, in cui sono stati inseriti l'algoritmo di power control e le funzioni per la raccolta dei risultati.

“...il miglior risultato si ottiene quando ogni componente del gruppo farà ciò che è meglio per sé e per il gruppo...”

Dal film “Beautiful mind”

CAPITOLO 1

RETI WIRELESS AD HOC

La terminologia “wireless ad hoc network” indica un insieme di dispositivi, presenti in un’area geografica limitata, in grado di comunicare tra loro usando l’etere come mezzo fisico senza far ricorso ad una pre-esistente infrastruttura di rete. Letteralmente, l’espressione latina “ad hoc” significa “per questo”, sottintendendo “per questo scopo solamente” e facendo, quindi, intuire anche il carattere solitamente spontaneo e temporaneo di questi sistemi di comunicazione.

1.1 Caratteristiche generali

Queste reti, conosciute anche come “mobile ad hoc network” (MANET), possono essere costituite da gruppi di terminali indipendenti, liberi di muoversi e di organizzarsi arbitrariamente (fig. 1.1), ma alcuni di essi possono anche essere connessi ad un sistema cellulare o ad una rete cablata fissa, ad esempio per l’accesso ad Internet: in tal caso si parla di rete ibrida (fig. 1.2). La mancanza di infrastruttura, inevitabilmente, fa sì che siano i mobili stessi a doversi occupare del controllo e della gestione della rete, mediante algoritmi che, in generale, devono essere distribuiti. Sono loro stessi, quindi, che devono procurarsi le informazioni necessarie per l’instaurazione e il mantenimento dei collegamenti, ed eseguire tutte le procedure per la scoperta vicini e lo scambio di informazioni di controllo e routing. I nodi costituenti una rete wireless ad hoc possono agire non solo da end-system, ossia come entità terminali che fungono da sorgente o da destinatari di informazione, ma anche come sistemi intermedi, che si occupano di inoltrare i pacchetti provenienti da un certo nodo e diretti verso una certa destinazione.

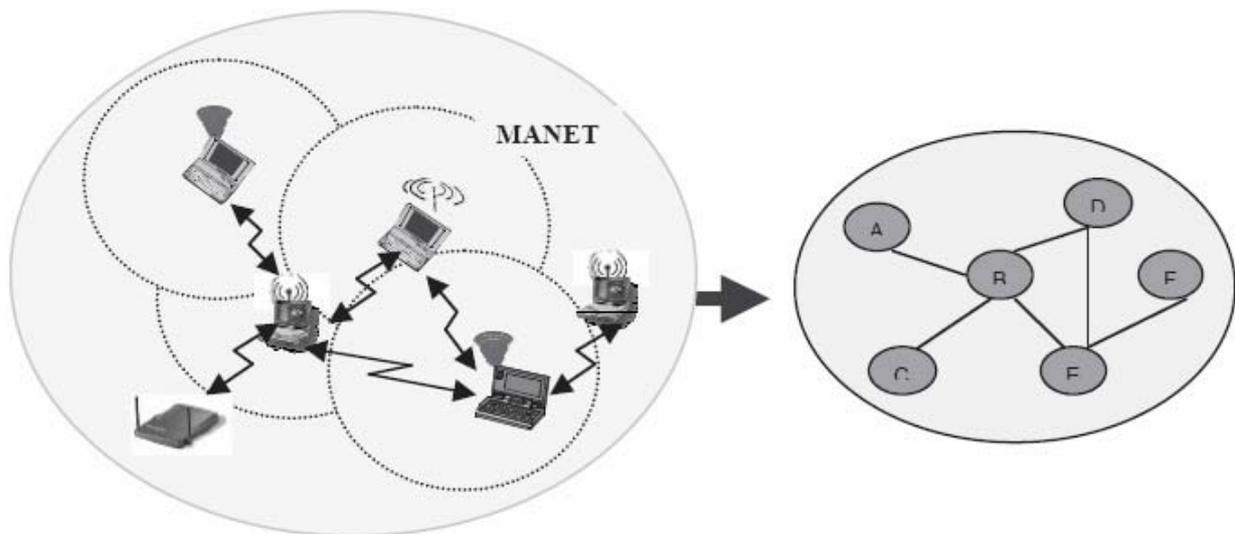


Figura 1.1 – Esempio di ad hoc network

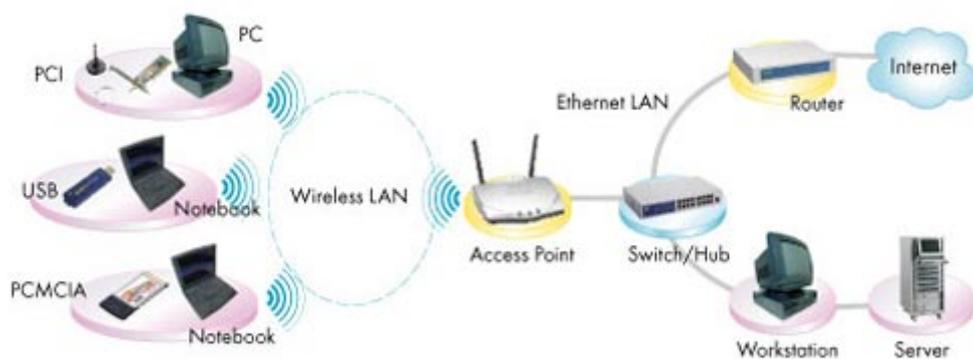


Figura 1.2 – Esempio di ad hoc network ibrida

È, quindi, possibile che due nodi comunichino tra loro anche quando essi sono fuori dal rispettivo range trasmissivo, in quanto nodi intermedi possono funzionare da ripetitori o da router. Questo è il motivo per cui queste reti sono conosciute anche come “multi-hop wireless network”. Un nodo sorgente, pertanto, ha a disposizione più percorsi per giungere allo stesso nodo e può selezionarne uno sulla base di diversi parametri, quali il numero di hop o la potenza necessaria per effettuare il collegamento. La figura 1.1 mostra questo aspetto ed evidenzia anche il fatto che esse possono eventualmente esser costituite da dispositivi di diverso tipo, come PC, notebook, telefoni cellulari, palmari, access point, etc. (fig. 1.3).



Figura 1.3 – Alcuni dispositivi per reti ad hoc

La caratteristica fondamentale di queste reti è che, dunque, i terminali sono (o meglio, devono essere) in grado di configurare se stessi *on-the-fly* senza l'intervento di un'unità di amministrazione centrale. La principale, grossa differenza con una rete cellulare sta, dunque, nell'assenza di base station e la conseguente indistinguibilità tra uplink e downlink. Rispetto a una rete cellulare, le ad hoc risultano, quindi, più flessibili e meglio adattabili a cambiamenti di traffico e di condizioni fisiche: esse possono “nascere” dovunque e in qualunque momento e possono, eventualmente, mantenere la connettività anche quando uno più terminali vengono spenti o rimossi, in quanto nessuno di essi è, in teoria, fondamentale per la vita della rete stessa, al contrario di una base station in un sistema cellulare. Robustezza, flessibilità, facilità e rapidità di formazione e supporto della mobilità sono, pertanto, le caratteristiche principali di queste reti, che le rendono particolarmente adatte a quelle situazioni in cui, per ragioni di costo, tempo di realizzazione e di complessità, non sarebbe altrimenti possibile instaurare un sistema di comunicazione. Già da questa breve descrizione, si può intuire l'enorme potenziale che questa tecnologia intrinsecamente possiede ed il gran numero di scenari di applicazione che si aprono per essa.

1.2 Applicazioni

Storicamente, il primo ambiente in cui si è cominciato a parlare di ad hoc network è stato quello militare, in cui era sentita, in particolare, l'esigenza di migliorare le comunicazioni strategiche sui campi di battaglia: la natura dinamica delle operazioni militari rendeva, e rende, impossibile fare affidamento sull'accesso ad un'infrastruttura di comunicazione preposta e fissa. Le comunicazioni radio tradizionali, inoltre, potevano facilmente esser soggette ad interferenza o esser captate e le radio frequenze oltre 100 MHz difficilmente si propagano oltre la *line of sight* (LOS). Un rete mobile ad hoc, invece, costituisce una struttura adatta a soddisfare, anche solo per un breve periodo di tempo, tutte queste occorrenze, senza il bisogno di una infrastruttura preesistente. Attualmente, sono comunque le applicazioni non militari ad avere la maggior attenzione sia da parte della ricerca

sia da parte del settore commerciale in generale. Una rete ad hoc può, in pratica, svolgere il ruolo di *estensione* o di *alternativa* ad una tratta di rete dotata di infrastruttura.

1.2.1 Legami con reti *infrastructure-based*

Il primo ruolo può esser espletato quando:

- l'infrastruttura non è presente per motivi economici; costruire un'infrastruttura è, infatti, economicamente svantaggioso quando:
 - la richiesta di servizio è, conseguentemente, la rendita è troppo bassa;
 - il traffico è inconsistente, ossia è presente solo per brevi intervalli;
 - costruire fisicamente i mezzi infrastrutturali risulta troppo costoso (a causa, ad esempio, di situazioni geografiche o terrestri impervie).
- l'infrastruttura è presente, ma inadeguata, in quanto:
 - incapace di far fronte a un improvviso (poi non prolungato nel tempo) aumento del traffico, come in situazioni di emergenza o in caso di cambiamenti della distribuzione geografica della densità del traffico;
 - inaccessibile a causa di limitazioni ambientali, come in una zona remota all'interno di un edificio.
- l'infrastruttura non è necessaria, in quanto:
 - il traffico è locale, ossia non ha bisogno di esser instradato attraverso l'infrastruttura di rete che è esterna a quella locazione: è il caso di comunicazioni effettuate in una meeting room, che non hanno necessariamente bisogno di esser collegate ad Internet.

Il secondo ruolo, quello di alternativa a una rete *infrastructure-based*, emerge quando, ad esempio, si pensa di usare una rete ad hoc per formare un sistema simil-cellulare. Commercialmente, la tecnologia ad hoc può esser usata per estendere il range delle WLAN mediante multipli radio hop. Possono esser costruite reti che coprono aree di diversi chilometri quadrati. Su scala più piccola, tecnologie come Bluetooth [13], che già si basano sull'ad hoc networking, possono esser sfruttate in svariati interessanti modi (ad esempio, in combinazione con altre tecnologie basate su 802.11 [12]) per dar vita a indoor o *embedded* network. In questo senso, l'*ad hoc networking technology* funge da soluzione intermedia o da passo economicamente conservativo prima di un eventuale costoso *upgrade* all'esistente infrastruttura.

1.2.2 Scenari d'uso

Gli scenari reali d'uso per le ad hoc network sono:

- Sensor Network, ossia un insieme di dispositivi sensoriali *embedded* (fig. 1.3 e 1.4), usati per raccogliere dati real-time, allo scopo, ad esempio, di automatizzare azioni e funzioni (sensori remoti per la meteorologia o le attività della Terra, sensori per macchine industriali, sensori per il monitoraggio del traffico veicolare in una città o su un'autostrada, sensori usati in ambito militare per rilevare gli spostamenti del nemico o la presenza di armi chimico/batteriologiche): si possono avere, ad esempio, 1000 - 100.000 nodi che collezionano campioni e poi li inoltrano, usando rate bassi e omogenei, ad un'entità centrale per l'elaborazione;

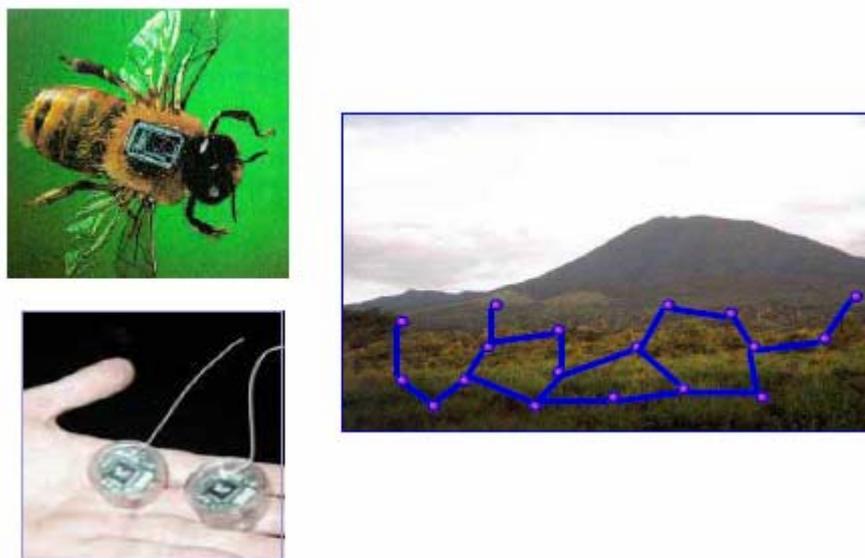


Figura 1.4 – Esempi di sensori e Sensor Network

- Reti per servizi di emergenza (fig. 1.5), ossia reti instaurate per operazioni di ricerca e salvataggio o per la sostituzione della rete fissa in caso di danneggiamento causato da calamità naturali (terremoti, uragani, incendi, etc.), magari per comunicazioni immediate con gli ospedali; uno scenario potrebbe essere una rete formata dai dispositivi di comunicazione radio posseduti dai pompieri e da eventuali “*breadcrumb router*”, piccoli dispositivi wireless a batteria posti intorno all'edificio a fuoco per fornire connettività;

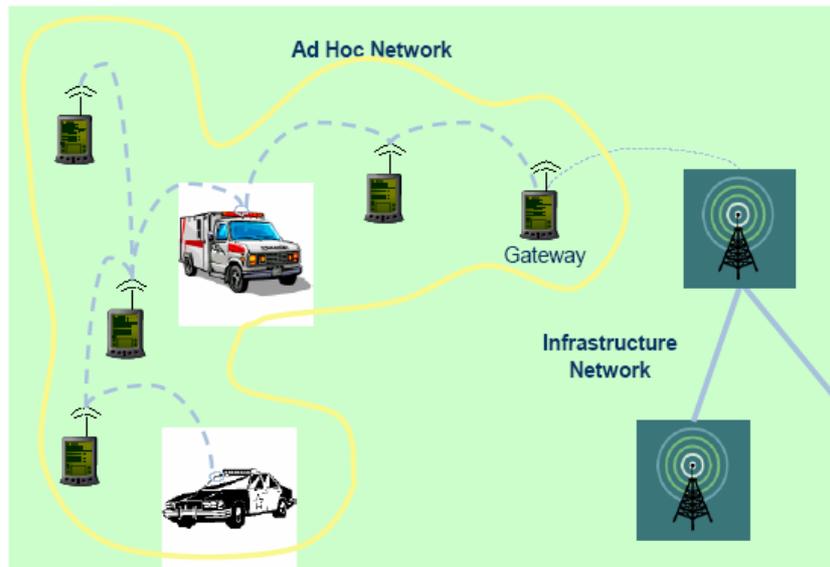


Figura 1.5 – Esempio di Emergency application

- Applicazioni commerciali, come E-Commerce (pagamento elettronico possibile ovunque, ad esempio in un taxi), Business (accesso dinamico a informazioni sul cliente immagazzinate in una locazione centrale *on-the-fly*);
- Reti Inter-Veicolari (fig. 1.6), formate tra veicoli in movimento e strutture facenti parte del sistema di trasporto, come semafori e cartelli stradali elettronici; tali reti possono essere utili per dirottare il traffico veicolare fuori da un'area congestionata mediante percorsi alternativi; quest'uso è particolarmente adatto in autostrada, in cui c'è una copertura di rete molto limitata; oppure tali reti possono essere usate per la trasmissione di notizie e previsioni del tempo;

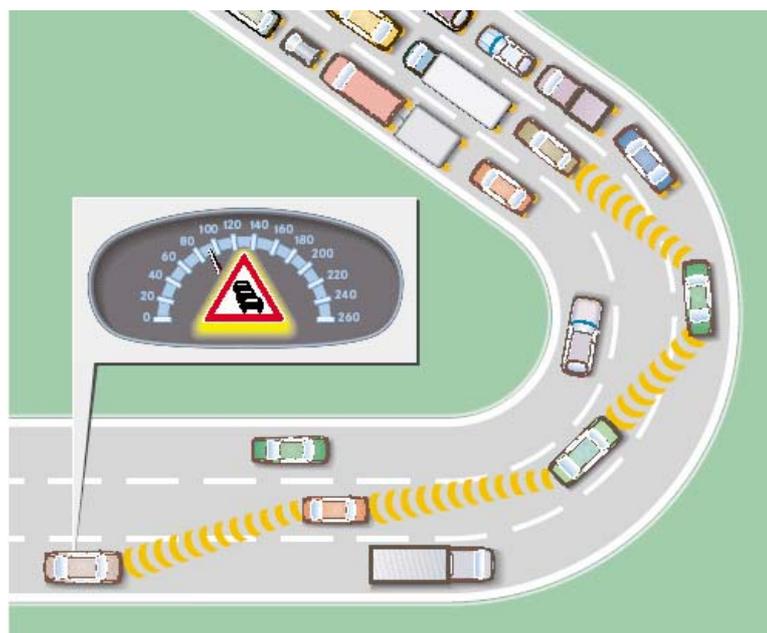


Figura 1.7 – Esempio di Inter-Vehicular Network

- Home Network, che rendono l'accesso ad Internet libero dai limiti di accessibilità, permettendolo virtualmente in qualsiasi punto della casa; una *home ad hoc network* può essere formata anche dai vari dispositivi elettronici presenti in casa per applicazioni come lo scambio di dati tra dispositivi portatili (notebook, PDA, etc.) o il comando a distanza di apparecchi (elettrodomestici, PC, stampanti, etc.), tutte con lo scopo di migliorare la qualità della vita domestica (fig. 1.8);



Figura 1.8 – Esempio di Home Network

- Enterprise Network, che vengono realizzate all'interno di un edificio di un'azienda o anche tra più edifici e fungono da mezzo *low cost* e *easy-to-install* in grado di facilitare la comunicazione e la mobilità degli impiegati;
- Applicazioni didattiche: si possono creare aule o stanze virtuali per videoconferenze; oppure, si possono instaurare vere e proprie comunicazioni ad hoc durante conferenze, meeting o lezioni: un presentatore può inviare in multicast slide e audio ai destinatari desiderati; gli spettatori a loro volta possono fare domande o interagire su una lavagna condivisa;
- Intrattenimento: giochi multi-player, animali-robot, accesso ad Internet all'aperto, musica.

Un esempio di scenario reale è mostrato in figura 1.9.

1.2.3 Vantaggi

È possibile, a questo punto, evidenziare i benefici introdotti da questa tecnologia, sia per gli utenti, sia per gli operatori di rete e i fornitori di servizi.

Per gli utenti si ha:

- supporto della mobilità: con una tecnologia ad hoc disponibile, un utente sarebbe in grado di comunicare con gli eventuali utenti vicini, dovunque egli si rechi;
- minimo supporto infrastrutturale richiesto;

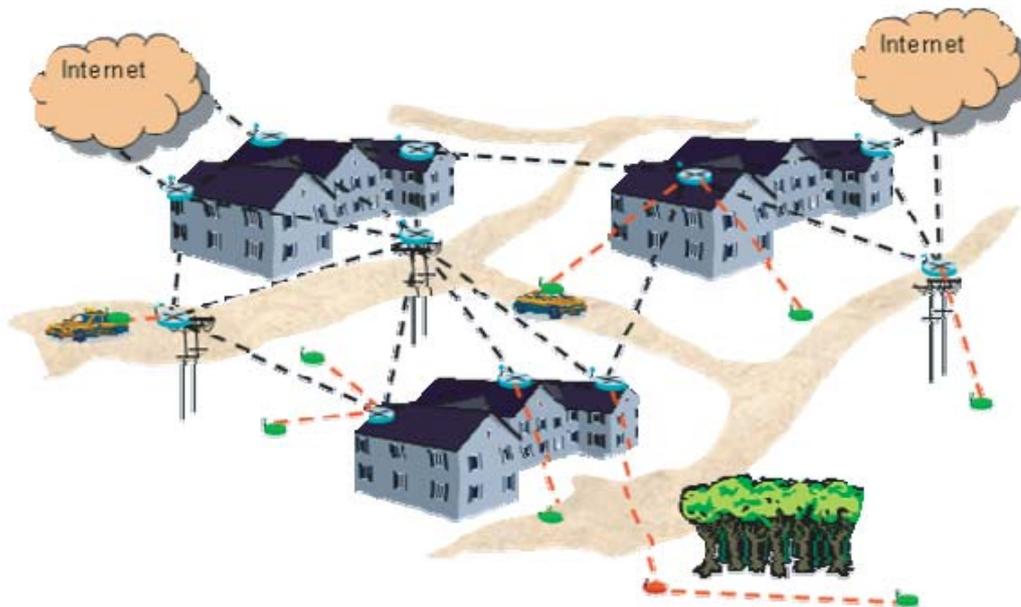


Figura 1.9 – Esempio di scenario reale per un'ad hoc network

- riduzione dei costi: con la tecnologia ad hoc, il convenzionale modello di comunicazione *infrastructure-based* distrutto dal punto di vista della rendita economica, poiché viene praticamente eliminata la dipendenza dall'infrastruttura;
- possibilità di instaurare una comunità di utenti locale che promuove la cooperazione e la condivisione di risorse (informative, computazionali, di connessione);
- possibilità di effettiva organizzazione tra persone vicine (in caso di emergenze e disastri);

Per i fornitori di servizi si ha invece:

- estensione della copertura di rete: un ad hoc network può costituire, come già detto, un elemento complementare a reti convenzionali, come la rete cellulare o Internet, in grado, cioè, di aumentarne l'estensione della rete in zone in cui l'infrastruttura non è disponibile oppure difficile e/o costosa da costruire;
- riduzione dei costi: le reti ibride eliminano i costi di installazione e mantenimento e anche di riconfigurazione in caso cambiamenti geografici della densità del traffico di rete;
- la riduzione dei costi legati all'infrastruttura consente l'entrata nel mercato di più operatori, in grado di offrire nuovi servizi wireless;
- il traffico locale può rimanere tale, ossia non è necessario che questo passi attraverso i mezzi infrastrutturali, riducendo per questi il carico di lavoro e garantendo, in tal modo, una più elevata capacità;

- la tecnologia ad hoc può costituire un'economica soluzione intermedia, che può precedere eventuali costosi miglioramenti all'infrastruttura esistente.

1.2.4 Svantaggi e limiti

I vantaggi, che finora sono stati messi in evidenza, devono però fare i conti con nuove problematiche, che nascono proprio in virtù delle caratteristiche innovative di questa tecnologia e che lanciano molte nuove sfide alla ricerca. Quest'ultima, nel passato, ha quasi sempre considerato le ad hoc network come una parte integrante di Internet, in un'architettura modulare *IP-centric*. Una tale architettura ha due grossi vantaggi: essa semplifica l'interconnessione con Internet e garantisce l'indipendenza da tecnologie wireless eterogenee. Il paradigma *layered* di Internet, che ne ha significativamente semplificato il progetto e ha facilitato la nascita di protocolli robusti e scalabili, può non essere altrettanto efficiente quando applicato alle MANET. In esse, infatti, diverse funzionalità (ad esempio gestione dell'energia, *security*, QoS e cooperazione tra nodi) difficilmente possono essere isolate in un unico livello protocollare: la loro implementazione efficiente è, piuttosto, il risultato dello sforzo coordinato dei vari livelli, che pertanto non possono essere trattati separatamente, bensì ne devono essere sfruttate efficacemente le interdipendenze.

Gli aspetti innovativi delle reti ad hoc rispetto alle tradizionali reti wireless sono:

- assenza di infrastruttura: il contesto operativo della rete risulta profondamente diverso rispetto a quello di una rete cablata o anche di una rete cellulare; di conseguenza i protocolli di rete sviluppati per queste ultime potranno non essere adatti per le reti ad hoc;
- topologia di rete dinamica: la mobilità che può caratterizzare gli utenti di queste reti fa sì che la topologia di rete possa cambiare frequentemente e imprevedibilmente, causando cambiamenti di route, partizioni di rete ed, eventualmente, perdita di pacchetti; la figura 1.10 mostra una tassonomia delle ad hoc network, basata sul tasso di connessioni/ disconnessioni;
- limiti del livello fisico: la maggior parte delle applicazioni delle reti ad hoc fanno uso della banda ISM (Industrial, Scientific, Medical), che non ha bisogno di licenze per l'impiego. In generale, il canale wireless è un mezzo strettamente controllato dalle normative poiché ha una banda utilizzabile molto piccola. Inoltre, il mezzo wireless è intrinsecamente “*error prone*”, poiché l'interfaccia radio di ogni nodo trasmette in broadcast e le collisioni tra più trasmissioni divengono inevitabilmente un aspetto intrinseco del canale: esse fanno sì che la probabilità di perdita di pacchetti a causa di errori sia molto più elevata rispetto a un sistema cablato, cosa che riduce la capacità dei link; inoltre, il canale wireless è caratterizzato da fattori come *fading*, rumore e interferenza che peggiorano le prestazioni delle trasmissioni;

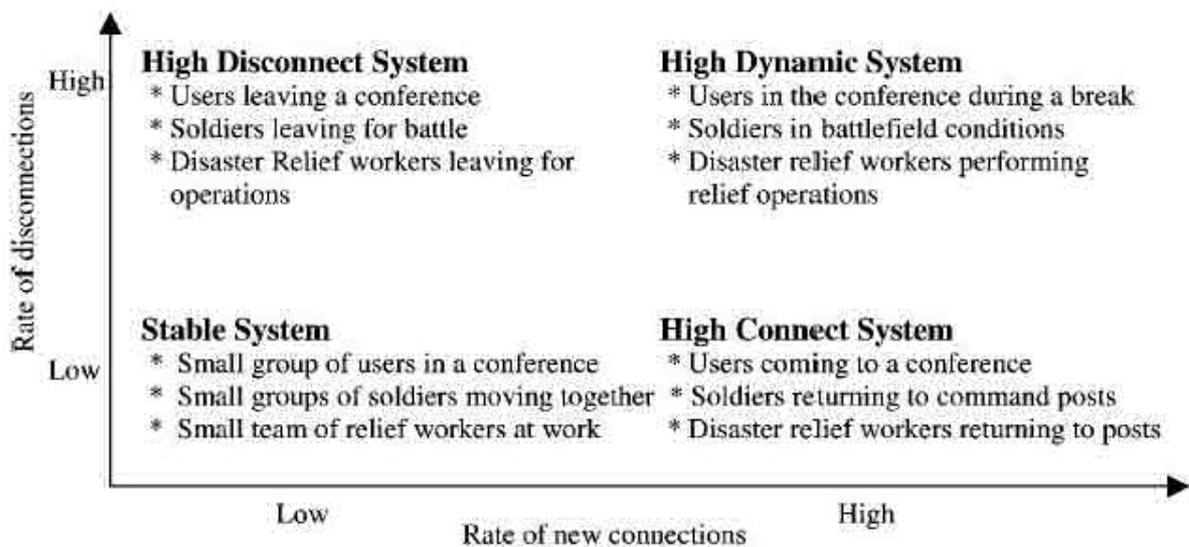


Figura 1.10 – Classificazione delle ad hoc network basata su tasso di connessioni/disconnessioni

- variabilità delle capabilities dei nodi: ogni nodo può essere dotato di una o più interfacce radio con capacità di trasmissione e di ricezione diverse e variabili, che possono lavorare su bande frequenziali differenti. Questa eterogeneità può eventualmente dar vita a link asimmetrici; in più, si possono avere differenze in termini di configurazione hardware o software, che danno luogo a diverse capacità di elaborazione;
- sicurezza di rete: dal punto di vista della *security*, le MANET presentano i problemi di base che affliggono qualunque infrastruttura di rete wireless, problemi che sono connessi con l'intrinseca debolezza di questo tipo di mezzo di comunicazione, come la facilità di interagire con la rete e di intercettare le informazioni in transito; però ad essi, ne vanno aggiunti altri, che nascono proprio in virtù delle caratteristiche innovative di questa tecnologia;
- robustezza ed affidabilità di rete: la mancanza di un monitoring di rete centralizzato e di elementi per la gestione globale, sebbene aumentino la flessibilità del sistema, finiscono per renderlo più debole dal punto di vista dell'affidabilità, in quanto, comunque, eventuali malfunzionamenti possono non essere rilevati, con la conseguente impossibilità per alcuni nodi di inoltrare pacchetti;
- scalabilità: in determinati contesti (come le *sensor network*), il numero dei nodi costituenti il sistema ad hoc può risultare molto grande (dell'ordine delle decine di migliaia), per cui l'aspetto della scalabilità è estremamente critico per lo sviluppo di queste reti e, tuttavia, esso si scontra con tutti quelli che sono stati finora elencati;

- potenza disponibile limitata: le prestazioni dei computer mobili sono fortemente limitate dalla durata delle batterie, per cui limitate sono anche le funzionalità e i servizi che possono venire forniti da ciascun nodo. Nelle reti ad hoc, questo aspetto è particolarmente sentito, dal momento che ogni nodo agisce sia come end-system, sia come router per l'inoltro di pacchetti provenienti da altri nodi.

Tutti questi aspetti si scontrano con quelle che, in teoria, sono le caratteristiche peculiari di una rete ad hoc, come l'auto-configurabilità e flessibilità a vari livelli (ad esempio, routing dinamico e controllo di accesso al mezzo distribuito): per questo progettare algoritmi e protocolli di rete per un contesto così particolare può risultare estremamente difficile.

1.3 Sfide per la ricerca

Attualmente, come sarà più chiaro nei paragrafi successivi, sono ancora molte le problematiche da risolvere prima che reti ad hoc su larga scala possano diffondersi ampiamente, mentre su piccola scala stanno già iniziando a diffondersi diversi esempi di applicazione di questa tecnologia. Sul mercato, infatti, sono già presenti da diversi anni schede di rete per sistemi single-hop, i quali costituiscono i blocchi elementari per costruire reti ad hoc su piccola scala.

In questa sezione vengono brevemente presentati gli aspetti delle MANET ancora bisognosi di ricerca: da esso, si potrà intuire che ancora si deve fare molto prima di rendere commercialmente proficua questa tecnologia.

1.3.1 Medium Access Control

Le varie tecniche di accesso al mezzo applicate alle reti ad hoc presentano vantaggi e svantaggi che fanno sì che ancora non ne esista una capace di soddisfare tutte le esigenze caratterizzanti questa tecnologia; il capitolo 2 spiegherà in dettaglio la maggior parte delle problematiche relative.

1.3.2 Routing

Il routing è uno dei più importanti aspetti in un'ad hoc network dal momento che la topologia varia frequentemente ed è necessaria una comunicazione basata su multi-hop. Esistono due tipi di protocolli di routing: pro-attivi e reattivi. I primi sono detti *table driven*, in quanto ogni nodo possiede una tabella, che viene aggiornata periodicamente e che contiene le informazioni di routing su tutti i nodi conosciuti. I secondi sono detti anche *on demand*, in quanto i percorsi di routing vengono stabiliti solo su richiesta. Utilizzare i protocolli pro-attivi fondamentalemente garantisce un ritardo end-to-end più piccolo poiché l'informazione sulla route è sempre disponibile e aggiornata in confronto a quella presente nei protocolli reattivi. Lo svantaggio, ovviamente, consiste nel fatto

che è necessario un maggior consumo di risorse (maggior overhead) per l'aggiornamento periodico, per cui in un ambiente altamente dinamico sarebbero difficilmente applicabili. Esempi di protocolli proattivi sono: *Destination Sequenced Distance Vector (DSDV)*, *Clusterhead-Gateway Switch Routing (CGSR)*, *Wireless Routing Protocol (WRP)*, *Optimized Link State Routing (OLSR)*, *Topology Broadcast Based on Reverse Path Forwarding (TBRPF)*, *Fisheye State Routing (FSR)* e *Landmark Routing (LAN-MAR)*. Esempi di protocolli reattivi sono: *Dynamic Source Routing (DSR)*, *Ad Hoc On Demand Distance Vector (AODV)*, *Temporally Ordered Routing Algorithm (TORA)*, *Associativity Based Routing (ABR)* e *Signal Stability Routing (SSR)*.

I ricercatori stanno cercando un equilibrio ottimo tra i due meccanismi di protocollo, lavorando su routing di tipo *power-aware* e *scalable*, in modo che un protocollo continui a funzionare anche in presenza di un gran numero di nodi e di un carico di traffico elevato. Si lavora anche per il supporto della QoS (IRTF RRG Ad hoc Network Systems Research Subgroup). Allo scopo di prendere le decisioni di routing ottime da punto di vista della QoS, il protocollo ha bisogno anche di costanti aggiornamenti delle informazioni sullo stato del link (per quanto riguarda il ritardo, la banda disponibile, il tasso di perdita di pacchetti, il tasso di errore), che causano un enorme quantità di overhead di controllo, che può risultare proibitivo per un contesto ad hoc. In più, la natura dinamica di questi sistemi rende estremamente difficile il recupero di precise informazioni sullo stato di un link.

1.3.3 Security

Il problema della *network security* è spesso considerato il maggior ostacolo per l'applicazione commerciale della tecnologia ad hoc wireless. I principali scopi a livello di *security*, per le reti wireless in generale, sono:

- evitare che le comunicazioni vengano intercettate (*confidentiality*);
- limitare e proteggere l'accesso alla rete (*access control*);
- proteggere l'integrità dei dati, evitando che il traffico venga alterato in qualche modo (*data integrity*);
- proteggere i nodi da attacchi di tipo *denial of service*.

Si è già accennato al fatto che le reti ad hoc presentano alcune caratteristiche che rendono l'aspetto della *security* più complicato rispetto al caso di una rete dotata di infrastruttura. Queste caratteristiche sono:

- vulnerabilità del canale: la natura aperta e broadcast dei canali wireless permettono una facile intercettazione e immissione “velenosa” di dati;
- vulnerabilità dei nodi: i nodi non si trovano in posti fisicamente protetti, per cui sono facilmente soggetti ad attacchi;
- assenza di infrastruttura: autorità che forniscono servizi di certificazione/autenticazione sono inevitabilmente assenti;
- topologia di rete variabile dinamicamente: la sicurezza dei protocolli di routing è minacciata da attacchi anche non intenzionali;
- limiti energetici e computazionali: algoritmi troppo complessi di crittografia non possono esser usati.

Il canale radio condiviso da tutti i nodi è un ambiente potenzialmente insicuro, che li rende esposti a molti attacchi. Volendo fare una sintetica classificazione degli attacchi alla sicurezza, si possono anzitutto distinguere attacchi passivi e attacchi attivi. In un attacco passivo, un *malicious node* o ignora operazioni che tutti gli altri nodi assumono che svolga (ad esempio può nascondere alcune informazioni di routing o scartare dati da inoltrate) o “ascoltare” il canale, tentando di recuperare informazioni cambiate (*eavesdropping*). In entrambi i casi, il nodo “cattivo” non inserisce alcun messaggio nuovo in rete. È molto difficile rintracciare un comportamento di questo tipo dai normali malfunzionamenti che possono presentarsi in un ad hoc network. In un attacco definito “attivo”, viene effettivamente inserita dell’informazione “velenosa” nella rete, il cui funzionamento complessivo (o solo di alcuni nodi) può risultare compromesso. Esempi di questi attacchi sono *impersonation/spoofing, modification, fabrication* e *disclosure attack*.

Il malfunzionamento del processo di routing-forwarding può esser causato da nodi che sono *malicious* o, semplicemente, “egoisti”. Nel primo caso, lo scopo del nodo è effettivamente creare danno al funzionamento della rete. Un nodo *selfish*, invece, mira a risparmiare energia per usarla solo per le sue comunicazioni. Semplicemente non partecipando al protocollo di routing o non eseguendo l’inoltro dei pacchetti, un nodo può danneggiare la rete, sebbene quello non sia realmente il suo scopo. Per evitare questa situazione, è necessario *forzare* la cooperazione tra nodi: un modo per farlo è introdurre un meccanismo di supervisione (*watchdog*) e di reputazione in grado di rivelare e isolare nodi che non si comportano correttamente. Questi ultimi vengono identificati mediante un processo di monitoring dei vicini che ogni nodo può fare. Sulla base delle informazioni raccolte da questo meccanismo, il sistema di reputazione mantiene un valore per ogni nodo che rappresenta il suo grado di reputazione. Un altro modo per incoraggiare la cooperazione tra nodi è l’introduzione di un sistema di *incentivi*. Un esempio è il Nuglet system, ossia un sistema basato su

una “moneta virtuale”, che appunto è il *nuglet*, che non è altro che un contatore mantenuto da ogni nodo e decrementato di 1 quando il nodo intende inviare uno dei suo pacchetti: è come se il nodo “pagasse” per la sua trasmissione. Viceversa, il contatore è incrementato quando il nodo inoltra un pacchetto per gli altri, cioè viene “ricompensato”.

Per quanto riguarda la protezione dei dati, i tradizionali metodi crittografici, che abbiamo detto esser anche molto pesanti dal punto di vista computazionale, si scontrano anche con l’aspetto della distribuzione e *refresh* delle chiavi segrete, in assenza di un autorità centrale. Uno degli approcci correnti ai problemi di sicurezza prevede la costruzione di una *self-organized public-key infrastructure* per la crittografia in una rete ad hoc.

1.3.4 Valutazione delle prestazioni

Devono ancora essere sviluppati e implementati strumenti che modellino, analizzino e controllino il funzionamento di queste reti, strumenti che, in tal modo, permettano il raggiungimento e la gestione di certi livelli prestazionali. Una rete ad hoc introduce notevoli difficoltà nell’analisi e nella gestione delle risorse radio, dal momento che le perturbazioni introdotte da un nodo possono influenzare fortemente le azioni degli altri nodi.

Sono fondamentalmente due gli approcci che possono essere intrapresi per valutare le prestazioni di un’ad hoc network: effettuare misure su un sistema reale o utilizzare modelli. Nel primo caso, deve esser costruito un banco di prova, cosa che di solito risulta molto costosa, senza contare che alcuni parametri, come la scalabilità di un protocollo o la sensibilità alla mobilità e alla velocità degli utenti, sono difficili da investigare su un sistema reale; non sono da trascurare, poi, la difficoltà nel riprodurre certe misurazioni e le dimensioni inevitabilmente limitate del banco di prova stesso. Il vantaggio che questo approccio possiede è che, ovviamente, si è in grado di scoprire problemi che non potrebbero esser notati con un modello. L’uso di modelli, d’altra parte, rende più semplice e flessibile lo studio della rete perché, grazie ad essi, è possibile variare agilmente certi parametri, considerando sistemi anche di grandi dimensioni; il limite dei modelli è, ovviamente, il fatto che spesso non sono in grado di far emergere alcuni problemi che si incontrano nell’implementazione reale del sistema. Esistono due tipi di modello: analitico e simulativo. Il primo è utile in particolare per lo studio della scalabilità del protocollo. Attualmente c’è una grossa mancanza di analisi teorica nell’*ad hoc networking*, dovuta in parte alla mancanza di fondamenti su cui basare i paragoni teorici, in parte all’astrusa natura dei problemi delle MANET. Al contrario, la simulazione costituisce uno strumento maturo, standardizzato e flessibile per modellare vari tipi di scenario, però i simulatori in commercio (OPNET, NS2, GloMoSim) non sono ancora del tutto testati e validati, poiché hanno fornito risultati discordanti tra loro, sia quantitativamente, che, in particolare, qualitativamente.

Ma è il funzionamento real-time della rete che ha soprattutto bisogno di strumenti in grado di effettuare un monitoraggio corretto, strumenti che, invece, allo stato attuale dell'arte mancano.

1.3.5 Scalabilità

La scalabilità di una rete, in generale, si riferisce alla sua capacità di fornire un accettabile livello di servizio anche in presenza di un gran numero di nodi nella rete. Principalmente le ad hoc network soffrono, per loro natura, di problemi di scalabilità legati alla capacità di rete. In una rete non-cooperativa, in cui vengono usate antenne omni-direzionali, il *throughput* decresce come $1/\sqrt{N}$, dove N è il numero di nodi [64]. Cioè, in una rete con 100 nodi, un singolo nodo ottiene approssimativamente al massimo un decimo del rate teorico della scheda di rete. Questo problema, comunque, non può esser risolto se non con miglioramenti del livello fisico, come l'introduzione di antenne *smart*. In secondo luogo, anche i protocolli di routing, come già accennato in precedenza, introducono alcuni limiti per la scalabilità delle MANET. Funzionalità come la *route acquisition*, il *service location* e lo scambio delle chiavi per l'*ecryption* dei dati sono solo alcuni esempi che richiedono un considerevole overhead, il quale crescerà rapidamente al crescere delle dimensioni della rete. Se le già scarse risorse di rete venissero sprecate per un abbondante controllo del traffico, le ad hoc network non vedrebbero mai la luce. Spesso al problema della scalabilità del routing, in generale, si cerca di rispondere con l'introduzione di gerarchie, nel routing stesso e in funzioni di rete di altri livelli. Un approccio usuale per implementare routing gerarchico in un ad hoc network è usare algoritmi di *clustering*, in grado di raggruppare nodi vicini in sottoinsiemi e, di conseguenza, limitare l'espansione dei messaggi di update al cluster in cui si verifica il cambiamento. Diverse soluzioni di questo tipo sono state proposte negli ultimi anni, tra cui ZRP (Zone Routing Protocol), OLSR (Optimized Link State Routing), CEDAR (Core Extraction Distributed Ad hoc Routing), CBRP (Cluster Based Routing Protocol). C'è ancora molto lavoro da fare per ottimizzare il compromesso tra capacità e scalabilità nei vari scenari e applicazioni, senza considerare una soluzione generale.

1.3.6 Power conservation

La potenza è una risorsa preziosa per i dispositivi di rete mobili. I terminali mobili traggono energia dalle batterie, la cui durata è piuttosto limitata: questo rappresenta uno dei maggiori vincoli nel progetto di algoritmi che coinvolgono questi dispositivi. In più, un nodo nella rete deve anche fare da router, per l'instradamento di messaggi provenienti da altri nodi. È, quindi, vitale che la potenza sia gestita in maniera efficiente, individuando i modi per usarne meno possibile, senza ovviamente avere impatto sulle applicazioni. La conservazione dell'energia non è un problema che coinvolge un

unico livello protocollare, bensì richiede uno sforzo coordinato di tutti i livelli, in particolare del livello fisico, delle applicazioni e del sistema operativo.

Le principali fonti di consumo energetico per i terminali mobili sono il processing dei dati e le comunicazioni radio. L'elaborazione dei dati coinvolge l'uso della CPU, della memoria, dell'hard drive, ecc., ma l'energia spesa per questi componenti è, di solito, trascurabile confrontata con quella necessaria per le radio comunicazioni: in esse, viene spesa potenza sia in fase di trasmissione, sia in fase di ricezione da parte di ogni componente coinvolto in tutti i nodi lungo il percorso da sorgente a destinazione. Esiste, comunque, un trade-off tra le due fonti di consumo: ad esempio, in [55], vengono introdotte tecniche di compressione dei dati che riducono la lunghezza dei pacchetti e, quindi, consentono un risparmio energetico in fase di trasmissione, e che però introducono anche una maggior pesantezza computazionale, che può finire per compensare il risparmio ottenuto.

In una rete basata su infrastruttura, per risparmiare energia, si può pensare di spostare gli sforzi computazionali e comunicativi verso l'infrastruttura fissa, mantenendo l'interfaccia di rete dei dispositivi mobili in stato inattivo più a lungo possibile. Si tratta di un metodo ovviamente non praticabile in una rete ad hoc, in quanto essa è priva dell'infrastruttura *energy-unconstrained* su cui poter riversare la maggior parte delle operazioni ed, inoltre, in un tale contesto operativo di solito esiste uno scopo addizionale implicito che è la massimizzazione della vita della rete stessa, per raggiungere la quale i nodi devono monitorare continuamente il mezzo per captare eventuali trasmissioni.

La ricerca in questo ambito si è focalizzata su diversi aspetti, inclusi lo studio del consumo energetico a livello di interfaccia di rete nei dispositivi portatili wireless [11] e il confronto tra diversi protocolli di routing e di livello MAC in termini di capacità di conservazione dell'energia. Per quanto riguarda il primo, in una rete ad hoc, come si sa, il consumo di energia legato alla comunicazione radio comprende la potenza consumata al trasmettitore, al ricevitore e ai nodi intermedi presenti nel percorso da sorgente a destinazione; in realtà, poi, un nodo in una MANET deve in ogni istante trovarsi in uno dei 4 seguenti stati: trasmissione, ricezione, *idle listening* (ossia il nodo non trasmette, né riceve bensì fa "monitoring" del canale) e *sleep* (il nodo non può comunicare). Anche durante lo stato *idle* si ha consumo di energia in quanto, in pratica, il nodo sta "ascoltando" il canale allo scopo di rilevare l'arrivo di un pacchetto indirizzato a lui ed, eventualmente, passare nello stato di ricezione. Allora, è stato scoperto che l'interfaccia di rete wireless, ossia la scheda di rete, consuma quasi lo stesso ammontare di energia in fase di ricezione, trasmissione e stato *idle*, mentre nello *sleep state* il consumo energetico viene altamente ridotto. La tabella 1.1 mostra il consumo energetico della scheda di rete 2.4GHz DSSS Lucent IEEE 802.11WaveLAN PC "Bronze" (2Mbps) nei diversi modi di funzionamento.

Si può notare che, come da molte altre misurazioni è emerso, l'energia spesa per l'*idle listening* è di poco inferiore a quella necessaria per la ricezione vera e propria: questo significa che in una MANET si ha un significativo consumo di energia anche quando non è presente traffico in essa. Tuttavia, il semplice massimizzare il periodo in cui l'interfaccia si trova in modalità *power-saving* (ossia in *sleep state*), apparentemente non è un approccio perseguibile in sistemi come le MANET, in quanto, come abbiamo già accennato, esse si basano sulla cooperazione tra i nodi partecipanti per fornire connettività di rete, nodi che pertanto devono teoricamente sempre stare "in ascolto".

Modes	Energy Consumption
Sleep Mode	14 mA
Idle Mode	178 mA
Receive Mode	204 mA
Transmit Mode	280 mA

Tabella 1.1

Un nodo "avaro", che rimane per la maggior parte del tempo nello *sleep state*, senza contribuire al routing e al forwarding, massimizzerà il tempo di vita della propria batteria, ma comprometterà la vita della rete. Per superare questo problema, sono state sviluppate strategie in base alle quali l'interfaccia di rete può essere messa in modalità *power-saving* con un impatto minimo sulle operazioni di trasmissione e ricezione. Queste *polices* tipicamente operano a livello fisico e a livello MAC. Ad esempio, a livello fisico, alcuni autori hanno proposto e analizzato procedure (basate sul monitoraggio delle frequenze di errore di trasmissione), che evitano trasmissioni inutili quando il rumore di canale riduce la probabilità di successo nella ricezione dei dati [37,38]. A livello MAC, è fondamentale notare che esiste una componente significativa di consumo energetico legata ad aspetti quali la ritrasmissione di pacchetti, il *node overhearing* e l'overhead protocollare in generale. La ritrasmissione è causata dalle collisioni (l'arrivo contemporaneo ad un ricevitore) tra più trasmissioni, che non consentono la ricezione corretta dei pacchetti, i quali, pertanto, devono essere scartati e trasmessi nuovamente. La ritrasmissione, ovviamente, aumenta il consumo energetico, per cui si capisce che, se si vuole risparmiare, c'è bisogno di ridurre il più possibile le collisioni. *Node overhearing*, invece, sta a indicare che, a causa della natura broadcast del canale wireless, un nodo può udire trasmissioni a lui non destinate provenienti da nodi vicini: a priori, il nodo non può sapere con certezza se è il destinatario delle trasmissioni che "sente", pertanto è costretto ad avviare i processi di ricezione che, nel caso in cui tali pacchetti non siano effettivamente a lui destinati, comportano un inutile spreco di potenza. L'overhead protocollare è,

invece, dovuto all'esistenza di quei pacchetti che servono per il controllo di rete e ai bit d'intestazione dei pacchetti stessi.

Per eliminare, o quanto meno abbassare il tasso delle collisioni, in modo da ridurre anche l'energia richiesta per trasmettere *con successo* un pacchetto, si può, ad esempio, fare in modo di:

- evitare di trasmettere quando il canale è congestionato: mentre un nodo trasmette un pacchetto, gli altri nodi presenti all'interno dello stesso range di interferenza e di *carrier sensing* devono rimanere in silenzio per non avere collisioni (il capitolo successivo entrerà più in dettaglio riguardo questa modalità, su cui si basa lo standard IEEE 802.11);
- sincronizzare l'istante di comunicazione (per un'ad hoc network di tipo single-hop sono presenti esempi in [41,42,43]);
- stabilire gli intervalli temporali durante i quali le interfacce di rete non hanno bisogno di stare in ascolto [40].

Si può notare (e sarà analizzato più in dettaglio nel capitolo successivo) che questi ultimi due aspetti consentono, in teoria, di minimizzare il problema del *node overhearing* e si possono ritrovare nell'uso di una tecnica di accesso basata su TDMA, che potrebbe, dunque, in teoria, consentire una forte riduzione del periodo di *idle listening*.

Da non trascurare non è neanche l'energia spesa da un nodo per effettuare una transizione da uno stato all'altro, poiché il tempo di transizione non è infinitesimo: ad esempio, la transizione dallo stato di trasmissione a quello di ricezione può tipicamente richiedere da 6 a 30 μs , mentre per passare da *sleep* a trasmissione o ricezione di solito serve molto più tempo (250 μs).

Sono state, poi, sviluppate altre strategie per un risparmio di energia a livello globale di rete, oltre a quelle a livello di nodo menzionate finora. Ad esempio, in una regione ad elevata densità di nodi, è necessario che solo un piccolo numero di loro "si accenda" per inoltrare il traffico, in modo che il tempo di vita della rete venga ottimizzato.

Ovviamente, non è soltanto il periodo di *idle listening* a dover essere efficientemente gestito se si vuole risparmiare energia: un altro importante metodo consiste ovviamente nel controllare la potenza dei nodi trasmettitori. La potenza consumata dall'amplificatore di potenza a RF della scheda d'interfaccia di rete è direttamente proporzionale alla potenza del segnale trasmesso. Una ridotta trasmissione di potenza permette, quindi, di allungare la durata delle batterie di un dispositivo e, al tempo stesso, di diminuire l'interferenza tra utenti, cosa che, oltre che consentire il riuso spaziale delle frequenze, può aiutare a incrementare il *throughput* totale di rete. Inoltre, agendo in questo modo sulla potenza del segnale, la probabilità di intercettazione e rivelazione

viene abbassata, aspetto utile in particolare per le applicazioni militari. Dall'altro lato, l'abbassamento del livello di potenza trasmessa significa anche un numero più piccolo di link instaurabili tra nodi e, quindi, una connettività single-hop inferiore. Questi due effetti hanno un impatto opposto sul consumo energetico. Gran parte del recente lavoro sull'efficienza energetica nelle ad hoc network si concentra sui protocolli di routing *energy-efficient* [44,45,46,47], nei quali il livello di potenza trasmessa è un'ulteriore variabile nel progetto del protocollo [10]. Tra questi lavori, Minimum Power Routine (MPR) seleziona, tra un dato nodo sorgente e una data destinazione, il percorso che richiede il minimo ammontare di potenza totale previsto, mantenendo un accettabile valore di rapporto segnale-rumore-interferenza ad ogni ricevitore [48,49]. Esso, inoltre, prevede l'uso di statistiche a livello di link e fisico per risparmiare in potenza e, al tempo stesso, compensare gli effetti di *shadowing*, *fading*, *propagation path loss* e interferenza.

Si possono anche introdurre *power-saving policies* a livello di sistema operativo che includono sistemi di *scheduling* della CPU [51,52] e di gestione ottimale dell'hard-disk [53]. A livello di applicazione, le strategie convenzionali per il risparmio energetico non sono applicabili nel contesto ad hoc [54,55]: si tratterebbe di tecniche che cercano di far sì che i dispositivi mobili trascorrono il maggior tempo possibile nello *sleep state*; tuttavia, in un'ad hoc network, come più volte ribadito, i nodi sono implicati anche nell'inoltro dei pacchetti per gli altri nodi.

1.3.7 Hardware-Software

La tecnologia che riguarda l'hardware e il software deve ancora maturare, al punto in cui il livello fisico possa essere adattato all'implementazione di qualsiasi forma d'onda, soggetta ovviamente a ragionevoli vincoli frequenziali e spettrali. Questa area di ricerca è già attiva in molte università e sostenuta dal SDR Forum e dal JTRS (Joint Tactical Radio System). Nonostante che rimangano ancora molti ostacoli nello sviluppo di un tale hardware-software, alcune tecnologie emergenti come MEMS, trasmettitori radio avanzati [26] e CCM [27], fanno pensare che ben presto tali scogli saranno superati.

È necessario, inoltre, sviluppare una struttura standard generale d'azione per i dispositivi wireless, tale che essi siano in grado di valutare le proprie capacità, i requisiti dei servizi, le potenziali forme d'onda e le condizioni ambientali, per poi stabilire, sempre comunque limitatamente alla propria conoscenza, le strategie necessarie per soddisfare i bisogni contingenti. Un tentativo di formalizzare questa struttura si trova in [28]: in esso, è stato sviluppato un ciclo d'azione, detto *cognition cycle*, che consiste in una macchina a stati finiti che riassume i cambiamenti di strategia che devono esser seguiti dal terminale, in risposta a variazioni delle condizioni contestuali e ambientali.

CAPITOLO 2

MAC LAYER IN AD HOC NETWORK

Uno dei maggiori ostacoli per l'implementazione pratica delle reti ad hoc è la mancanza di un protocollo per il livello MAC che ben si adatti ad esse. Il livello MAC nelle reti wireless è quello responsabile della gestione della condivisione del canale da parte degli utenti. Esso comprende le procedure e le funzioni necessarie per trasferire dati tra due o più nodi presenti nella rete; in particolare esso ha il compito di mettere in pratica specifici algoritmi per funzionalità come *framing*, indirizzamento fisico, *flow* ed *error control*, ovvero la correzione degli errori causati da anomalie che si verificano a livello fisico. È responsabile, inoltre, della risoluzione degli eventuali conflitti che si verificano tra nodi per l'accesso al canale. Lo scopo del livello MAC è, dunque, consentire un'efficiente condivisione del canale wireless tra i nodi che hanno bisogno di accedervi. Per garantire questa efficienza, il MAC deve servirsi del riuso spaziale del canale, ossia deve fornire il maggior numero di comunicazioni simultanee possibile. Dal momento che il livello MAC ha una diretta relazione sull'affidabilità e l'efficienza con cui i dati possono essere trasmessi tra due nodi, esso influenza anche la QoS complessiva di rete.

2.1 Livello MAC nelle reti ad hoc

In virtù di quelle che sono le caratteristiche di una rete ad hoc, che sono state presentate nel capitolo precedente, un protocollo di livello MAC per una rete ad hoc dovrebbe avere le seguenti proprietà:

- funzionamento distribuito;

- utilizzo efficiente e *fair* della banda a disposizione;
- ritardo di accesso (in pratica il tempo necessario ad un pacchetto per esser trasmesso) più basso possibile;
- overhead di controllo più basso possibile;
- scalabilità;
- meccanismi di *rate control* e di *power control*;
- supporto della QoS per traffico *real-time*;
- minimizzazione degli effetti di *hidden terminal problem* e *exposed terminal problem*.

Come è noto, questi ultimi 2 sono i problemi principali che si presentano al livello MAC nelle reti wireless in generale.

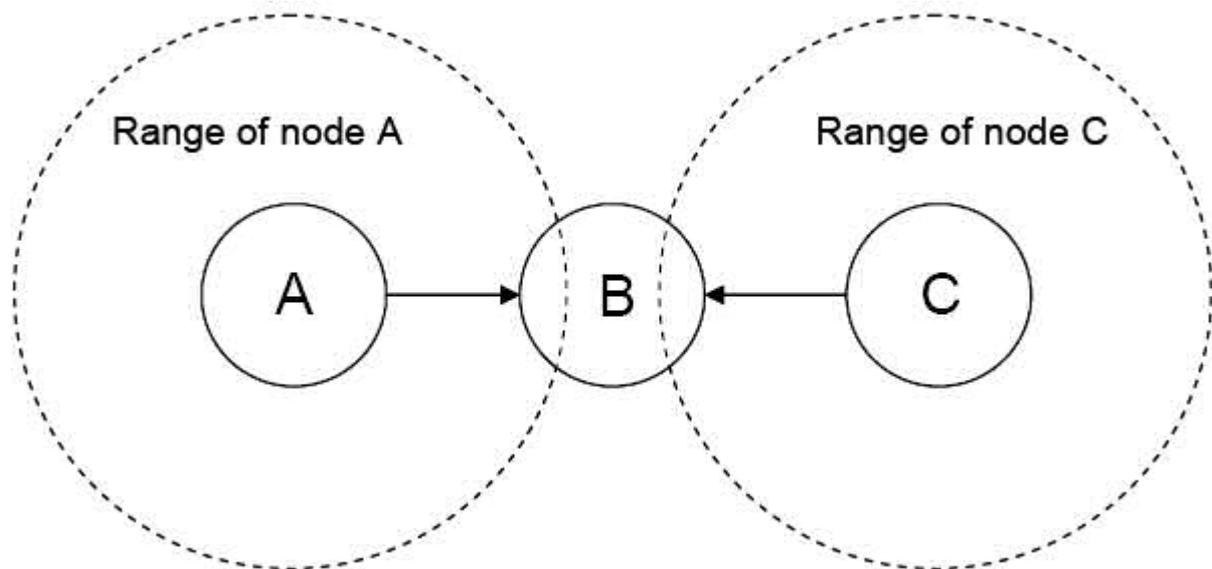


Figura 2.1 – Hidden terminal problem

Il primo di essi (fig. 2.1) consiste nel fatto che il nodo *B*, come mostrato in figura, è all'interno dei range trasmissivi sia di *A* sia di *C*, però *A* non si trova nel range di *C* e viceversa (cioè sono *nascosti* l'uno rispetto all'altro). Allora *C*, ad esempio, non può rendersi conto se *A* sta trasmettendo a *B*, quindi *C* ed *A* potrebbero trasmettere contemporaneamente a *B*, causando una collisione presso *B*.

Il secondo problema (fig. 2.2), invece, si presenta se, nella situazione mostrata in figura, è *B* che trasmette ad *A*. Dal momento che si trova all'interno del range di *B*, *C* sente la presenza di una trasmissione (è *esposto* a *B*) e decide di ritardare la propria verso *D*; questo, tuttavia, non è necessario perché l'eventuale trasmissione di *C* non potrebbe collidere col ricevitore *A*.

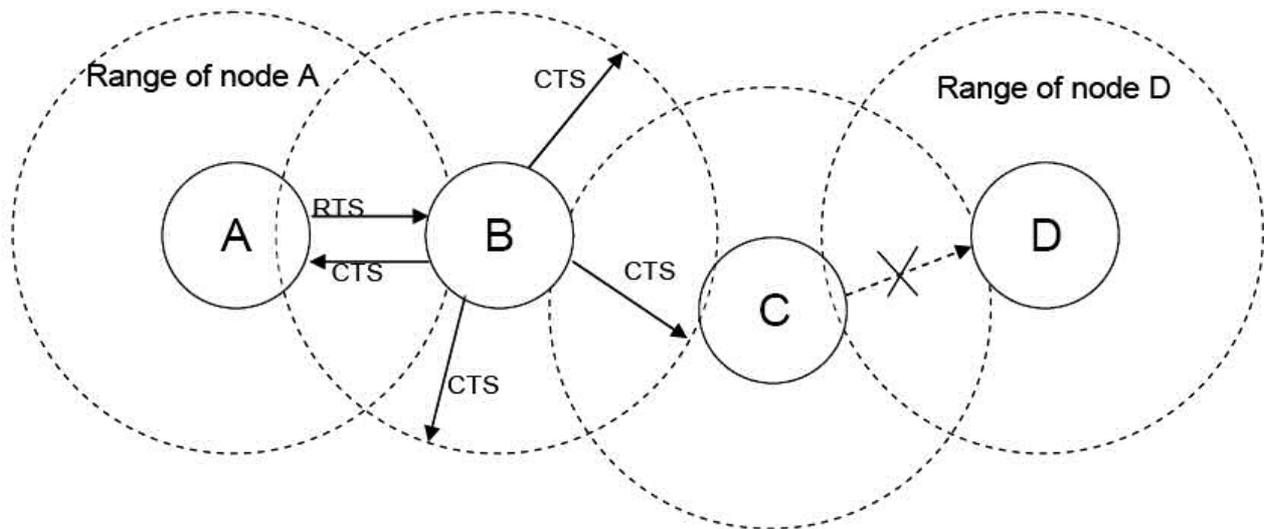


Figura 2.2 – Exposed terminal problem

Non esiste un unico criterio per classificare tutti i vari protocolli di livello MAC che sono stati proposti fino ad oggi. In [80] è presente una tassonomia che suddivide i vari protocolli MAC per ad hoc network in base a 6 criteri, che sono:

- la tecnica di accesso e la suddivisione del canale;
- la topologia per la quale i protocolli sono progettati;
- la potenza, ossia i protocolli sono classificati in base alla loro efficienza energetica;
- l'iniziazione della trasmissione: la maggior parte dei protocolli per reti centralizzate adottano un approccio di tipo *sender-initiated*, ma le ad hoc network supportano un esteso insieme di applicazioni, per alcune delle quali un approccio *receiver-initiated* può risultare più adatto;
- la scalabilità e il carico di traffico, ossia si classificano i protocolli in base alle loro prestazioni, identificando le rispettive ideali condizioni di rete, inclusa la densità dei nodi e il carico di traffico;
- il range trasmissivo, che è un aspetto altamente legato alle prestazioni, alla flessibilità e alla mobilità di ciascun protocollo: ognuno di essi, infatti, è progettato per uno scenario applicativo specifico, per cui il range operativo varia ampiamente da un protocollo all'altro.

L'obiettivo dell'elaborato è analizzare l'applicazione di un particolare algoritmo di power control in una rete ad hoc basata su tecnica d'accesso CDMA. In questo capitolo, pertanto, si è voluto classificare e confrontare vari protocolli MAC sulla base del primo degli aspetti prima elencati, ossia la tecnica di accesso, allo scopo di giungere, alla fine, alla descrizione del "CDMA in ad hoc",

ossia del contesto operativo cui verrà applicato il power control, esaminato in dettaglio nel capitolo successivo.

2.2 Tecniche di accesso multiplo in reti ad hoc

Il metodo usato dagli utenti per accedere al canale è detto *tecnica di accesso multiplo* e costituisce uno degli aspetti chiave per un protocollo di livello MAC.

Le tecniche di accesso multiplo possono essere distinte in 2 grossi gruppi: le tecniche basate sulla contesa o *contention-based* e le tecniche *contention-free*. Le prime, in generale, sono caratterizzate dal fatto di essere prive di una qualche forma di allocazione *a priori* delle risorse: in pratica, quando un pacchetto deve essere trasmesso, un nodo compete coi suoi vicini per l'accesso al canale condiviso. Dei protocolli basati su queste tecniche, le collisioni tra trasmissioni simultanee sono un aspetto intrinseco. Le seconde, al contrario, non hanno bisogno che i nodi competano tra loro, ma fanno in modo che questi abbiano a disposizione una "porzione" del totale delle risorse disponibili.

2.2.1 Tecniche Contention-Based

I protocolli appartenenti a questa categoria, possono essere suddivisi a loro volta in due sottocategorie: protocolli *random access* e protocolli *dynamic reservation/collision resolution*. Dei primi fanno parte schemi come ALOHA, in cui un nodo può accedere al canale quando vuole. Ovviamente, più di un nodo può trasmettere nello stesso istante, per cui si possono verificare collisioni tra le trasmissioni. ALOHA è, pertanto, più adatto a sistemi con bassi traffici poiché offre un *throughput* relativamente basso. Una variante di ALOHA è chiamata "Slotted ALOHA", la quale introduce slot temporali per trasmissioni sincronizzate, similmente a TDMA. Si dimostra che tale modifica raddoppia il *throughput* rispetto all'ALOHA puro, al prezzo però della necessità di sincronizzazione tra nodi.

Tra i secondi, quello di gran lunga più usato per le reti ad hoc è CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) presente nell'IEEE 802.11, che attualmente costituisce lo standard *de facto* per il livello MAC della corrente generazione di ad hoc network, sebbene esso non sia stato originariamente progettato per lavorare in un tale contesto operativo, e per questo motivo, è stato descritto piuttosto dettagliatamente nel seguito.

2.2.1.1 CSMA/CA dell'IEEE 802.11

L'idea alla base di questa tecnica consiste nell'informare, mediante brevi frame, i nodi vicini a trasmettitore e ricevitore che sta per cominciare una comunicazione e, in tal modo, evitare che tali nodi inizino a trasmettere. La tecnica si basa sull'uso di 4 frame per il trasferimento di ogni

pacchetto dati. Quando il sender A vuole trasmettere un frame al receiver B , prima invia un frame RTS (Request To Send), al quale B risponde con un frame CTS (Clear To Send). Alla ricezione di CTS, A può cominciare a trasmettere i DATA. Infine, B completa il trasferimento con l'invio di un Acknowledgment (ACK) (fig. 2.3).

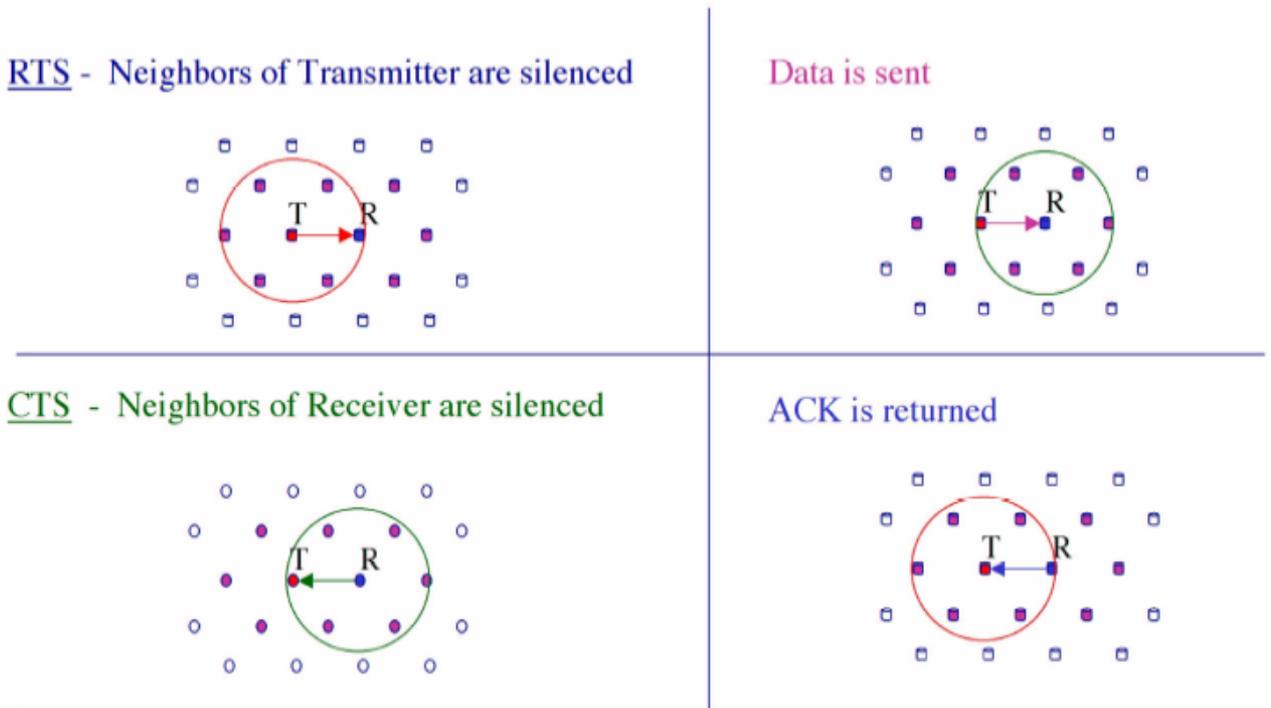


Figura 2.3 – Handshake RTS/CTS

Entrambi i frame RTS e CTS contengono il tempo mancante alla fine della trasmissione, per cui ogni stazione nel raggio d'azione di A o B riceverà, a sua volta, uno o entrambi i frame, ed imposterà un particolare indicatore, chiamato NAV (*Network Allocation Vector*), con la durata indicata dal frame. Questo NAV non è altro che un contatore che viene decrementato nel tempo, fino a 0; quando il NAV di un nodo è diverso da zero, significa che una trasmissione è in atto nelle vicinanze. Il comportamento seguito dai nodi vicini è detto *Virtual Carrier Sensing* (VCS).

Questa procedura, così com'è descritta, funziona bene solo nel caso teorico in cui le stazioni abbiano tutte lo stesso raggio d'azione e i frame RTS e CTS possano essere scambiati in tempo infinitesimo. In caso contrario si possono avere collisioni frequenti dovute a tentativi di acquisizione contemporanea del mezzo; per ridurre la probabilità, al *carrier sensing virtuale* ne viene affiancato uno *reale*, dando luogo al vero e proprio CSMA/CA che è descritto di seguito:

- 1) La stazione trasmittente A cerca di determinare lo stato del mezzo valutando il contenuto di NAV ed *ascoltando* il mezzo. Il canale è considerato libero, quando sia il *Carrier Sensing Virtuale* che quello *Reale* non rilevano attività. I casi sono due:

- a. se il canale rimane libero per un intervallo di tempo detto DIFS (successivamente sarà spiegato con maggior precisione), salta al punto 3;
 - b. se invece il canale è occupato (o viene occupato durante l'intervallo DIFS), prosegue al punto 2;
- 2) *A* avvia la procedura di backoff (discussa più avanti);
 - 3) *A* emette un RTS;
 - 4) se entro un intervallo di tempo ben definito, *A* non riceve il CTS da *B*, vuol dire, molto probabilmente, che l'RTS ha colliso con un altro frame; spesso ciò significa che due stazioni hanno scelto lo stesso slot nella finestra di backoff. Per questo motivo, prima di ritentare la trasmissione, *A* raddoppia la dimensione della finestra di backoff (*CW*) e poi ripete dal punto 2. Lo scopo di tale raddoppio è quello di adattare la dimensione della finestra al numero di contendenti, in considerazione del fatto che le collisioni sono indice di "affollamento";
 - 5) quando *B* riceve l'RTS, risponde con un CTS;
 - 6) ricevuto il CTS, *A* può cominciare a trasmettere il frame contenente i dati veri e propri.
 - 7) se entro un intervallo di tempo ben definito, *A* non riceve un ACK da *B*, vuol dire che il frame *Dati* non è stato ricevuto correttamente, e quindi *A* deve ritrasmetterlo ripetendo tutta la procedura;
 - 8) una volta che *B* ha ricevuto correttamente il frame *Dati*, risponde con un ACK concludendo il protocollo.

La figura 2.4 mostra il comportamento di 4 stazioni durante la trasmissione da *A* a *B*; la stazione *D* si trova nel raggio d'azione di *B*, ma non in quello di *A*, e quindi aggiorna il proprio NAV solo dopo aver ricevuto il CTS proveniente da *B*.

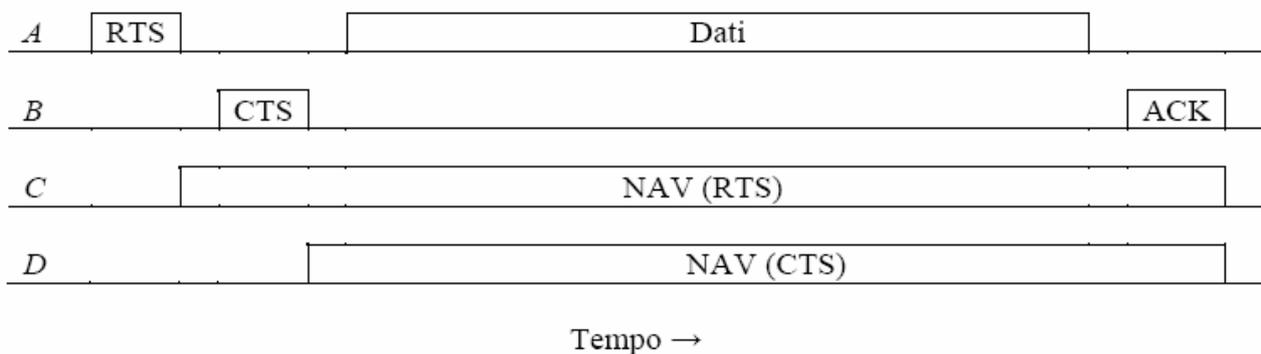


Figura 2.4 – Esempio di funzionamento CSMA/CA

2.2.1.2 Procedura di backoff

La procedura di backoff è una procedura di attesa per un tempo casuale e limitato, basata sull'algoritmo di *binary exponential backoff*. Grazie ad essa, si può evitare che più stazioni, che attendono che il canale si liberi, tentino di acquisire contemporaneamente il canale nell'istante in cui questo viene rilasciato. L'algoritmo è il seguente:

- a. A aspetta che il canale si liberi, e che rimanga libero per tutto un intervallo DIFS;
- b. solo se il contatore di backoff di A è a 0, sceglie un numero a caso compreso fra 0 e CW (larghezza corrente della finestra di contesa) altrimenti lo lascia al valore attuale (che è il residuo di una procedura di backoff precedentemente interrotta);
- c. per ogni intervallo *SlotTime* in cui il canale rimane libero, il contatore viene decrementato; se il mezzo viene occupato, torna al punto 1;
- d. quanto il contatore giunge a zero, la procedura di backoff termina.

Un esempio in merito alla dimensione della finestra di backoff (CW): inizialmente CW è pari a 7 (CW_{min}); le ritrasmissioni, come visto, implicano un raddoppio del numero di slot, portando CW a 15, 31, 63, 127, 255. Arrivata a 255 (CW_{max}), la dimensione non cresce più. Al corretto completamento di una trasmissione, CW viene riportato a 7.

Nello standard 802.11 sono previste due modalità di funzionamento: la prima, chiamata DCF (*Distributed Coordination Function*), prevede che siano le stazioni a gestire, in modo distribuito, l'accesso al mezzo, secondo il protocollo CSMA/CA. La seconda, chiamata PCF (*Point Coordination Function*) affida ad un Access Point il coordinamento di tutte le stazioni nella sua cella. In DCF è prevista una tecnica di frammentazione dei frame: un frame può essere scomposto in più frammenti, ognuno numerato e riscontrato (ACK) separatamente. La motivazione è presto detta: le trasmissioni radio sono affette da un'elevata rumorosità e, se un frame è troppo grande, la probabilità che venga danneggiato diventa elevatissima. La frammentazione consente di restringere le ritrasmissioni ai soli frammenti danneggiati, anziché all'intero frame. Notare che una volta acquisito il canale con RTS e CTS, più frammenti possono essere inviati in sequenza, dando origine a quello che viene chiamato *fragment burst* (fig. 2.5).

In PCF, la stazione base interroga ciclicamente le altre stazioni (procedura detta *polling*), le quali possono trasmettere solo quando è il loro turno; siccome l'ordine delle trasmissioni è completamente controllato dalla stazione base, non si possono mai verificare collisioni. Tutte le stazioni devono supportare DCF, mentre PCF è opzionale, e raramente implementato.

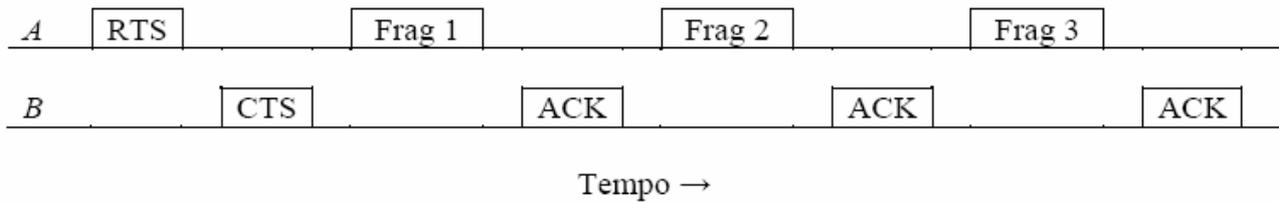


Figura 2.5 – Fragment burst

DCF e PCF possono coesistere all'interno di una stessa cella, grazie ad un'intelligente assegnazione dei tempi d'attesa.

Lo standard definisce 4 intervalli di tempo, i quali forniscono diversi livelli di priorità ai vari protocolli:

- SIFS (Short Inter Frame Space): è l'intervallo più breve, usato per separare i frame appartenenti ad una singola trasmissione. Ad esempio, gli intervalli tra i frame RTS, CTS, Frag ed ACK di un fragment burst, hanno tutti questa durata. Notare che una sola stazione ha il permesso di rispondere dopo un intervallo SIFS. In questo modo si impedisce alle altre stazioni, le quali devono attendere per un periodo più lungo, di interrompere una trasmissione in corso;
- PIFS (PCF Inter Frame Space): trascorso un intervallo più lungo, PIFS, è il turno dall'AP (chiamato anche Point Coordinator, in questo contesto);
- DIFS (DCF Inter Frame Space): se l'AP non ha nulla da dire, trascorso l'intervallo DIFS, ogni stazione può tentare di acquisire il canale per iniziare una nuova trasmissione; è a questo punto che possono verificarsi le collisioni per accesso contemporaneo, ed infatti la finestra di contesa (*slotted backoff window*) comincia proprio dopo questo intervallo;
- EIFS (Extended Inter Frame Space) - usato al posto di DIFS dalle stazioni che hanno ricevuto un frame incomprensibile, dal quale, quindi, non è stato possibile estrarre l'informazione necessaria all'aggiornamento dell'indicatore NAV. Questo intervallo è stato studiato in modo da consentire ad un'altra stazione di rispondere al frame ignoto, risincronizzando anche questa stazione.

2.2.1.3 Svantaggi e limiti

Il meccanismo del VCS, sebbene allevi in una certa misura il problema del terminale nascosto, ha due grossi svantaggi: anzitutto, esso ha un impatto negativo sull'utilizzazione del canale, poiché non consente a due trasmissioni simultanee di aver luogo, dal momento che i nodi nelle vicinanze sono bloccati. La situazione è mostrata in figura 2.6, in cui è evidente l'assunzione che i nodi siano

dotati di antenne omnidirezionali. Il nodo A utilizza la sua massima potenza trasmissiva per inviare i suoi pacchetti a B ; i nodi C e D sentono il messaggio CTS inviato da B , per cui rinunciano a trasmettere. È facile intuire che entrambe le trasmissioni, $A \rightarrow B$ e $C \rightarrow D$, potrebbero aver luogo contemporaneamente, se i nodi fossero in grado di selezionare opportunamente la loro potenza di trasmissione.

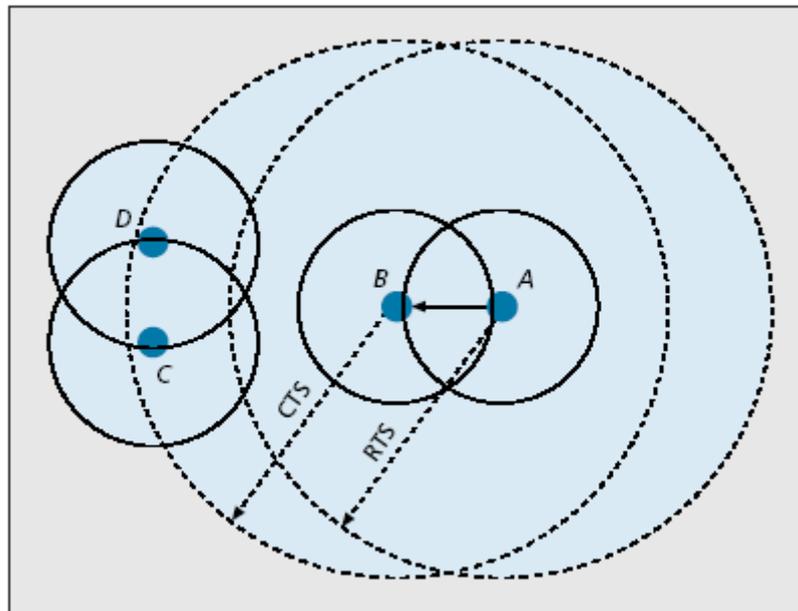


Figura 2.6 – Sottoutilizzazione del canale

In secondo luogo, se si estende l'area in cui il *carrier sensing* è effettivo, si peggiora il problema della stazione esposta, comportando un inevitabile sottoutilizzo della banda disponibile. Non esiste praticamente nessuno schema che riesce a risolvere questa problematica per l'802.11: l'unica soluzione avanzata di una certa efficacia per l'*exposed station problem* sarebbe l'uso di antenne direzionali [66]. In ogni caso, usando CSMA/CA, i nodi devono rimanere nello stato attivo (*idle listening*) per un tempo di solito maggiore rispetto a una tecnica *contention free*, come sarà chiaro successivamente, cosa che peggiora le prestazioni energetiche del sistema; senza contare, poi, che il problema delle collisioni, sebbene alleviato, non risulta eliminato del tutto, cosa che, oltre a portare inefficienza energetica, rende imprevedibile il ritardo trasmissivo e può aumentare il tasso di scarto dei pacchetti. In particolare, sono gli aspetti del multicast e del broadcast che risentono del problema delle collisioni, per difendersi dalle quali un nodo o è costretto a trasmettere lo stesso pacchetto multicast più volte (una per ogni gruppo multicast vicino), oppure lascia che i pacchetti vengano inviati con probabilità di corretta ricezione bassa come quella del protocollo ALOHA. In sostanza, l'*handshake* RTS/CTS riesce a contrastare bene il problema della *hidden station* in una

rete wireless in cui tutti i nodi possono sentire le trasmissioni di ogni altro nodo. La condizione sufficiente per non avere nodi nascosti è che un qualsiasi nodo che può eventualmente interferire con la ricezione di un pacchetto dal nodo A a B si trovi all'interno del *sensing range* di A . Questa condizione, ovviamente, può benissimo non esser soddisfatta in una rete multi-hop. Il grosso vantaggio che CSMA/CA possiede rimane, comunque, il fatto che è autonomo e non richiede alcun controllo esterno, cosa che lo rende tutto sommato piuttosto semplice da implementare.

2.2.1.4 Estensioni ed alternative

Esistono moltissime altre varianti ed estensioni all'IEEE 802.11: nel seguito vengono menzionate quelle ritenute più interessanti. In [81] è proposto MARCH, Multiple Access with Reduced Handshake, che tenta di ridurre la segnalazione di controllo, pur mantenendo il paradigma RTS/CTS. Lo scambio di informazioni di controllo impiegato in MARCH è mostrato in figura 2.7.

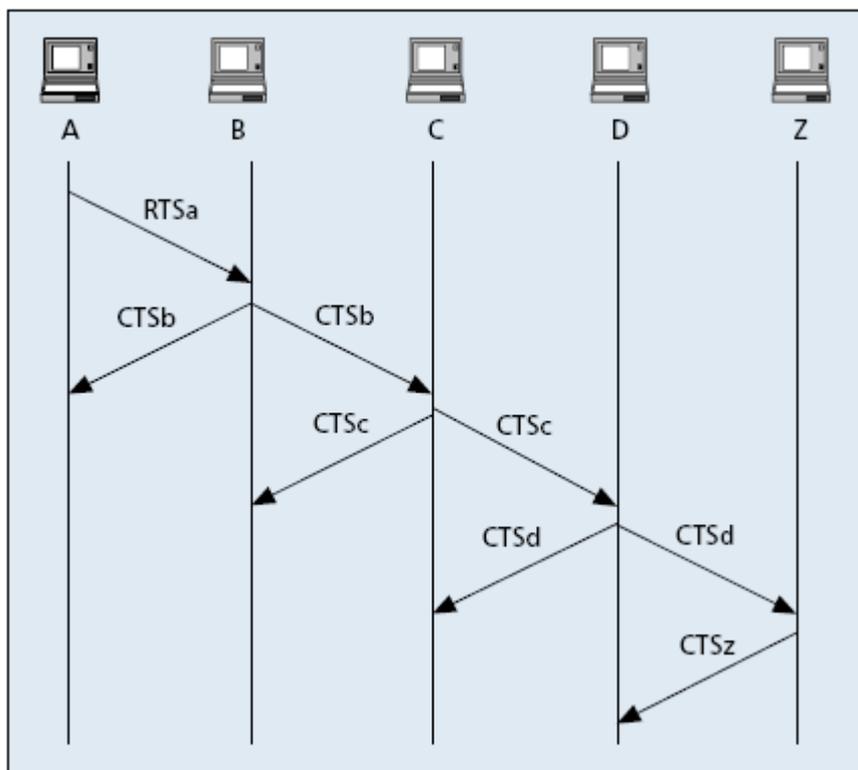


Figura 2.7 – Handshake in MARCH

Si supponga che il nodo A abbia dati da inviare al nodo Z , usando il percorso A, B, C, D, Z . A invia RTS_A al next hop sul percorso, B . Quando B risponde ad A con CTS_B , C sente quel messaggio e viene a sapere che B invierà dati per A , così dovrà rispondere a B con un CTS_C all'istante appropriato. Lo stesso processo è ripetuto al nodo D e Z . In questo modo, MARCH propone un

meccanismo con un unico RTS nel primo hop del percorso, mentre per ogni successivo hop è richiesto solo un CTS.

In [82] è proposto DPC/ALP (Distributed Power Control with Active Link Protection/Adaptive Probing), anch'esso basato sul paradigma RTS/CTS. In esso, il sender invia un RTS a un livello di potenza tale da farlo sembrare rumore e continua a inviarlo incrementandone progressivamente la potenza, fintantoché il receiver non risponde con un CTS. Se la potenza usata per un RTS eccede una certa soglia senza risposta dal ricevitore, il sender interrompe il tentativo. Questo meccanismo fa sì che un RTS interferisca minimamente con le altre trasmissioni in atto, dal momento che supererà di poco la potenza di rumore.

Infine, in [83] viene proposto PS-DCC (Power Save with Distributed Contention Control), che costituisce un meccanismo di backoff probabilistico per ogni canale di comunicazione in reti IEEE 802.11. Il metodo prevede che i nodi misurino costantemente l'utilizzazione del canale e adattivamente calcolino la probabilità di invio in base al traffico di rete misurato: l'algoritmo fornisce prestazioni migliori rispetto a uno schema di backoff statico.

Quelli menzionati finora sono protocolli che considerano il mezzo come un unico canale che i nodi devono spartirsi; esistono, tuttavia, molti altri protocolli che separano il piano di controllo dal piano dati, assegnando un canale per la segnalazione di controllo e uno o più distinti canali per le trasmissioni dati. Alcuni tra questi metodi descrivono uno schema generalizzato di separazione del canale, in cui lo schema di accesso è basato sull'avere a disposizione canali multipli, però essi non specificano come separare questi canali. In qualche modo, quindi, questi protocolli sono una sorta di via di mezzo tra quelli *contention-based* e quelli *contention-free*. Il protocollo BTMA (Busy Tone Multiple Access) [84], ad esempio, è un protocollo usato in un'architettura di rete centralizzata, in cui una base station che sente un canale dati occupato, trasmette un'onda sinusoidale (il *busy tone*) sul canale di controllo per impedire agli altri nodi di trasmettere. Dual Busy Tone Multiple Access (DBTMA) [85] è una recente estensione del precedente schema, che prevede l'utilizzo di due *busy tone* su canali separati da quelli dati, allo scopo di proteggere gli RTS ed evitare che i nodi vicini al ricevitore inizino a trasmettere. Il carattere distribuito che differenzia questo metodo dal precedente, lo rende applicabile alle reti ad hoc per la soluzione dell'*hidden terminal problem*. Power Aware Multiple Access with Signaling (PAMAS) [40], propone l'uso di un canale di controllo per lo scambio dei frame RTS/CTS e un canale separato per i dati. In entrambi i frame di controllo, è inclusa la durata delle trasmissioni che avverranno: in tal modo, i nodi vicini che sentono i messaggi RTS o CTS possono passare allo *sleep mode* per il periodo previsto per quella trasmissione, riducendo il consumo di energia e la probabilità di collisioni. Dynamic Channel Assignment with Power Control (DCA-PC) si basa sulla stessa filosofia dei

precedenti, ossia su una suddivisione generalizzata del canale, poiché prevede un canale di controllo e N canali dati. Il protocollo prevede che il sender controlli se alcuni dei canali dati appaiono liberi e, in caso affermativo, ne scelga uno ed invii un segnale di RTS alla destinazione sul canale comune di controllo con massima potenza. Se la destinazione accetta la scelta del canale fatta dal sender, essa risponde con un CTS avente un livello di potenza appropriato per raggiungere il sender; altrimenti, invia la propria lista di canali liberi al sender in modo che questo possa scegliere un canale più adatto.

2.2.2 Tecniche Contention-Free

Le tecniche *contention-free* sono, invece, meccanismi di accesso basati sullo *scheduling* delle risorse; le principali, sono: *Frequency-Division Multiple Access (FDMA)*, *Time-Division Multiple Access (TDMA)*, e *Code-Division Multiple Access (CDMA)*. Questa classificazione è basata sul dominio cui appartengono le risorse che vengono suddivise. FDMA divide lo spettro frequenziale disponibile in più sotto-bande non sovrapposte, ciascuna assegnata a un nodo diverso: in tal modo essi accedono al canale contemporaneamente l'un con l'altro, ma operano a frequenze distinte. Dall'altro lato, TDMA assegna distinti slot temporali a ogni utente, i quali adesso condividono la stessa banda frequenziale, però accedono al mezzo a differenti intervalli temporali. CDMA si basa, invece, su codici ortogonali assegnati a ciascun utente: se a ciascuno è assegnato un codice diverso, possono coesistere trasmissioni simultanee sulle stesse frequenze. Queste tre tecniche sono state largamente studiate in letteratura. FDMA è stata applicata ai sistemi mobili di prima generazione, mentre le altre due tecniche vengono ampiamente usate come metodi di accesso multiplo nei sistemi di seconda generazione. Con l'introduzione di esse, i link wireless occupati da distinte trasmissioni possono esser mantenuti attivi contemporaneamente e senza collisioni; per questo lo *scheduling* in frequenza, nel tempo o nel dominio dei codici, diviene essenziale per la coesistenza e il coordinamento di trasmissioni da parte di utenti indipendenti, in quanto esso elimina (o, quanto meno, riduce nella pratica) l'interferenza. In tal modo, si elimina anche lo spreco di energia dovuto a ritrasmissioni e, in più, si può aumentare il *throughput* di rete e l'effettiva utilizzazione del canale. Rispetto, quindi, ai protocolli MAC ad accesso casuale, nei quali i nodi si contendono le risorse di canale ogni volta che devono inviare un pacchetto, quelli ad accesso "programmato" consentono un più efficiente sfruttamento complessivo dell'energia. Il prezzo che si paga con l'uso di queste tecniche di accesso consiste nella complessità che caratterizza il controllo necessario per la fase di setup e di mantenimento dello *scheduling*, complessità che può (e non deve), tuttavia, controbilanciare il risparmio energetico ottenuto con la riduzione delle collisioni, se non addirittura peggiorare il consumo energetico complessivo. Un'altra restrizione all'applicazione di queste

tecniche sta nella dinamicità della topologia: essi richiedono che quest'ultima rimanga piuttosto stabile o, comunque, vari piuttosto lentamente, in modo che l'effetto dello *scheduling* abbia durata tale da compensare l'elevato overhead della fase di setup. Tutto questo fa sì che esse non possano essere direttamente utilizzate in contesti caratterizzati da una troppo elevata dinamicità. Da un punto di vista puramente teorico, comunque, i tre suddetti approcci *scheduling-based* presentano le stesse identiche prestazioni, in termini di efficienza spettrale ed energetica. Tuttavia, è soprattutto CDMA che attrae la maggiore attenzione della ricerca, in quanto esso, per la maggior parte dei sistemi wireless, offre vantaggi maggiori. In un sistema TDMA, ad esempio, la sincronizzazione dei nodi è un requisito molto più stringente e critico rispetto a quella necessaria in un sistema CDMA e per questo un protocollo basato su TDMA deve comprendere meccanismi di controllo del sincronismo molto complicati e pesanti. In CDMA è, invece, sufficiente praticamente la sola sincronizzazione tra sender e receiver, raggiunta la quale gli utenti possono trasmettere quando vogliono, condividendo lo stesso dominio temporale e frequenziale, perché tanto i loro segnali sono distinti mediante codici di spreading. Inoltre, è stato dimostrato [54] che la capacità complessiva di rete raggiunta con questa tecnica può arrivare ad essere superiore a quella ottenuta con soluzioni basate su TDMA/FDMA.

2.2.2.1 TDMA

Nei sistemi basati su TDMA, l'asse temporale viene suddiviso in un certo numero di trame di lunghezza fissa e ogni trama viene a sua volta divisa in un certo numero di intervalli temporali, detti slot. Ciascun utente può trasmettere solo durante lo slot temporale che gli è stato preventivamente assegnato, in modo da ridurre l'interferenza tra nodi vicini. Durante tale intervallo, il nodo ha a completa disposizione l'intera banda del canale. Ogni trama è divisa in due parti (fig. 2.8), una di controllo e una per i dati. La dimensione della parte di controllo è molto minore rispetto a quella dati.

Il problema fondamentale per l'applicazione della tecnica TDMA ad un'ad hoc network è il protocollo di assegnamento degli slot temporali a ciascun nodo. Infatti, in un tale contesto, oltre a non essere presente un'entità centrale di riferimento, responsabile dell'assegnazione dei vari slot ai vari nodi, che, pertanto, sono costretti ad agire in maniera distribuita (come, del resto, lo sono per molti altri aspetti), l'assegnazione degli slot deve sapersi adattare ai continui cambiamenti topologici che si possono verificare nella rete. In sostanza, non si può parlare di assegnamento statico o fisso: esso deve necessariamente essere dinamico e distribuito.

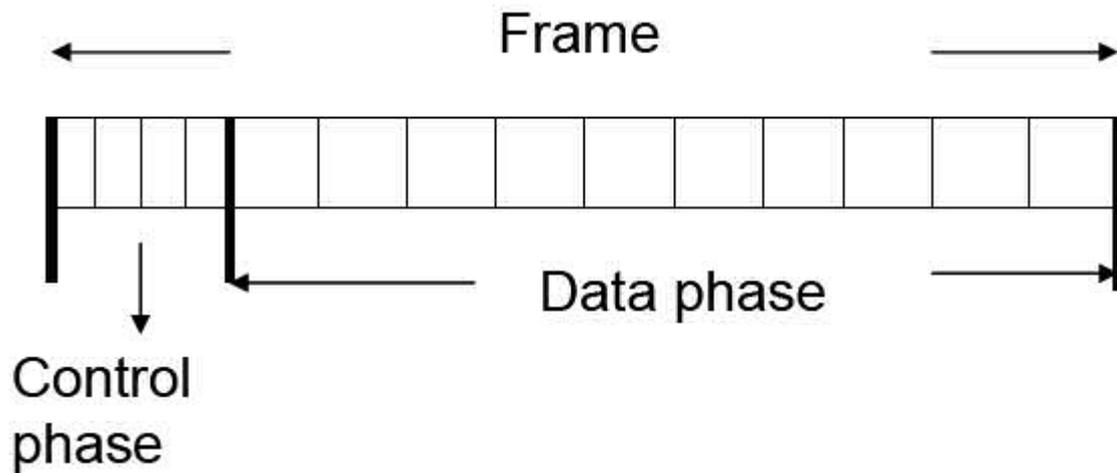


Figura 2.8 – Frame TDMA

TDMA, pertanto, richiede che ciascun nodo monitori l'assegnamento degli slot temporali: quando un nodo sente la trasmissione di un suo vicino, verrà a conoscenza degli slot da lui usati, per cui cancellerà tali slot dalla sua *table list*. Ad esempio, in figura 2.9, il nodo *S* ha bisogno di uno slot per inviare un pacchetto al nodo *D*. Anzitutto, gli slot 1 e 2 non possono essere considerati perché il primo è usato da *S* per trasmettere e il secondo è usato da *D* per ricevere; poi, gli slot 3 e 4 non possono essere usati perché provocherebbero collisione ad *A* e *B*. Infine, gli slot 5 e 6 non possono essere usati perché *B* e *C* stanno trasmettendo in quegli slot: solo gli slot 7 e 8 possono essere usati.

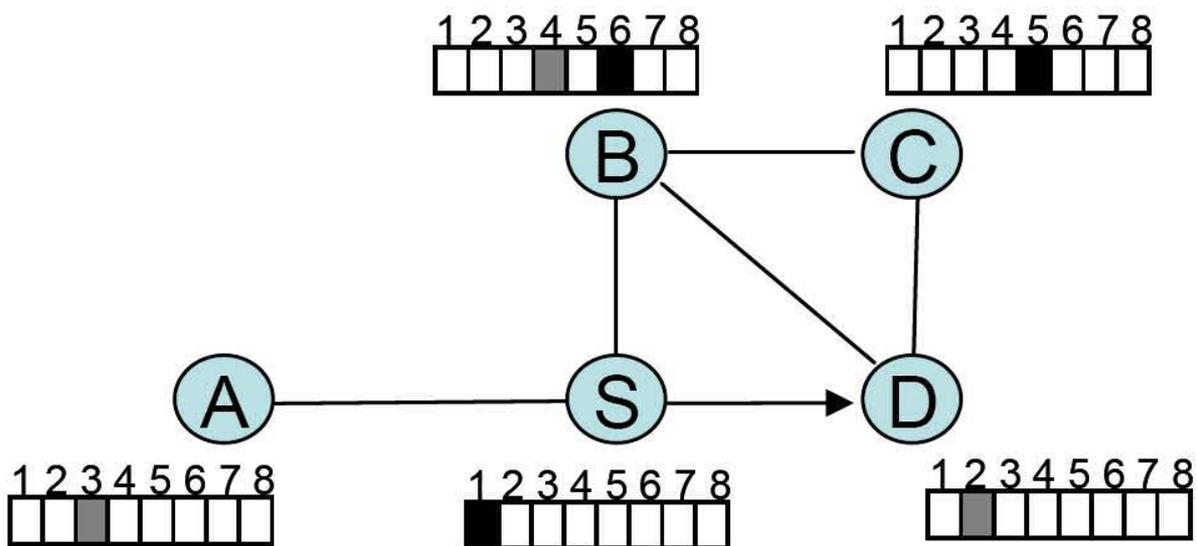


Figura 2.9 – Assegnamento degli slot in TDMA

Uno schema TDMA può risolvere i problemi del terminale nascosto e del terminale esposto. Si consideri la figura 2.10: la sorgente *A* trasmette pacchetti al nodo di destinazione *D* e tutti gli slot

nella trama sono assunti liberi. Supponiamo che *A* riservi gli slot 1 e 2 per trasmettere pacchetti a *B* e quest'ultimo, a sua volta, utilizzi gli slot 3 e 4 per inviare pacchetti a *C*. Poiché *A* e *C* sono nascosti l'uno rispetto all'altro, *C* può voler inviare pacchetti a *D* usando gli slot 1 e 2: in tal caso, si genera una collisione in *B*, perché *B* può ricevere pacchetti da *A* e *C* simultaneamente. Quindi, *C* può usare gli slot 5 e 6 per inoltrare pacchetti a *D*, per evitare la collisione in *B*. Per risolvere i suddetti problemi, dunque, si devono poter usare più slot temporali.

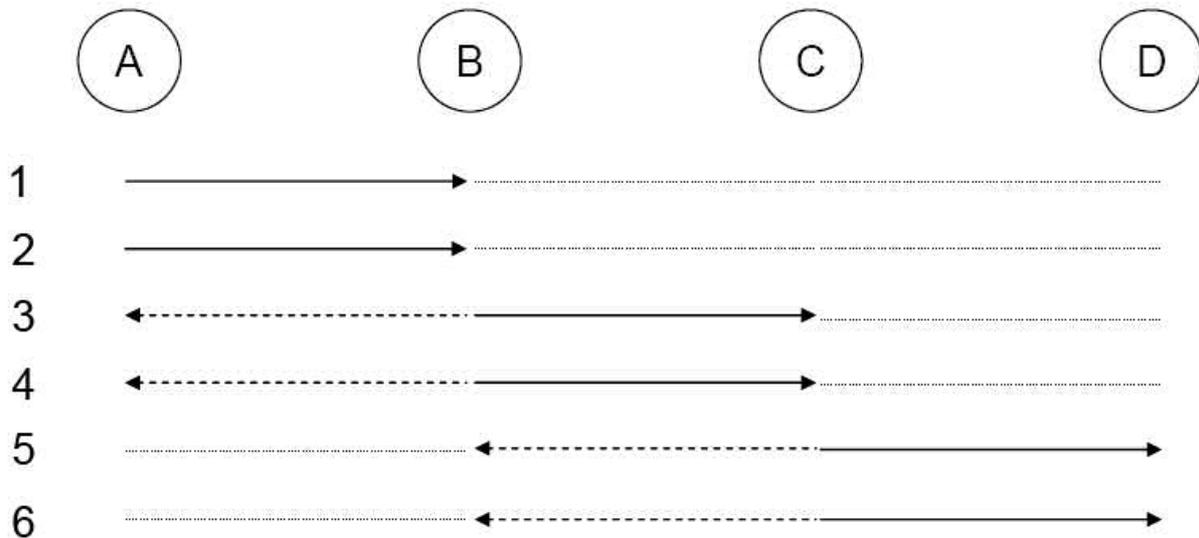


Figura 2.10 – Soluzione all'hidden terminal problem

La natura “multi-hop” delle MANET fa sì che qualsiasi schema di accesso al canale basato su TDMA abbia a che fare con problemi che non sono presenti in una convenzionale rete totalmente connessa, per la quale sono stati progettati numerosi protocolli di assegnamento degli slot TDMA. Infatti, in una multi-hop network:

- i nodi che sono vicini l'uno all'altro non necessariamente condividono lo stesso insieme di vicini; quindi, un qualsiasi schema di assegnamento degli slot deve essere accettato non solo tra nodi immediatamente vicini, ma deve propagarsi per almeno due hop;
- qualsiasi protocollo deve tener conto che la topologia è dinamica e, pertanto, deve essere in grado di adattarsi in maniera relativamente rapida ad ogni cambiamento nella connettività tra nodi.

Sono stati proposti, nel recente passato, diversi protocolli in grado di fornire un'allocazione dinamica degli slot, senza la necessità della presenza di base station. Questi protocolli possono essere classificati in *topology-independent* (o *topology-transparent*) e *topology-dependent*.

I primi, [71,72,73,74,75,76,77], prevedono l'assegnamento dinamico e distribuito ad un nodo di una *transmission schedule*, una sorta di elenco temporale in cui sono specificati gli intervalli durante i quali il nodo può trasmettere e ricevere senza interferire con i nodi distanti da lui 1 o 2 hop, aumentando, così, il riuso del canale al diminuire del numero di vicini per nodo. L'assegnazione degli slot in questi protocolli è basata fondamentalmente su contese tra nodi. Tali meccanismi, inoltre, si basano sulla divisione di ogni slot in diversi mini-slot e la contesa avviene per ciascuno di questi mini-slot. Tutto ciò, limita la minima durata che gli slot possono avere. Questi algoritmi hanno lo svantaggio che la loro efficienza e robustezza è vulnerabile in un contesto altamente mobile, in quanto:

- il processo di calcolo per lo *scheduling* può non convergere mai, se la velocità di variazione della topologia supera la velocità con cui il protocollo riesce a ricalcolare e distribuire le nuove *schedule*. Se D è il numero medio di nodi vicini, ossia distanti un unico hop da qualsiasi nodo (detto anche *grado* della rete grado della rete), allora il numero medio di nodi distanti 2 hop da qualsiasi nodo è proporzionale a D^2 (poiché ogni nodo ha mediamente D vicini, ciascuno con altrettanti nodi vicini). Con un protocollo *topology-dependent*, la *transmission schedule* deve essere ricalcolata ogni volta che due nodi, precedentemente non vicini (cioè distanti più di 1 hop), lo diventano; dal momento che un nuovo vicino può essere ottenuto solo dall'insieme dei nodi distanti 2 hop, la frequenza di ricalcolo è proporzionale a D^2 , così come al grado di mobilità della rete. Pertanto, per reti che sono altamente mobili e moderatamente dense, la topologia è notevolmente mutevole, cosa che, per *scheduling topology-dependent*, rende altrettanto instabili le *transmission schedule*;
- nel processo di riassegnamento degli slot in un insieme di nodi distanti 2 hop, può essere necessario un overhead elevato, che causa un eccessivo consumo di risorse preziose, come banda e potenza;
- a seconda della successione degli eventi, della densità della rete e della particolare connettività, il processo di riassegnamento degli slot, all'interno di un dato insieme di nodi con distanza 2 hop, può innescare il riassegnamento in un altro insieme di nodi adiacente, causando una sorta di "effetto *ripple*" che può potenzialmente propagarsi anche a grande distanza dalla zona d'origine;
- durante il periodo di transizione nel processo di riassegnamento degli slot, le trasmissioni sono perse; quanto più è instabile la topologia, tanto maggiore sarà il numero di trasmissioni perse. Se i pacchetti persi contengono informazioni necessarie per la "riparazione" della rete,

o informazioni riguardanti lo *scheduling* per trasmissioni point-to-point, le prestazioni di rete possono subire un duro colpo durante questi transienti;

- per la maggior parte di questi tipi di protocollo, il limite superiore sul numero di slot temporali tra due trasmissioni consecutive provenienti dallo stesso nodo è proporzionale a D^2 . Per cui, il *throughput* di un'individuale trasmissione broadcast decade quadraticamente in funzione del grado della rete, provocando una forte degradazione prestazionale in reti da moderatamente ad altamente dense.

Nei secondi [68,69], a ciascun nodo viene pre-assegnata (in base all'ID, ad esempio) o, comunque, in base a criteri prefissati, ognuno di essi adotta una *transmission schedule*, che viene resa pubblica e in cui sono specificati gli slot utilizzati dal nodo per la trasmissione. Questi protocolli garantiscono, o comunque forniscono un'elevata probabilità, che almeno un intervallo temporale nella *schedule* di un nodo non sia in conflitto con qualche altro nodo distante uno o due hop. Nei metodi presentati in [68,69], i nodi non sono in grado di determinare quali trasmissioni avverranno, complicando, così, il lavoro dei protocolli di livello superiore. Gli algoritmi presenti, inoltre, richiedono la conoscenza di due valori di ingresso, che sono il numero totale di nodi nella rete e il numero massimo di vicini per ciascun nodo. Il metodo proposto in [68], poi, ha prestazioni peggiori di CSMA in termini di ritardo e *throughput*. L'approccio seguito in [70], assume che ogni nodo sia sincronizzato coi propri vicini (eliminando, così, a priori il problema delle collisioni), abbia la conoscenza delle loro *schedule* e sia capace di ricevere da più trasmettitori vicini simultaneamente. Quest'ultima assunzione, richiede un hardware molto sofisticato. I limiti che hanno questi tipi di protocollo sono:

- il sender non è in grado di conoscere quali nodi vicini siano in grado di ricevere correttamente il pacchetto inviato in un particolare slot temporale; per cui, tali metodi non possono esser accoppiati con un meccanismo come l'*handshake* RTS/CTS, usato per coordinare l'attivazione senza collisioni di un link tra un particolare sender e un particolare receiver, o un qualsiasi altro meccanismo che richiede la presenza di un unico trasmettitore attivo in una zona estesa 2 hop;
- il numero di slot temporali tra due trasmissioni consecutive (di nuovi pacchetti) dallo stesso nodo è proporzionale a D_{max}^2 (con D che è il grado della rete); quindi l'efficienza di questi metodi decresce in maniera quadratica al crescere della densità;
- questi metodi non sono adatti per reti che presentano elevata variabilità di densità e/o dimensioni.

In [67], si propone un approccio unificato per l'assegnamento degli slot per un accesso TDMA e per il routing distribuito: il protocollo proposto per la *dynamic reservation* degli slot fornisce anche informazioni sulla connettività di rete, che possono essere usate per costruire un meccanismo di routing in grado di rispondere ai cambiamenti topologici della rete. In questo protocollo, una parte degli slot di una trama viene usata per il broadcast di un numero sufficiente di richieste di *reservation*: lo scopo è far sì che ad ogni nodo venga fornita una sufficiente informazione sull'uso degli slot nei nodi a lui vicini, in modo che esso ne possa scegliere uno in maniera efficiente. Questa informazione viene anche usata per costruire le tabelle di routing.

Un accesso al mezzo *time-based* sarebbe teoricamente in grado di offrire molti vantaggi dal punto di vista dell'ottimizzazione dell'energia: risulterebbero, infatti, minimizzati fattori come l'energia sprecata per l'*overhearing*, le collisioni, il periodo di *idle mode* e le transizioni tra i diversi stati funzionali (rif. capitolo 1). In più, si migliorerebbe la capacità di prevedere i ritardi e si limiterebbero gli scarti dei pacchetti causati da interferenza e buffer overflow. Gli svantaggi, non banali, di una tale tecnica sono il fatto che si tratta di un problema NP-hard, cosa che ne rende ancor più ardua la scalabilità, e il fatto che essa introduce un eccessivo overhead protocollare. Senza dimenticare l'essenziale necessità di mantenere il sincronismo di clock tra i vari nodi. Queste problematiche compensano i vantaggi teorici propri del TDMA e lo rendono particolarmente inefficiente dal punto di vista dell'utilizzo delle risorse. Si è cercato di superare il problema della sincronizzazione in [79], in cui si propone un TDMA *asincrono*, con un approccio del tutto simile a quello seguito dalla tecnologia Bluetooth. Seguendo questo approccio, ogni link usa un riferimento temporale fornito dal clock di sistema di uno dei due endpoint; slot temporali vengono persi per l'allineamento del riferimento temporale, per cui ovviamente il periodo minimo necessario per realizzare una data allocazione degli slot sarà maggiore rispetto di quello richiesto da un sistema perfettamente sincronizzato. Questo aumento può esser visto come l'overhead legato all'asincronia del sistema. In [79], si cerca di trovare il modo di minimizzare questo overhead. Il limite principale che affligge questo metodo sta nella scarsa scalabilità. In alternativa, per aumentare l'efficienza prestazionale di TDMA, un'altra possibilità è quella di usare Spatial-TDMA (STDMA) [65], che tiene conto del fatto che le unità radio sono di solito piuttosto sparpagliate geograficamente, per cui unità con una sufficiente separazione spaziale possono usare lo stesso slot temporale per la trasmissione. In STDMA, lo spazio è diviso in celle geografiche virtuali (*space slot*) che vengono raggruppate in trame virtuali che si ripetono periodicamente (*space frame*) allo scopo di facilitare il riuso spaziale, e gli slot temporali vengono assegnati agli *space slot*. Ogni nodo, però, deve esser dotato di un mezzo per determinare la propria locazione istantanea (ad es. GPS) per essere a conoscenza dello *space slot* cui appartiene e, di conseguenza, dello slot ad esso assegnato.

2.2.2.2 CDMA

CDMA è basato su tecniche di espansione dello spettro (*spread spectrum* – SS), mediante le quali ogni utente utilizza l'intera banda a disposizione. Al trasmettitore, un segnale digitale di banda B_1 viene spettralmente espanso usando (ossia “moltiplicando per” o, meglio ancora, “effettuando un'operazione di XOR bit a bit tra pacchetto da inviare e”) una sequenza di codice pseudo-casuale di banda B_2 , con $B_2 \gg B_1$. Questa sequenza, detta anche di pseudo-rumore, è binaria e statisticamente soddisfa i requisiti di una sequenza puramente casuale, solo che essa può essere esattamente riprodotta al ricevitore mediante precise regole matematiche. Usando questa sequenza riprodotta localmente, il ricevitore “ricomprime” (operazione detta *despreading*) spettralmente il segnale ricevuto, per recuperare da esso l'informazione originale. Mediante questa espansione-compressione spettrale, si possono far coesistere contemporaneamente distinti e indipendenti segnali codificati con la stessa procedura all'interno dello stesso canale frequenziale, a patto che per ogni segnale venga usata una sequenza di codice diversa. Questo aspetto è cruciale per reti con traffico *bursty*, elevata densità di nodi e range trasmissivi asimmetrici. Nel paragrafo precedente, si è già accennato ai grossi vantaggi che porta con sé l'uso di questa tecnica; ad essi vanno aggiunti la resistenza al *multipath*, un'intrinseca diversità in frequenza e una forte riduzione dell'impatto dell'interferenza. L'espansione della potenza del segnale su un intervallo frequenziale molto più ampio rispetto alla banda del segnale originario da trasmettere, fa sì che un sistema CDMA goda di immunità da:

- fading selettivo, eventualmente localizzato nella banda originaria del segnale trasmesso, in quanto, adesso, un eventuale *notch* di attenuazione non colpisce più gran parte delle frequenze del segnale, ma solo una piccola parte, in quanto, ora, la banda è molto più estesa;
- interferenza a banda larga, in quanto al ricevitore il segnale viene riportato in banda base e filtrato e l'interferenza su quel range ristretto di frequenze è minima (effetto del *despreading*);
- interferenza a banda stretta, in quanto, con lo *spreading*, anche questa viene spalmata su una banda molto più larga ma in ricezione non viene riportata in banda base, rimanendo ‘spalmata’ su un range molto più grande, per cui risulta essere minima nella banda base del segnale (effetto del *despreading*);
- intercettazioni, in quanto solo il ricevitore corrispondente conosce le sequenze di *despreading* relativa al segnale che si vuole ricevere.

È per tutti questi motivi che questa tecnica di accesso è stata la scelta per i sistemi cellulari, inclusi i sistemi 3G recentemente adottati, e si propone anche come tecnica da usare nelle reti ad hoc. Lo standard IEEE 802.11 (peraltro ormai obsoleto) utilizza tecniche SS a livello fisico (nelle specifiche sono presenti sia la Direct Sequence – DSSS, sia la Frequency Hopping – FHSS, ma è preferita la DS per alcune desiderabili proprietà), però solo per mitigare l’impatto che ha il canale wireless. Più specificatamente, nel protocollo 802.11 tutti i segnali trasmessi vengono espansi usando una stessa sequenza di pseudo-rumore, precludendo la possibilità di trasmissioni multiple contemporanee nei pressi del ricevitore.

2.2.2.2.1 CDMA in ad hoc network

I vantaggi derivanti dall’uso del CDMA in un ad hoc network possono esser riassunti così:

- *Capacity improvement*, in quanto si aumenta il numero di corrette ricezioni a livello datalink;
- *Energy saving*, in quanto si riduce l’eventuale intervallo temporale che un nodo deve attendere prima di trasmettere, perché, senza CDMA, il canale deve esser sentito *idle*: in tal modo per un nodo si riduce la fase di *idle listening* (che, comunque, non può essere eliminata completamente) che, come è stato visto nel capitolo precedente, consuma energia; dall’altro lato, il consumo energetico è ridotto in quanto l’accesso *scheduled* consente di evitare collisioni;
- *Routing overhead reduction*, in quanto è più semplice implementare il routing multi-path nelle CDMA ad hoc network.

Utilizzare CDMA come tecnica di accesso in una rete ad hoc è, tuttavia, un obiettivo non banale da raggiungere. Non sono, infatti, ancora risolti diversi importanti problemi, tra cui la sincronizzazione, l’implementazione di un protocollo di *coding* e il *near-far effect*, che nascono essenzialmente a causa del comportamento non coordinato dei nodi e dal fatto che manca un controllo centralizzato, presente, invece, in una rete cellulare grazie alla base station.

Sebbene meno stringenti rispetto a TDMA, in cui in teoria tutti i nodi devono essere sincronizzati, anche CDMA possiede dei requisiti di sincronismo, in particolare necessita una accurata fase di sincronizzazione tra sender e receiver. Mentre in un sistema cellulare, il problema è risolto abbastanza agevolmente mediante l’uso di canali globali di controllo usati dalla base station, in una rete ad hoc esso a prima vista sembra aver la stessa natura del problema “prima l’uovo o la gallina?”, in quanto come fanno gli utenti a sincronizzarsi se non possono comunicare? Sono, in realtà state, proposte varie soluzioni, che spaziano da un complesso schema pseudo-casuale a tecniche di

accesso casuale più standard come ALOHA. Una tecnica innovativa prevede la creazione di un canale basato su CSMA a banda stretta per l'instaurazione della connessione, che consente ai nodi di effettuare la sincronizzazione dei codici ed può aiutare anche il power control e la stima del canale.

L'assenza di un controllo centralizzato, inoltre, rende di non facile soluzione altre due problematiche strettamente legate tra loro. La prima consiste nello stabilire un *protocollo di assegnamento del codice*, in modo che a distinti terminali siano associati codici distinti. Si tratta di un problema banale in una rete piccola, ma diventa complicato in reti di grandi dimensioni, in cui il numero di sequenze di *spreading* diviene più piccolo del numero di terminali e la conseguente necessità di un riuso spaziale di esse (il numero di codici è di solito vincolato dallo spettro disponibile e dal rate informativo richiesto). Sono stati proposti diversi protocolli di assegnamento [56,57,58]: in essi, in pratica, si cerca di assegnare codici con il vincolo che a tutti i nodi vicini a un certo nodo siano assegnati codici diversi. La seconda problematica, dipendente strettamente dalla prima, consiste nel determinare un protocollo per decidere quali codici propriamente usare per la trasmissione di un certo pacchetto verso una certa destinazione e per monitorare il canale prima della ricezione di un pacchetto [59]. Un tale protocollo può essere *receiver-based*, *transmitter-based* o ibrido. In un protocollo *receiver-based*, il trasmettitore, per lo *spreading* del pacchetto, usa il codice del ricevitore desiderato, mentre un terminale *idle* monitora costantemente il proprio codice. Questo approccio semplifica la circuiteria del ricevitore, poiché quest'ultimo non deve monitorare l'intero insieme di codici presente nella rete. Sfortunatamente, anche sotto un corretto assegnamento dei codici, sono ancora possibili collisioni *primarie*, ossia che avvengono tra segnali che sono stati espansi con lo stesso codice. Ad esempio, due nodi "non vicini" tra loro (cioè fuori da entrambi i rispettivi range trasmissivi) *A* e *C*, con due codici differenti, possono avere un vicino comune, il nodo *B*, che ha un suo codice: se *A* e *C* contemporaneamente tentano di trasmettere a *B*, avviene una collisione perché usano lo stesso codice, ossia quello di *B* (problema analogo all'*hidden terminal problem*). Un modo per garantire che non avvengano collisioni primarie è usare codici diversi associati ai segnali trasmessi simultaneamente o, in alternativa, far precedere la trasmissione da una fase di controllo del canale. Un altro svantaggio dell'approccio *receiver-based* è il fatto che inviare un messaggio in broadcast imporrebbe a un trasmettitore di inviare tale messaggio in unicast ad ogni ricevitore, usando i rispettivi codici. In un protocollo di *spreading-code* di tipo *transmitter-based*, invece, deve essere assegnato ad ogni terminale un codice per la trasmissione e i ricevitori devono essere in grado di monitorare l'attività dell'intero insieme di codici pseudo-casuali. Con un tale metodo, teoricamente le collisioni primarie non possono accadere. In più, il broadcast è intrinsecamente supportato. Il grosso svantaggio è che la circuiteria dei ricevitori deve essere molto

complessa e costosa e, soprattutto, questo approccio causa un consumo inutile di energie da parte dei ricevitori non destinatari dei pacchetti. Ciascun nodo a cui arriva una trasmissione, infatti, deve risalire all'indirizzo di destinazione, presente nell'header del pacchetto, per capire se è uno dei receiver desiderati. Tuttavia sono possibili anche approcci ibridi: in [59] sono proposti due schemi ibridi, *common transmitter-based protocol* e il *receiver-transmitter-based protocol*. Nel primo, il campo dell'header del pacchetto che contiene gli indirizzi di sorgente e destinazione sono espansi usando un codice comune, mentre il resto del pacchetto viene espanso mediante il codice del trasmettitore. Un terminale *idle* costantemente monitora il codice comune: se riconosce il proprio indirizzo nel campo di destinazione, il nodo seleziona il codice del trasmettitore per ricevere il resto del pacchetto. Il problema nasce quando più trasmettitori trasmettono contemporaneamente allo stesso ricevitore. Il secondo protocollo ibrido prevede l'assegnamento a ogni nodo di due codici di espansione, uno per la ricezione e uno per la trasmissione. I campi dell'header di un pacchetto che contengono gli indirizzi di sorgente e di destinazione, vengono espansi usando il codice del ricevitore desiderato, mentre il resto del pacchetto è espanso col codice del trasmettitore. Un terminale *idle* monitora costantemente il suo codice di ricezione: se riconosce il proprio indirizzo e possiede il codice della sorgente, usa questo per recuperare il resto del pacchetto. Uno svantaggio è che se due nodi trasmettono contemporaneamente allo stesso ricevitore, questo non può decodificare completamente l'informazione proveniente da ciascuna sorgente. Inoltre, il numero di codici da assegnare con questo metodo sarebbe doppio. Esiste anche un altro approccio, detto *pairwise-based*, in cui a ciascuna coppia di nodi, viene assegnato un unico codice: in tal modo il trasmettitore avrà una tabella di assegnamento in cui cercherà il codice da usare per trasmettere ad un certo ricevitore; quest'ultimo dovrà comunque monitorare un insieme di codici simultaneamente, cosa che, come si è detto, rende complessa l'implementazione circuitale del dispositivo. Inoltre il numero di codici necessari per un tale schema cresce eccessivamente col numero di nodi presenti.

A causa del dinamismo che caratterizza la topologia, è poi necessario che la distribuzione dei codici venga aggiornata nel tempo. Questa fase di *recoding* determina un ulteriore overhead per l'acquisizione del canale da lato ricevitore e la necessità di risincronizzazione tra sender e receiver. Si tratta di un processo pesante e costoso in termini di risorse, che deve esser ridotto il più possibile [63].

Finora sono stati proposti diversi protocolli MAC basati su CDMA per le MANET [58,59,60,61,62]. Tutti si basano su un accesso casuale al canale, ossia un nodo che ha pacchetti da trasmettere può farlo quando vuole (possibilmente con un preventivo scambio di RTS/CTS), senza tener conto dello stato del canale: in tal caso, si parla infatti di *random access* CDMA (RA-CDMA). Sotto un corretto assegnamento dei codici e un appropriato schema di *spreading-code*, questi protocolli RA-

CDMA garantiscono assenza di collisioni primarie. È purtroppo presente un altro tipo di problematica, che è legata alla non perfetta ortogonalità che può caratterizzare i codici di espansione: tale problematica è nota come *near-far effect*.

2.2.2.2.2 Near-far effect

Quando un ricevitore CDMA effettua il *despreading* sul segnale, quello che fa è calcolarsi la cross-correlazione tra il segnale e la sequenza di pseudo-rumore generata localmente: se questa sequenza locale è identica a quella usata per espandere il segnale al trasmettitore (ossia il messaggio è effettivamente indirizzato a *questo* ricevitore), il calcolo consente di risalire ai dati informativi originari. Altrimenti, tale calcolo dà come risultato un valore pari a 0 o diverso da 0, a seconda che il sistema sia *sincrono* o *asincrono*. Un sistema è detto *sincrono*, se tutti i segnali sono generati dallo stesso trasmettitore, come nel caso del downlink di una rete CDMA cellulare. In teoria, è possibile avere un sistema sincrono anche in presenza di più trasmettitori, però, nella pratica, è difficilissimo raggiungere una perfetta sincronizzazione tra tutti. La sincronia che caratterizza un tale sistema si manifesta in due modi. Per prima cosa, si ha che distinte trasmissioni destinate a distinti ricevitori hanno un riferimento temporale comune; poi, dal punto di vista di un dato terminale mobile, tutti i segnali si propagano attraverso gli stessi percorsi e, in tal modo, sperimentano gli stessi ritardi. In un sistema sincrono, è possibile progettare codici di *spreading* completamente ortogonali: infatti, nello standard IS-95 per i sistemi cellulari basati su CDMA, ad ogni utente del canale è assegnato un codice di Hadamard (o di Walsh). Viceversa, un sistema è detto *asincrono* se i segnali sono prodotti da più trasmettitori contemporaneamente, come nel caso dell'uplink di un sistema cellulare o in una MANET. Anche in questo caso, ci sono due motivi da cui trae origine l'asincronia. Primo, dal momento che i segnali sono prodotti da distinti e indipendenti trasmettitori, non è possibile avere un riferimento temporale comune per tutte le trasmissioni che arrivano al ricevitore. Secondo, questi segnali si propagano lungo diversi percorsi, cosicché sperimentano ritardi differenti. Tutto questo fa sì che, in un sistema asincrono, non sia possibile progettare codici di espansione che risultino ortogonali per ogni intervallo temporale: in tal caso, la cross-correlazione tra i codici può non essere trascurabile. Infatti, codici che sono ortogonali in un sistema sincrono hanno un valore elevato di cross-correlazione quando non sono perfettamente sincronizzati; per questo, in un sistema asincrono, vengono usati codici che hanno sempre un limitato valore di cross-correlazione. La cross-correlazione non nulla tra codici può, infatti, introdurre la cosiddetta *interferenza da accesso multiplo* (*multi-access interference* – MAI) presso un ricevitore, durante le fase di *despreading*: un ricevitore che, infatti, sta cercando di rivelare il segnale proveniente dall'*i*-esimo trasmettitore, può esser fisicamente molto più vicino ad

un altro trasmettitore, indicato con la lettera j . Se tutte le potenze trasmissive sono uguali, il segnale proveniente dal j -esimo trasmettitore arriverà al ricevitore in questione con un potenza sufficientemente più grande rispetto a quella del segnale proveniente dall' i -esimo, causandone, a meno che i codici non siano perfettamente ortogonali, una decodifica non corretta: quest'ultimo fenomeno è detto anche collisione *secondaria*.

2.2.2.2.3 Impatto dell'interferenza da accesso multiplo

Si consideri la ricezione di un pacchetto al terminale indicato dalla lettera i e si indichi con $P_0^{(i)}$ la potenza media ricevuta del segnale desiderato al nodo stesso. Si supponga che ci siano K trasmissioni interferenti con potenze ricevute P_j , con $j = 1, \dots, K$. La qualità della ricezione desiderata è adeguatamente misurata dal rapporto tra energia per bit e densità spettrale di rumore efficace (*effective bit energy-to-noise spectral density ratio*) al ricevitore, indicata con $E_b/N_{0\text{eff}}$. Per un sistema *direct-sequence* asincrono BPSK, si ha che

$$\left(\frac{E_b}{N_{0\text{eff}}} \right)^{(i)} = \left(\frac{2 \sum_{j=1}^K P_j}{3WP_0^{(i)}} + \frac{1}{\mu_0} \right)^{-1}$$

dove W è lo *spreading factor* e μ_0 è il valore di $E_b/N_{0\text{eff}}$ al ricevitore in assenza di interferenza. Maggiore è la potenza interferente, minore sarà il valore del rapporto e maggiore di conseguenza sarà la probabilità di errore. Come esempio, si consideri un sistema CDMA che usa una modulazione BPSK e un codice convoluzionale con rate $1/2$ e *constraint length* 7, e decodifica soft di Viterbi. Si prenda $W = 100$. Allora, per ottenere una probabilità di errore pari a 10^{-6} , il rapporto $E_b/N_{0\text{eff}}$ richiesto è 5.0 dB. Trascurando il rumore termico e utilizzando l'espressione suddetta, la potenza totale interferente deve soddisfare

$$\frac{\sum_{j=1}^K P_j}{P_0^{(i)}} \leq 47.43$$

I trasmettitori sono, di solito, localizzati a distanze differenti dal ricevitore; si supponga che le potenze trasmissive siano tutte fisse e uguali e si consideri il caso di un unico interferente ($K = 1$) a distanza d_1 dal ricevitore. Sia d_0 , invece, la distanza del trasmettitore dal ricevitore desiderato; allora, usando il modello di propagazione a due raggi per comunicazioni terrestri (attenuazione $\sim 1/d^4$), è facile dimostrare che per ottenere il *bit error rate* voluto, si deve avere $d_1 \geq 0.38d_0$. Per cui, se c'è anche solo un interferente a distanza minore di $0.38d_0$ dal ricevitore, non è possibile instaurare una comunicazione affidabile, poiché si ha una collisione secondaria.

2.2.2.2.4 Caratteristiche necessarie per l'applicazione di un protocollo MAC basato su CDMA

Il *near-far effect* che è stato appena descritto può, pertanto, causare una significativa riduzione nel *throughput* di rete e, ovviamente, non può esser trascurato nel progetto di un protocollo MAC basato su CDMA per una rete ad hoc. In un sistema cellulare, questo problema è stato praticamente risolto mediante un algoritmo di power control, che è brevemente illustrato nella sezione successiva e che, in pratica, garantisce che ogni terminale mobile provochi una *stessa* potenza di segnale ricevuto alla base station. Questo tipo di algoritmo non è, tuttavia, applicabile al contesto delle MANET per una serie di ragioni. Anzitutto, esso è implementabile grazie alla presenza della base station e, cioè, è centralizzato, mentre un contesto come quello di un ad hoc network richiede necessariamente un approccio distribuito. In secondo luogo, si osservi la figura 2.11 e si indichi con d_{ij} la distanza tra il nodo i e il nodo j . Come si può notare, A trasmette a B usando un determinato codice e C fa la stessa cosa con D , però usando un codice diverso. Si supponga che $d_{AB} \approx d_{CD}$, $d_{CB} \ll d_{AB}$ e che $d_{AD} \ll d_{CD}$. Allora, la MAI causata da C rende impossibile a B ricevere la trasmissione di A . Allo stesso modo, la MAI causata da A rende impossibile a D ricevere la trasmissione di C .

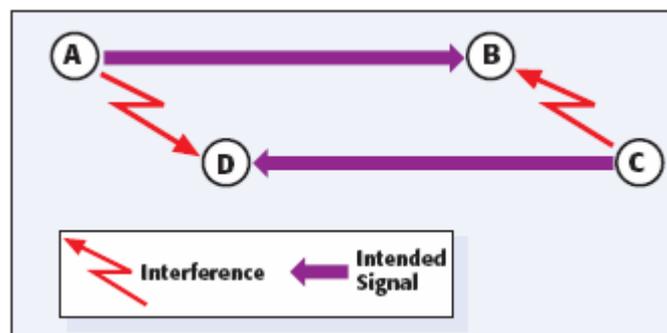


Figura 2.11 – Near-far effect

È importante notare che le due trasmissioni non possono aver luogo contemporaneamente e senza tener conto di quali potenze trasmissive sono state scelte: un eventuale comando di power control emesso da A per B , come ad esempio quello di innalzare la potenza trasmissiva di 10dB, può infatti scontrarsi con i bisogni di C , che sta già ricevendo troppa interferenza da B e che vorrebbe piuttosto che questi abbassasse la potenza.

Questo esempio, quindi, mette in luce due questioni: per prima cosa, può non esser possibile per due trasmissioni che usano due codici di *spreading* anche distinti avvenire nello stesso tempo; è questo, ovviamente, un problema di *medium access*. In secondo luogo, le due trasmissioni possono avvenire simultaneamente se i terminali trasmettitori aggiustano le potenze dei segnali inviati in modo che l'interferenza causata da una trasmissione non sia così elevata da impedire la corretta

ricezione da parte di altri terminali. Pertanto, la soluzione al problema del *near-far effect* deve comprendere senz'altro questi due aspetti: *power control* e *medium access*.

È importante distinguere il protocollo di *spreading code* e quello di livello MAC. Il primo regola la scelta dei codici di pseudo-rumore da usare per espandere il segnale, ma non risolve la contesa per l'accesso, della quale è, invece responsabile il protocollo di livello MAC, che ha lo scopo di migliorare l'utilizzazione della banda disponibile, riducendo o eliminando le collisioni. L'uso di un protocollo di livello MAC implica che se anche un terminale è dotato di un codice di espansione disponibile, può non avere il permesso di trasmettere.

A livello MAC, dunque, si deve avere che:

- il protocollo deve essere asincrono, distribuito e scalabile per grandi reti; deve, inoltre, richiedere il minimo scambio di informazioni di overhead, essere adatto per implementazione real-time e sapersi adattare a variazioni del canale e della topologia;
- la circuiteria del ricevitore non deve essere “troppo complessa”, nel senso che non deve essere necessario il monitoraggio dell'intero insieme di codici di espansione;
- il protocollo deve minimizzare (o, ancor meglio, eliminare) le collisioni anche se l'assegnamento dei codici non è “corretto”: questo perché è, di solito, difficile garantire un assegnamento costantemente corretto nonostante la mobilità dei nodi.

2.3 Power Control a livello fisico

Si tratta di un problema di gestione efficiente di una risorsa, che appunto è la potenza usata per la trasmissione dei segnali, che, come si è visto nel capitolo precedente, è particolarmente critica nelle reti wireless ad hoc. In pratica, il *power control* consiste in un insieme di algoritmi real-time implementati nei nodi di una rete, allo scopo di utilizzare la potenza in maniera globalmente “ottima”, ad esempio massimizzando (o minimizzando) una certa (o più) metrica prestazionale. Ogni schema di *power control* è progettato per una particolare applicazione e per certi dispositivi previsti. A livello propriamente fisico, si possono distinguere due schemi principali:

- Open Loop Power Control
- Closed Loop Power Control

2.3.1 Open Loop Power Control

In uno schema di tipo Open Loop, un nodo, *A*, misura la potenza ricevuta di un segnale di prova inviato dall'altro nodo, *B*, e, sulla base del confronto tra tale valore e quello nominale, stima

l'attenuazione di canale sperimentata dal segnale in quella trasmissione. Il nodo A può, così, variare il livello di potenza da usare per la sua trasmissione collegando il circuito responsabile del controllo automatico di guadagno (AGC) al guadagno del trasmettitore, come mostrato in figura 2.12. In questo modo, quando l'amplificatore a guadagno variabile al ricevitore incrementa il suo guadagno, esso incrementa anche il suo guadagno dell'amplificatore al trasmettitore. Il sistema CDMA IS-95 [55], lo standard americano per i sistemi cellulari di seconda generazione, implementa uno schema di power control di questo tipo. Lo scopo principale dell'introduzione del PC in questo sistema è, come già accennato, l'eliminazione del *near-far effect*, in modo che tutti gli utenti nel sistema possano controllare la propria potenza trasmessa così che dalla base station venga ricevuto approssimativamente lo stesso livello di potenza. L'Open-Loop Power Control cerca, quindi, di mantenere la potenza ricevuta vicina a un livello di target, *senza richiedere* alcun feedback da parte dell'altro nodo (la base station nel sistema cellulare), eliminando, così, anche il ritardo associato ad esso.

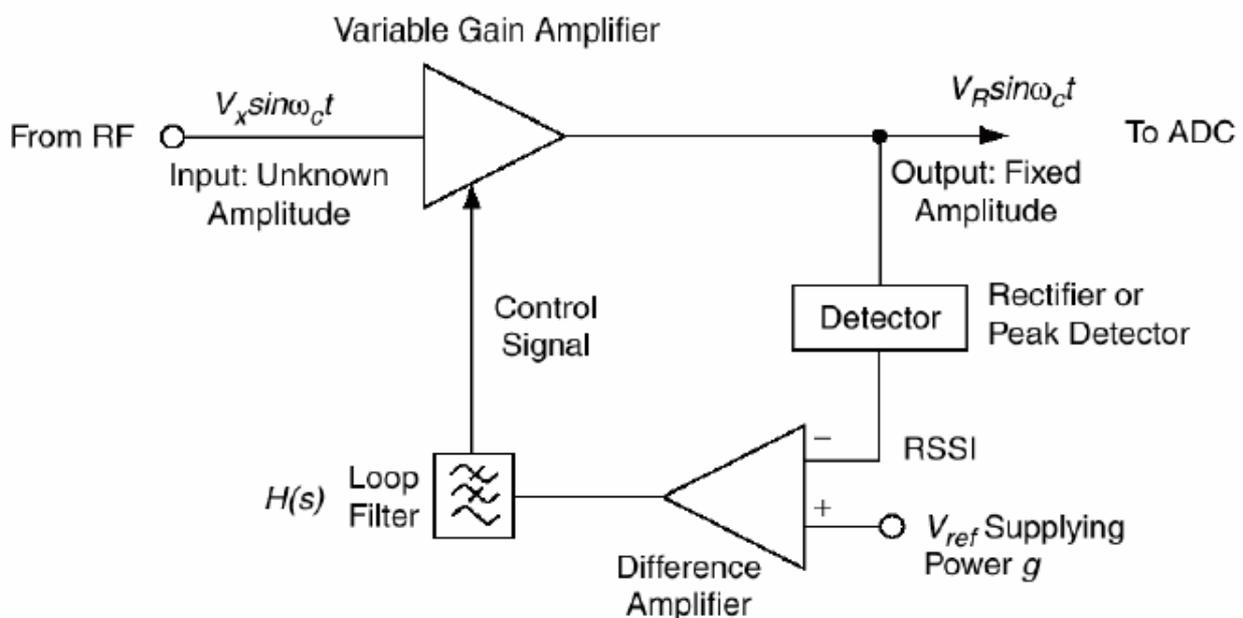


Figura 2.12 – Open Loop Power Control

Nello schema in figura 2.12 è il trasmettitore che svolge una misura sulla qualità del canale, per poi assumere che le caratteristiche da lui trovate siano simili a quelle nei pressi del nodo ricevitore. Questa assunzione di reciprocità si mantiene valida quando si effettuano stime su fenomeni come l'attenuazione da spazio libero (*path loss*), lo *shadowing* e il *fading* a lungo termine, ma non lo è più quando entra in gioco il *fading* a breve termine. La ragione di questo fatto è che la separazione frequenziale tra l'uplink e il downlink è superiore alla banda di coerenza del processo di *fading*, cosa che rende incorrelate le componenti su piccola scala. Gli effetti del fading a breve termine possono, perciò, differire notevolmente tra i due canali, in quanto essi generalmente sperimentano

differenti percorsi multipli (*multipath*). In tal modo, i due canali risultano non reciproci e l'informazione su di uno di essi fornisce scarsa informazione sull'altro. Un esempio di link non reciproco è mostrato in figura 2.13, in cui RSS sta per *received signal strength*.

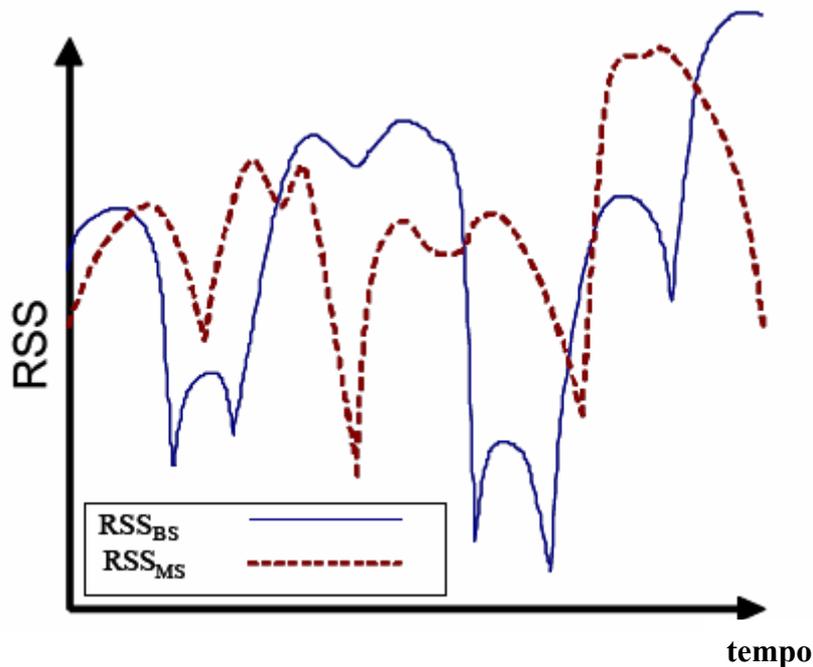


Figura 2.13 – Esempio di link non reciproco

2.3.2 Closed Loop Power Control

Per rimediare al problema della non-reciprocità del canale, deve esserci una comunicazione tra il nodo ricevitore, che effettua la stima del canale, e il trasmettitore, che deve aggiustare la potenza usata in base alle indicazioni fornitegli. Questo approccio è detto Closed Loop Power Control.

Nello standard IS-95, ad esempio, questo schema di power control esso consiste, a sua volta, in un *Inner Loop* e un *Outer Loop*. Col primo, si tenta ancora di mantenere un certo *SNIR target* ricevuto dalla base station. Quest'ultima, ogni 1.25 ms misura lo SNIR e comunica al mobile le indicazioni per modificare il proprio livello di potenza di ± 1 dB, tramite un bit di feedback. Questa modifica si aggiunge a quella calcolata mediante l'Open Loop, andando così a determinare l'effettivo livello di potenza trasmesso dal mobile. L'*Outer Loop* aggiorna ogni 10 ms lo SNIR target in funzione del tasso di errore sulla trama. In questo modo, nel IS-95 coi due Loop Power Control (Open e Closed) si tiene conto di entrambi i tipi di fading. Sebbene la presenza di questo feedback introduca necessariamente overhead, le prestazioni continuano ad esser buone, poiché la perdita è relativamente limitata (800 Hz nell'IS-95), fintantoché il canale continua a variare lentamente.

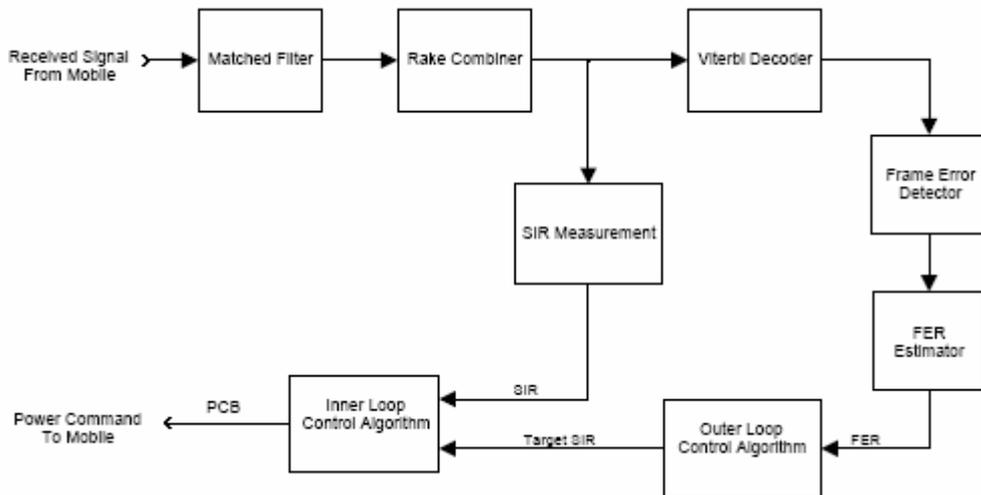


Figura 2.14 – Closed Loop Power Control

Tuttavia, se il contesto di rete è caratterizzato da un'elevata mobilità, le prestazioni calano drasticamente, a causa delle forti variazioni del canale. Un esempio di quello che può essere l'andamento dello SNIR ricevuto è in figura 2.15.

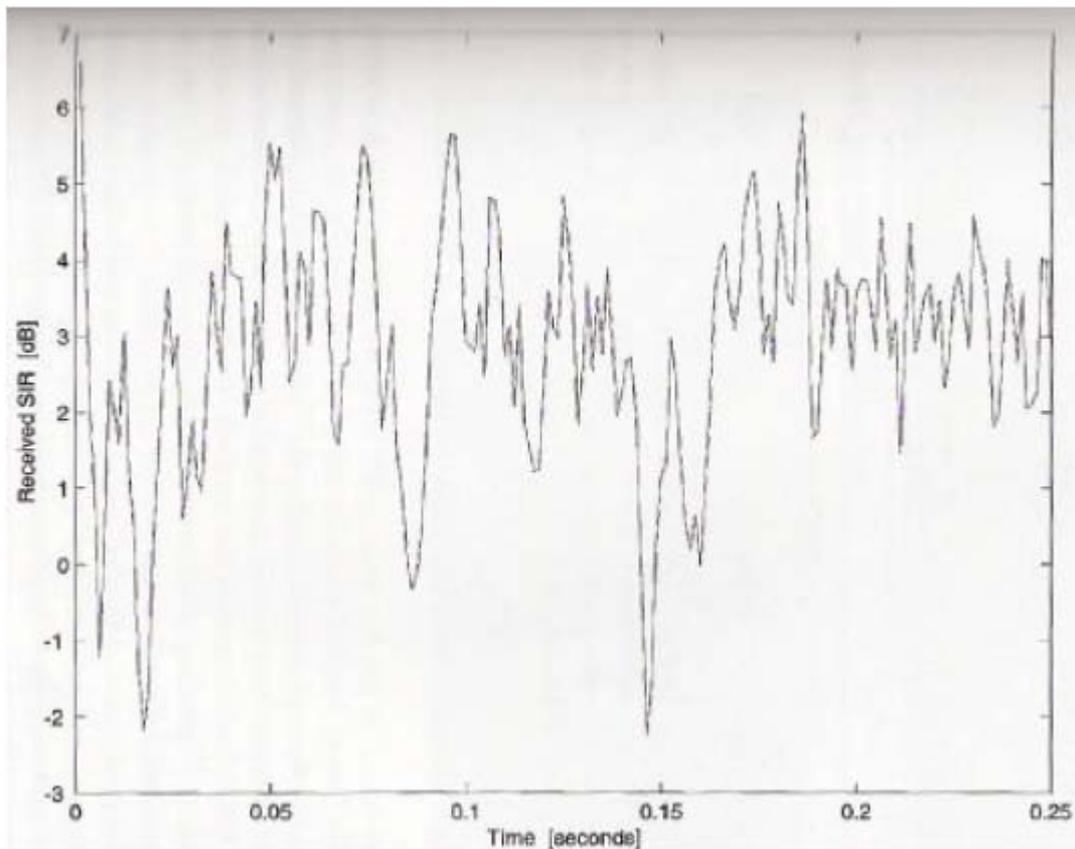


Figura 2.15 – Esempio di canale fortemente variabile

In [33] si può trovare una più precisa descrizione della variazione dello SNIR sperimentato usando lo schema Closed Loop in un sistema cellulare IS-95: in tale documento, si precisa che la potenza del segnale ricevuto ha una distribuzione log-normale, con una varianza di 1-2 dB. Questa varianza risulta legata alla banda del segnale di power control e la banda dell'effetto Doppler: in generale, si può dimostrare che maggiore è il rapporto tra banda del segnale di power control e la banda Doppler, minore è la varianza della potenza del segnale ricevuto.

In generale, dunque, lo schema Closed Loop deve lavorare ad una frequenza maggiore rispetto alla banda di fading del canale: se questo subisce cambiamenti prima che le potenze vengano aggiornate, il Closed Loop risulta inefficace.

CAPITOLO 3

GAME THEORY E POWER CONTROL

Per poter venire incontro alla crescente ed esigente domanda di servizi wireless, le future architetture di rete “senza fili” devono garantire banda larga, integrazione di servizi, flessibilità e scalabilità, oltre all’integrazione dei dispositivi mobili nell’ambiente Internet. Queste esigenze si scontrano con la particolare ed avversa natura dei canali radio.

Il canale radio è fondamentalmente un mezzo di comunicazione broadcast: i segnali inviati da un utente possono potenzialmente esser ricevuti da tutti gli altri utenti all’interno del range del trasmettitore. Questo aspetto può risultare molto utile per alcune applicazioni broadcast, come la trasmissione del segnale radio o televisivo, mentre necessita di un attento controllo nel caso di sistemi di comunicazione personale.

Un canale wireless è, infatti, fortemente condizionato dai parametri di propagazione del segnale, i quali possono variare molto rapidamente, rendendo quindi tempo-variante la natura del mezzo, e dall’interferenza co-canale (CCI), che è dovuta alla condivisione (per ragioni di efficienza spettrale) dello stesso canale da parte dei vari utenti. I vari fenomeni cui è soggetto il segnale che si propaga, come il *fading* dovuto ai percorsi multipli (multipath) seguiti dal segnale stesso, lo *shadowing*, l’attenuazione dovuta alla propagazione (*path loss*), e il rumore, fanno sì che il SNIR all’ingresso del ricevitore possa subire fluttuazioni dell’ordine delle decine di dB, cosa che in generale degrada le prestazioni di una rete wireless. Un’intelligente gestione delle risorse radio (Radio Resource Management, RRM) costituisce una strategia generale per contrastare questi dannosi effetti e migliorare le prestazioni delle reti wireless. In generale, il suo scopo consiste nell’utilizzare il

canale in maniera più efficiente e nel ridurre l'interferenza mediante la scelta di parametri trasmissivi ottimali, come la potenza del segnale inviato, il rate, la modulazione, la codifica, la banda, ecc., in dipendenza da quelle che sono le condizioni del canale.

3.1 Radio Resource Management

Il progetto di un algoritmo di gestione intelligente delle risorse radio richiede, quindi, la comprensione di un gran numero di processi complessi e interconnessi tra loro. In un contesto operativo come quello di una rete ad hoc, gli utenti in gioco, ovvero le entità che usufruiscono delle risorse a disposizione, possono esser caratterizzati da obiettivi, priorità, requisiti di tempo e vincoli di potenza piuttosto diversi. Allo stesso modo, le entità che forniscono le risorse per il sistema possono risultare diversamente caratterizzate dal punto di vista della capacità, della disponibilità, dei costi e delle misure per la sicurezza. Stabilire e allocare le risorse di sistema appropriate per il soddisfacimento di tutte le esigenze prestazionali di ciascun utente, mantenendo i costi entro i rispettivi "budget", costituisce un problema estremamente complicato.

Il problema del RRM può esser pensato e formalmente descritto come un problema di ottimizzazione probabilistica vincolato, in cui, cioè, si vogliono idealmente massimizzare (o minimizzare) alcune grandezze operazionali, andando a scegliere opportunamente certi parametri (che sono le risorse disponibili), i quali sono soggetti a certi vincoli. Il carattere probabilistico del problema deriva dal fatto che le quantità in gioco variano in maniera casuale in quanto casuali sono le variazioni che subiscono le condizioni del canale (sia a causa di fenomeni fisici, sia per la mobilità degli utenti): per questo si deve lavorare a livello statistico; ciò lo rende molto diverso rispetto alla maggior parte dei comuni problemi di ottimizzazione matematica. Il problema è, inoltre, reso più complicato dall'assoluta complessità delle interazioni tra le stesse grandezze considerate e soprattutto dal fatto che, spesso, i vari obiettivi da raggiungere, ovvero le quantità da ottimizzare, risultano in contrasto tra loro; il che significa che, il più delle volte, è necessario raggiungere delle situazioni di compromesso tra esigenze conflittuali. Le risorse a disposizione possono esser classificate in base al livello dello stack protocollare cui appartengono: ad esempio, al livello fisico appartengono la potenza trasmessa, il rate (source rate, channel rate, symbol rate), le dimensioni dell'antenna; al livello MAC appartengono la dimensione del buffer, il tempo di attesa, il rate di servizio; al livello di rete appartengono i percorsi da sorgente a destinazione; al livello applicazione appartengono il rate di codifica per i codificatori di sorgente. Queste variabili, poi, nella pratica, sono soggette a vincoli imposti dai limiti fisici dei dispositivi, dai costi e/o dalle dimensioni: ad esempio, un terminale mobile può trasmettere un livello di potenza massimo e con un rate limitato; oppure le trasmissioni non devono superare un valore massimo di ritardo, pena una qualità del

servizio non adeguata. A livello di prestazioni di rete, gli obiettivi che si vogliono raggiungere possono essere diversi e, come già accennato, conflittuali: ad esempio, si può voler massimizzare il *throughput* di rete complessivo e/o la qualità del servizio globale, minimizzare la potenza trasmessa e/o la distorsione media di segnale e/o la probabilità di fuori servizio. Queste quantità sono funzioni delle risorse suddette e possono avere profili diversi (lineari, convessi o non-lineari del tutto) a seconda delle diverse tipologie di rete e condizioni.

3.1.1 Classificazione schemi RRM

Gli schemi di RRM possono essere classificati in diversi modi. Si parla di schemi *fissi*, quando le scelte dei vari parametri da gestire vengono fatte una volta per tutte, tipicamente prima che il sistema entri propriamente in funzione. Una volta che le decisioni sono prese, entro certi limiti, tali risorse non possono essere riallocate. Si tratta dello schema progettuale di RRM più semplice, adatto al caso in cui la rete risulti un'entità statica e deterministica. In un tale schema, i progettisti devono determinare parametri generali di sistema sulla base di altre grandezze operazionali previste e desiderate e di criteri prestazionali ed economici. Alcune di queste grandezze possono essere la banda totale di sistema, il numero di access point o base station, le forme d'onda usate, il fattore di riutilizzo di frequenza, le dimensioni delle antenne, i livelli di potenza di segnale usati e altri criteri per specifiche prestazionali. Tutti questi valori hanno un impatto significativo sulle prestazioni globali della rete e, inoltre, definiscono i vincoli che devono "guidare" il processo di allocazione fissa di risorse. Come sappiamo, però, la mobilità degli utenti è una delle (se non la) caratteristica fondamentale delle reti wireless: essa è tale per cui le distribuzioni di traffico previste, le locazioni dei terminali, i profili di fading e virtualmente ogni altra assunzione fatta in fase di progetto e di allocazione di risorse fissa variano durante il funzionamento del sistema. Questo, quindi, fa sì che quasi ogni parametro debba, poi, nella pratica, adattarsi ai cambiamenti avvenuti nel sistema per poter risultare ancora "ottimale". All'attuale stato dell'arte, nelle reti wireless, la maggior parte dei parametri scelti in fase di progetto rimangono fissi: la ricerca promette che, nel prossimo futuro, diversi fattori fondamentali, come la modulazione, la codifica, la banda e la potenza possano essere modificati dinamicamente. Fondamentalmente esistono due approcci per un RRM *dinamico*, approcci che sono detti *centralizzato* e *distribuito*. Nel primo, una singola autorità centrale, come una base station, raccoglie informazioni dai vari nodi della rete, calcola un eventuale cambiamento nell'allocazione delle risorse e ordina tale cambiamento ai nodi. Nel secondo, invece, non sono presenti entità di controllo, ovvero ciascun nodo agisce in maniera indipendente ed è, quindi, come se ciascun nodo fosse un'autorità a se stante. Intuitivamente, si capisce come in un approccio centralizzato, la conoscenza globale di rete che l'autorità centrale possiede, faciliti il compito che le

viene richiesto e permetta di ottenere un'allocazione di risorse sicuramente "più ottimale" rispetto al caso decentralizzato; in quest'ultimo caso, infatti, di solito ciascun nodo può basarsi solo su informazioni di carattere locale e/o parziale, o comunque su un insieme di dati più limitato rispetto a uno schema centralizzato. L'entità centrale di controllo, come una base station in una rete cellulare, può inoltre possedere un'elevata capacità di calcolo, che le consente di trattare e implementare algoritmi anche molto complicati. Inoltre, il funzionamento di algoritmi distribuiti può risultare arduo da prevedere, a causa della forte interdipendenza tra le decisioni prese da ciascuna entità in gioco. È per questo che, di solito, vengono usate simulazioni, piuttosto che l'analisi teorica, per la progettazione della rete. Inoltre, per un algoritmo distribuito devono essere valutati molti più aspetti, come la stabilità, la convergenza, la complessità e l'interazione con gli altri livelli protocollari. Un grande vantaggio che, però, hanno gli algoritmi distribuiti è che, normalmente, richiedono un minor *overhead* rispetto agli altri e, quindi, meglio si prestano a trattare reti con un gran numero di utenti, limitando il problema della scalabilità. Ma al di là dei vantaggi e svantaggi che presentano i due metodi, è necessario osservare che l'applicabilità dell'uno o dell'altro metodo è strettamente legata al tipo di rete con cui si ha a che fare. È evidente che un'architettura come quella di una rete cellulare si adatta perfettamente a una gestione di tipo centralizzato (sempre che, comunque, il numero complessivo di utenti gestito dalla base station rimanga limitato), mentre per una rete ad hoc l'approccio distribuito appare come una scelta quasi obbligata. In tale contesto, il problema dell'allocazione delle risorse radio presenta un'intrinseca natura distribuita, dal momento che non esiste un'autorità centrale che possiede tutte le informazioni necessarie per formulare e risolvere un tale problema. È necessario, invece, che le singole entità effettuino le varie scelte in maniera distribuita e indipendente, ma volta, comunque, al benessere comune: uno strumento adatto per affrontare una tale problematica è la teoria dei giochi.

3.2 Introduzione alla *game theory*

La teoria dei giochi è uno strumento sviluppato per elaborare modelli matematici in grado di analizzare e descrivere le interazioni tra entità che si trovano in un contesto di interdipendenza strategica, ovvero in una situazione in cui il risultato delle azioni di una entità dipende anche dalle azioni intraprese dalle altre, a causa della conflittualità reciproca tra gli interessi delle singole entità. In altre parole, la *game theory* analizza il processo delle decisioni, ossia delle azioni intraprese dalle varie entità, il cui singolo "guadagno" dipende anche dalle scelte fatte da tutte le altre; per cui, la preferenza di una certa entità per una certa azione da svolgere dipende da quali scelte hanno fatto le altre e, quindi, in particolare, da ciò che quella entità *crede* che le altre facciano. Questa teoria è stata originariamente concepita e usata in ambito economico, ma negli ultimi anni ci si è resi conto

che essa risulta anche particolarmente adatta per affrontare e risolvere problemi legati ai sistemi di telecomunicazione. Anzi, la teoria dei giochi, per il rigore matematico che caratterizza le assunzioni su cui si fonda, meglio si presta a trattare problematiche in cui le varie entità in gioco agiscono in maniera perfettamente razionale (ovvero coerentemente al perseguimento di un dato obiettivo), come macchine e computer, piuttosto che avere a che fare con comportamenti umani. Essa, non a caso, può costituire uno strumento adeguato per trovare soluzioni a problemi di ottimizzazione distribuita come quelli che si presentano nell'ambito delle reti di telecomunicazioni, in particolare nelle reti wireless ad hoc: in pratica, grazie ad essa, si riescono a *decentralizzare* gli algoritmi creati per un certo scopo, in modo che le singole entità in gioco possano eseguirli senza dover esser controllate da un'entità centrale di riferimento. In sintesi, quindi, con l'applicazione della teoria dei giochi, si può fare in modo che ciascuna singola entità in gioco, pur agendo "egoisticamente" per sua natura (ovvero perseguendo un proprio obiettivo, indipendente da quello altrui, e senza lo scambio di informazioni con le altre entità) si comporti allo stesso tempo in modo tale da raggiungere una sorta di "bene comune", cioè anche quello di tutte le altre entità: è così che il singolo riesce ad ottenere il massimo profitto per sé e, allo stesso tempo, per il gruppo.

3.2.1 Concetti fondamentali di *game theory*

La *game theory* ha, dunque, come scopo, quello di studiare il modello creato per l'analisi della situazione conflittuale in questione: tale modello è appunto detto *gioco*. In un *gioco in forma canonica* (o *normale*), si possono individuare 3 componenti fondamentali:

- 1) un insieme di giocatori, ciascuno dei quali può esercitare la propria scelta su
- 2) un insieme A_i di possibili azioni, dove il pedice i indica l' i -esimo giocatore; l'azione intrapresa è detta *strategia* e l'insieme delle strategie intraprese da tutti i giocatori è detto *profilo di strategia*;
- 3) un insieme di funzioni utilità (una per giocatore), che mappano i profili di strategia in numeri reali.

L'insieme di tutti i possibili profili strategici è detto *spazio d'azione del gioco* e si indica con A ed è dato dal prodotto cartesiano degli insiemi delle strategie di tutti i giocatori. Ad ogni profilo di strategie è associata una "conseguenza" per ciascun giocatore, conseguenza che nella realtà può esser costituita da una somma di denaro o altro e che viene semplicemente quantificata con un numero reale dalla funzione utilità. Da un punto di vista prettamente matematico, una funzione che assegna un valore numerico agli elementi dell'insieme delle azioni possibili A ($u: A \rightarrow \mathfrak{R}$) è una

funzione utilità se, per tutti gli $x, y \in A$ (cioè x e y sono profili di strategie), x è preferito almeno quanto y se e solo se $u(x) \geq u(y)$. Lo scopo di ciascun giocatore è massimizzare la propria funzione utilità, che è una funzione di molte variabili poiché dipende dalla strategia del corrispondente giocatore, ma anche dalla strategia attuata dagli altri giocatori. Ciascun partecipante sa comunque che gli altri giocatori sono di fronte allo stesso problema della massimizzazione della loro funzione utilità e che tutti agiscono in maniera perfettamente razionale, ovvero coerentemente al perseguimento di un dato obiettivo. Queste ultime due assunzioni sono piuttosto forti e, in ambito economico, sarebbero anche poco realistiche, mentre bene si adattano ad affrontare i problemi presentati in questo elaborato.

3.2.2 Classificazione dei giochi

I giochi possono essere classificati sulla base di diversi aspetti:

- in base al livello di cooperazione tra i giocatori: nel caso in cui siano stati preventivamente stabilite relazioni di cooperazione tra giocatori, si parla di gioco *cooperativo*, altrimenti si dice *non cooperativo*;
- in base alla quantità di informazioni che ciascun giocatore ha a disposizione riguardanti le strategie che sono state attuate dagli altri giocatori nel corso del gioco fino a quel momento: nel caso in cui conosca esattamente tutte le mosse (come ad esempio gli scacchi), si parla di gioco *ad informazione perfetta*, altrimenti ad informazione *imperfetta*;
- si parla di gioco *ripetuto*, quando i giocatori fanno le proprie scelte strategiche sulla base delle osservazioni delle azioni intraprese dagli altri giocatori e di previsioni sulle loro future strategie, e tenendo anche a mente le proprie passate strategie; altrimenti il gioco è detto *statico*; in tal caso, i giocatori scelgono la propria strategia senza conoscere minimamente quelle scelte dagli altri, tant'è che il gioco viene anche detto *simultaneo*, perché le decisioni potrebbero anche essere prese tutte contemporaneamente, vista la loro indipendenza.

3.2.3 Concetto di soluzione

La soluzione di un gioco è costituita da un profilo di strategie che, con elevata probabilità, sarà attuato dai giocatori. La teoria dei giochi si propone di:

- 1) stabilire quali caratteristiche deve avere un profilo di strategie per poter essere la soluzione del gioco (quindi quali sono le condizioni necessarie che deve soddisfare);

- 2) stabilire se la soluzione del gioco esiste, ovvero se c'è una conclusione del gioco più probabile di tutte le altre.

Il secondo dei due aspetti risulta molto delicato e non può essere affrontato in generale, bensì va valutato caso per caso, tenendo presente che, spesso, per stabilire se esiste la soluzione di un gioco è necessario disporre di maggiori informazioni rispetto a quelle contenute nella rappresentazione strategica. Per quanto riguarda il primo problema, invece, in letteratura sono state avanzate varie risposte ad esso, o meglio, sono stati proposti vari *concetti di soluzione*. Nelle ipotesi che sono state fatte sul comportamento dei giocatori (ovvero perfetta razionalità e stesso obiettivo perseguito da tutti, assunzioni che abbiamo detto essere assolutamente plausibili per i problemi presi in esame nel seguito), si può dimostrare che, se esiste la soluzione di un gioco, essa deve essere un *equilibrio di Nash*.

3.2.4 Equilibrio di Nash

Un equilibrio di Nash di un gioco è un profilo di strategie in cui la strategia di ogni singolo giocatore dà un risultato (rappresentato dal valore della funzione utilità) non migliorabile rispetto alla combinazione delle strategie altrui nel profilo stesso, ossia la strategia di ogni giocatore è una risposta ottima alla combinazione di strategie altrui descritta dal profilo. In altri termini, un profilo di strategie è un equilibrio di Nash quando non ammette deviazioni unilaterali vantaggiose per alcun giocatore, ammettendo però che i corrispondenti avversari seguano, a loro volta, strategie di equilibrio di Nash, cioè che tutti ragionino secondo lo stesso criterio.

Formalmente, un equilibrio di Nash può essere espresso come quel profilo di strategie \mathbf{x} , tale che per tutte le singole strategie x'_i ammissibili, si ha che

$$u_i(x_i, \mathbf{x}_{-i}) \geq u_i(x'_i, \mathbf{x}_{-i}) \quad \forall i = 1, \dots, N \quad \forall p'_i \in A_i$$

dove con \mathbf{x}_{-i} si indica il vettore \mathbf{x} senza la componente i -esima.

Con una scelta strategica complessiva corrispondente a un equilibrio di Nash, nessun giocatore ha interesse e motivo di cambiarla, perché sarebbe svantaggioso per ciascuno di loro: ecco perché con tale profilo si è in una situazione di equilibrio. In generale, non è detto che per uno stesso gioco esista un punto di equilibrio di Nash e, nel caso in cui esista, esso sia unico. E non è detto, inoltre, che una situazione identificata da un equilibrio di Nash sia la migliore possibile in assoluto: un concetto molto importante per misurare la “bontà” di una soluzione di equilibrio è quello di *efficienza Paretiana*.

3.2.5 Efficienza Paretiana

Un profilo di strategie è detto *Pareto efficiente* (o *Pareto ottimale*) se non ne esiste un altro tale da migliorare contemporaneamente il risultato ottenuto da tutti i giocatori; in altre parole, con una scelta strategica Pareto efficiente, è impossibile migliorare il risultato ottenuto da un giocatore senza ridurre quello di un altro. Formalmente, il profilo \mathbf{x}^* è Pareto efficiente se non esiste un altro profilo \mathbf{x} tale che

$$u_i(\mathbf{x}) \geq u_i(\mathbf{x}^*) \quad \forall i \in N \quad \text{e} \quad u_i(\mathbf{x}) > u_i(\mathbf{x}^*) \quad \text{per qualche } i \in N$$

La differenza fondamentale col concetto di equilibrio di Nash è che in quest'ultimo si cerca la miglior scelta strategica del singolo giocatore *in risposta* alle strategie di tutti gli altri che sono fissate, quindi è una sorta di massimo “vincolato” alle scelte altrui; in generale, se non ci fossero vincoli di questo tipo, si potrebbe anche trovare un profilo di strategie tale da migliorare (o lasciare al più inalterato) il risultato ottenuto da tutti i giocatori. Un profilo Pareto efficiente non è detto che costituisca una situazione di equilibrio: anche se i giocatori scegliessero un tale profilo, il gioco dovrebbe comunque ritornare in una situazione di equilibrio, cioè dovrà terminare con un equilibrio di Nash, ammesso che questo esista. È bene precisare inoltre che, di profili Pareto efficienti, ne possono esistere molti e che il concetto di efficienza non comporta alcun confronto interpersonale di utilità o di benessere ed è quindi slegato da qualsiasi giudizio di equità o giustizia sociale. Un profilo di strategie è, invece, detto *Pareto dominante* se con esso è possibile aumentare l'utilità di alcuni giocatori, senza danneggiare gli altri. Formalmente, un profilo \mathbf{x}' “Pareto domina” un altro profilo \mathbf{x} se

$$u_i(\mathbf{x}') \geq u_i(\mathbf{x}) \quad \forall i \in N \quad \text{e} \quad u_i(\mathbf{x}') > u_i(\mathbf{x}) \quad \text{per qualche } i \in N$$

La figura 3.1 chiarisce la differenza tra Pareto dominanza e Pareto efficienza, che sono due concetti che non vanno confusi.

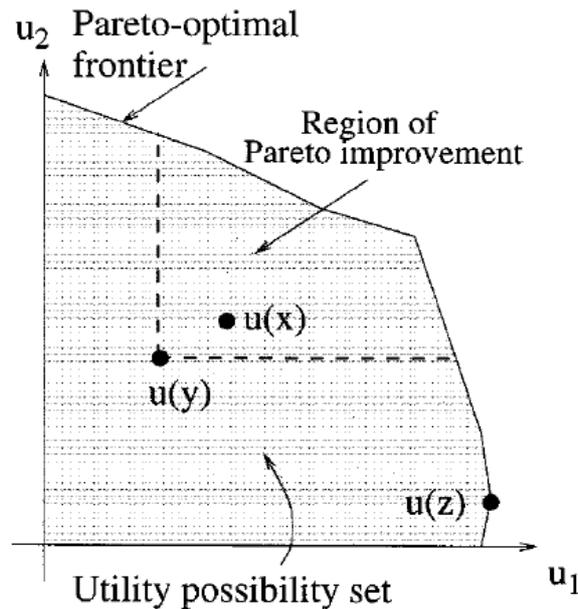


Figura 3.1 – Efficienza Paretiana

Si fa riferimento a un gioco in cui ci sono due giocatori, i cui insiemi strategici sono mappati nell'insieme dei valori possibili della funzione utilità (*utility possibility set*), rappresentato dall'area ombreggiata. Un qualsiasi profilo strategico che produce una variazione non decrescente di entrambe le utilità individuali rispetto al profilo y , produce un *miglioramento paretiano*: un tale profilo è ad esempio indicato nella figura con x e si trova nella regione indicata con *region of Pareto improvement*; tutti i profili appartenenti a quell'area sono Pareto dominanti rispetto a y . Nella figura 3.1 è anche indicata la frontiera della Pareto efficienza: è da notare che profili Pareto ottimali non necessariamente dominano tutti gli altri profili: z , ad esempio, è un profilo Pareto ottimale, ma comunque esso non domina y in quanto $u_2(z) < u_2(y)$, al di là del fatto che $u_1(z) > u_1(y)$. L'applicazione pratica, che vedremo nei prossimi paragrafi, di tutti questi concetti al problema del power control in una rete wireless ad hoc fornirà anche un aiuto per chiarirne meglio il significato.

3.3 Applicazione della teoria dei giochi allo studio di una rete wireless

Le condizioni che deve soddisfare il modello di rete wireless per poter essere trattato come un gioco possono essere suddivise in:

- Condizioni di Razionalità

- 1) i processi decisionali devono essere ben definiti, ossia ciascun terminale, nella selezione di una certa azione in risposta a cambiamenti “ambientali”, deve seguire un ben preciso e deterministico insieme di regole;
- 2) la scelta di un terminale di cambiare strategia deve essere motivata da una ragionevole aspettativa di migliorare la “propria condizione”.

Tali condizioni possono essere soddisfatte facendo in modo che ogni nodo voglia massimizzare una ben definita funzione-obiettivo (l'utilità) o implementando in ognuno di essi una sorta di macchina a stati finiti che determini i parametri radio operativi in risposta ai cambiamenti di stato.

- Condizioni di Gioco Non Banale

- 1) deve esserci più di un'entità decisionale nella rete;
- 2) almeno 2 entità decisionali devono avere uno spazio d'azione formato da più elementi.

Se la prima condizione non è soddisfatta, come frequentemente avviene nelle reti wireless (ad esempio una rete cellulare, in cui decide tutto la base station), il sistema può costituire un interessante problema di ottimizzazione, però non è un gioco, dal momento che il processo di decisioni non è legato alle interazioni fra le varie entità. La seconda condizione formalizza la prima, specificando che un'entità che deve prendere decisioni senza poterle scegliere da un insieme formato da più opzioni non è di fatto un'entità decisionale.

3.3.1 Esempi di applicazioni

Per illustrare quando e se una rete wireless soddisfi tali condizioni, sono stati riportati alcuni esempi di situazioni e problematiche in cui è appropriato modellare una rete come un gioco, e altri in cui non lo è.

Esempi di applicazioni inappropriate

- Site Planning: durante questa fase, vengono svolti molti importanti studi in modo che la locazione del sito, l'assegnamento del canale e i fattori di riuso vengano scelti in modo tale da massimizzare l'area di copertura, la capacità e la QoS degli utenti, e da minimizzare i costi (che tipicamente si concentrano nel numero e nella complessità delle base station).

Senza dubbio, si tratta di un complicato problema di RRM, a causa delle intricate considerazioni che devono esser fatte e delle relazioni inverse tra i primi tre obiettivi e quello finale. Tuttavia, questo problema non è modellabile come un gioco, dal momento che è presente solo un'unica entità che prende decisioni, che è il progettista del sito, e, quindi, manca la parte delle interazioni conflittuali. È da notare, comunque, che, ai fini dell'applicabilità della teoria dei giochi, la non-dinamicità del processo decisionale non costituisce un problema: ad esempio, un gioco in forma normale non è dinamico.

- Schemi di Accesso Casuale al Mezzo: in una tale problematica, ci sono ovviamente numerose entità decisionali, le decisioni sono legate alle interazioni tra le entità e possono esser chiaramente definiti degli obiettivi (ad es. minimizzare la BER); nonostante ciò, la maggior parte degli schemi di accesso casuale non possono esser modellati come giochi, dal momento che la scelta di un'azione non è deterministica, cioè essa può avere o meno un ragionevole aspettativa di migliorare le prestazioni. Ad esempio, si consideri uno schema di accesso di tipo "slotted ALOHA", in cui la probabilità di ritrasmissione è fissata; ci sono, quindi, un certo numero di slot differenti che ciascun terminale può scegliere di ritrasmettere e la scelta di uno di essi è completamente casuale: la prima condizione di razionalità è violata. Questo fa sì che, senza la minima idea su come ogni terminale possa agire, non si possano fare previsioni sulla soluzione del gioco. È da notare però che, se i terminali sono liberi di adattare la propria probabilità di ritrasmissione, lo schema può senza dubbio esser modellato come un gioco (come è stato fatto in [18]), in cui la scelta di tale probabilità costituisce l'azione del giocatore.

Esempi di applicazioni appropriate

- Power Control Distribuito: la teoria dei giochi non è applicabile universalmente a tutti gli algoritmi di power control. Ovviamente gli schemi centralizzati, in cui è presente un'unica entità che prende decisioni (come una base station), costituiscono situazioni non adatte per la *game theory*, la quale, invece, risulta uno strumento appropriato quando le decisioni divengono *distribuite*. Scenari con tali proprietà includono open loop power control, closed loop power control in un'ad hoc network e il closed loop power control in una rete cellulare, in cui, però, a prendere decisioni sono le coppie base station-mobile.
- Channel Assignment Dinamico e Distribuito: in un tale processo, i canali assegnati per l'uso in ciascuna cella o cluster vengono dinamicamente alterati in base al cambiamento percepito di fattori "ambientali", come l'interferenza. I canali selezionati all'interno di una cella (o di

un cluster) hanno impatto sull'assegnamento dei canali nelle celle (o cluster) vicini, dando luogo così a un processo di decisioni conflittuali. Esempi di questa problematica vengono forniti in [19] e [20]. Un concetto strettamente collegato è *l'adaptive interference avoidance*, in cui ciascun link adatta la propria forma d'onda impiegata per la trasmissione in risposta all'interferenza percepita. All'interno di una cluster network, tale scenario può presentarsi ed esser trattato come un problema di assegnamento di canale distribuito e dinamico, se i vari adattamenti della forma d'onda vengono considerati come l'insieme dei "canali" disponibili per l'uso.

3.4 Applicazione della teoria dei giochi a rete wireless ad hoc

In generale, molti altri algoritmi di gestione distribuita delle risorse radio possono correttamente venire modellati come giochi. Quello che verrà applicato ed esaminato in questo elaborato, come vedremo più in dettaglio nel seguito, sarà un tipo di gioco *non cooperativo*, in cui cioè i giocatori non possono collaborare l'uno con l'altro, scambiandosi informazioni utili, bensì agiscono indipendentemente l'uno dall'altro: in maniera intuitiva, già si può dedurre che, presumibilmente, il risultato di un tale gioco sarà un'allocazione meno efficiente, rispetto a quella che potrebbe esser ottenuta attraverso una cooperazione tra terminali o un'ottimizzazione centralizzata. La scelta di un gioco non cooperativo, tuttavia, appare necessaria nel contesto delle reti wireless ad hoc, in cui effettivamente i terminali non hanno la conoscenza delle condizioni e delle scelte degli altri e non possono tanto meno cooperare tra loro; le decisioni da loro prese, ovviamente finalizzate al miglioramento della loro condizione, non possono che basarsi su certi parametri che essi sono in grado di misurare durante il loro funzionamento.

Un sistema di comunicazione wireless ad hoc, basato su CDMA, in cui si vuole "ottimizzare" la potenza irradiata da ciascun utente appare come uno scenario "ideale" per l'applicazione della teoria dei giochi, in quanto sono presenti tutte le caratteristiche e gli elementi fondamentali di un gioco. Si ha, infatti, un insieme di giocatori, che è costituito dagli utenti, eventualmente mobili e con interessi reciprocamente conflittuali: ciascuno di essi trasmette la propria informazione via radio, basandosi su una certa tecnica di accesso al mezzo; le frequenze radio, dunque, costituiscono il mezzo condiviso da tutti i segnali e per questo motivo il segnale emesso da un utente diviene una potenziale sorgente di interferenza per i segnali ricevuti da tutti gli altri utenti. D'altra parte ciascun utente vuole ottenere per sé le massime prestazioni, che possono essere espresse in termini di rapporto segnale-interferenza o BER, e allo stesso tempo risparmiare sul consumo energetico. Ciascun giocatore, poi, ha a disposizione un insieme di azioni attuabili, che è costituito dal range dei possibili livelli di potenza che egli può scegliere di trasmettere; il range risulta fissato dai limiti

fisici degli strumenti, che sono digitali e dunque i livelli saranno distribuiti in maniera discreta. Ridurre il consumo di potenza non solo prolunga la vita delle batterie (aspetto molto importante per i terminali mobili) e riduce le emissioni elettromagnetiche, ma fa diminuire anche l'interferenza nei confronti degli altri utenti. Quest'ultimo aspetto risulta particolarmente significativo nei sistemi CDMA. Lo scopo pratico del power control è, quindi, fare in modo che il segnale trasmesso da ciascun trasmettitore abbia una qualità "adeguata", senza però causare agli altri terminali ricevitori interferenza non necessaria.

Il problema del power control nelle reti wireless ad hoc, dunque, può essere schematizzato come un gioco che è classificabile come *ripetuto, ad informazione imperfetta e non cooperativo*, in quanto ciascun giocatore (terminale) deve cercare di massimizzare la propria utilità comportandosi in maniera "egoistica", dal momento che può agire solo sulla potenza da lui usata per la trasmissione, perseguendo una strategia basata su informazioni locali e/o parziali, relative cioè solo a un numero limitato di utenti e non alla globalità, come succederebbe in un algoritmo centralizzato. Allo stesso tempo, però, la funzione utilità deve essere definita in modo tale che gli utenti siano comunque spinti verso un'implicita collaborazione, o quanto meno tendano ad ostacolarsi a vicenda il meno possibile, in modo da ottenere un risultato globalmente buono ed equo.

3.4.1 Algoritmo di aggiornamento delle potenze

Nella realtà pratica, ad un terminale ricevente facente parte di una rete ad hoc, per poter realizzare il "gioco", serve un algoritmo in grado di calcolare il valore di potenza che deve essere trasmessa dal suo corrispondente terminale sorgente, al fine di massimizzare la propria funzione utilità. L'algoritmo, come più volte specificato, deve basarsi solo su informazioni che il nodo ricevitore è in grado di procurarsi da solo, altrimenti non sarebbe implementabile in maniera distribuita. Quello che verrà usato in questo elaborato è uno dei più conosciuti algoritmi distribuiti, originariamente proposto in [29] e poi ulteriormente studiato in molti documenti, tra cui [30] e [31].

L'algoritmo è dato dalla seguente equazione di aggiornamento:

$$P_i(k+1) = P_i(k) \frac{\gamma_i}{SNIR_i(k)} \quad \forall i = 1, \dots, N$$

dove γ_i è il valore di SNIR minimo desiderato. Si tratta di un algoritmo che è stato analizzato e applicato per sistemi cellulari ma che, tuttavia, possiede molte interessanti proprietà che lo rendono attraente anche per l'applicazione alla problematica in esame: oltre alla sua natura autonoma e distribuita, esso, in particolare, risulta asincronicamente convergente all'assegnamento di potenze

Pareto ottimale, a patto che il sistema sia “realizzabile”, ovvero effettivamente esista una distribuzione di potenze tale da soddisfare i vincoli imposti sulla QoS (ovvero quello sullo SNIR). Se così non è, l’algoritmo diverge: questo è l’unico limite che esso ha. La divergenza è causata dal fatto che il vincolo imposto sullo SNIR ($>\gamma_i$) è rigido e deve esser rispettato da tutti gli utenti ad ogni costo. Intuitivamente, applicando l’algoritmo suddetto, ogni utente “aggiusta ciecamente” la potenza usata dal suo trasmettitore, in modo che lo SNIR misurato si avvicini a quello di soglia al passo successivo; in tal modo, però, non può rendersi conto se il sistema è “irrealizzabile”, cioè se per tutti simultaneamente è impossibile soddisfare quel requisito sullo SNIR: come conseguenza, in tale circostanza le potenze trasmissive crescono smisuratamente, perché l’obiettivo di ciascuno è raggiungere un valore di SNIR non inferiore a quello di soglia. La causa di questo comportamento, dunque, sembrerebbe la “cecità” degli utenti, legata al fatto che l’algoritmo è distribuito e dipende da sole informazioni locali; tuttavia, è ragionevole pensare che un utente potrebbe dedurre un’eventuale situazione di irrealizzabilità del sistema proprio a partire da un’informazione locale come, ad esempio, un’elevata interferenza ricevuta. Il trucco per evitare l’irrealizzabilità del sistema, allora, potrebbe essere quello di inserire, per ogni utente che vuol trasmettere, una sorta di *admission control* basato su informazioni locali e su requisiti di QoS non così rigidi. Prima di spiegare il metodo usato per questo scopo, è opportuno illustrare altre importanti proprietà dell’algoritmo suddetto.

Questo algoritmo gode anche della notevole proprietà di essere “standard”, nel senso definito in [32]: in generale, un algoritmo di power control della forma $\mathbf{P}(k+1) = A(\mathbf{P}(k))$ è detto “standard” se la funzione A soddisfa le seguenti proprietà per ogni valore di potenza appartenente al range usato:

- 1) Positività: $A(\mathbf{P}) > 0$;
- 2) Monotonicità: se $\mathbf{P}' > \mathbf{P}$, allora $A(\mathbf{P}') > A(\mathbf{P})$;
- 3) Scalabilità: $\mu A(\mathbf{P}) > A(\mu\mathbf{P})$, $\forall \mu > 1$.

$\mathbf{P}' > \mathbf{P}$ significa $p_i > p_i' \forall i \in N$ e $A(\mathbf{P})$ è detta anche *funzione interferenza*. L’essere standard è, per l’algoritmo, un aspetto molto importante in quanto, se esiste un punto fisso di convergenza, tale punto fisso è unico. Per poter esistere quest’ultimo, è sufficiente che $A(\mathbf{P}) \geq 1$ per qualche \mathbf{P} appartenente al range. In tal modo si è sicuri della convergenza dell’algoritmo.

Rimane da capire il legame che c’è tra un algoritmo di questo tipo e la teoria dei giochi, di cui tanto si è parlato in precedenza. La teoria dei giochi entra in ballo nella realizzazione di quel controllo di ammissione che prima è stato solo accennato. Anziché ammettere alla trasmissione utenti che rispettano il rigido vincolo $\text{SNIR} > \gamma_i$, ciascun utente si calcola, in maniera distribuita, quel valore di

rapporto segnale-rumore-interferenza che massimizza la propria funzione utilità e utilizza tale valore come soglia, decidendo *autonomamente* se è il caso di trasmettere o meno. Lo SNIR di soglia dovrà ovviamente tener conto anche di una specifica sullo SNIR reale, che corrisponde a un requisito minimo di qualità del servizio. Tale valore, comunque, essendo calcolato in maniera distribuita grazie alla teoria dei giochi, grazie alla quale ogni utente riesce a tener conto anche della presenza degli altri, porta con sé una maggior valenza di equità e efficienza.

3.4.2 Criteri-guida e obiettivi del power control

Progettare e mettere in pratica uno schema per un efficiente utilizzo delle risorse radio in un contesto come quello delle reti ad hoc, richiede necessariamente la comprensione di un gran numero di processi, molto complessi ed interconnessi fra loro. Si è detto nei paragrafi precedenti che una rete wireless ad hoc, in cui si voglia “ottimizzare” la potenza irradiata da ciascun utente, costituisce uno scenario ideale per l’applicazione della teoria dei giochi. Ma che significa “ottimizzare”? Cioè, qual è lo scopo dei vari algoritmi di ottimizzazione, quali sono i parametri di sistema che si vogliono ottimizzare?

Prima di definire questi concetti, è necessario, anzitutto, stabilire quale criterio di fondo viene seguito per la costruzione complessiva dell’algoritmo; in particolare, si deve comprendere la differenza che esiste tra un approccio che possiamo chiamare user-centric e uno detto network-centric: il primo mira a massimizzare i parametri che rappresentano gli interessi di ciascun singolo utente, indipendentemente dagli altri; il secondo tende, invece, a massimizzare lo sfruttamento delle potenzialità della rete nella sua globalità, come, ad esempio, il *throughput* complessivo, dato dalla somma dei *throughput* di tutti gli utenti, o la capacità informativa (di Shannon) complessiva. Essendo, dunque, motivati da obiettivi sostanzialmente diversi, i due tipi di approccio determinano allocazioni di risorse di rete altrettanto differenti, in particolare con gradi di equità anche molto diversi. È stato dimostrato, in particolare, che, seguendo un metodo di tipo user-centric si tende a distribuire la qualità del servizio in maniera più equa rispetto a uno di tipo network-centric, il quale, invece, tende a fornire la maggior parte delle risorse a quegli utenti con il miglior stato del canale, un risultato analogo al cosiddetto waterfilling. Considerando che il contesto operativo è quello di una rete ad hoc, la quale esiste proprio in virtù della presenza e del funzionamento degli utenti, una distribuzione di risorse di tipo waterfilling appare del tutto inadeguata, in quanto tale da limitare in maniera eccessiva il numero di utenti attivi e, di conseguenza, la vita stessa della rete. La precarietà e “debolezza” di una rete ad hoc (ma anche wireless in generale), dovute, come sappiamo, specialmente alle caratteristiche del canale wireless e alla mobilità, rendono necessario un elevato grado di equità e la quasi totale assenza di privilegi tra coloro che la formano: per questo, come

criterio-base da seguire va senza dubbio scelto quello più equo, di tipo user-centric. Oltretutto, in prima analisi, un approccio di tipo network-centric non sarebbe neanche banale da mettere in pratica in una rete ad hoc: per la sua attuazione, esso generalmente richiede una visione e una conoscenza globale della rete che un terminale appartenente a un sistema ad hoc di solito non può avere; tale metodo potrebbe già meglio adattarsi ad un'architettura centralizzata, come quella di una rete cellulare: un elemento come la base station, dotata in aggiunta di elevata capacità di calcolo e di memoria, nonché di tutte le informazioni che caratterizzano la cella, sarebbe in grado di gestire e attuare algoritmi anche piuttosto complessi.

Tuttavia, quando il numero di utenti in gioco risulta piuttosto "limitato" e questi non sono eccessivamente interessati a risparmiare potenza, un approccio che, in maniera distribuita, tende a massimizzare la capacità informativa complessiva di rete, può essere interessante da studiare e da implementare. In tale situazione, essendo presenti pochi terminali, l'interferenza prodotta può non risultare un fattore così degradante le prestazioni e, se per gli utenti non risulta troppo dispendioso, può risultare opportuno cercare di ottenere una maggiore capacità informativa, per poter aumentare la velocità trasmissiva. Si tratta, ad ogni modo, di un obiettivo che non può esser al tempo stesso compatibile con l'aspetto *power saving*: in quest'ultimo caso, si cerca, in generale, il vettore di potenze globalmente più piccolo che, comunque ci garantisce un certo livello di QoS, ovvero di SNIR; quando si vuole massimizzare la capacità informativa di rete, essendo questa proporzionale allo SNIR secondo la relazione di Shannon, non si vuole più che lo SNIR di ciascun utente sia semplicemente maggiore di una certa soglia, ma che sia il massimo SNIR, compatibilmente con la presenza di altri utenti e con il vincolo di potenza massima. Ma allora, se si vuole che lo SNIR sia il massimo possibile in tale situazione, non si può al tempo stesso pretendere di risparmiare in potenza!

Si potrebbe però obiettare che, se tutti i nodi abbassano le proprie potenze trasmissive, si riduce il range di interferenza introdotta da ciascuno e, in tal modo, ciascuno di essi può raggiungere uno SNIR più elevato. Tuttavia, ridurre il range trasmissivo può rendere necessari più hop per arrivare alla destinazione e ogni hop ovviamente "mangia" parte delle risorse di rete presenti nell'area in cui esso è localizzabile: ecco perché un criterio *power saving* non è compatibile con la massimizzazione della capacità informativa.

Inoltre, per poter mantenere la connettività di rete, un nodo può dover trasmettere con potenza elevata, ad esempio, verso un nodo distante 1 solo hop, però con distanza effettiva grande, impattando ovviamente sulla comunicazione degli altri nodi più vicini. Questo conflitto può esser riassunto introducendo un parametro che è la *transport capacity*, che misura il rate trasmesso su una certa distanza e che si misura in bit-metro/sec: un link a 10 bit/sec su 1m ha la stessa capacità di

trasporto di un link a 1 bit/sec su 10m. In [78] si dimostra che se la dimensione di una rete ad hoc aumenta di N nodi, la capacità di trasporto per ciascun nodo viene scalata approssimativamente di $1/\sqrt{N}$, dal momento che la capacità di rete varia come \sqrt{N} . In altre parole, reti ad hoc di grandi dimensioni sembrano fornire una bassa capacità a tutti gli utenti, dal momento che essi devono competere per le limitate risorse di sistema: ecco perché prima si è specificato che il numero di utenti deve essere “limitato”.

Quando, comunque, si guarda esclusivamente alla soddisfazione del singolo utente (ad esempio andando a massimizzare le singole funzioni utilità, trascurando una buona globale gestione della rete), si può finire per ottenere una soluzione complessiva che risulta eccessivamente inefficiente, ossia lontana da quelle che sono le reali capacità della rete [22]. Esistono alcuni metodi, tra cui [21, 23], che cercano di trovare una sorta di “via di mezzo” tra i due possibili approcci prima descritti, metodi che però sono stati pensati per una rete cellulare, o comunque, per un’architettura con fisionomia centralizzata, che in una rete ad hoc va esclusa a priori. In questo elaborato, come vedremo più in dettaglio nel seguito, si è cercato comunque anche di mediare i due criteri, senza far ricorso alla presenza di una supervisione centralizzata, come risulta necessario in un contesto come quello delle reti ad hoc.

In generale, gli obiettivi principali che si possono voler perseguire con l’introduzione del power control possono essere:

- riduzione della potenza usata per la trasmissione da parte di ciascun terminale, *mantenendo una certa prefissata QoS* per il collegamento
- riduzione dell’interferenza complessiva nella rete
- aumento del *throughput* complessivo di rete
- aumento capacità informativa di rete

I primi tre scopi sono quelli classici di ogni algoritmo di power control e, come già accennato in precedenza, non sono affatto indipendenti tra loro. Risparmiare energia per inviare i pacchetti, oltre al beneficio di allungare la vita delle batterie e, quindi, del terminale stesso, comporta ovviamente anche una minore interferenza nei confronti degli altri utenti che, pertanto, per ottenere la stessa QoS, possono abbassare la potenza trasmissiva. Tipicamente, i dispositivi terminali mobili sono capaci di trasmettere un segnale con un livello di potenza che, in base a normative internazionali e a causa di limiti fisici, non può superare un certo livello massimo. Qualsiasi trasmissione causa interferenza nella regione circostante la sorgente, dovuta alla condivisione del canale: abbassando il range trasmissivo o, che dir si voglia, il livello di potenza, l’area interessata da interferenza viene

ridotta. Per cui, eventuali utenti che vogliono trasmettere trovano la possibilità di farlo, in quanto non sono costretti a superare il loro livello massimo di potenza per raggiungere la QoS richiesta, contribuendo, così, ad aumentare il *throughput* complessivo di rete. In tal modo, si riesce a tener conto anche delle prestazioni globali di rete, “mediando” i due approcci suddetti.

In alternativa, sotto certe condizioni, come già detto, si può voler ottenere la massima capacità informativa di rete.

Fondamentale, per poter raggiungere gli scopi suddetti, è la scelta della funzione utilità.

3.4.3 Scelta della funzione utilità

La funzione utilità, anzitutto, deve certamente ben rappresentare il grado di soddisfazione sperimentato da ciascun utente. Prima di tutto è, dunque, necessario identificare i criteri di preferenza che sono specifici del problema esaminato e poi determinare una funzione utilità che soddisfi tale struttura; ovviamente non esiste un'unica funzione utilità in grado di descrivere un particolare insieme di regole di preferenza. Inoltre, per lo scopo che ci siamo prefissati, ossia quello di mediare tra soddisfazione dell'utente e buona gestione della rete, l'utilità deve, in qualche modo, dare la possibilità di non trascurare anche le prestazioni globali del sistema.

Nelle telecomunicazioni, in particolare in quelle wireless, la soddisfazione di un utente è, di solito, legata alla qualità del servizio (QoS) che riesce ad ottenere. Una misura della QoS, in un sistema digitale wireless come quello in questione, è data dalla probabilità di errore su bit (BER), che si mappa direttamente nel rapporto segnale-rumore-interferenza (SNIR); in alternativa, come parametro di qualità, si può considerare la capacità informativa di Shannon che, tanto, a sua volta, è funzione proprio dello SNIR:

$$C = B \cdot \log_2(1 + SNIR)$$

dove B è la banda del segnale utile. Sarà interessante confrontare i risultati ottenuti utilizzando l'una o l'altra grandezza come parametro di qualità.

Inoltre, avendo gli utenti wireless tipicamente una batteria con potenza e durata piuttosto limitate, la potenza trasmessa, come già detto, costituisce un altro elemento da non trascurare e da legare alla funzione utilità. È chiaro che un terminale wireless desidererebbe ottenere un'elevata qualità di ricezione (alto SNIR o capacità di Shannon) e allo stesso tempo spendere poca energia per ottenerla; un utente che, invece, ottiene un'elevatissima QoS, pagando, però, un prezzo altrettanto elevato, può voler rifiutare il servizio, così come farebbe un utente con scarsa QoS.

Tutte queste considerazioni si traducono in una funzione utilità che, in generale, dovrebbe soddisfare le seguenti proprietà, in cui il parametro SNIR può eventualmente esser sostituito dalla capacità informativa di Shannon:

- 1) per ipotesi, l'utilità (u_i) è assunta come funzione non negativa dello SNIR e, quando SNIR = 0, si assume che l'utilità sia nulla; si assume inoltre che l'utilità tenda a zero se la potenza trasmessa (p_i) tende a zero, cioè

$$\lim_{p_i \rightarrow 0} u_i = 0$$

questo perché, intuitivamente, non ha senso avere un'utilità diversa da zero se non si trasmette nulla;

- 2) fissata la potenza trasmessa di un utente (p_i), l'utilità deve essere una funzione monotona crescente dello SNIR del corrispondente utente, cioè

$$\frac{\partial u_i(p_i, SNIR_i)}{\partial SNIR_i} > 0, \quad \forall SNIR_i, p_i$$

questo perché, se si fissa la potenza trasmessa, il terminale sperimenta tassi di errore più bassi al crescere dello SNIR, fatto che si traduce in una maggiore soddisfazione derivante dall'uso delle risorse di sistema;

- 3) fissata la potenza trasmessa di un utente, l'utilità marginale, ovvero la derivata dell'utilità, deve tendere a zero per valori dello SNIR sempre più grandi, cioè

$$\lim_{SNIR_i \rightarrow \infty} \frac{\partial u_i(p_i, SNIR_i)}{\partial SNIR_i} = 0 \quad \forall p_i$$

questo, in altre parole, significa che per elevati valore dello SNIR, cioè per un tasso di errore prossimo allo zero, l'utilità tende asintoticamente a un valore costante;

- 4) fissato un valore di SNIR (ovvero fissato il tasso d'errore), l'utilità deve essere una funzione monotona decrescente della potenza trasmessa (p_i), cioè

$$\frac{\partial u_i(p_i, SNIR_i)}{\partial p_i} < 0, \quad \text{per } \forall SNIR_i, p_i$$

questo perché, come già accennato, l'aumento della potenza trasmessa accelera l'esaurimento della batteria, fatto che riduce la "soddisfazione" del terminale mobile; ma, soprattutto, questa proprietà è fondamentale perché, dal momento che si cerca di massimizzare lo SNIR su ogni collegamento, si corre il rischio di creare maggiore interferenza agli altri utenti i quali, a loro volta, aumenteranno la potenza per raggiungere lo SNIR ottimo creando maggior disturbo, e così via: se questa proprietà non è soddisfatta, dunque, non si è immuni da un possibile effetto retroattivo;

- 5) fissato un valore di SNIR, se la potenza trasmessa tende all'infinito, allora l'utilità deve tendere a zero, cioè

$$\lim_{p_i \rightarrow \infty} u_i = 0$$

questo per lo stesso motivo accennato per la proprietà precedente.

3.4.4 Segnale dati vs segnale vocale

È necessario, anzitutto, distinguere la QoS per un segnale vocale da quella per un segnale dati: per il primo, è essenziale un ritardo basso, mentre gli errori di trasmissione sono tollerati (ovviamente fino ad un certo limite); il secondo invece può tollerare un po' di ritardo, ma non errori trasmissivi. Questo si riflette nelle diverse caratteristiche che deve avere la funzione utilità, a seconda che essa sia relativa ad un utente appartenente ad un sistema di comunicazione wireless per segnali vocali o a uno per segnali dati. Nel primo caso, infatti, l'obiettivo essenziale di QoS consiste nell'ottenere un ritardo complessivo limitato, mentre gli errori di trasmissione sono ben tollerati, almeno fino a un certo punto; raggiunto e superato un determinato valore di SNIR, quindi, l'utilità sperimentata dall'utente non presenta alcuna crescita; al di sotto di tale valore, però, la qualità del segnale ricevuto risulta troppo bassa per essere intelligibile, per cui l'utilità risulta praticamente nulla. Queste considerazioni portano alla costruzione di una funzione utilità che possiede semplicemente un gradino in corrispondenza del valore di soglia del SNIR, come mostrato in figura 3.2.

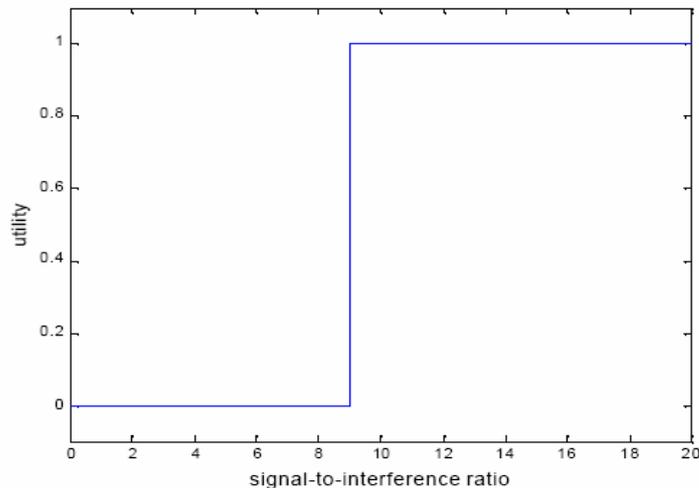


Figura 3.2 – Funzione utilità per sistema vocale

Tale funzione evidenzia che, in un sistema vocale, un utente è sostanzialmente indifferente a piccole variazioni dello SNIR. Un algoritmo di controllo ottimo per sistemi wireless vocali, dunque, è quello che massimizza il numero di conversazioni che possono simultaneamente ottenere una certa qualità del servizio, ovvero un fissato valore dello SNIR.

Nel secondo caso, cioè in un sistema per dati, il ritardo è un fattore che, entro certi limiti, può essere ben tollerato, mentre non è così per gli errori di trasmissione che, oltretutto causano la ritrasmissione dei pacchetti e, di conseguenza, riducono il *throughput* del sistema. L'intolleranza agli errori fa sì che in un sistema per dati lo SNIR costituisca una quantità molto più importante rispetto al caso di un sistema vocale, per via della diretta relazione tra SNIR e la probabilità di errore di trasmissione; più elevato è lo SNIR, più basso sarà il tasso di errore (e quindi anche il numero di ritrasmissioni) e maggiore sarà il *throughput*. Tutto questo implica che la funzione utilità per un utente dati debba essere una funzione continua del SNIR ottenuto dall'utente stesso: l'elaborato tratterà esclusivamente funzioni utilità adatte per segnali dati.

Con i vincoli da rispettare stabiliti finora, di funzioni utilità se ne possono trovare infinite. Essa, comunque, deve anche essere caratterizzata da un certo grado di semplicità e trattabilità dal punto di vista matematico, in modo da poter ricavare un algoritmo che non risulti troppo pesante per i terminali mobili.

3.4.5 Costruzione della funzione utilità

Esistono diversi approcci per la costruzione di una funzione utilità adatta allo sviluppo di un algoritmo di power control distribuito, ma nessuno di essi possiede una base teorica matematica tale da garantirne la superiorità rispetto agli altri. L'unica maniera per confrontarli è utilizzare

simulazioni. Come punto di partenza per la costruzione di una tale funzione, è stato preso in considerazione l'approccio di D. Goodman in [24], dal quale poi, come vedremo, si sono prese le distanze.

Il contesto cui si fa riferimento per la ricerca di un algoritmo di power control è quello di un sistema basato su CDMA, in cui vi sono N link attivi (dunque i "giocatori" sarebbero i link o, se preferiamo, sono i nodi ricevitori in quanto sono loro che attuano le strategie, cioè cercano di modificare la potenza trasmessa dal corrispondente trasmettitore, in base all'algoritmo stabilito) e in cui il rate dei trasmettitori sia R bit/s su una banda di W Hz. Se si indica con p_i la potenza usata dal trasmettitore dell' i -esimo link, G_{ij} il guadagno di canale tra il trasmettitore del j -esimo link e il ricevitore dell' i -esimo link e σ_N^2 la potenza di rumore termico (che si assume uguale per tutti i link), il rapporto segnale-rumore-interferenza (SNIR) al ricevitore del link i -esimo è definito come:

$$SNIR_i = \frac{W}{R} \frac{G_{i,i} p_i}{I_i + N} = \frac{W}{R} \frac{G_{i,i} p_i}{\sum_{\substack{j=1 \\ j \neq i}}^N G_{i,j} p_j + \sigma_N^2}$$

In un sistema dati wireless, i terminali si scambiano pacchetti contenenti L bit di informazione. La codifica di canale fa sì che la dimensione complessiva di ogni pacchetto sia $M > L$. Indichiamo con R il rate di trasmissione, espresso in bit/s. Al ricevitore del terminale i -esimo, il rapporto segnale interferenza è indicato con $SNIR_i$ e la probabilità di corretta ricezione del pacchetto è $q(SNIR_i)$, dove la funzione $q(\cdot)$ dipende da vari fattori della trasmissione, quali la modulazione, la codifica, il tipo di propagazione e la struttura del ricevitore. Il numero di trasmissioni necessario per ricevere un pacchetto correttamente è una variabile aleatoria, indicata con K . Se si assume che tutte le trasmissioni siano statisticamente indipendenti, K ha una distribuzione geometrica con funzione massa di probabilità data da

$$P_K(k) = q(SNIR_i)[1 - q(SNIR_i)]^{k-1} \quad \text{per } k = 1, 2, 3, \dots$$

Il valor medio di K è $E[K] = 1 / q(SNIR_i)$. La durata di ciascuna trasmissione è M/R secondi, per cui il tempo totale di trasmissione necessario per avere una corretta ricezione è la variabile aleatoria KM/R , espressa ovviamente in secondi. Indicata con p_i la potenza usata per la trasmissione, espressa in watt, l'energia spesa corrisponde alla variabile aleatoria $p_i KM/R$, espressa in joule, il cui valor medio è dato da $E[K]p_i M/R = p_i M/[R q(SNIR_i)]$. Quello che interessa al terminale è comunque solo

l'informazione contenuta nel segnale, cioè gli L bit di informazione. Quindi per il terminale, l'utilità è esprimibile come

$$\frac{\text{bit utili}}{\text{energia media}} = \frac{LRq(SIR_i)}{Mp_i} \quad \text{bit/joule}$$

cioè l'utilità per un terminale può essere interpretata come il numero di bit di informazione ricevuti per ciascun joule di energia spesa.

Come punto di partenza per determinare un algoritmo di power control, questa definizione di utilità ha diversi vantaggi, ma anche alcuni svantaggi non banali. I vantaggi stanno nella sua interpretazione fisica (*bit per joule*) e nella sua semplicità matematica. Gli svantaggi derivano anzitutto dall'assunzione semplificativa che tutti gli errori di trasmissione possano essere rilevati in ricezione. I sistemi di trasmissione dati contengono potenti codici di correzione d'errore che rendono questa assunzione vera "per tutti gli scopi pratici". Il vero svantaggio, in realtà, è un problema matematico, dato dal fatto che la probabilità che un pacchetto arrivi correttamente non è zero anche se la potenza trasmessa risulta nulla. In un sistema di trasmissione binario con pacchetti di M bit e p_i nulla, infatti, un ricevitore semplicemente può tirare a indovinare il valore di ciascuno degli M bit che sono stati trasmessi. La probabilità di indovinarli tutti è 2^{-M} : questo significa che, se $p_i = 0$, il numeratore nell'espressione per la funzione utilità è positivo, per cui la funzione è infinita. Questo suggerirebbe che il miglior approccio per il power control sarebbe quello che prevede di spegnere tutti i trasmettitori e aspettare che il ricevitore indovini quello che si vorrebbe trasmettere. Ovviamente questa strategia ha due difetti: il primo è che il tempo di attesa di ricezione corretta del pacchetto potrebbe essere dell'ordine di diversi mesi, il secondo è che ci potrebbero essere ricezioni non corrette che non vengono rilevate dai codici a protezione d'errore.

Per eliminare, dunque, questa soluzione degenera, Goodman ha proposto di modificare la funzione utilità sostituendo a $q(SNIR_i)$ un'altra funzione $f(SNIR_i)$ che ne "segua" l'andamento e che allo stesso tempo sia tale che $f(SNIR_i)/p_i = 0$ per $p_i = 0$. Per quanto detto in precedenza riguardo alla soluzione degenera, questa modifica può essere interpretata con un'inclusione implicita nella funzione utilità di un vincolo sul ritardo. Se si assume una perfetta rivelazione degli errori senza alcuna correzione e un $SNIR_i$ costante su tutta la durata del pacchetto, si può esprimere

$$q(SNIR_i) = (1 - BER_i)^M$$

La BER in generale è una funzione decrescente dello SNIR e dipende, come sappiamo, dal tipo di modulazione usata. La funzione $f(SNIR_i)$ che sostituisce $q(SNIR_i)$ è data da

$$f(SNIR_i) = (1 - 2BER_i)^M$$

La figura 3.3 mostra quanto le due funzioni abbiano un andamento simile nel caso di modulazioni BPSK e FSK non coerente.

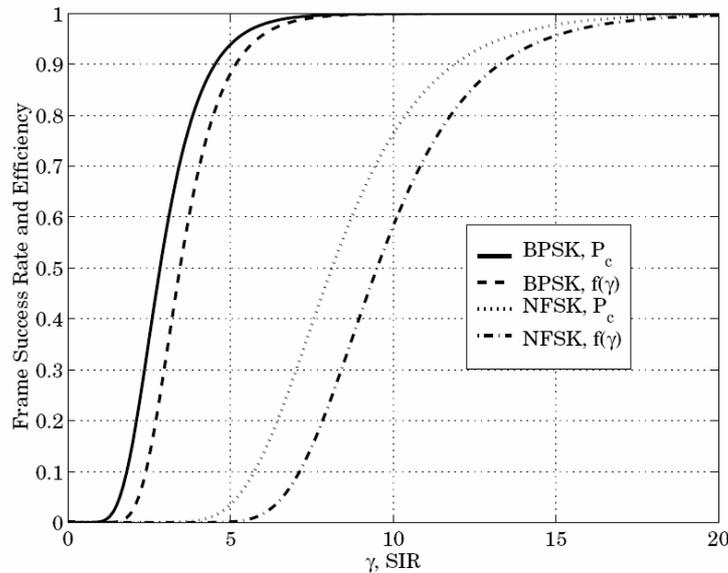


Figura 3.3 – Funzione utilità di Goodman

Dunque la funzione utilità che è stata costruita è data da

$$u_i = \frac{LR}{Mp_i} f(SNIR_i) \quad \text{bit/joule}$$

Il rigore e, tutto sommato, la semplicità e la chiarezza con cui questa funzione è stata costruita, la rendono particolarmente attraente agli occhi del progettista di rete; tuttavia, la sua dipendenza da parametri di sistema come modulazione, codifica e dimensione della trama, la rendono anche difficilmente analizzabile e anche l’algoritmo di power control, che è stato costruito mediante essa, risulta molto complicato, senza contare che esso possiede una bassa efficienza Paretoiana. Si è cercato di risolvere quest’ultimo problema in [25], dando luogo ad un algoritmo sì più efficiente, ma praticamente impossibile da implementare in una rete ad hoc.

La forma della funzione utilità che allora è stata presa in considerazione è, comunque, non troppo dissimile da quella appena vista e presenta proprietà matematiche che la rendono particolarmente adatta per progettare un relativamente “semplice” algoritmo di power control; si tratta della funzione sigmoide (fig. 3.4), data dall’equazione

$$U_i(SNIR_i = x) = \frac{1}{1 + e^{-a_i(x-b_i)}}$$

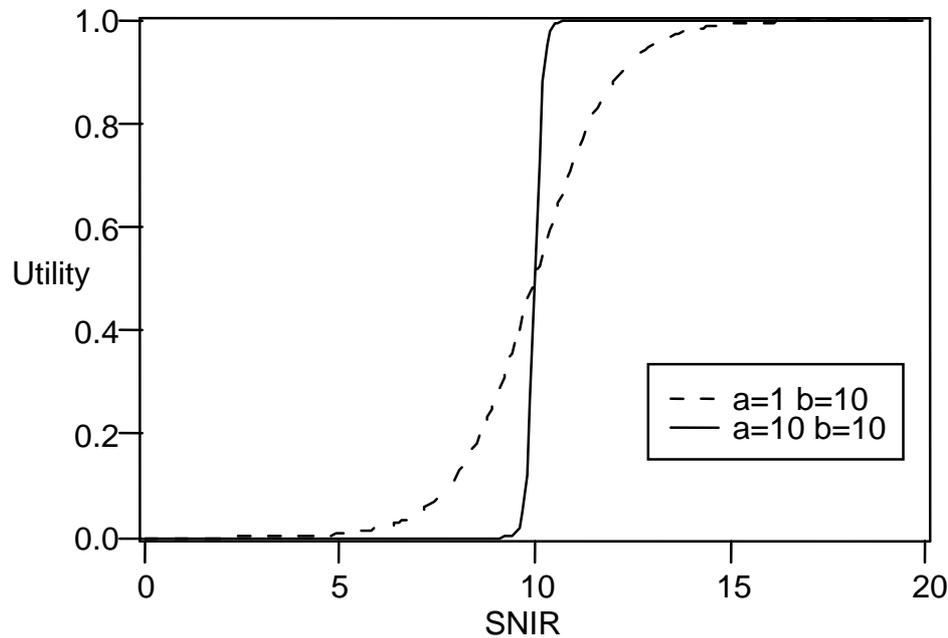


Figura 3.4 – Funzione Utilità

in cui il parametro a_i rappresenta l’inclinazione della curva e corrisponde alla pendenza massima della tangente alla curva; tale parametro può esser utile per distinguere il tipo di utente in base al servizio richiesto da lui richiesto (dati o voce), come è stato meglio specificato in precedenza. In questo elaborato verrà preso in esame solo il caso di servizio dati. Il parametro b_i è, invece, il “centro” della funzione utilità, cioè il punto in cui essa raggiunge il valore di 0.5. Dapprima, come indicato nella formula stessa, si è preso come parametro di qualità lo SNIR misurato dal ricevitore i -esimo, successivamente verrà assunta la capacità di Shannon.

La funzione, però, non rispetta le regole che sono state elencate in precedenza: innanzitutto non assume valore nullo per $SNIR = 0$, infatti $\frac{1}{1 + e^{-a_i b_i}} \neq 0$, ma se assumiamo che $a_i b_i \gg 1$ l’approssimazione è da considerarsi valida senza un grosso errore. Ad ogni modo, possiamo sempre risolvere questo problema, se proprio necessario, considerando la trasformazione lineare:

$$U(x) = \frac{\frac{1}{1 + e^{-a(x-b)}} - \frac{1}{1 + e^{ab}}}{1 - \frac{1}{1 + e^{ab}}}$$

Nello sviluppo dell'algorithmo, verrà utilizzata la funzione nella sua forma più semplice.

In secondo luogo, cosa più importante, la funzione, fissato lo SNIR, non tende a zero al crescere della potenza di trasmissione e ciò potrebbe causare un effetto di feedback indesiderato, come è stato precedentemente sottolineato. Per eliminare questo difetto, è stato introdotto il concetto di *pricing*, ossia la funzione utilità, così com'è, viene modificata mediante l'aggiunta di una funzione costo, che deve essere monotona crescente con la potenza trasmessa impiegata e deve assumere valore nullo solo nell'origine; la più semplice funzione costo è, ovviamente, quella lineare con la potenza, cioè

$$C_i(p_i) = \alpha_i p_i$$

dove α_i rappresenta il coefficiente di "prezzo". L'introduzione di un prezzo sulla potenza, oltre ad eliminare il problema suddetto, incoraggia un comportamento "più sociale" tra gli utenti, i quali sono implicitamente spinti a collaborare. Il *pricing* è anche il metodo utilizzato in [25] per migliorare l'efficienza Pareto dell'algorithmo. Quindi, sottraendo la funzione costo alla funzione utilità si ottiene l'utilità netta definita come:

$$NU_i(p_i) = U_i(SNIR_i) - \alpha_i p_i$$

È da notare che, in questo modo, si ottiene una funzione che, fissato lo SNIR, invece di azzerarsi per una potenza trasmessa infinita, tende a infinito negativo, cosa che, comunque, è accettabile.

3.4.6 Equilibrio di Nash

In generale, si cerca quella che è una delle più largamente usate soluzioni a un problema di teoria dei giochi, che è, come già detto, l'*equilibrio di Nash*. All'equilibrio di Nash, ciascun utente sceglie un livello di potenza che costituisce la miglior risposta alle potenze scelte dagli altri utenti in quel momento. Formalmente, un vettore (o profilo) di potenze $\mathbf{p} = (p_1, \dots, p_N)$ è un equilibrio di Nash per un gioco non cooperativo di power control se

$$u_i(p_i, \mathbf{p}_{-i}) \geq u_i(p_i', \mathbf{p}_{-i}) \quad \forall i = 1, \dots, N \quad \forall p_i' \in A_i$$

dove u_i è l'utilità (in generale) dell' i -esimo utente e \mathbf{p}_{-i} denota il vettore formato da tutti gli elementi di \mathbf{p} , escluso l' i -esimo. Tale notazione enfatizza che, come sappiamo, l' i -esimo utente ha il

controllo solo sulla propria potenza p_i . La miglior risposta del terminale i -esimo può anche venire indicata come la corrispondenza $r_i: A_{-i} \rightarrow A_i$ che assegna ad ogni $\mathbf{p}_{-i} \in A_{-i}$ l'insieme

$$r_i(\mathbf{p}_{-i}) = \{p_i \in A_i : u_i(p_i, \mathbf{p}_{-i}) \geq u_i(p_i', \mathbf{p}_{-i}) \quad \forall p_i' \in A_i\}$$

Con tale notazione, la definizione di equilibrio di Nash può essere riformulata in una forma più compatta come segue: il vettore di potenze $\mathbf{p} = (p_1, \dots, p_N)$ è un equilibrio di Nash per un gioco non cooperativo di power control se e solo se $p_i \in r_i(\mathbf{p}_{-i})$ per ogni $i \in N$. Da notare che se per ogni $p_i \in A_{-i}$, $r_i(\mathbf{p}_{-i})$ è composto esattamente da un elemento, allora $r_i(\cdot)$ può esser visto come una funzione nell'usuale concezione. Quella che prima abbiamo chiamato corrispondenza risulta dunque una generalizzazione del concetto di funzione.

In particolare, nel gioco in esame, la miglior risposta del terminale i -esimo a un dato vettore di interferenze \mathbf{p}_{-i} è data da

$$r_i(\mathbf{p}_{-i}) = \min(P_{max}, \underline{\mathbf{p}}) \quad \text{con} \quad \underline{\mathbf{p}} = \arg \max_{p_i \in \mathfrak{R}_+} u_i(p_i, \mathbf{p}_{-i})$$

3.4.7 Formalizzazione del gioco

Formalmente, dunque, il gioco non cooperativo di power control in un sistema di comunicazione wireless ad hoc formato da N utenti è espresso da

$$\max_{p_i} NU_i(p_1, p_2, \dots, p_i, \dots, p_N), \quad p_i \in A_i, \quad \forall i = 1, \dots, N$$

dove NU_i è la funzione utilità netta relativa all' i -esimo utente, A_i è lo spazio delle possibili strategie scelte dall'utente i -esimo, cioè l'insieme delle possibili potenze trasmissive che può scegliere. In questa fase di sintesi dell'algoritmo, assumeremo che lo spazio delle strategie A_i di ogni giocatore sia continuo nella potenza trasmessa corrispondente p_i (sebbene nella pratica, come abbiamo già accennato, lo spazio sia costituito da un insieme discreto di valori) e abbia il range $[0, P_{max}]$. Dall'espressione si può dedurre che la potenza trasmissiva che ottimizza l'utilità individuale dipende anche dalle potenze trasmesse di tutti gli altri terminali nel sistema.

Essendo la funzione utilità netta continua e derivabile in $p_i \in]0, +\infty[$, per determinare il valore di potenza a cui viene raggiunto il massimo di essa, è sufficiente risolvere il seguente problema:

$$\frac{\partial NU_i(p_i)}{\partial p_i} = 0$$

$$\frac{\partial^2 NU_i(p_i)}{\partial p_i^2} < 0$$

La condizione di negatività della derivata seconda dell'utilità netta è necessaria per ottenere il massimo relativo con valori di potenza per i quali l'utilità netta possiede la concavità rivolta verso il basso, ovvero con valori di potenza con cui si conseguono $SNIR_i > b_i$. Derivando, quindi, la NU_i e imponendo le condizioni precedenti, si ottiene

$$\frac{d(NU_i(SNIR_i, p_i))}{dp_i} = \frac{d(NU_i(SNIR_i) - \alpha_i p_i)}{dp_i} = \frac{d(NU_i)}{dSNIR_i} \frac{d(SNIR_i)}{dp_i} - \alpha_i = 0$$

Siccome

$$\frac{d(SNIR_i)}{dp_i} = \frac{W G_{ii}}{R I_i}$$

dove stavolta si è indicato con $I_i = \sum_{\substack{j=1 \\ j \neq i}}^N G_{i,j} P_j + \sigma_N^2$ l'interferenza totale al ricevitore del link i-esimo,

compreso il rumore gaussiano, si ha che

$$\frac{d(NU_i)}{dSNIR_i} = \alpha_i \frac{RI_i}{WG_{i,i}}$$

Graficamente, come è mostrato in figura 3.5, l'equazione precedente corrisponde a trovare il valore di $SNIR_i$, in cui si ha intersezione tra la derivata della funzione utilità e la retta parallela all'asse delle ascisse, che corrisponde alla costante $\alpha_i \frac{RI_i}{WG_{i,i}}$: delle due soluzioni quella maggiore di b_i corrisponde al massimo, mentre l'altra al minimo.

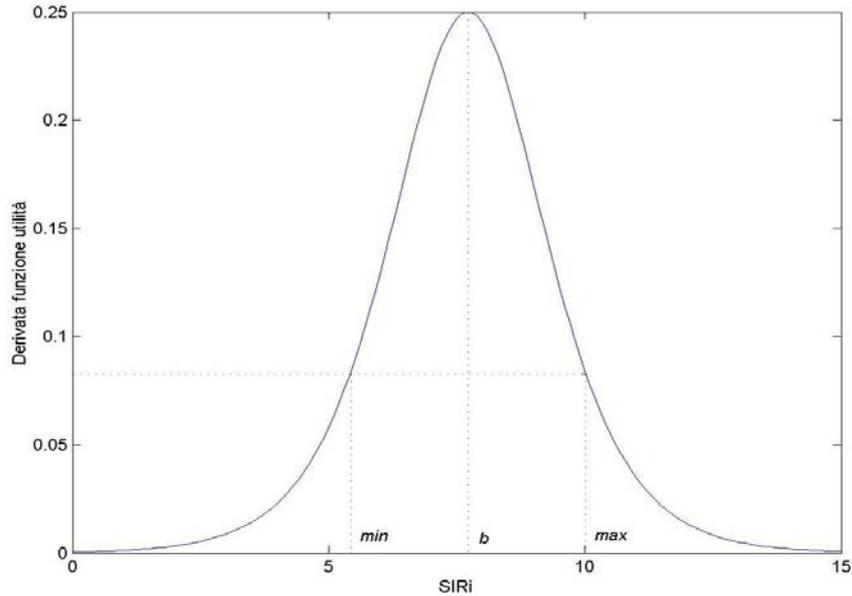


Figura 3.5 – Derivata della Funzione Utilità

Il valore dello SNIR che ci interessa può essere determinato in forma chiusa. Ponendo $A_i = \alpha_i \frac{RI_i}{WG_{ii}}$, che risulta essere una quantità nota o misurabile localmente, questo valore è ottenibile come

$$SNIR_i^{tar} = f_i^{-1}(A_i)$$

dove $SNIR_i^{tar}$ è appunto il rapporto segnale-rumore-interferenza che è stato definito di *target* e $f_i(SNIR_i) = NU'_i(SNIR_i)$ nella parte concava di U_i , cioè dove è possibile trovare un massimo locale.

Poiché

$$NU'_i(x) = \frac{d(NU_i(x))}{dx} = \frac{d(U_i(x))}{dx} = \frac{aU_i(x)}{1-U_i(x)}$$

l'equazione sopra può essere riscritta come

$$a_i e^{-a_i(x-b_i)} = A_i [1 + e^{-a_i(x-b_i)}]^2 \Rightarrow [e^{-a_i(x-b_i)}]^2 + \left(2 - \frac{a_i}{A_i}\right) e^{-a_i(x-b_i)} + 1 = 0$$

Per cui, ponendo $\lambda = e^{-a_i(x-b_i)}$, si ottiene un'equazione di secondo grado in λ che ammette le due seguenti soluzioni:

$$\lambda_{1,2} = -1 + \frac{a_i}{2A_i} \pm \sqrt{\left(1 - \frac{a_i}{2A_i}\right)^2 - 1} = -1 + \frac{a_i}{2A_i} \pm \sqrt{\frac{a_i}{2A_i} \left(\frac{a_i}{2A_i} - 2\right)}$$

Affinché le due soluzioni siano reali, si deve soddisfare la condizione:

$$\frac{a_i}{2A_i} > 2 \Rightarrow A_i < \frac{1}{4a_i}$$

In tal caso, ho due soluzioni reali equidistanti da $\frac{a_i}{2A_i} - 1$ (sempre maggiore di 1). Come si può notare dalla figura 3.6, solo la soluzione più piccola (λ_1) corrisponde al punto di massimo.

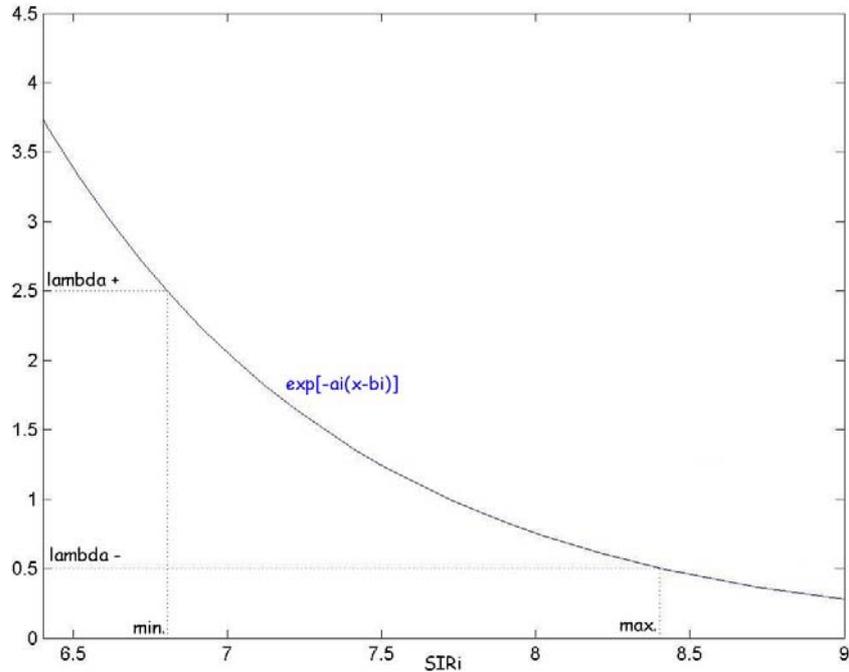


Figura 3.6 – Soluzioni dell’equazione

Pertanto, ponendo $\lambda = \lambda_1$ e invertendo l’esponenziale, si ottiene il valore dello $SNIR_i$ che si raggiunge all’equilibrio di Nash, detto *SNIR target*, e che vale:

$$SNIR_i^{tar} = b_i - \frac{1}{a_i} \ln \left[\frac{a_i}{2A_i} - 1 - \sqrt{\frac{a_i}{2A_i} \left(\frac{a_i}{2A_i} - 2 \right)} \right]$$

Una volta fissati i parametri a_i e b_i , dalla figura 3.7 si può notare come il valore dello *SNIR target* sia una funzione decrescente di A_i .

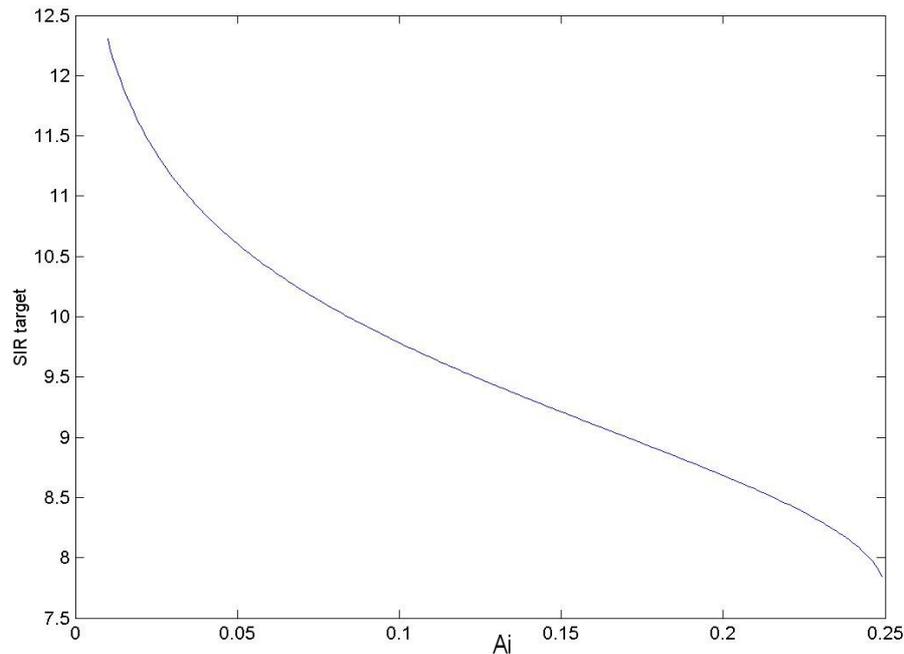


Figura 3.7 – Andamento dello $SNIR^{tar}$ al variare di A_i

Poiché l'aumento di A_i corrisponde a un aumento dell'interferenza e quindi a un peggioramento delle condizioni ambientali, si può affermare che il valore dello SNIR a cui ogni utente aspira, decresce in seguito a un incremento dell'interferenza introdotta dagli altri utenti. Può accadere, però, che, essendo lo $SNIR_{target}$ trovato in tal modo un massimo relativo, in esso l'utilità netta sia negativa e quindi convenga utilizzare potenza nulla in trasmissione, conseguendo un'utilità nulla, ma pur sempre maggiore.

3.4.8 Scelta dei parametri della funzione utilità

Come vanno scelti i parametri liberi presenti nell'espressione della funzione utilità, ovvero a_i , b_i e il coefficiente di prezzo α_i ? Per quanto riguarda quest'ultimo valore, le considerazioni sono rimandate al capitolo successivo. Il parametro b_i specifica il "centro" della funzione utilità; per comprendere che valore gli si deve assegnare, è necessario fare alcune considerazioni. In figura 3.8, sono rappresentate una funzione utilità e diverse funzioni costo al variare dello $SNIR_i$. Il coefficiente di linearità della $C_i(SNIR_i)$ su questo piano non è più α_i , ma $\frac{\alpha_i R I_i}{W G_{i,i}}$, cioè A_i .

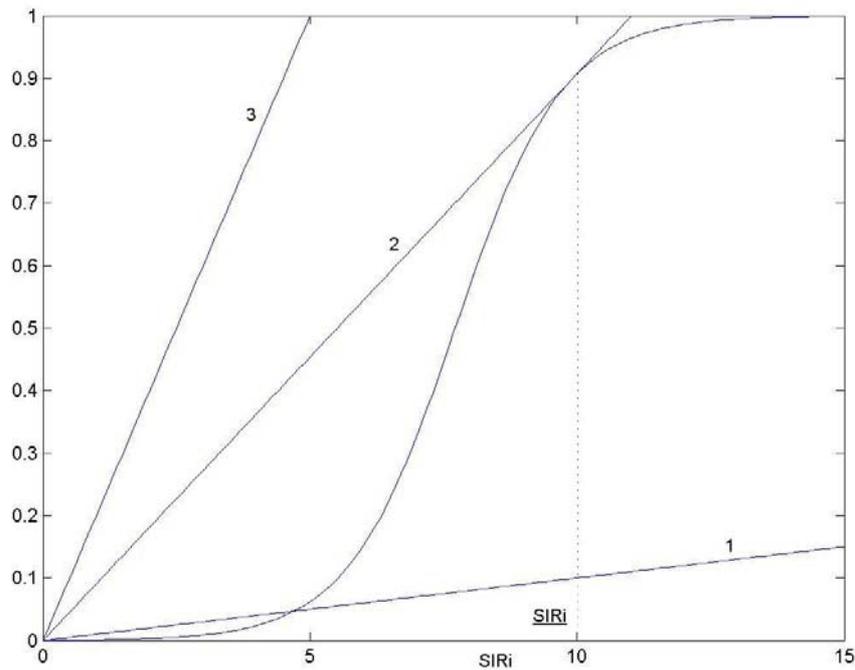


Figura 3.8

A_i può assumere tutti i valori reali positivi maggiori di $\frac{\alpha_i \sigma_N^2 R}{WG_{i,i}}$, in quanto l'interferenza complessiva I_i ovviamente non può valere meno del solo rumore termico σ_N^2 : il valore minimo di pendenza è, pertanto, rappresentato dalla retta numero 1. Aumentando la pendenza, si raggiunge un punto (retta numero 2) in cui la funzione costo è tangente alla funzione utilità. In questo caso, l'utilità netta è sempre negativa per ogni valore di $SNIR_i$ ad eccezione di due punti: l'origine e il punto indicato in figura come \underline{SNIR}_i , che può esser chiamato *turnoff SNIR*, ossia lo SNIR "di spegnimento", in quanto chiaramente, in questa situazione, è più conveniente, per l'utente i -esimo e per gli altri, che venga scelta l'origine come punto di equilibrio, ovvero potenza nulla, in modo da evitare di sprecare potenza inutilmente, recando oltretutto danno con l'interferenza per gli altri. Tale punto corrisponde al più basso valore di SNIR che l'utente i -esimo può raggiungere mentre si trova in trasmissione. Avendo, quindi, in base al tipo di segnale e di trasmissione, delle specifiche precise sullo SNIR (del tipo $SNIR_i > \gamma_i$), basta imporre $\underline{SNIR}_i = \gamma_i$ per soddisfare le richieste. Per determinare questo \underline{SNIR}_i per via analitica in funzione dei parametri della funzione utilità, si deve tener conto che in quel punto la retta del costo, che oltretutto è passante per l'origine, è tangente alla funzione utilità, per cui si ha che

$$U'_i(\underline{SNIR}_i) \cdot \underline{SNIR}_i = U_i(\underline{SNIR}_i)$$

Indi, dopo vari calcoli, si ottiene

$$\frac{e^{a_i SNIR_i}}{a_i SNIR_i - 1} = e^{a_i b_i}$$

Se, dunque, imponiamo $SNIR_i = \gamma_i$, il “centro” b_i della funzione utilità può essere calcolato così:

$$b_i = \gamma_i - \frac{\ln(a_i \gamma_i - 1)}{a_i}$$

Per quanto riguarda a_i , è stato già accennato che esso può consentire di distinguere il tipo di utente in base al servizio da lui richiesto: un utente-voce richiederà soprattutto basso ritardo e continuità nella comunicazione e il suo target SNIR potrà essere relativamente basso, per cui dovrà esser caratterizzato mediante una funzione utilità simile ad una funzione gradino, dunque con un grande valore di a_i e un piccolo γ_i , in modo che, specialmente per bassi valori di b_i , esso sia difficilmente “spugnabile”. Osservando l’andamento dello $SNIR_i^{tar}$ al variare del valore di a_i (fig. 3.9), si nota che questo decresce più lentamente in funzione di A_i quando il valore di a_i aumenta. Questo vuol dire che per un utente con a_i maggiore, anche per valori di A_i abbastanza alti, si mantiene uno SNIR target superiore al *turnoff* SNIR e, quindi, si ha una probabilità di spegnimento minore.

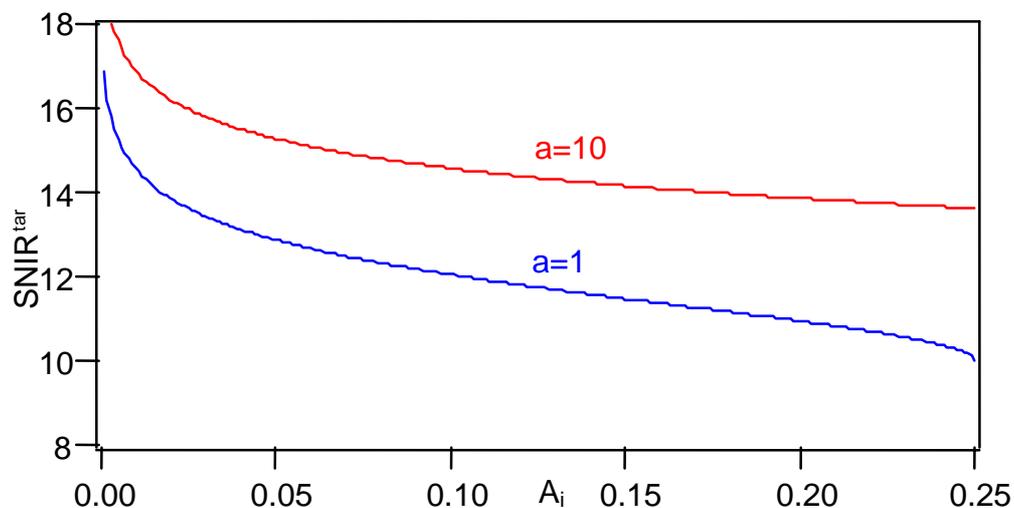


Figura 3.9 – Andamento $SNIR^{tar}$ al variare di A_i

Un utente-dati, invece, potrà accettare un po’ di ritardo nella ricezione dei pacchetti, ma non errori trasmissivi, per cui la sua funzione utilità dovrà avere a_i piccolo e γ_i grande in base alle specifiche sulla probabilità di errore che si vuole ottenere. Un tale utente potrà raggiungere uno SNIR elevato durante la trasmissione, però deve accettare di poter esser “spento” ed eventualmente ritardare la trasmissione. In questo elaborato, la trattazione è limitata agli utenti che trasmettono solo dati.

3.4.9 Algoritmo di power control

Una volta fissati i parametri della funzione utilità, ciascun utente può determinare in maniera distribuita il valore dello SNIR di target da usare, come già specificato, nell'algoritmo di aggiornamento delle potenze. La procedura iterativa che ognuno deve seguire risulta, pertanto specificata dai seguenti passi:

1) si misura l'interferenza ricevuta al passo k ($I_i(k)$) e il valore del guadagno di canale tra trasmettitore e ricevitore del link i -esimo ($G_{i,i}(k)$);

2) si calcola il valore del SNIR _{i} target utilizzando la formula trovata prima e si controlla se è maggiore o minore di γ_i :

- se è inferiore al γ_i , si azzerla la potenza di trasmissione ponendo $P_i(k+1) = 0$;

- se è superiore al γ_i , allora si pone $p_i(k+1) = \min \left\{ p_i(k) \frac{SNIR_i^{tar}(k)}{SNIR_i(k)}, p_{\max} \right\}$ e

$$p_i(k+1) = \max \left\{ p_i(k) \frac{SNIR_i^{tar}(k)}{SNIR_i(k)}, p_{\min} \right\}.$$

Per velocizzare l'algoritmo, e in accordo con la figura della derivata della funzione utilità rispetto ad A_i , poniamo la potenza di trasmissione a zero quando il valore di A_i è maggiore di 0.25. Per tali valori infatti sicuramente non troveremo un punto di equilibrio, poiché non ci sono intersezioni tra la retta e la derivata della curva.

3.4.10 Funzione utilità basata sulla capacità di Shannon

In alternativa al rapporto segnale-rumore-interferenza, è stato considerato un altro parametro di riferimento per la QoS, dato dalla capacità di Shannon che, come già detto, risulta proporzionale allo SNIR secondo la formula

$$C = B \cdot \log_2(1 + SNIR)$$

In tal modo, la funzione utilità risulta espressa da

$$U_i(C_i = x) = \frac{1}{1 + e^{-a_i(x-b_i)}}$$

Utilizzando la stessa funzione costo, quella lineare con la potenza, si ha che l'utilità netta è data da

$$NU_i(p_i) = U_i(C_i) - \alpha_i p_i$$

La maniera di trattare questa funzione è analoga a quella vista in precedenza, con le dovute sostituzioni. Il problema da risolvere è di nuovo

$$\frac{\partial NU_i(p_i)}{\partial p_i} = 0$$

$$\frac{\partial^2 NU_i(p_i)}{\partial p_i^2} < 0$$

Da cui

$$\frac{d(NU_i(SNIR_i, p_i))}{dp_i} = \frac{d(NU_i(C_i) - \alpha_i p_i)}{dp_i} = \frac{d(NU_i)}{dC_i} \frac{d(C_i)}{dp_i} - \alpha_i = 0$$

Solo che stavolta si ha che

$$\frac{d(C_i)}{dp_i} = \frac{1}{\left(p_i + \frac{1}{Z}\right) \ln(2)}$$

dove $Z = \frac{W}{R} \frac{G_{i,i}}{I_i}$.

Per cui, l'equazione che si ottiene è

$$\frac{d(NU_i)}{dC_i} = \alpha_i \ln(2) \left(\frac{1}{Z} + p_i \right)$$

Come sappiamo

$$\frac{d(NU_i)}{dC_i} = \frac{aU_i(x)}{1 - U_i(x)}$$

perciò, ponendo $A_i = \alpha_i \ln(2) \left(\frac{1}{Z} + p_i \right)$, si ottiene la stessa equazione del caso precedente, cioè

$$\left[e^{-a_i(x-b_i)} \right]^2 + \left(2 - \frac{a_i}{A_i} \right) e^{-a_i(x-b_i)} + 1 = 0$$

la quale, risolta nello stesso modo già visto, consente di trovare

$$C_i^{tar} = b_i - \frac{1}{a_i} \ln \left[\frac{a_i}{2A_i} - 1 - \sqrt{\frac{a_i}{2A_i} \left(\frac{a_i}{2A_i} - 2 \right)} \right]$$

Il valore di b_i stavolta va espresso in funzione di un vincolo minimo sulla capacità, vincolo che, comunque, possiamo legare ancora allo SNIR minimo γ_i che si vuole ottenere, ossia porre

$$C_{\min,i} = \log_2(1 + \gamma_i)$$

A quel punto si ha semplicemente

$$b_i = C_{\min,i} - \frac{\ln(a_i C_{\min,i} - 1)}{a_i}$$

Fin qui, praticamente nulla di nuovo; quello che cambia è, invece, l'algoritmo di aggiornamento della potenza, che adesso diviene, dovendo invertire la formula della capacità per trovare la potenza,

$$p_i(k+1) = \frac{e^{(C_i(k) \cdot \ln(2) - 1)}}{Z}$$

e che, con alcuni semplici passaggi, può trasformarsi in

$$p_i(k+1) = \frac{(1 + Zp_i(k))^{\ln(2)}}{e \cdot Z}$$

Anche questa funzione risulta *standard* secondo il criterio specificato in precedenza.

CAPITOLO 4

SIMULAZIONI E RISULTATI

4.1 Parametri fisici del sistema simulato

Per la scelta di molti parametri di livello fisico, si è fatto riferimento allo standard IEEE 802.11b, che è considerata l'evoluzione del DSSS 802.11, tant'è che è anche conosciuto come HR-DSSS (High Rate Direct Sequence Spread Spectrum). La banda di lavoro scelta è, pertanto, la cosiddetta *unlicensed* ISM a 2.4 GHz (Industrial Scientific Medical 2.4-2.4835 GHz), così chiamata perché per il suo uso non è richiesta una licenza, a patto che la modulazione sia a spettro allargato e la potenza trasmessa inferiore a 1 W. Essa è indicata anche come ISM2, dato che tra le possibili scelte *unlicensed* esistono anche 915 MHz e 5.8 GHz. L'ampiezza complessiva di questo intervallo frequenziale è 83.5 MHz e in essa sono allocati 13 canali a RF sovrapposti, ognuno ampio 22 MHz e con frequenze centrali separate 5 MHz. La frequenza portante dell'*i*-esimo canale può essere espressa così:

$$f_{ch_i} = 2.412MHz + (i - 1) \times 5MHz, \quad i = 1, \dots, 13$$

L'ambiente oggetto della simulazione è il classico ambiente urbano outdoor in cui, tra i fenomeni che caratterizzano la propagazione del segnale, sono da comprendere la riflessione, la diffrazione, la diffusione, etc. Se il sistema prevedesse la presenza di una base station a cui ogni mobile fa riferimento, le distanze di trasmissione potrebbero risultare anche molto estese e la visibilità ottica tra ricevitore e base station sarebbe presente solo raramente, a causa di costruzioni ed altri ostacoli. Sebbene andasse considerato anche il tipo di applicazione per cui viene usata, in una rete ad hoc, in

generale, le distanze tra mobili sono abbastanza più piccole rispetto a un sistema cellulare e, molto più frequentemente, è presente un percorso in LOS tra mobili, specialmente in caso di elevata densità di nodi.

È noto che l'attenuazione subita dal segnale che si propaga in spazio libero è descritta dalla formula (in dB)

$$A_{fs}|_{dB} = 20 \log \left(\frac{4\pi d}{\lambda} \right)$$

in cui il pedice *fs* sta proprio per *free space*. Essa descrive l'attenuazione subita dal segnale che si propaga in uno spazio "ideale", omogeneo, privo di ostacoli e riflessioni: essa è semplicemente proporzionale al quadrato della distanza. Ovviamente, nella realtà, come accennato in precedenza, sono presenti innumerevoli fattori di attenuazione, per tener conto dei quali è necessario utilizzare un'espressione modificata più veritiera per l'attenuazione:

$$A|_{dB} = A_{fs}|_{dB}(d_0) + 10\gamma \log_{10}(d/d_0)$$

Si introduce il parametro γ , detto *path loss exponent*, che rende più flessibile il modello di propagazione, in quanto può essere variato a seconda del contesto che si vuole analizzare (vedi tabella 4.1).

Environment	Path-loss exponent (γ)
Free space	2
Ideal specular reflection	4
Urban cells	2.7 - 3.5
Urban cells, shadowed	3 - 5
In building, line-of-sight	1.6 - 1.8
In building, obstructed path	4 - 6
In factory, obstructed path	2 - 3

Tabella 4.1

I valori tabulati sono stati determinati grazie a campagne di misure effettuate nei vari ambienti.

Il termine $A_{fs}|_{dB}(d_0)$ rappresenta un valore di *path loss* noto, assunto a una distanza di riferimento d_0 relativamente vicina al trasmettitore. Dai modelli proposti per comunicazioni di tipo “*microcellular*”, si nota che l’attenuazione di propagazione presenta, al variare della distanza, una transizione da un modello di tipo “*free-space loss*”, a uno con γ pari a 3.5 - 4. In particolare, si trova un andamento di tipo spazio libero, se la distanza trasmettitore-ricevitore è inferiore a $4 \frac{h_r h_t}{\lambda}$, dove h_r e h_t sono rispettivamente le altezze delle antenne in ricezione e trasmissione: ponendo $\lambda = .125m$ (portante a 2.4 GHz) e $h_r = h_t = 1.5m$, $4 \frac{h_r h_t}{\lambda}$ vale circa 72m.

È possibile modellare questa transizione con un andamento brusco, oppure graduale, per cui il modello di canale utilizzato è quello indicato in figura 4.1.

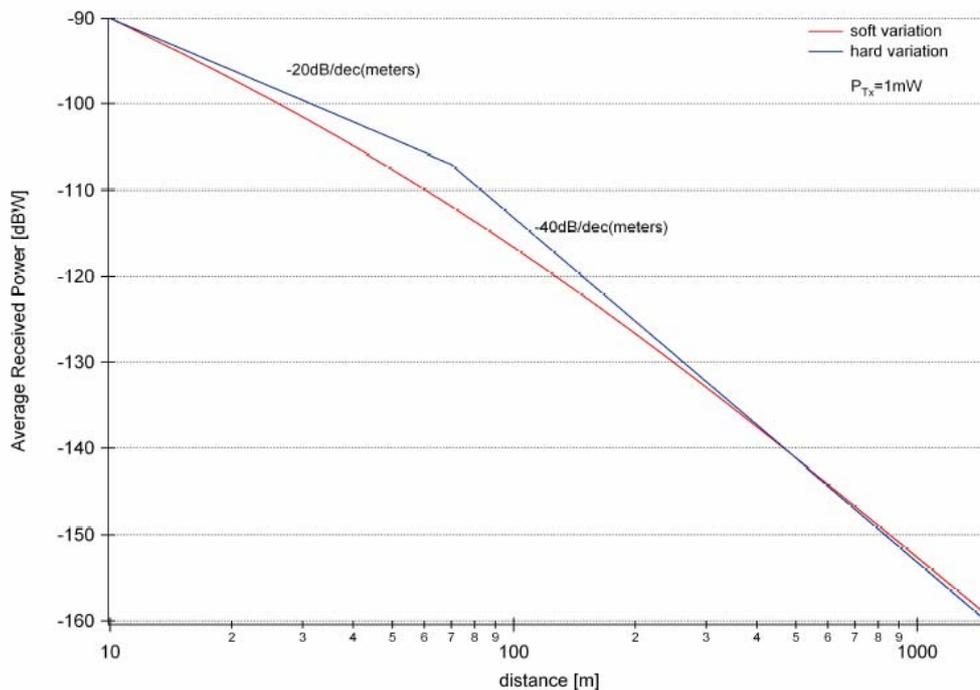


Figura 4.1 – Modello di canale usato

In fase di simulazione in realtà si è scelto 60m come distanza di transizione e andamento brusco per semplicità.

Si è supposto che il canale sia stazionario, ossia l’attenuazione introdotta sia costante durante tutta la trasmissione.

A livello MAC, si è supposto di usare una tecnica di accesso CDMA-based, in cui la banda del segnale espanso è proprio uguale a 22 MHz e *processing gain* pari a 64; pertanto il bit-rate informativo è dato da $22 \text{ MHz}/64 = 343.75 \text{ kbit/s}$ per ogni terminale.

Inoltre si è supposto che:

- i nodi siano dotati di antenne omnidirezionali;
- i dispositivi trasmettitori siano in grado di regolare la loro potenza trasmissiva a passi di 0.1, 1 o 10 dB, su eventuale comando dei rispettivi ricevitori;
- il range di potenze erogabili sia ampio 80 dB, con una potenza massima di 20 dBm;
- la cifra di rumore di ciascun ricevitore sia pari a 5 dB;
- la densità spettrale monolaterale di rumore sia KT , dove K è la costante di Boltzmann e T è la temperatura dell'ambiente, espressa in kelvin, scelta pari a 290 k.

4.2 Simulazione

La simulazione prevede la creazione di una cella, che può essere di forma circolare o rettangolare, in cui vengono distribuiti casualmente, per diverse iterazioni (ad esempio 1000), un numero fissato di nodi, tra cui si instaurano collegamenti unidirezionali, di tipo punto-punto, con trasmissione continua e rate costante nell'arco dell'intero tempo di simulazione. Un nodo può eventualmente fungere anche da ripetitore per un altro nodo, dando vita a una trasmissione multi-hop. Le simulazioni svolte non hanno introdotto l'aspetto della mobilità dei nodi, in quanto ci si è voluti concentrare sull'analisi del funzionamento dell'algoritmo di power control implementato. Una volta creata la cella e stabilite le connessioni iniziali, viene applicato l'algoritmo di power control descritto nel capitolo precedente; si è visto che 50 iterazioni sono più che sufficienti per arrivare alla convergenza. Implicitamente, comunque, si è supposto che la frequenza di ripetizione dell'algoritmo sia maggiore rispetto alla banda di coerenza del canale, in modo che la convergenza sia effettivamente garantita.

Si è detto che l'algoritmo è basato su di un gioco non cooperativo, in cui i giocatori sono i nodi ricevitori, che devono pertanto prendere delle decisioni sulla base di osservazioni e misure che devono essere in grado di effettuare; in particolare, si è assunto che essi siano in grado di misurare quantitativamente lo SNIR percepito, e quindi anche l'interferenza sentita, e che siano capaci di comunicare al trasmettitore i comandi per la variazione di potenza da usare in trasmissione. L'attivazione di un link, ad esempio, dipende fondamentalmente dal valore di *SNIR target* che il nodo può calcolarsi se è in grado di misurare l'interferenza percepita e che deve essere maggiore di un certo valore di soglia, che è stato scelto essere 6.78 dB. Tale scelta corrisponde al valore di SNIR necessario per ottenere una probabilità di errore pari a 10^{-3} con una modulazione BPSK.

Al termine delle 50 iterazioni di power control, vengono abbattuti i link caratterizzati da uno SNIR ricevuto più basso di quello minimo e, poi, si prova a riallocare un nuovo link: quando non si riesce ad instaurarne più, si passa all'immagazzinamento dei dati simulativi, che servono per il calcolo dei

parametri statistici che si vogliono misurare, e ad una nuova iterazione, cioè ad una nuova configurazione dei nodi nella cella e all'instaurazione di nuove connessioni.

A causa della lentezza delle simulazioni, per poter variare la densità dei nodi, si è quasi sempre reso necessario mantenere fisso il numero di nodi (20) e variare l'area della cella; fissando $N = 20$, la variazione del raggio della cella è riassunta dalla tabella 4.2.

DENSITY (nodes/km ²)	Radius (m)
10	797,8846
20	564,1896
30	460,6589
40	398,9423
50	356,8248
60	325,735
70	301,572
80	282,0948
90	265,9615
100	252,3133
200	178,4124
300	145,6731
400	126,1566
500	112,8379
600	103,0065
700	95,36545
800	89,20621
900	84,10442
1000	79,78846

Tabella 4.2

4.3 Coefficiente di prezzo

Nel capitolo precedente, è stato mostrato (fig. 3.9) che lo *SNIR target* decresce al crescere della quantità $A_i = \alpha_i \frac{RI_i}{WG_{ii}}$ e che quando tale quantità supera un certo valore (pendenza della retta 2),

l'utilità netta non potrà essere positiva, per cui l'utente i dovrà interrompere la trasmissione.

Chiaramente, il coefficiente di prezzo α_i ha, sul power control, la stessa influenza che ha $\frac{RI_i}{WG_{ii}}$. In

altre parole, le trasmissione viene scoraggiata sia da un grande valore di α_i , sia da una cattiva condizione "ambientale" (ossia un'elevata interferenza, espressa in termini di $\frac{RI_i}{WG_{ii}}$). Un grande

valore di α_i significa che l'utente deve pagare un prezzo più elevato per ottenere praticamente lo stesso "servizio". Le simulazioni hanno però mostrato che, dal punto di vista della rete nella sua

globalità, il coefficiente di prezzo può essere usato come un efficace mezzo per gestire le risorse e massimizzare certi parametri. Il power control è stato modellato mediante la teoria dei giochi come un gioco non cooperativo, in cui ciascun utente mira a massimizzare “egoisticamente” la propria funzione utilità: il coefficiente di prezzo si è dimostrato in grado di rendere le decisioni, prese dagli utenti in maniera decentralizzata, compatibili con l’efficienza complessiva del sistema, promuovendo un comportamento più “sociale” dei nodi stessi, ossia più teso ad un’efficiente condivisione delle risorse piuttosto che ad un’aggressiva competizione. È come se il *pricing* introducesse implicitamente una sorta di cooperazione all’interno del gioco che, per sua natura, rimane non cooperativo.

Nel seguito vengono riportati i risultati ottenuti con una vasta indagine simulativa, in cui l’algoritmo di power control, i cui fondamenti teorico-matematici sono stati trattati nel capitolo precedente, è stato applicato in diverse configurazioni di rete. In particolare, si evidenzia come lo stesso algoritmo si possa specializzare in base alla finalità/metrica di rete perseguita. Come già asserito, infatti, col controllo delle potenze in una rete CDMA ad hoc, si può, da un lato, cercare di ridurre la potenza usata per la trasmissione (mantenendo comunque una certa qualità del segnale ricevuto) e, in tal modo, aumentare il numero complessivo di link attivi (ossia, essendo a rate costante, il *throughput* medio) grazie alla riduzione dell’interferenza complessiva; dall’altro, si può tendere alla massimizzazione della capacità informativa di rete. Il primo criterio è stato chiamato “Power Saving” e per esso è stata utilizzata la funzione utilità basata sullo SNIR, come visto nel capitolo precedente; il secondo criterio è stato chiamato “Max Capacity” e per esso è stata utilizzata la funzione utilità basata sulla capacità di Shannon, che, in termini appunto di massimizzazione della capacità informativa totale di rete, ha prodotto risultati migliori.

Sono stati individuati i parametri che permettono di distinguere le due versioni dell’algoritmo e particolare enfasi è stata data al coefficiente di prezzo, per la scelta ottimale del quale sono state tracciate le linee guida, in dipendenza dal contesto di rete e dalla finalità perseguita dall’algoritmo. Questo coefficiente di prezzo, nelle simulazioni, è stato impostato uguale per tutti i nodi e viene selezionato da loro stessi in base, appunto, al contesto operativo in cui si trovano; in particolare si è visto che esso dipende dalla densità di nodi presenti nella cella e dalla potenza trasmittiva iniziale che viene utilizzata da essi, assunta anch’essa uguale per tutti. Successivamente, è stata presa in esame anche una versione “adattativa” di questo coefficiente, ossia, come sarà spiegato nel seguito, lo si è fatto dipendere da parametri direttamente legati alle condizioni dell’ambiente trasmissivo. Con entrambe le versioni dell’algoritmo, le metriche che sono state esaminate sono:

1. il numero medio di link attivi (o *throughput* medio)

2. la potenza media usata per la trasmissione da ciascun nodo trasmettitore
3. la capacità informativa totale di rete
4. lo SNIR medio percepito da ciascun ricevitore

4.4 Criterio *Power Saving*

Seguendo questo criterio, i parametri da considerare per la scelta del coefficiente sono stati la potenza media di trasmissione e il numero di link attivi.

Il valore ottimo di *alpha*, sulla base del criterio *power saving*, viene scelto all'interno dell'intervallo di valori con cui si ottiene congiuntamente il massimo numero medio di link attivi e il maggior risparmio di potenza mediamente trasmessa, col vincolo che questa non sia superiore a quella iniziale. La figura 4.2 mostra in maniera chiara l'approccio usato.

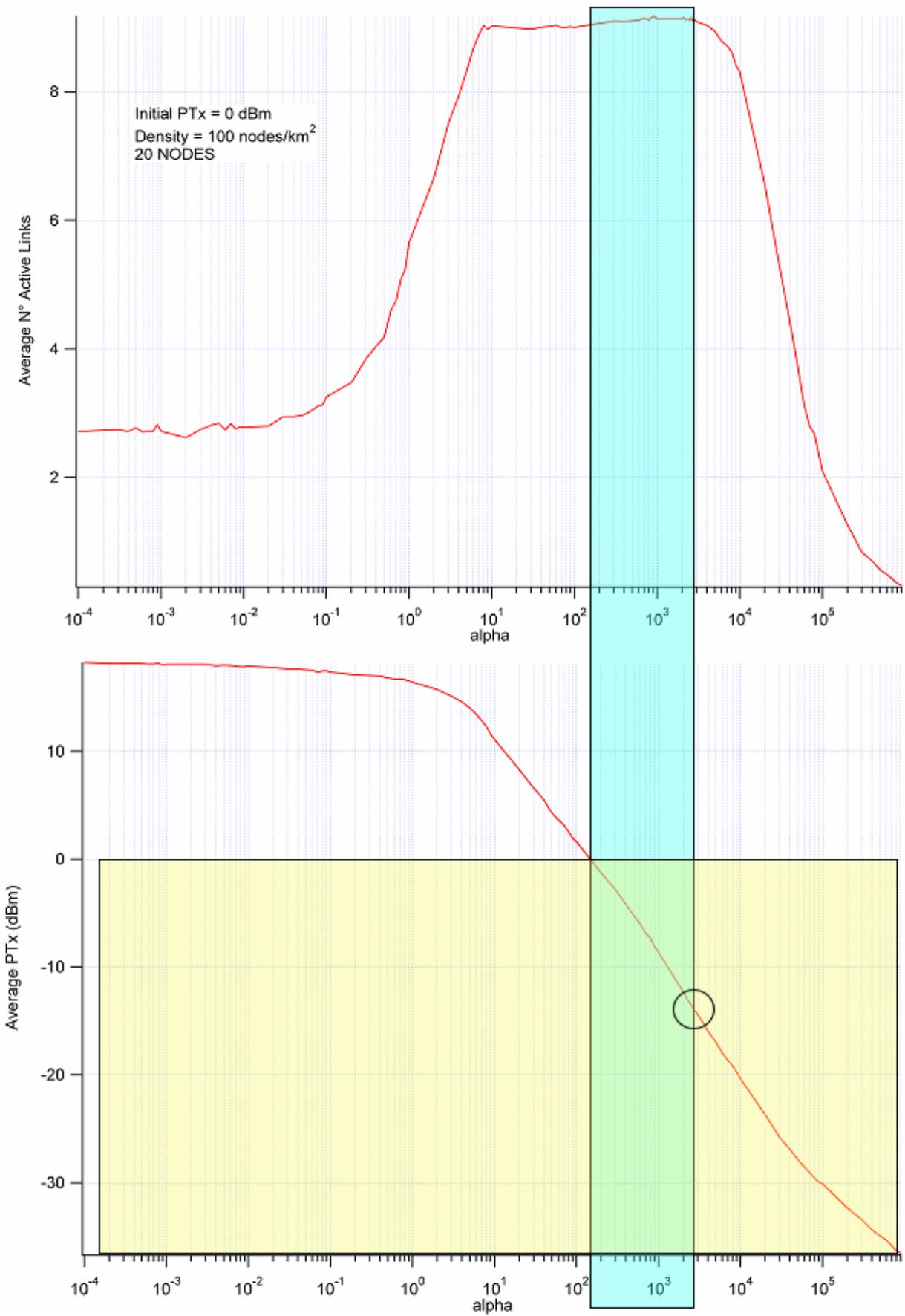


Figura 4.2 – Criterio usato per determinare coefficiente ottimo Power Saving

Le figure 4.3 – 4.6, 4.7 – 4.10, 4.11 – 4.14, 4.15 – 4.18, mostrano l'andamento di tutte e quattro le metriche prestazionali al variare del coefficiente di prezzo (che nei grafici è indicato con α), per diversi valori di densità di nodi.

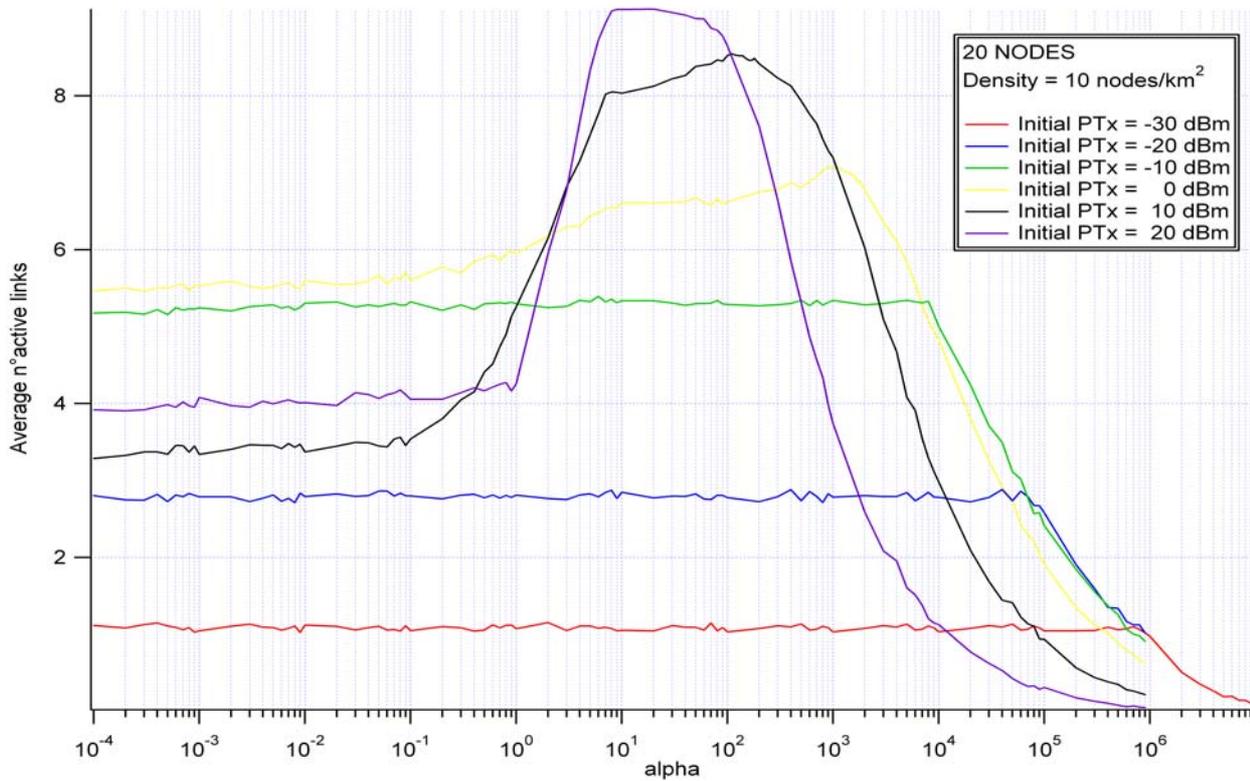


Figura 4.3

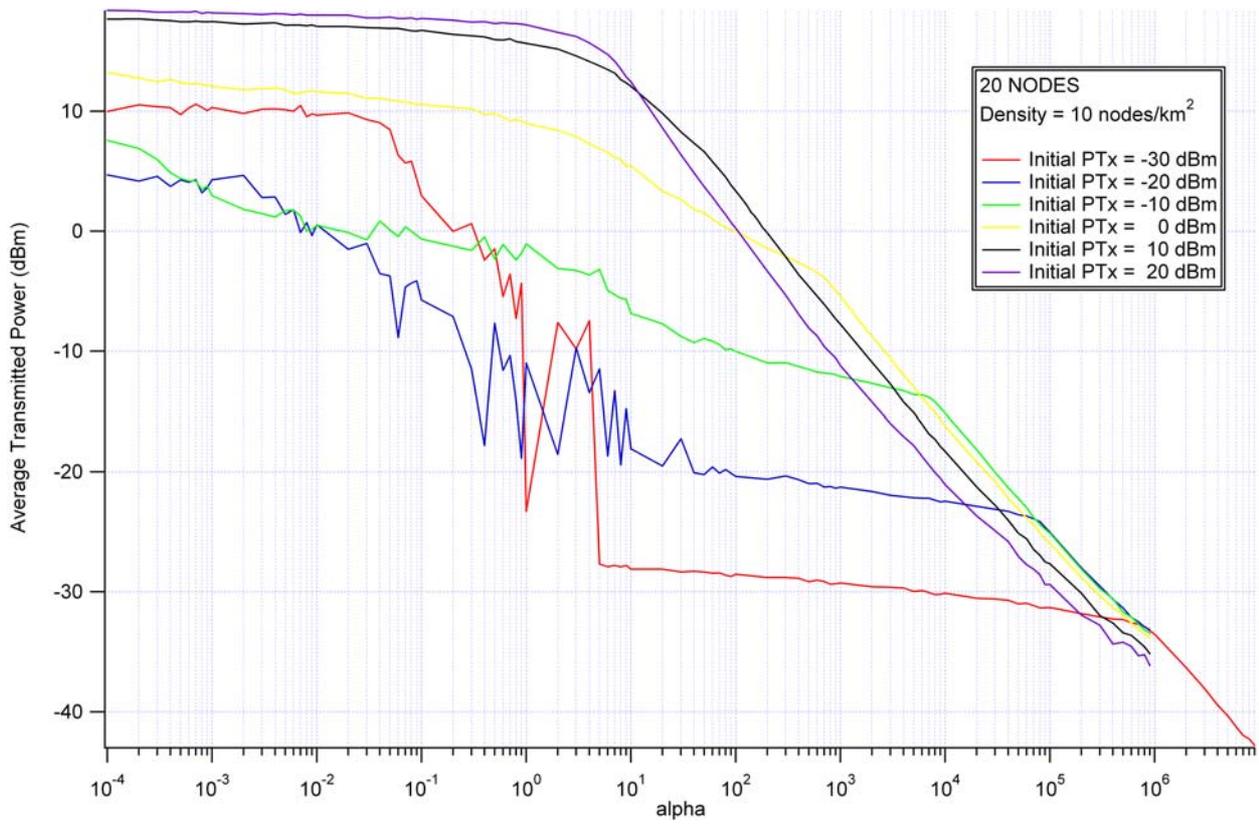


Figura 4.4

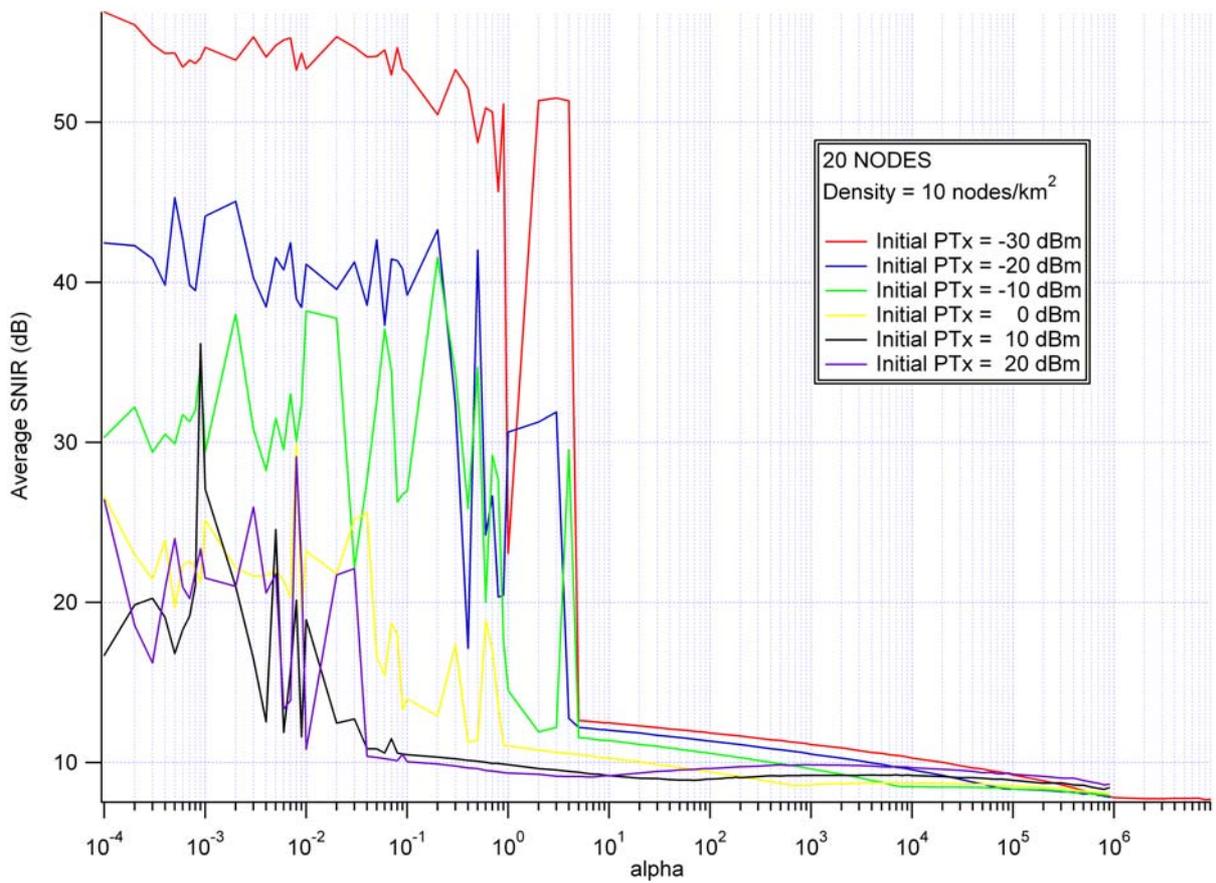


Figura 4.5

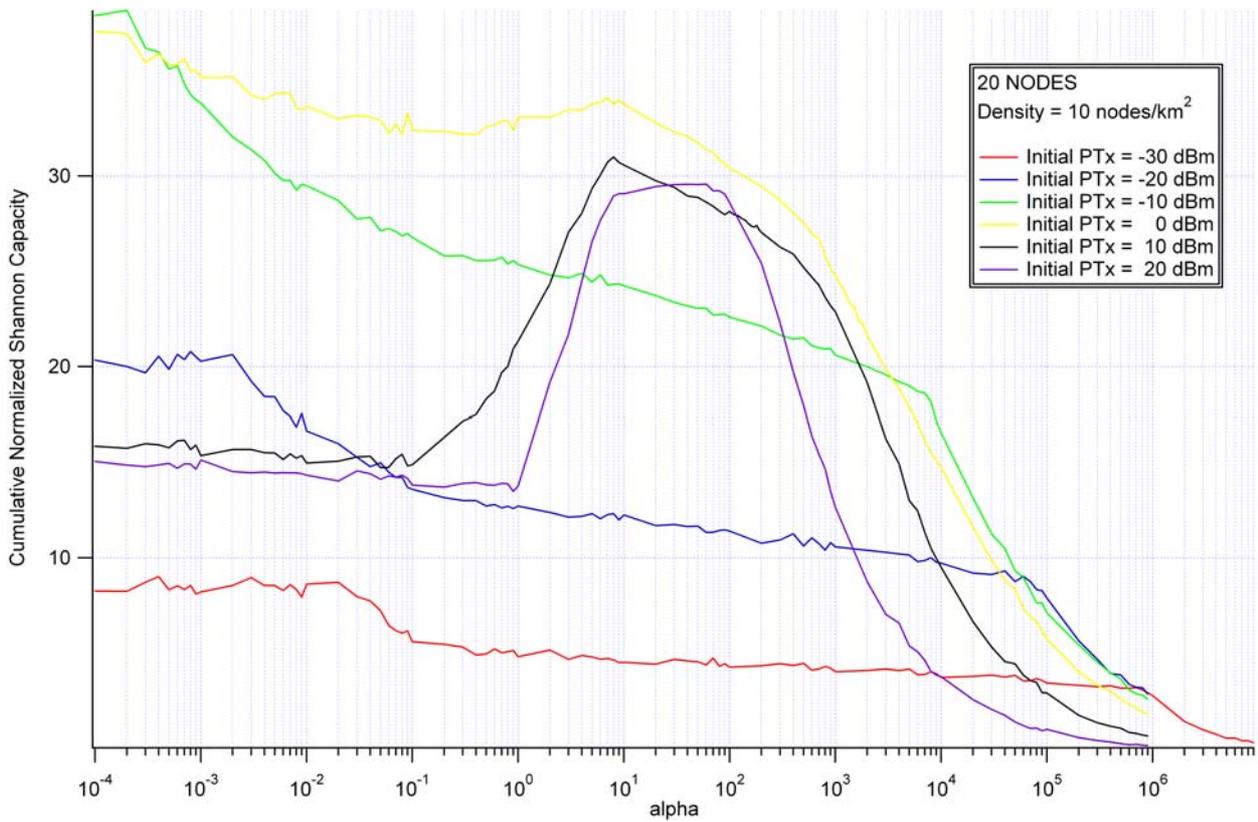


Figura 4.6

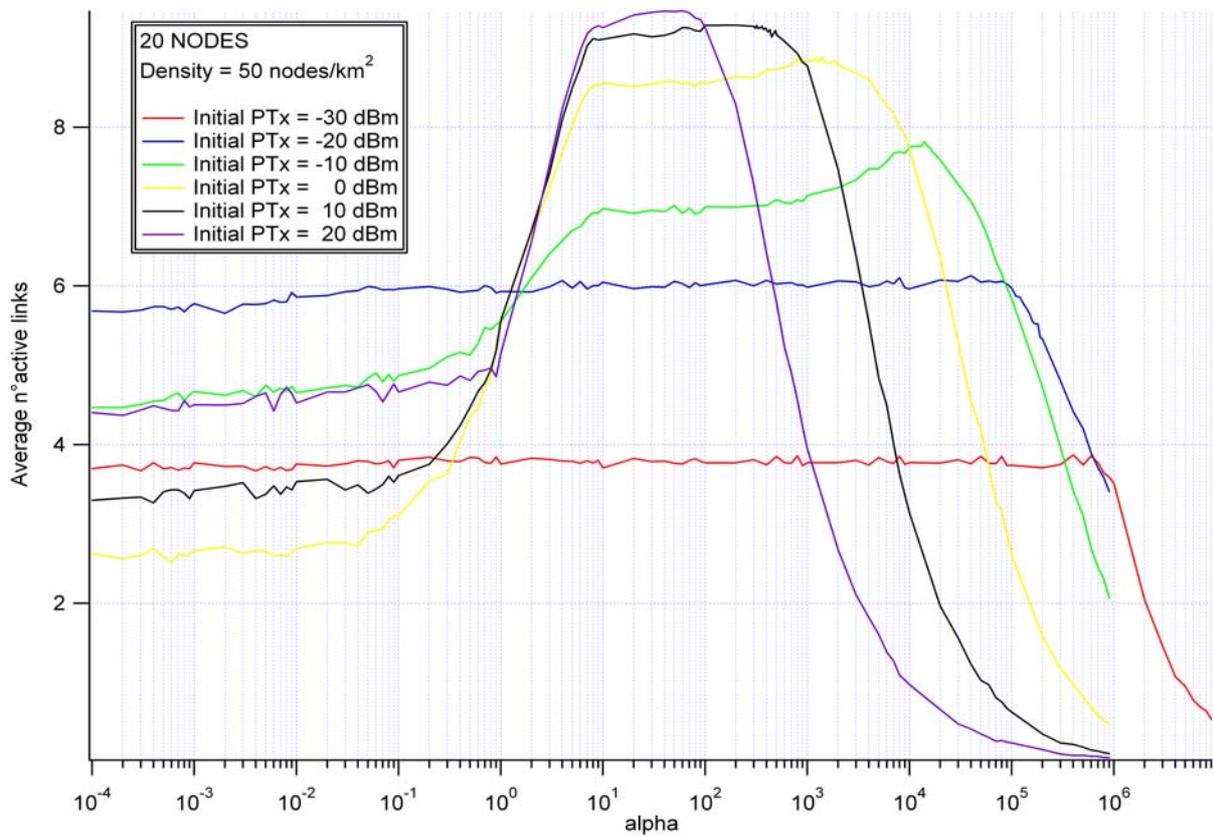


Figura 4.7

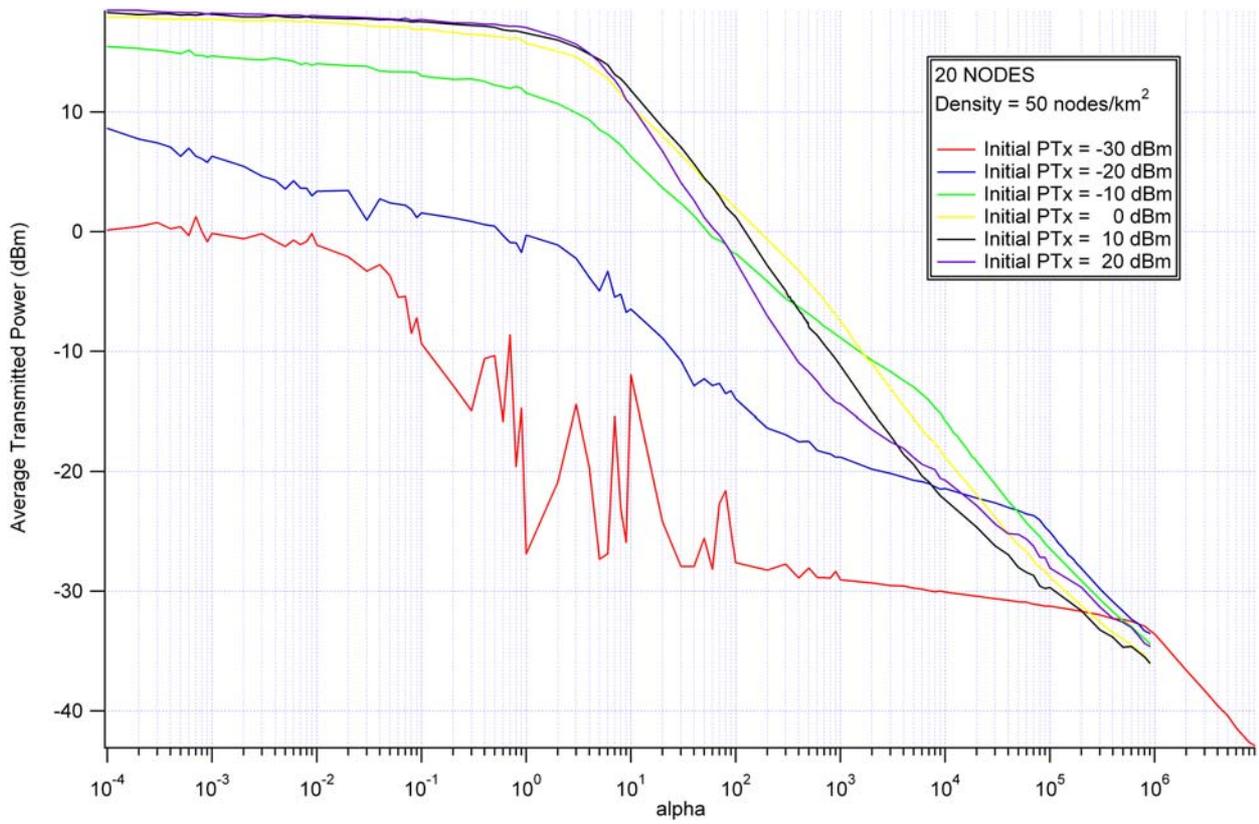


Figura 4.8

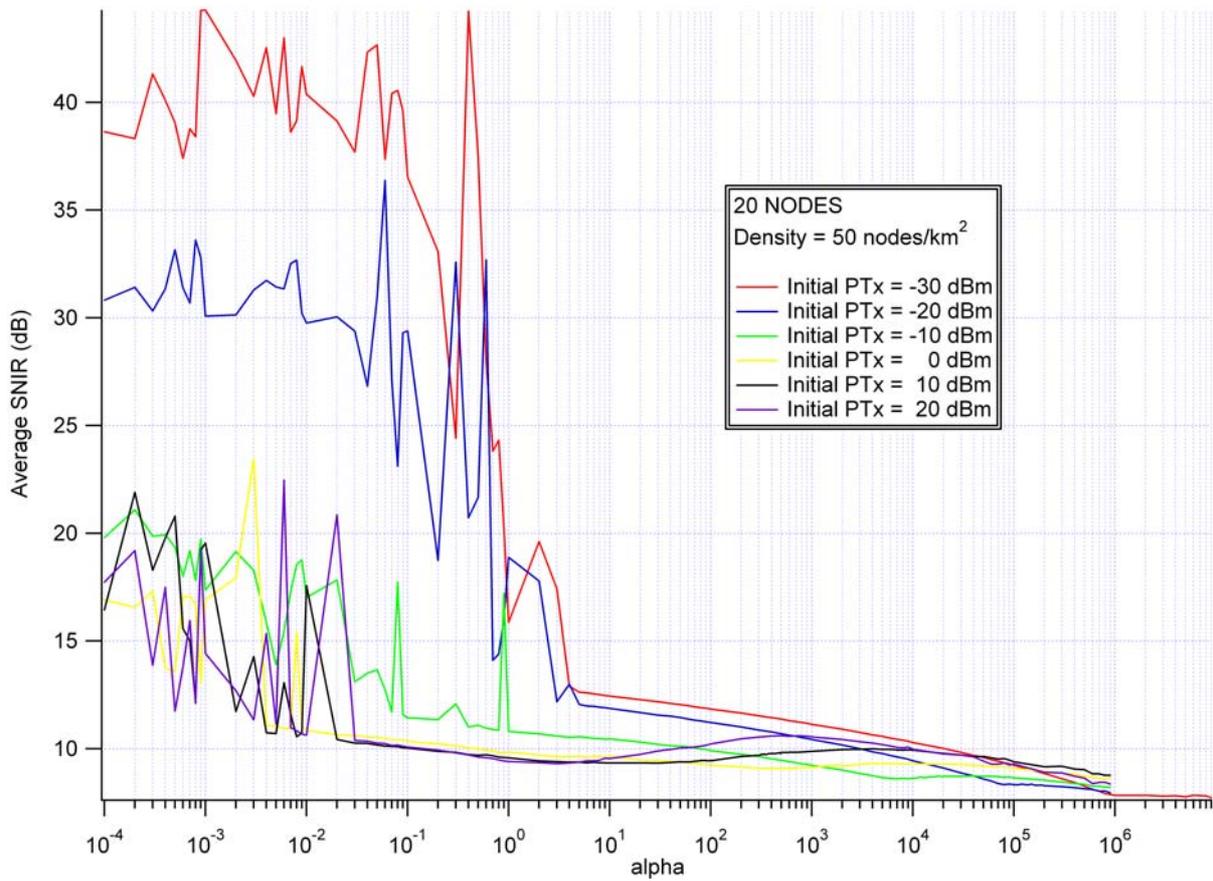


Figura 4.9

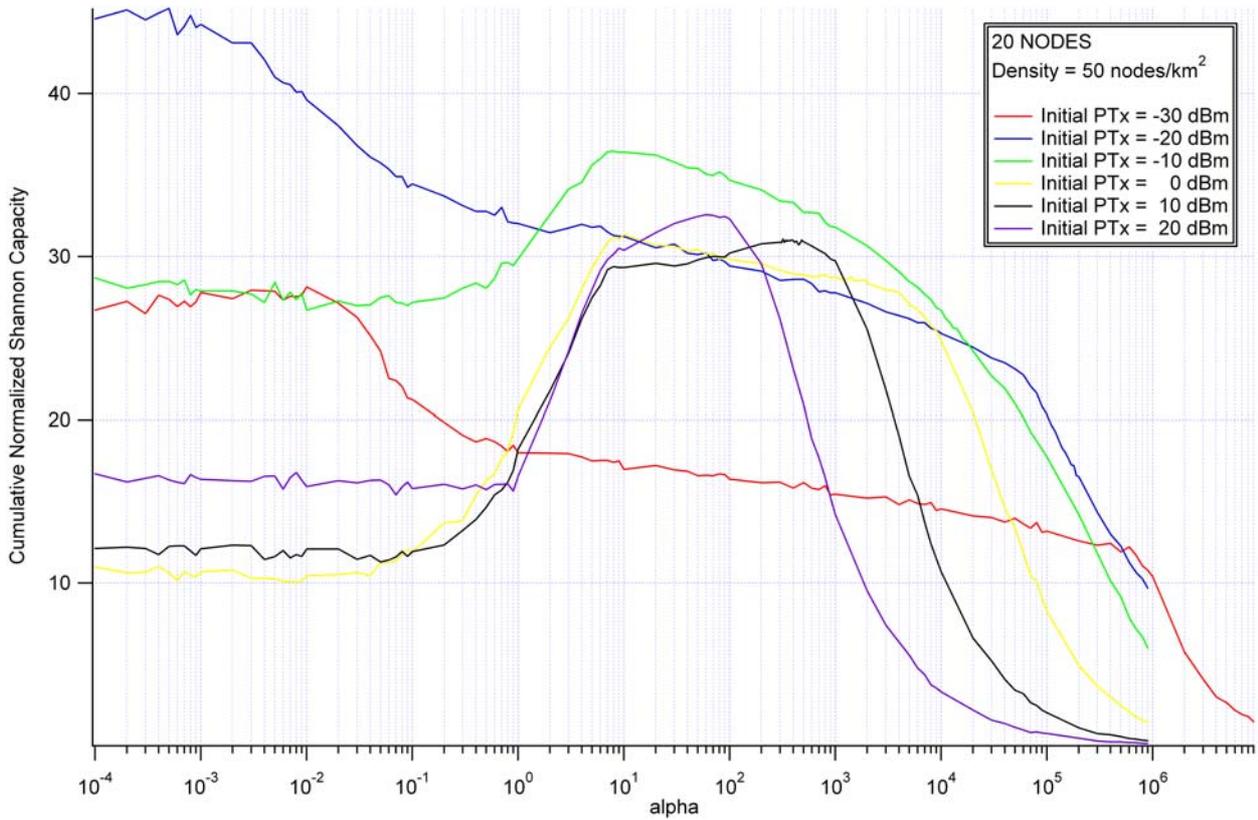


Figura 4.10

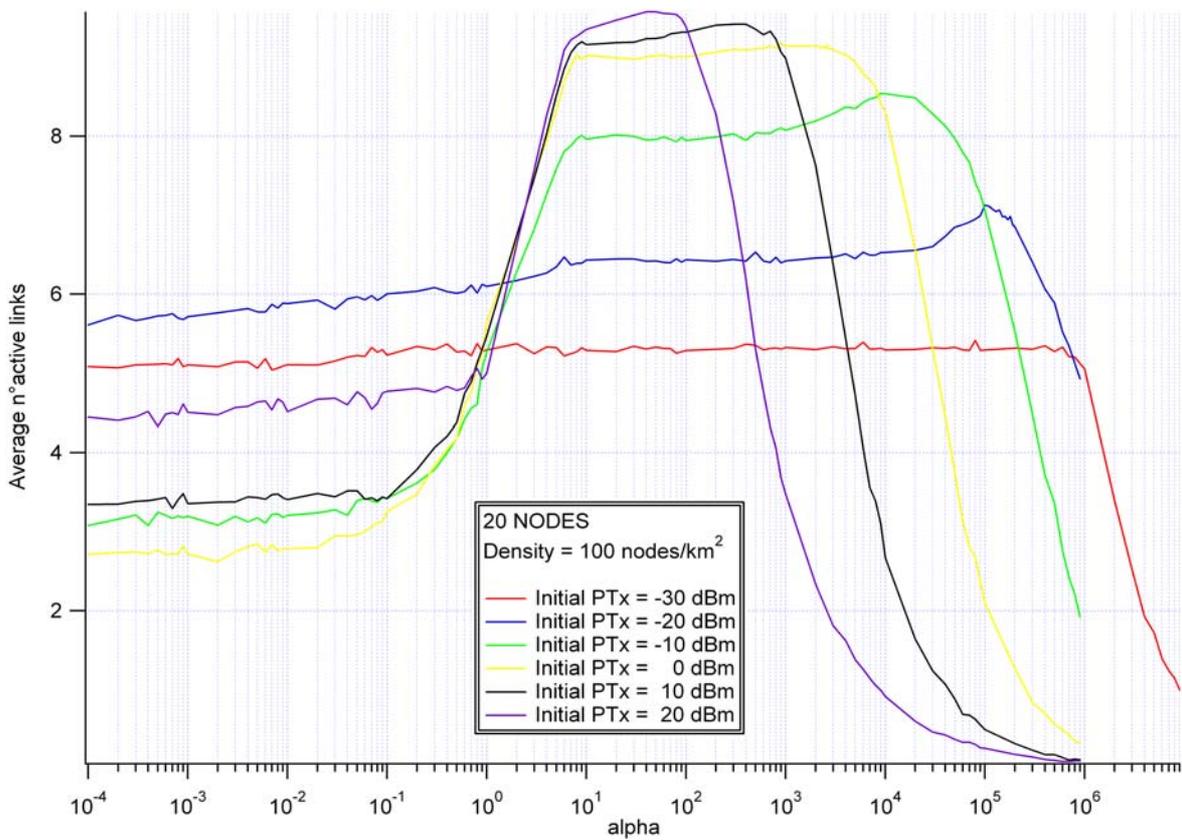


Figura 4.11

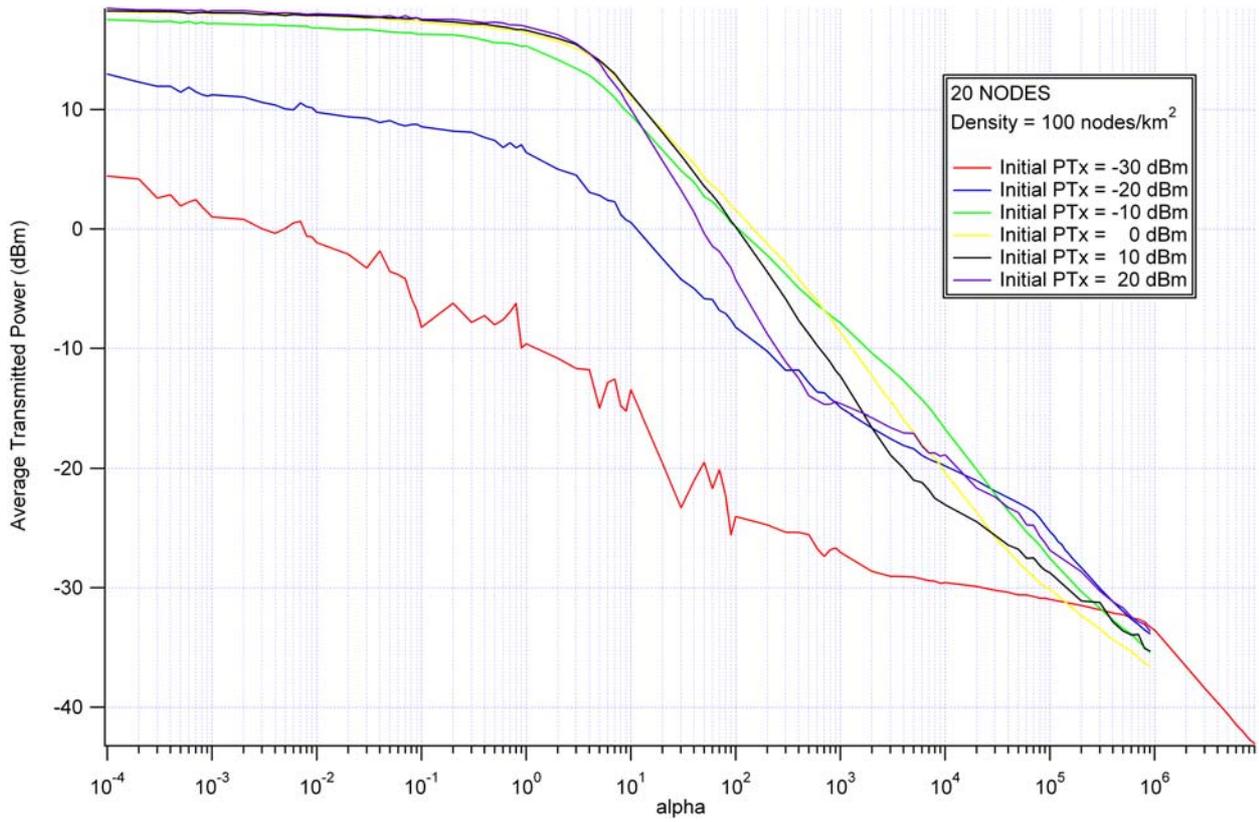


Figura 4.12

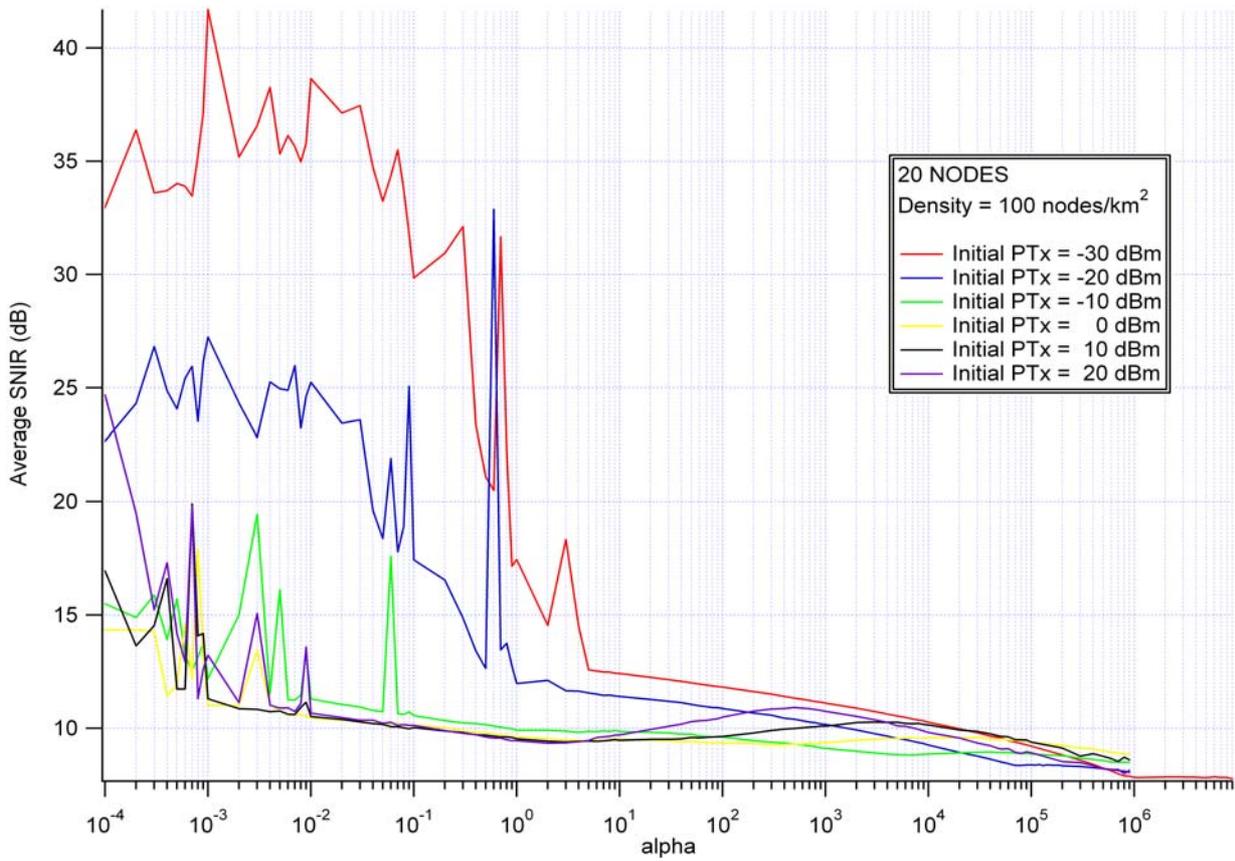


Figura 4.13

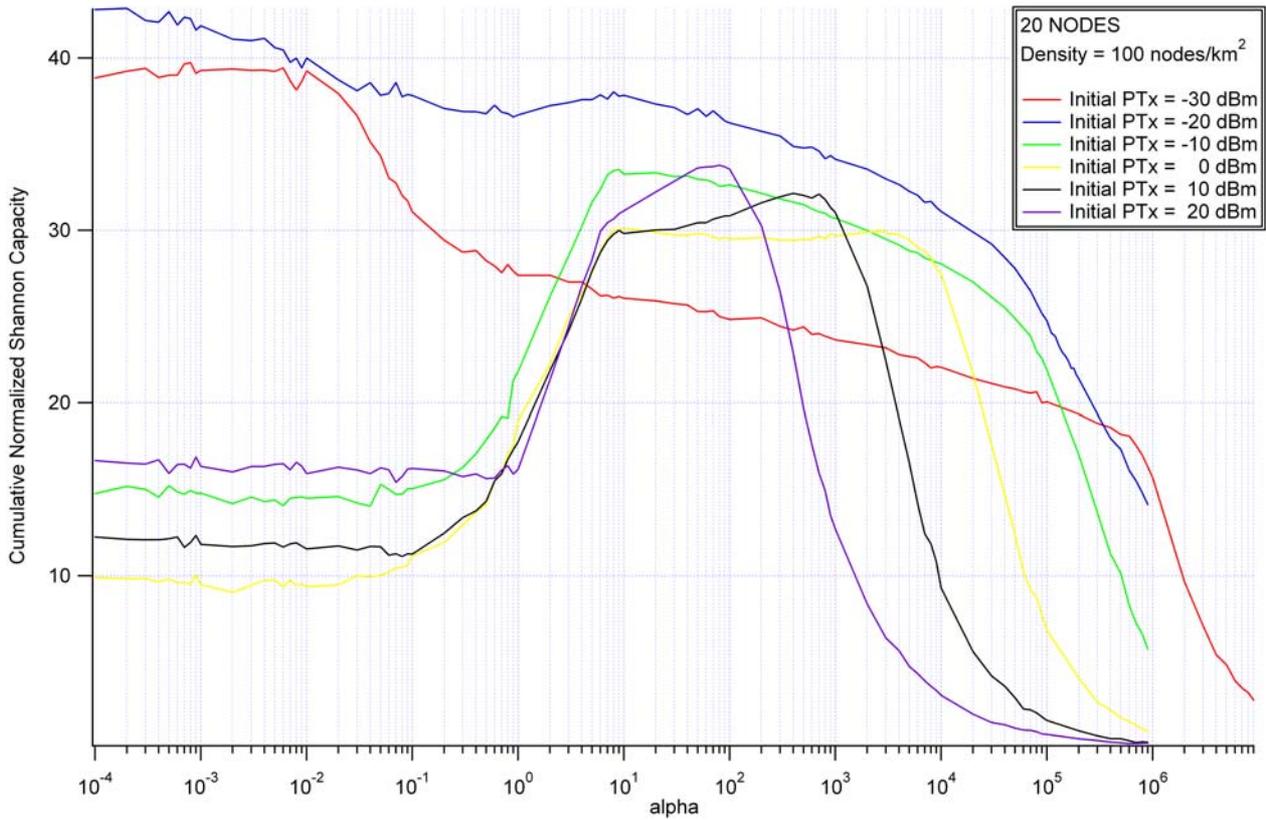


Figura 4.14

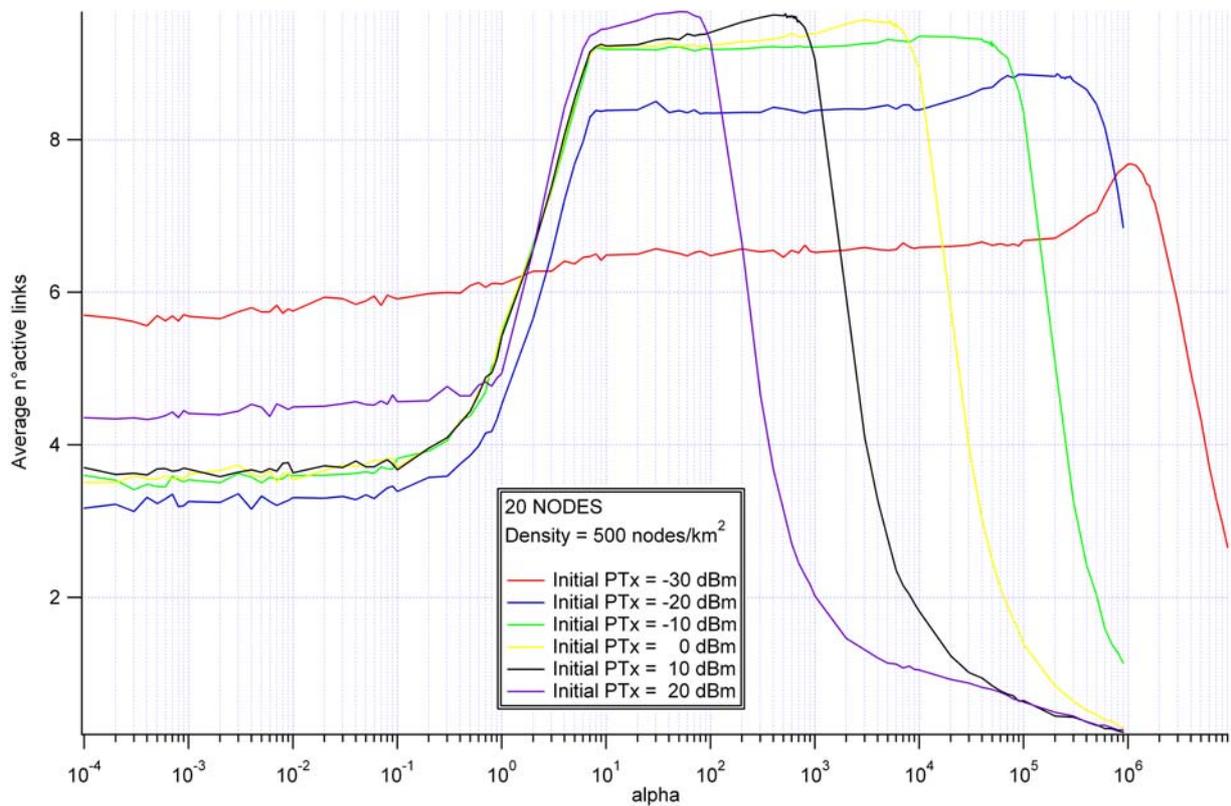


Figura 4.15

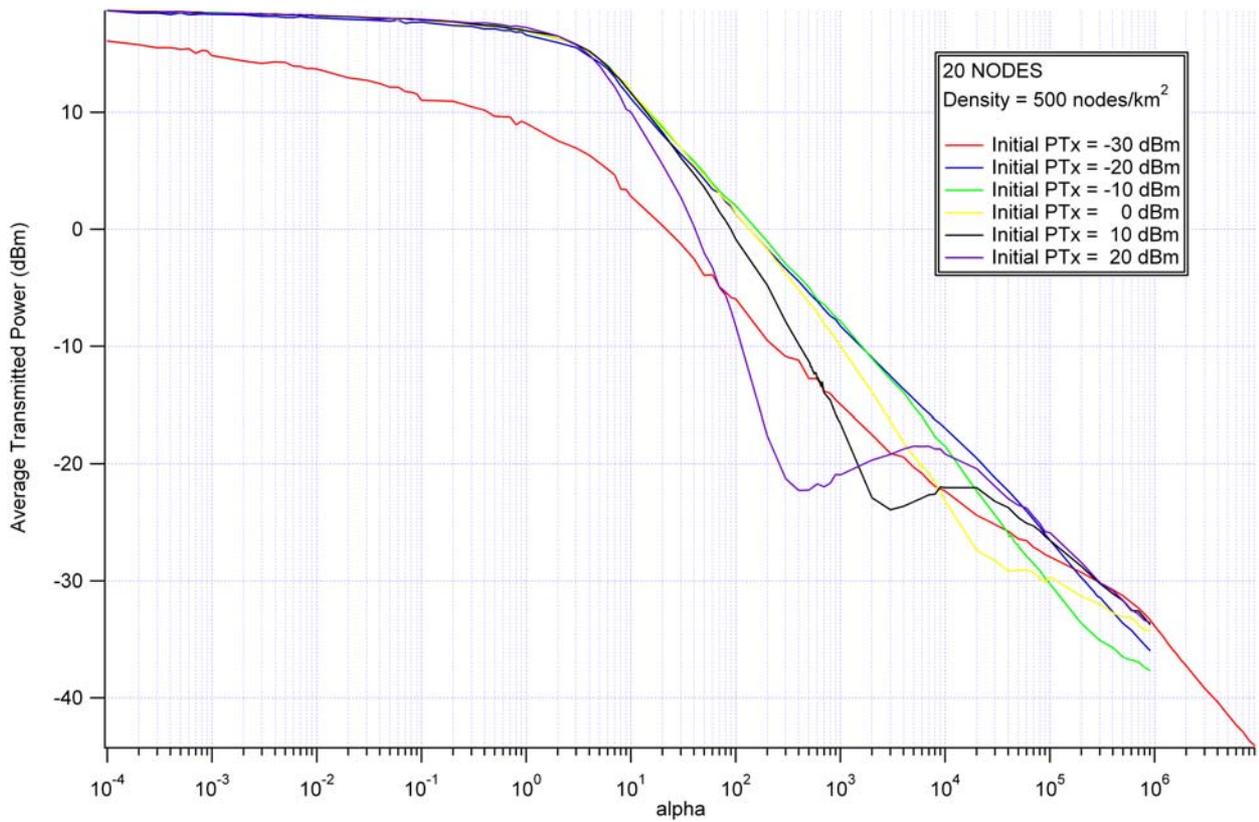


Figura 4.16

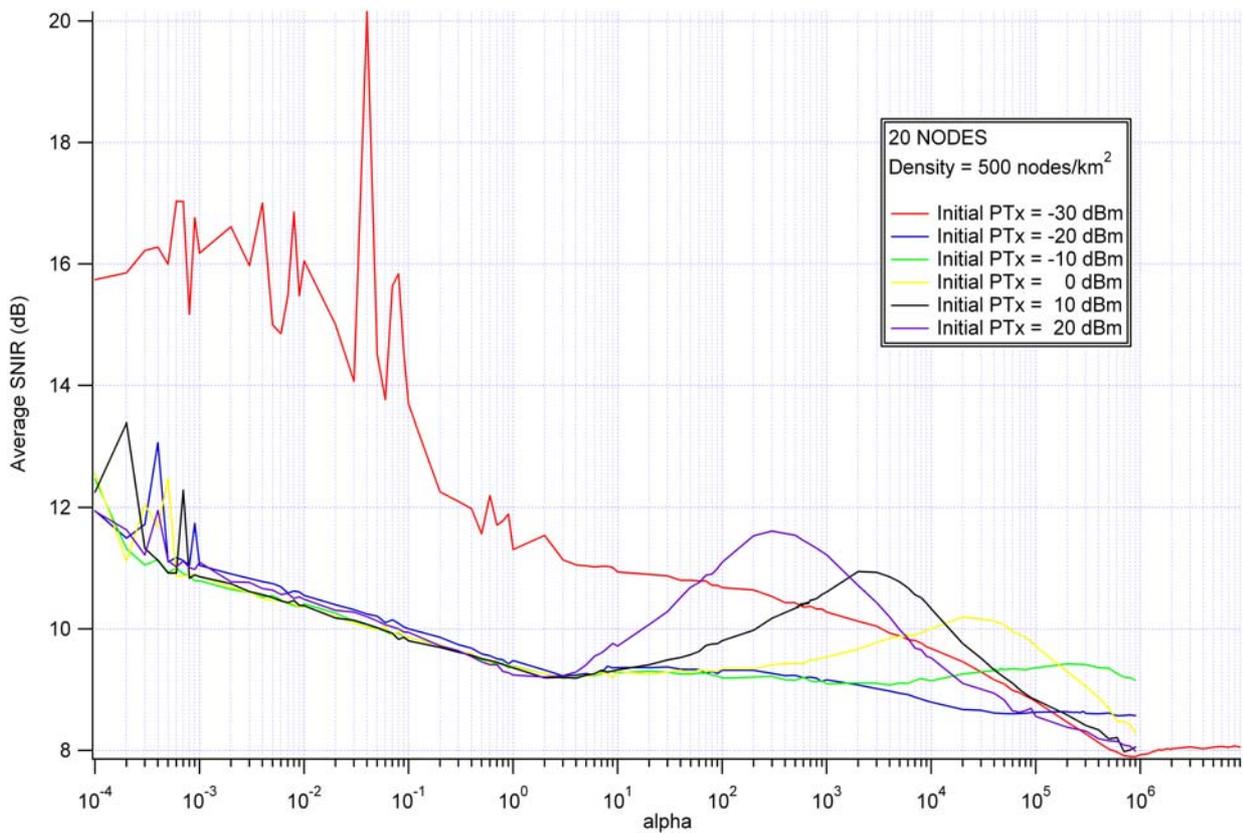


Figura 4.17

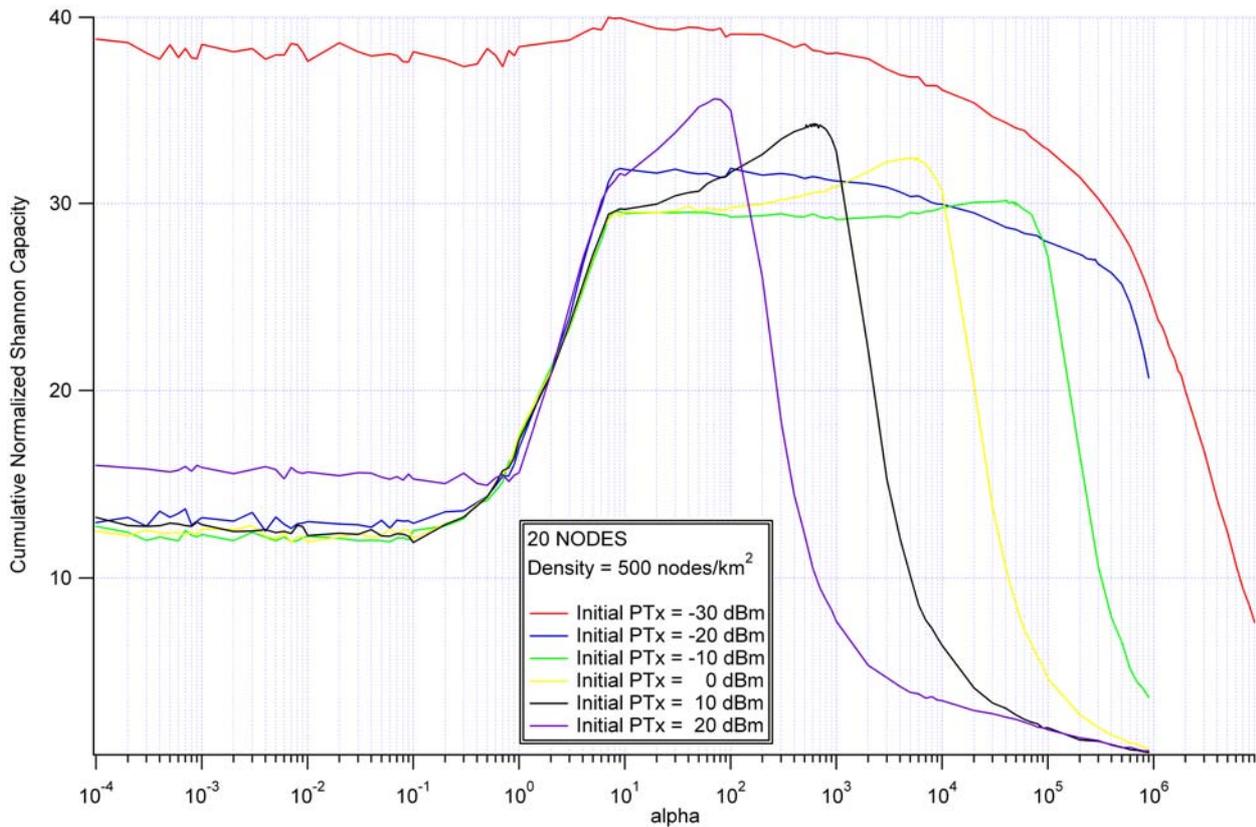


Figura 4.18

Le considerazioni che si possono fare a seguito dell'osservazione di questi grafici, sono:

- in generale, per tutti i valori di densità, più si aumenta la potenza di trasmissione iniziale, più la banda di valori “buoni” del coefficiente α (cioè quelli che consentono di ottenere il massimo numero di link attivi, con risparmio di potenza) diviene stretta, poiché si restringe l'intervallo di valori in corrispondenza del massimo numero di link attivi (fig. 4.3, 4.7, 4.11, 4.15);
- indipendentemente dalla densità dei nodi, l'intervallo $\alpha = (10,100)$ permette la massimizzazione del numero di link attivi per potenze iniziali comprese tra 10 e 20 dBm; per gli altri valori di potenze iniziali si ottiene, comunque, un numero medio di link attivi non lontano dal valore massimo (fig. 4.3, 4.7, 4.11, 4.15);
- in corrispondenza dello stesso intervallo di valori di α , cioè $\alpha = (10,100)$, si ottiene un valore di capacità di Shannon che è indipendente dalla densità di nodo e che non è molto più piccolo dei valori massimi che si sono ottenuti complessivamente con l'applicazione di questo criterio che, si è detto, non è rivolto alla massimizzazione della capacità di Shannon (fig. 4.6, 4.10, 4.14, 4.18);

- al variare della densità dei nodi, cambia l'intervallo di potenze iniziali in corrispondenza del quale si ottiene il massimo numero di link attivi (fig. 4.3, 4.7, 4.11, 4.15): per densità elevate (~ 500 nodi per km^2), tale intervallo spazia da 20 dBm a -10 dBm, mentre per densità più basse (~ 10 nodi per km^2) esso si riduce a 10 ÷ 20 dBm;
- se il livello di potenza di trasmissione iniziale è quello massimo, cioè 20 dBm, in pratica con l'applicazione del power control si ottiene risparmio di potenza per ogni valore di α , mentre con i valori più bassi di potenza iniziale si ha risparmio solo per α superiore ad un certo valore (fig. 4.4, 4.8, 4.12, 4.16);
- l'andamento dello SNIR medio, mentre per valori bassi di α presenta una forte varianza, oltre un certo valore del coefficiente tende ad assestarsi e a decrescere debolmente al crescere di α (fig. 4.5, 4.9, 4.13, 4.17).

Grazie alle simulazioni, si è, dunque, potuto determinare quale deve essere il valore del coefficiente da utilizzare in base alle condizioni operative in cui si trovano i nodi, in particolare in funzione della densità e della potenza iniziale impiegata (fig 4.19).

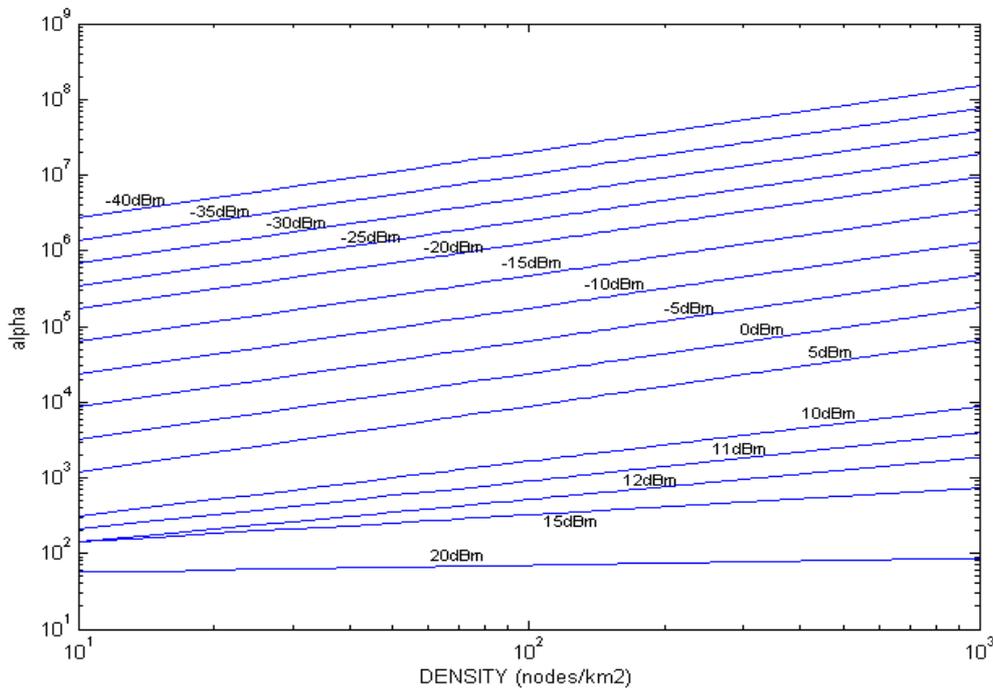


Figura 4.19

Le figure seguenti mostrano il guadagno che si ottiene in termini di *throughput* (fig. 4.20) e di potenza trasmessa (fig. 4.21) utilizzando l'algorithmo di power control con parametro α "ottimo" (dal punto di vista del criterio *power saving*), rispetto al caso in cui viene sempre mantenuto lo stesso livello di potenza.

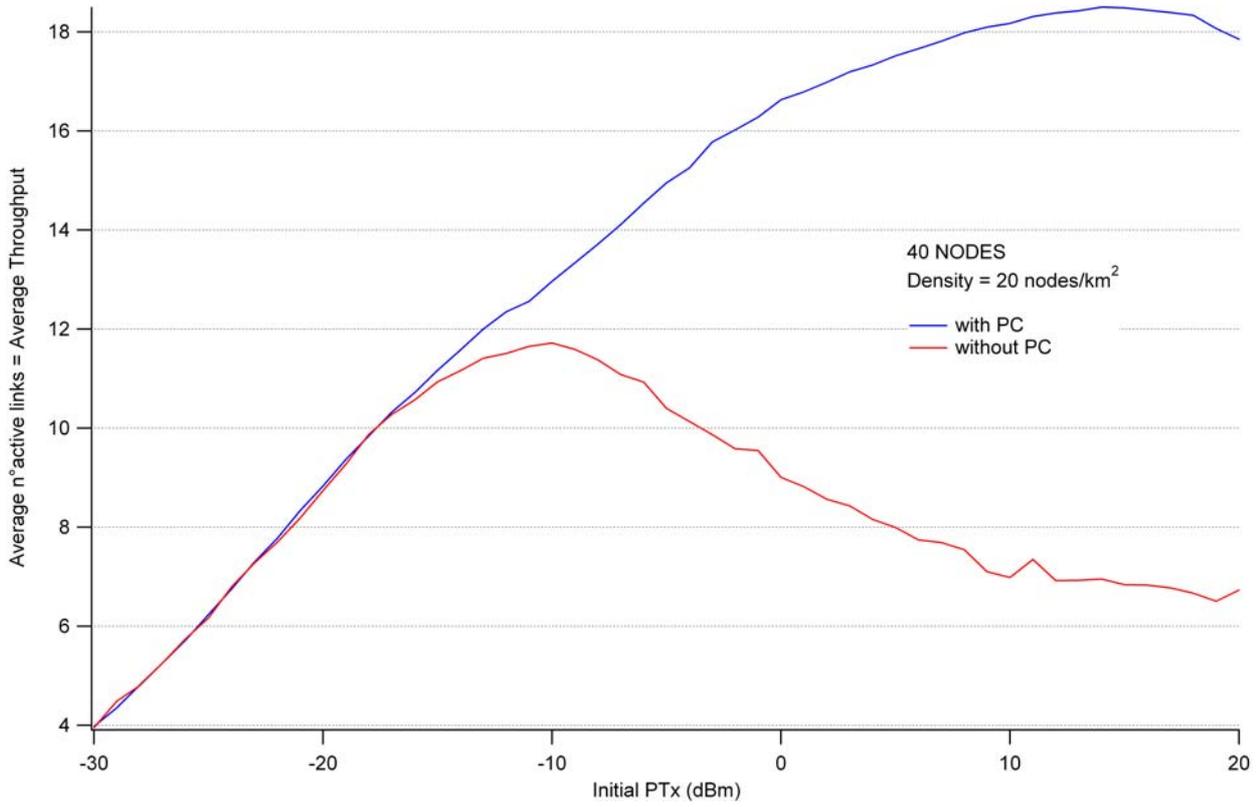


Figura 4.20

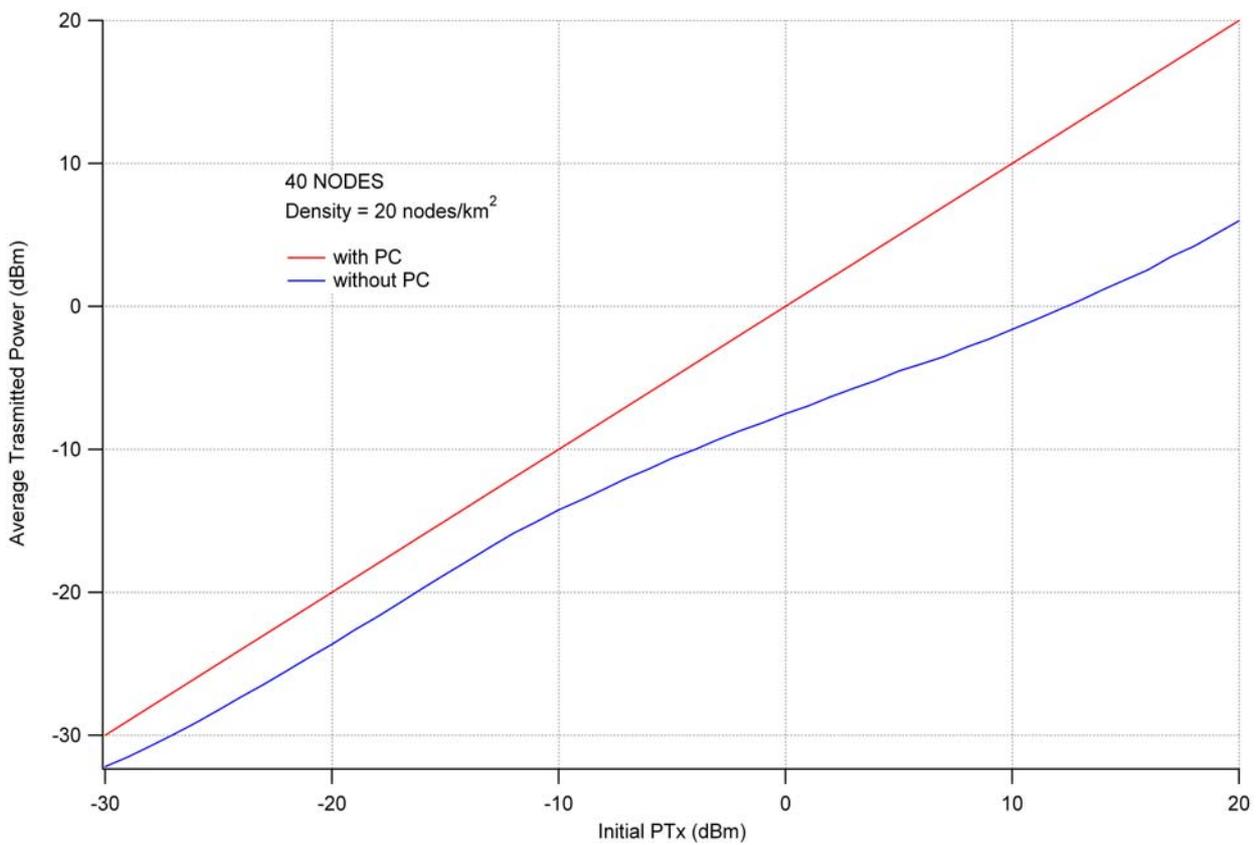


Figura 4.21

Per potenze iniziali comprese tra 10 e 20 dBm, in pratica vengono “accesi” quasi tutti i nodi, arrivando quasi al *throughput* massimo possibile (fig. 4.20), con un risparmio di potenza di quasi 15

dB (fig. 4.21). Per potenze iniziali più basse (nel caso in figura 4.20, inferiori a -15 dBm), non si riescono ad instaurare nuove connessioni, perché questi valori sono “troppo bassi” per poter raggiungere una qualità di segnale sufficiente; si riesce, comunque, ad ottenere un guadagno di potenza di circa 3-4 dB (fig. 4.21), mantenendo lo stesso *throughput*. Un effettivo incremento del numero di link si ottiene, comunque, solo per potenze iniziali superiori a -10 dBm. Per quanto riguarda lo SNIR, coerentemente col modo con cui è impostato l’algoritmo, si ha che, con l’applicazione di esso, i link instaurati sono tutti caratterizzati da valori di SNIR simili, tutti abbastanza vicini a quello minimo richiesto, come è mostrato nelle figure 4.22 e 4.23. La figura 4.23, in particolare, mostra la forte riduzione che subisce la varianza dello SNIR. D’altra parte l’algoritmo ha lo scopo di abbassare il più possibile le potenze, a patto di soddisfare la specifica minima sulla qualità del segnale, quindi c’era da aspettarsi un andamento dello SNIR come quello trovato.

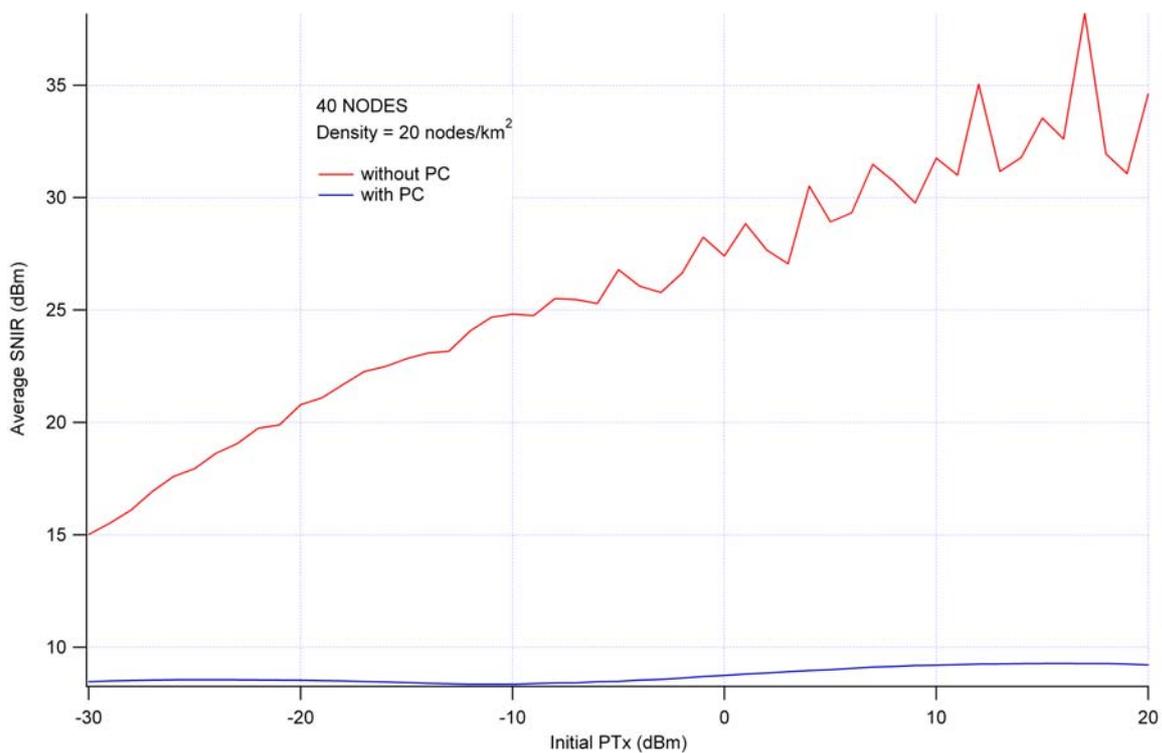


Figura 4.22

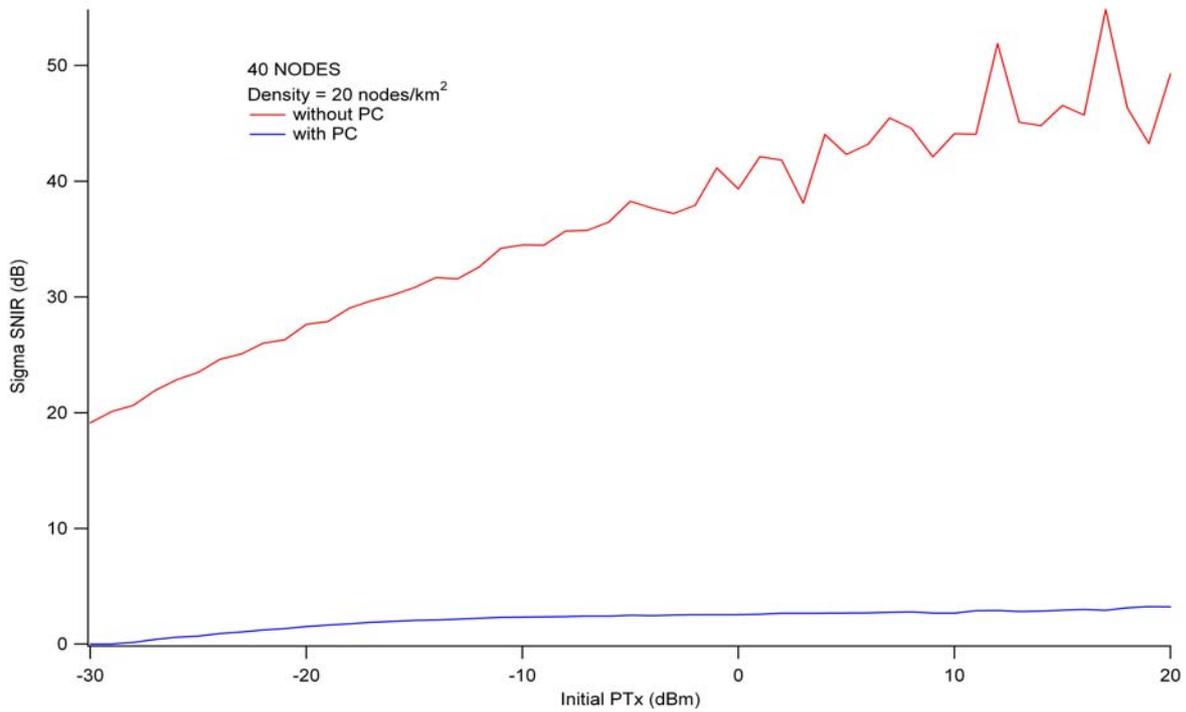


Figura 4.23

Le simulazioni sono poi state ripetute per vari valori di densità dei nodi, confrontando i risultati ottenuti applicando il power control con valore “ottimo” (nel senso *power saving*) di α e senza applicarlo (fig. 4.24, 4.25).

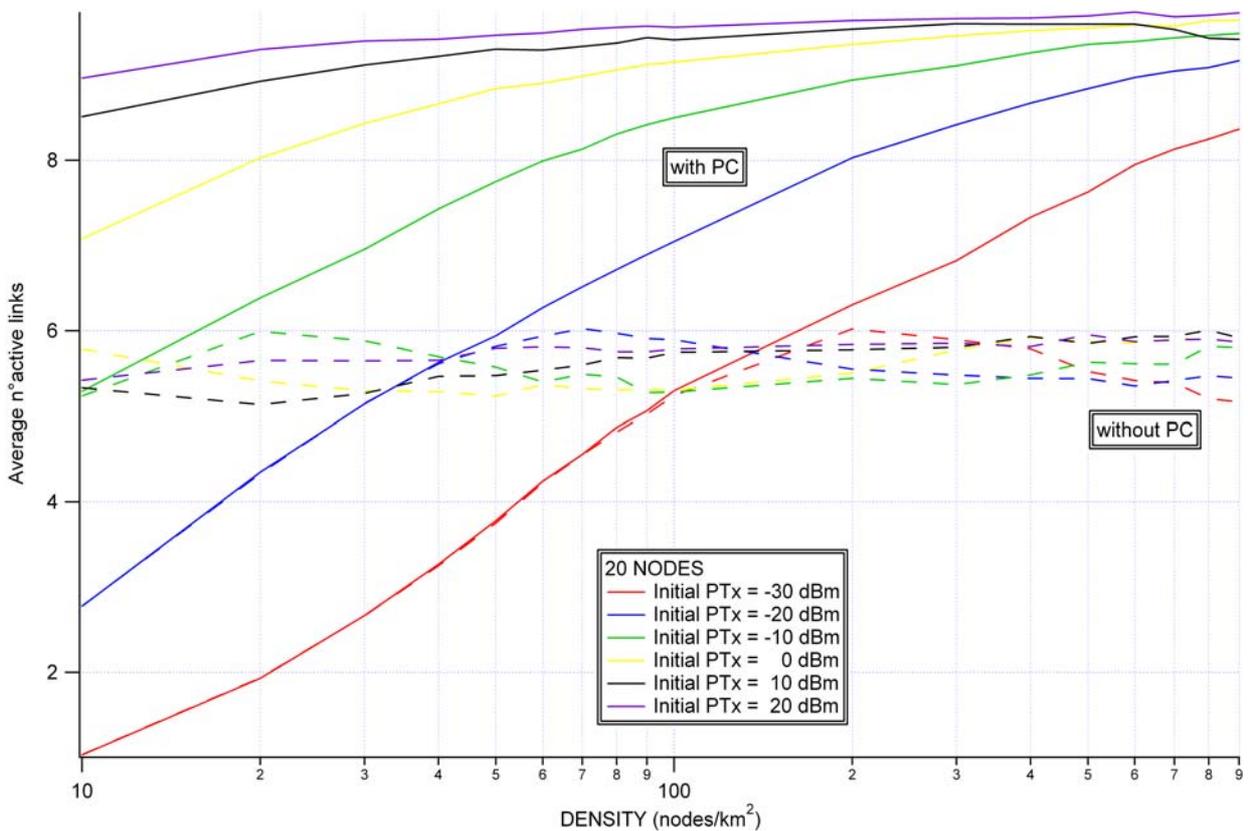


Figura 4.24

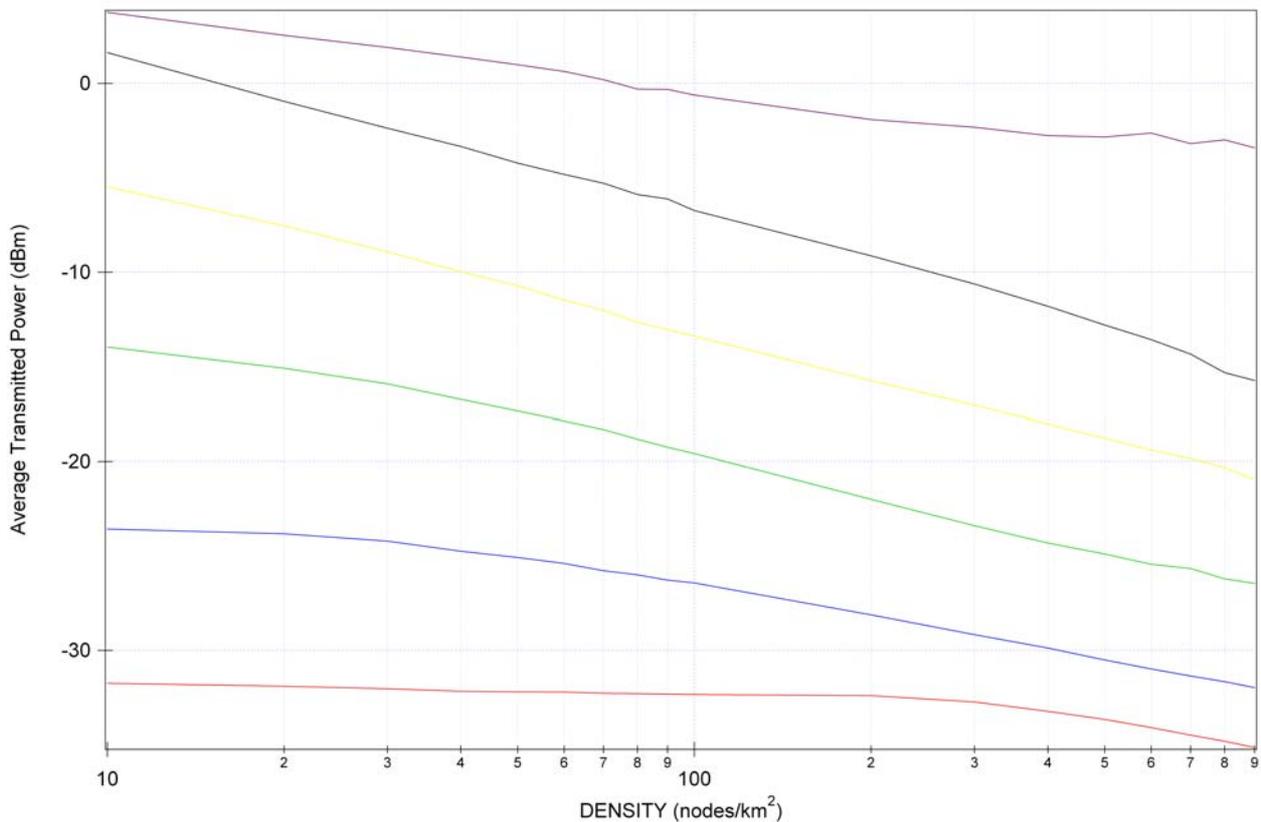


Figura 4.25

Dalla figura 4.24, si osserva che, per potenze superiori a -10 dBm, si ottiene un aumento consistente del throughput medio rispetto alla situazione in cui il power control non è presente, per ogni valore di densità dei nodi.

4.5 Criterio *Max-Capacity*

L'obiettivo che si cerca di raggiungere con l'applicazione di questo secondo criterio, è la massimizzazione della capacità informativa totale di rete, la quale può essere indicata con

$$C = \sum_{i=1}^N C_i = \frac{1}{2} \sum_{i=1}^N \log_2(1 + SNIR_i)$$

dove N è il numero di nodi.

Questa quantità, pertanto, costituisce l'unico parametro preso in esame ai fini della scelta del coefficiente di prezzo *alpha*. Le figure seguenti sono analoghe a quelle realizzate per il criterio precedente. Si può notare che per potenze iniziali relativamente basse, si riescono ad attivare un numero di link inferiore, e che, se si vuole massimizzare la capacità di Shannon, si deve aumentare il più possibile la potenza trasmessa allo scopo di aumentare il più possibile lo SNIR di ciascun

link: dal momento che i link sono pochi, l'interferenza causata dall'aumento delle potenze non riuscirà a peggiorare la qualità dei link attivi. In tal caso, dunque, il coefficiente di prezzo dovrà essere molto piccolo: per semplicità lo si è posto nullo, tanto il valore ottimo di esso si è visto che riuscirebbe a offrire un risparmio di potenza trascurabile. Per potenze iniziali più alte, invece, si riescono ad attivare più link, per cui il metodo precedente non può essere applicato, in quanto un aumento sconsiderato di potenza può causare un'interferenza non più trascurabile per la qualità dei link attivi. Si dovrà, piuttosto, gestire meglio la qualità dei link attivi (a meno che non si decida di spegnerli!), dosando opportunamente la potenza trasmittiva, ossia scegliendo un opportuno coefficiente di prezzo.

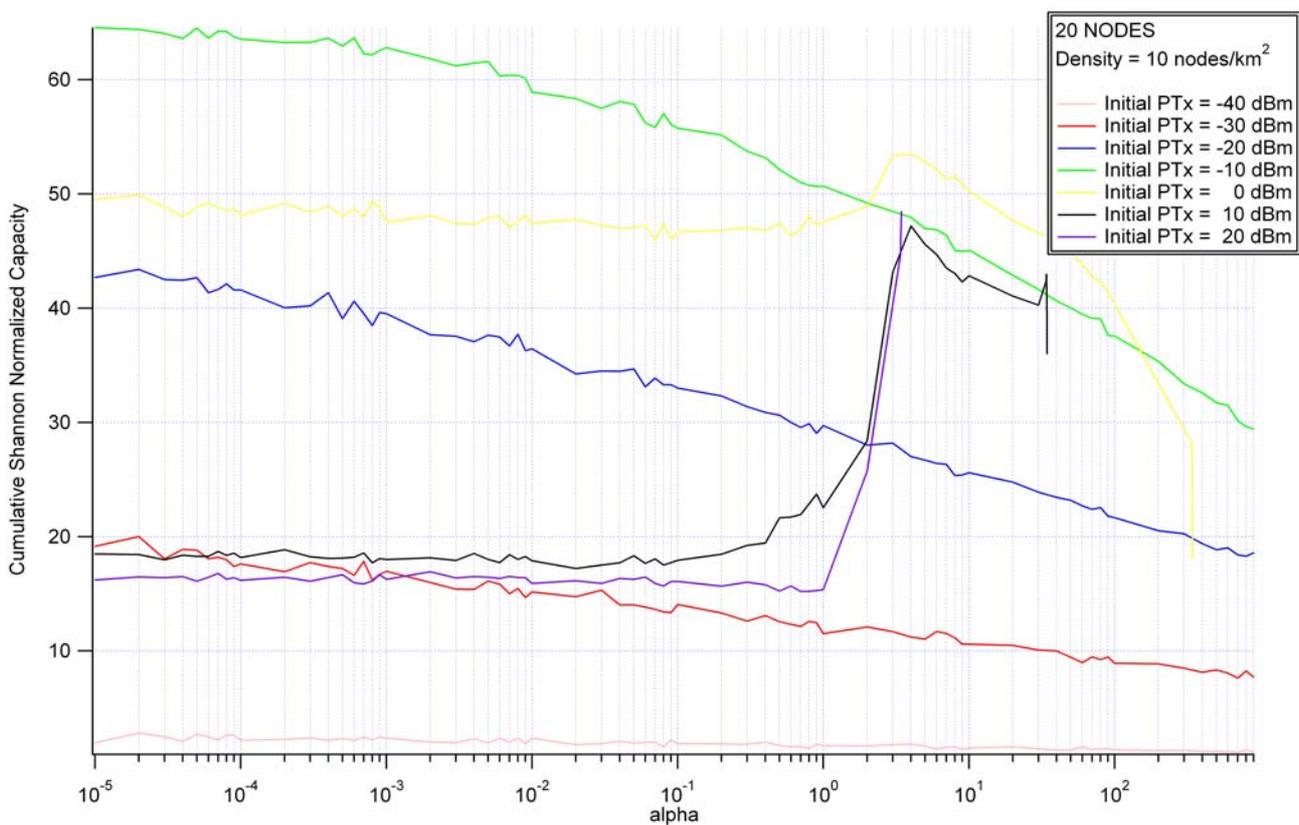


Figura 4.26

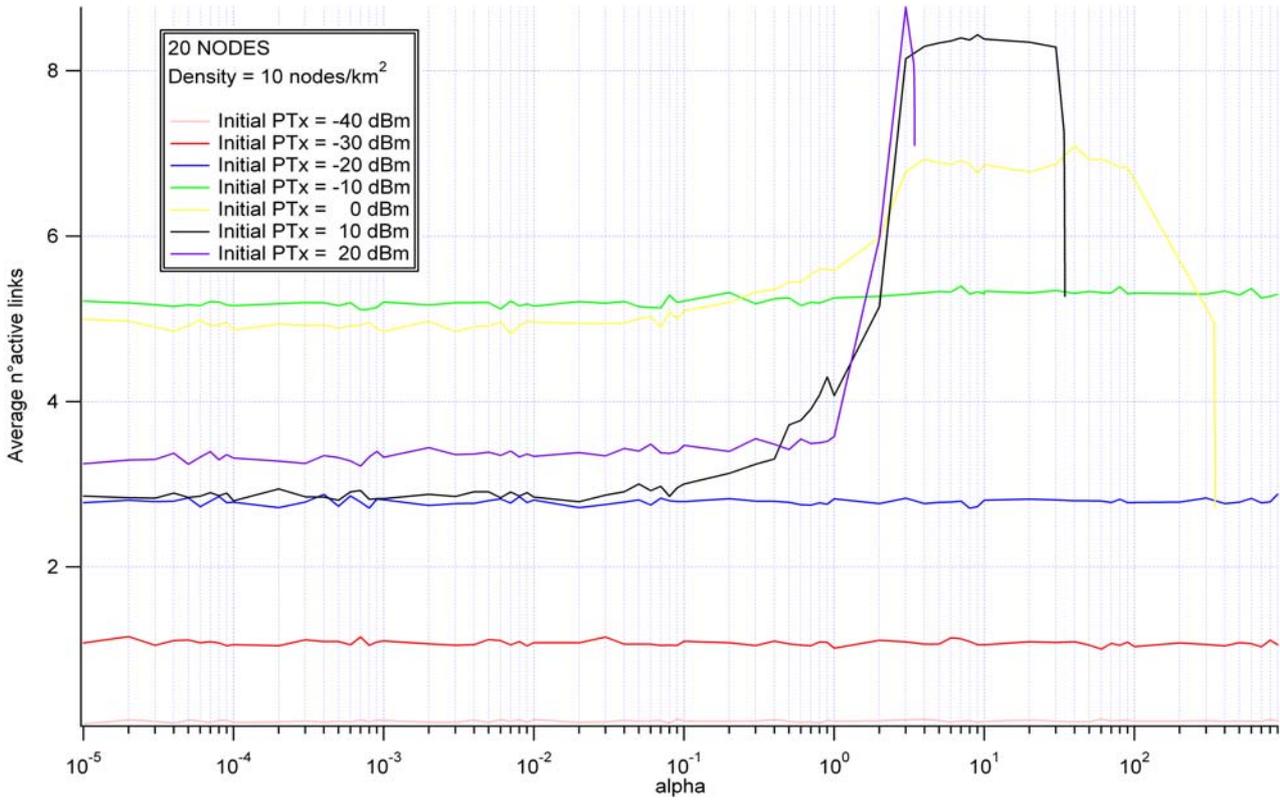


Figura 4.27

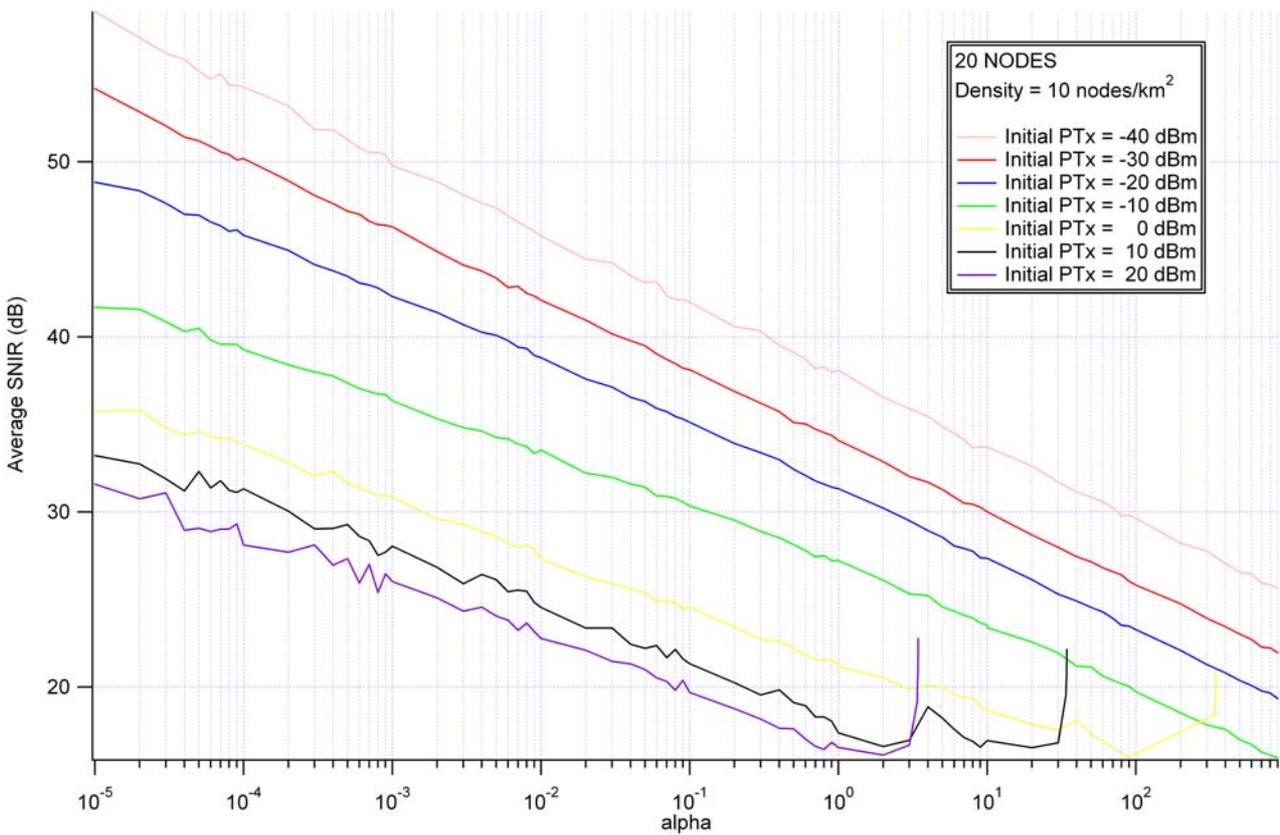


Figura 4.28

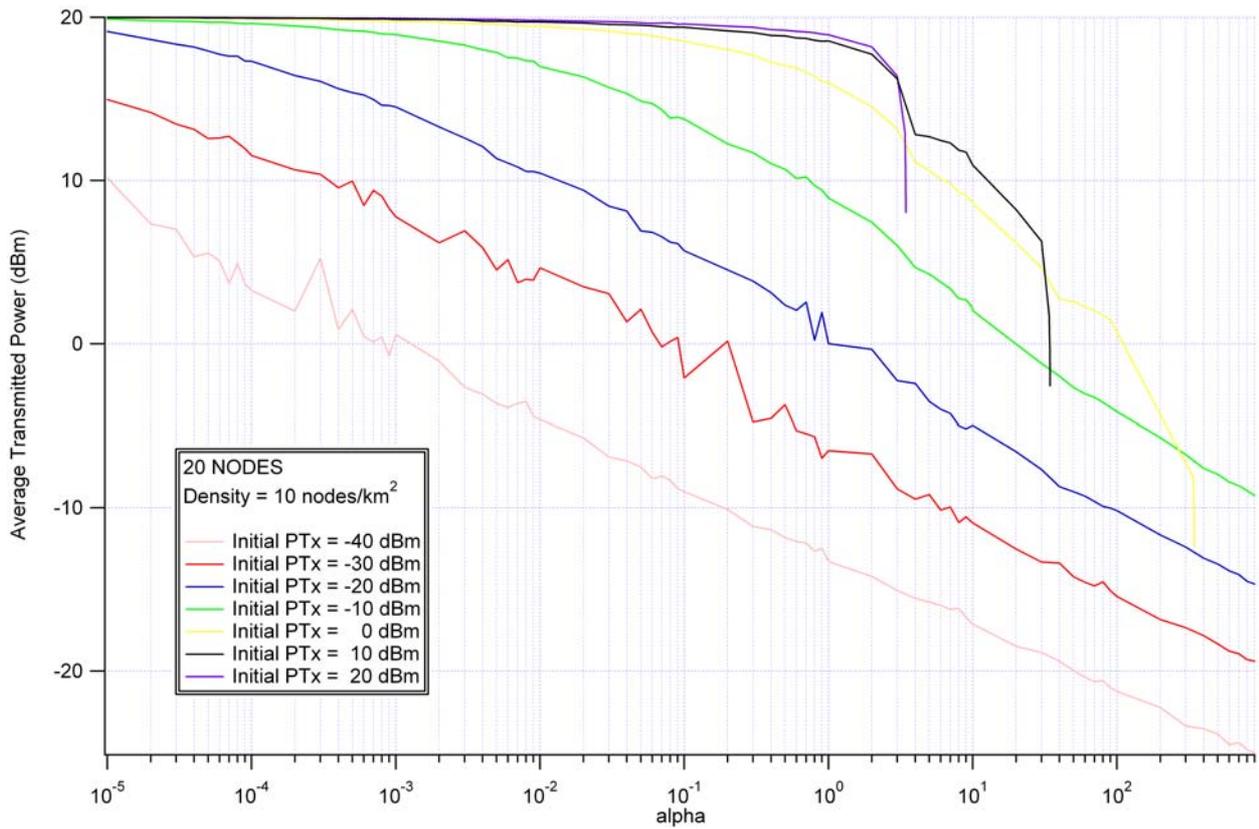


Figura 4.29

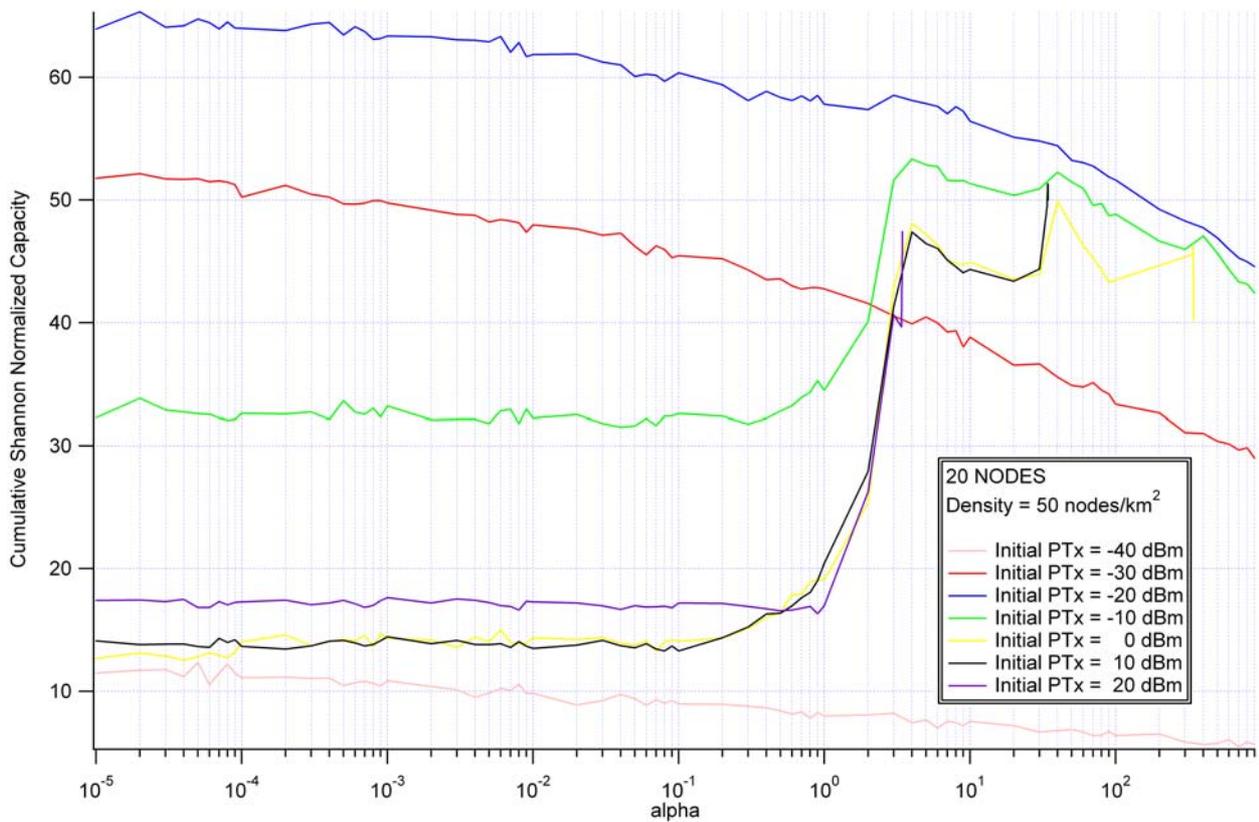


Figura 4.30

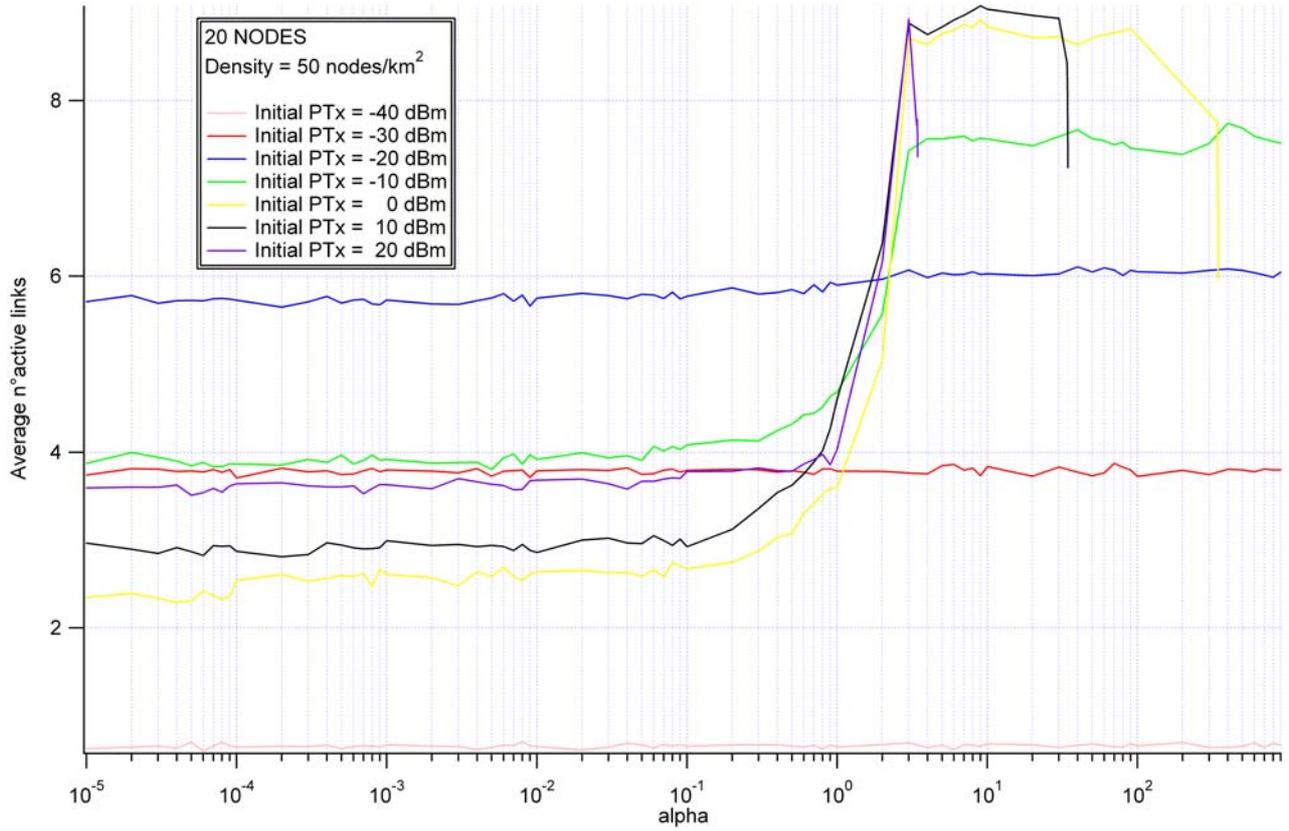


Figura 4.31

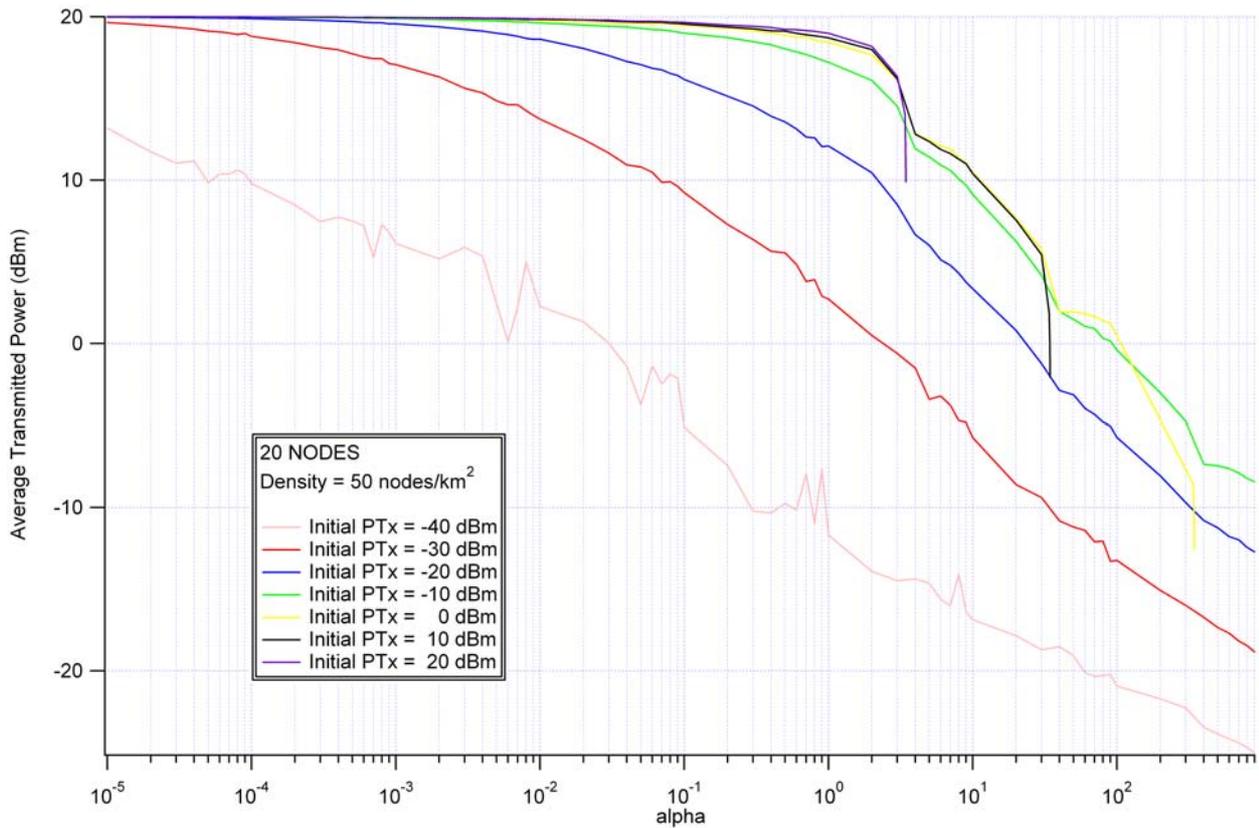


Figura 4.32

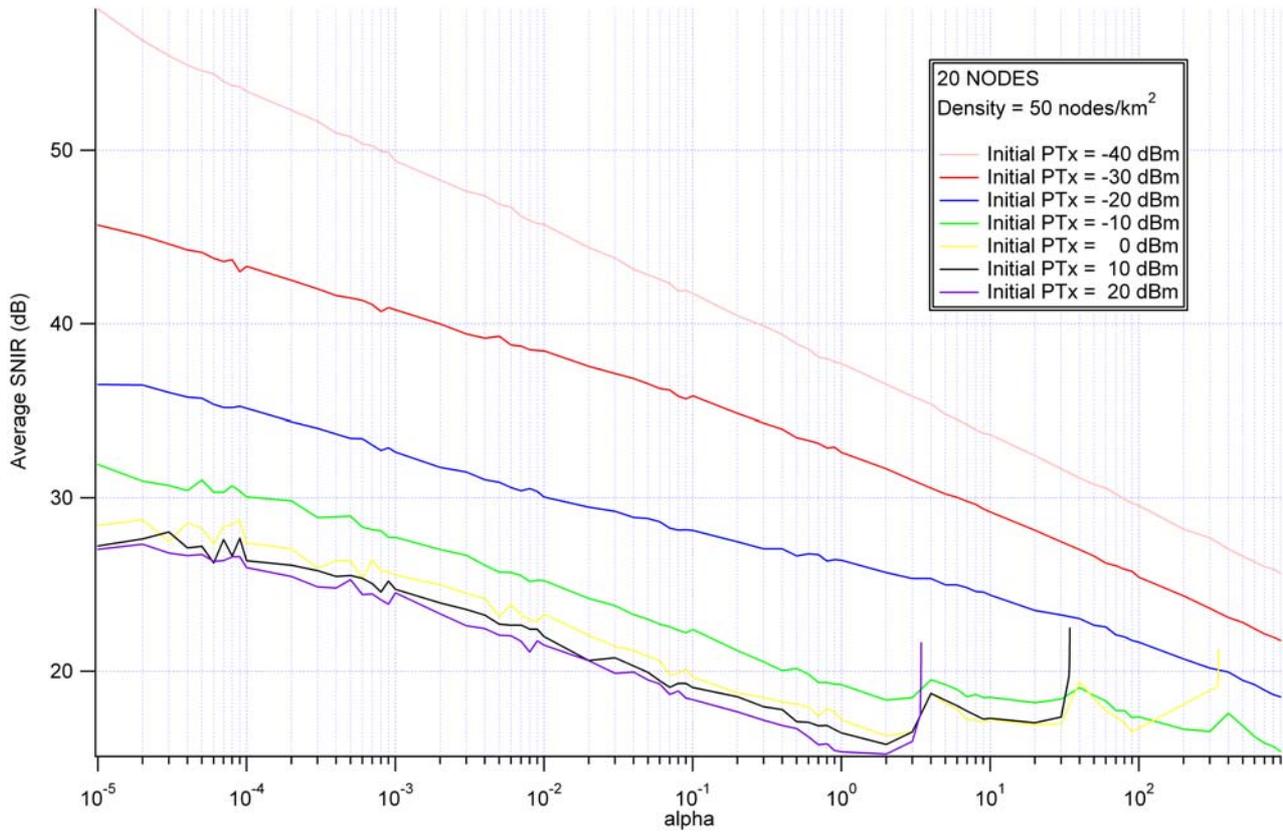


Figura 4.33

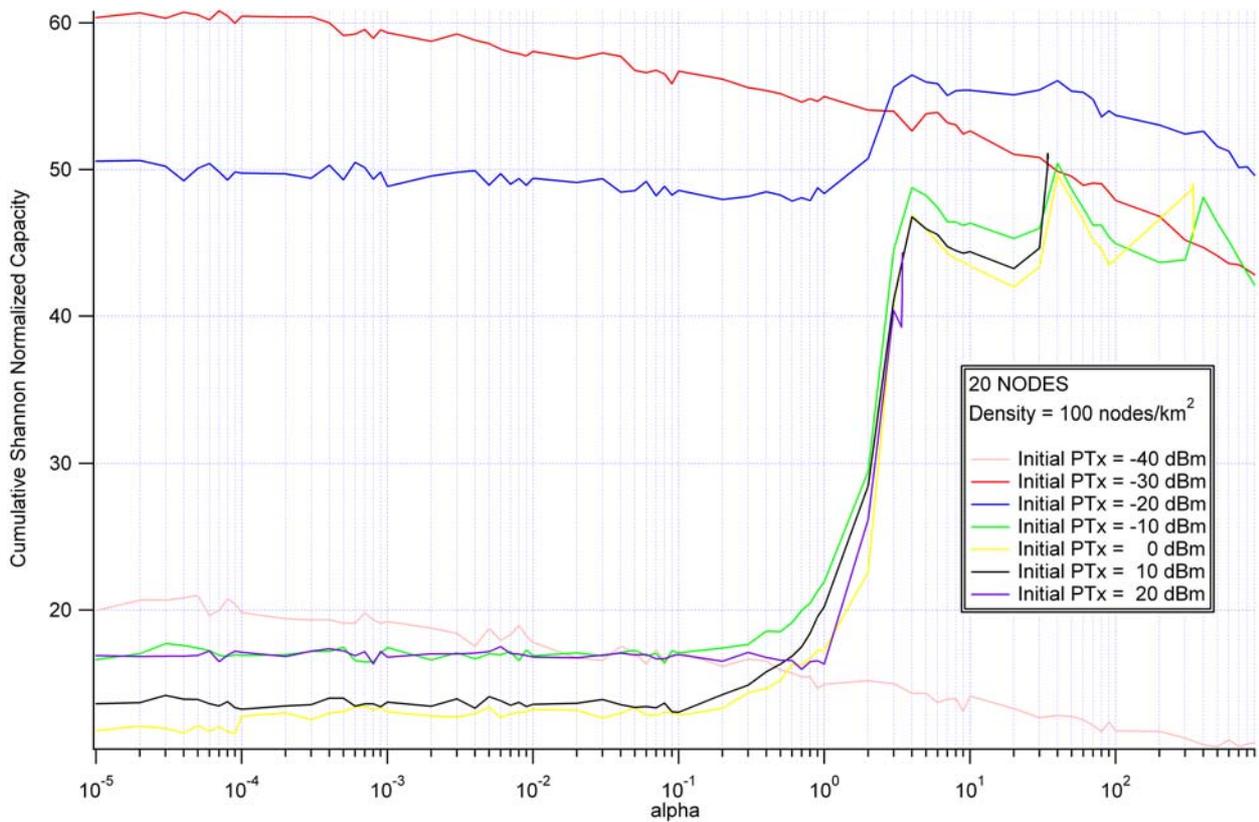


Figura 4.34

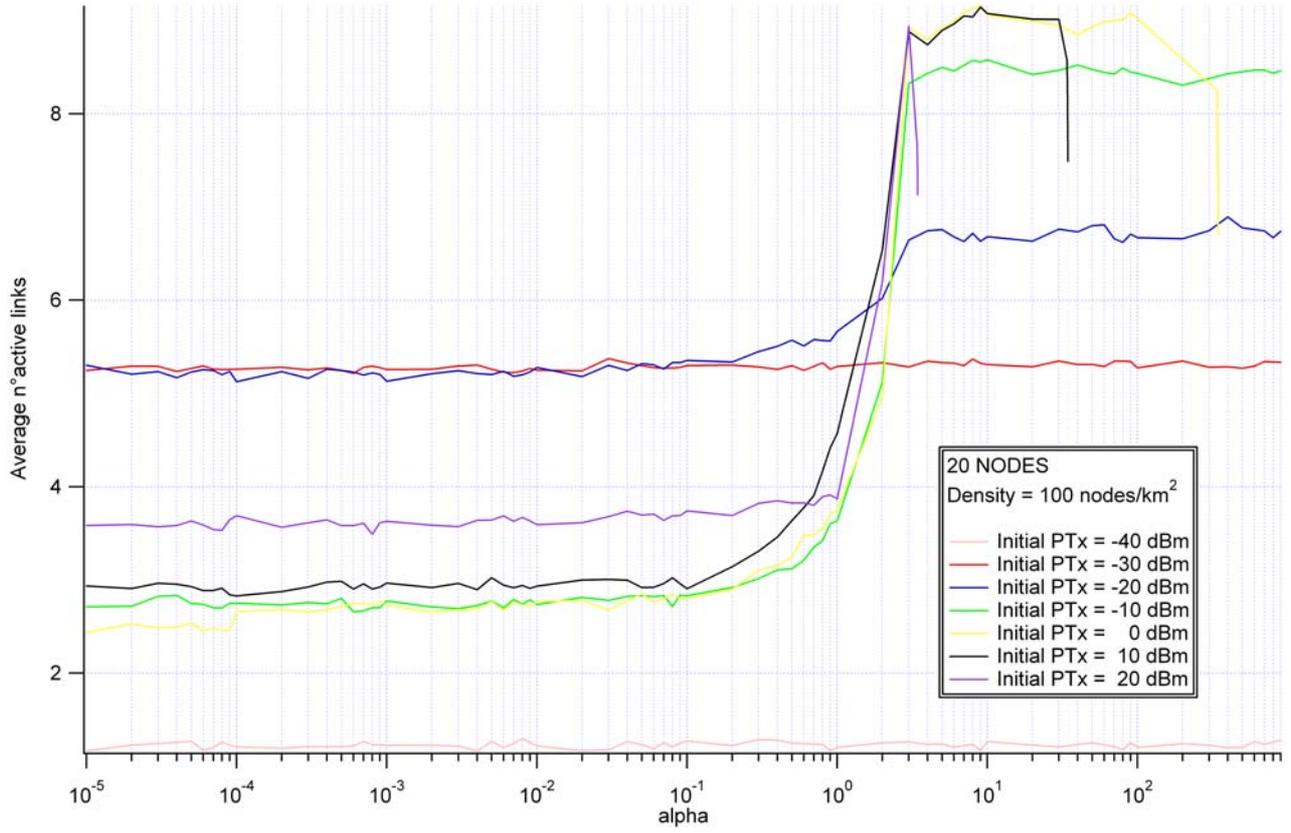


Figura 4.35

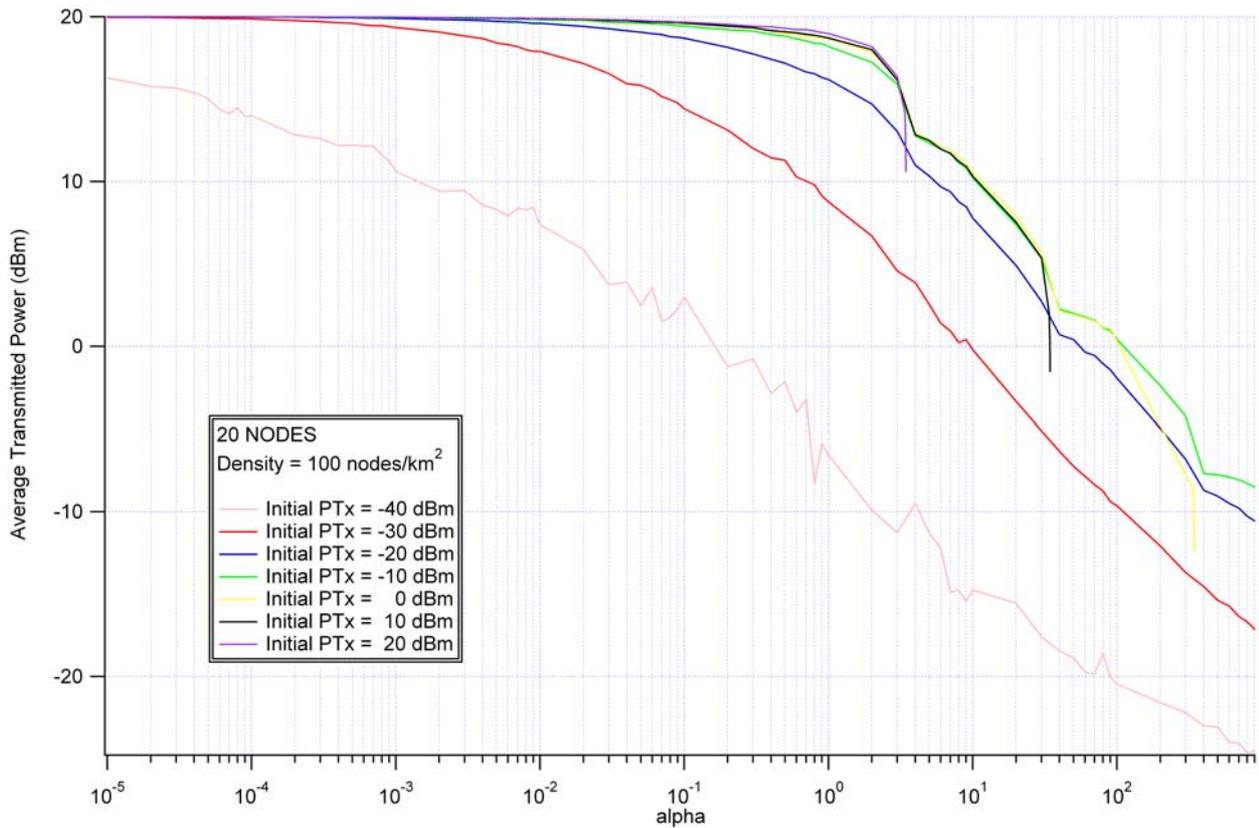


Figura 4.36

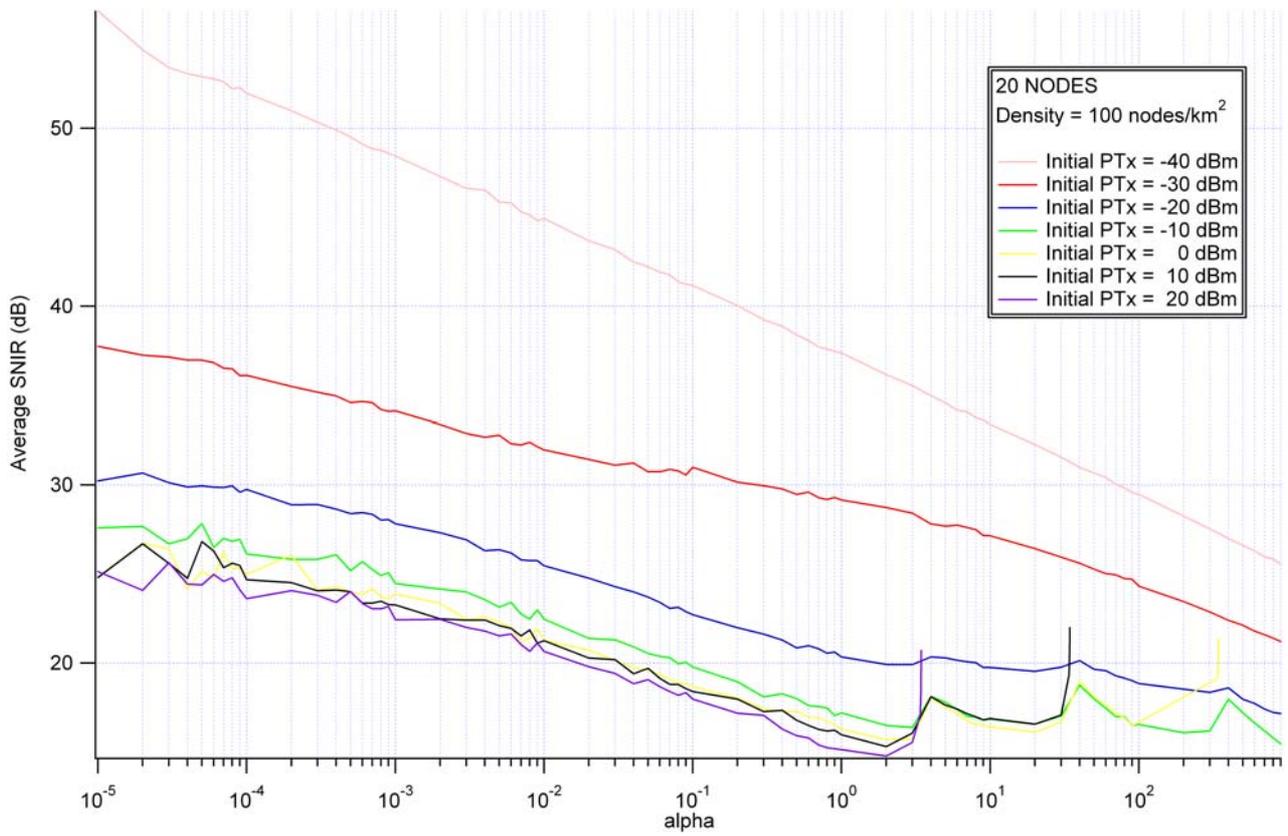


Figura 4.37

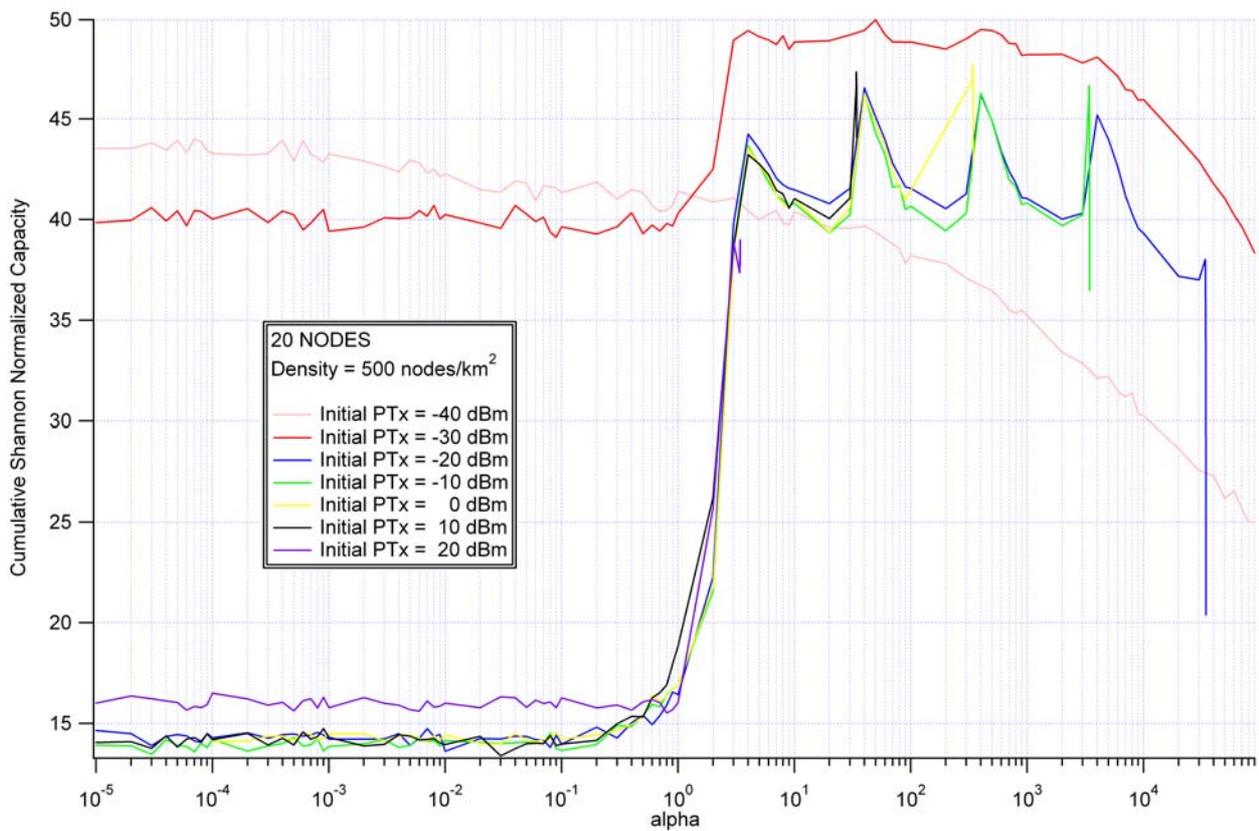


Figura 4.38

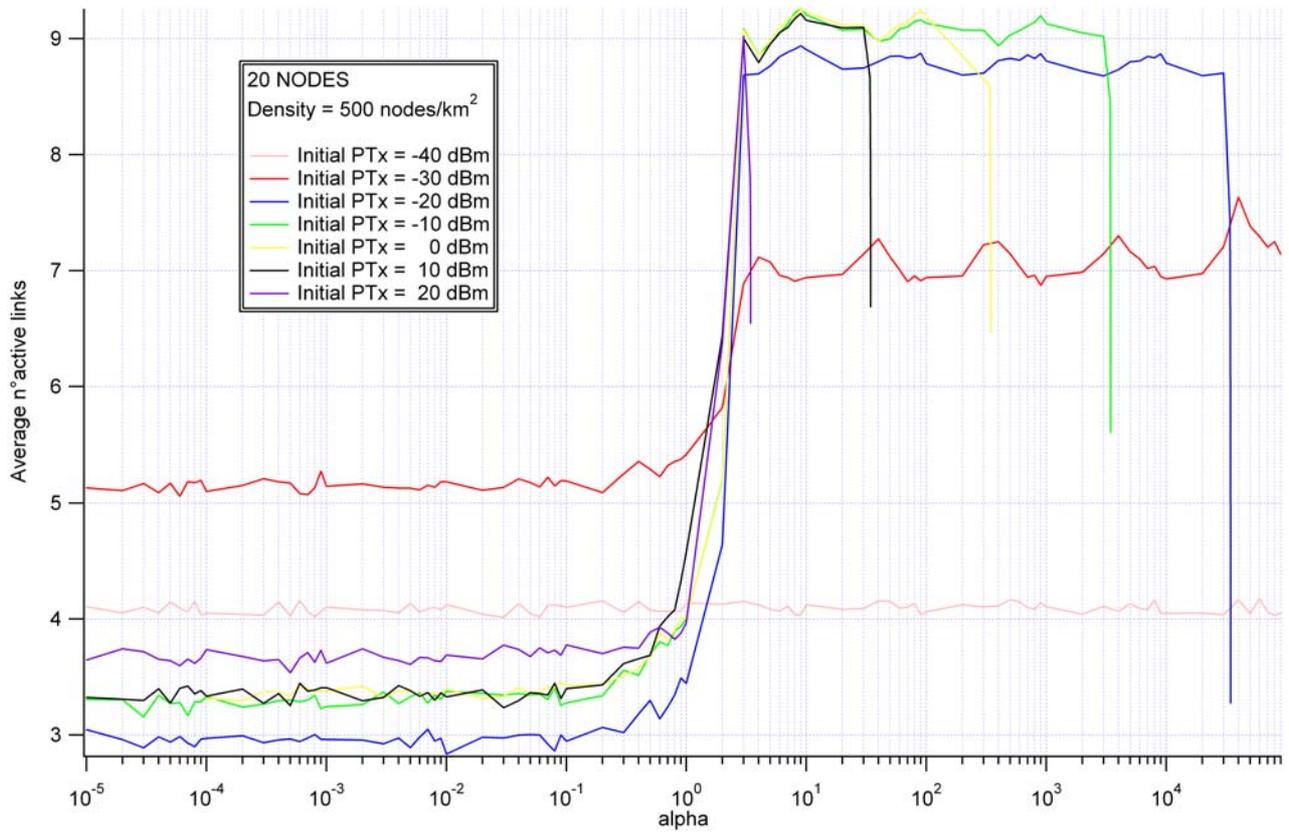


Figura 4.39

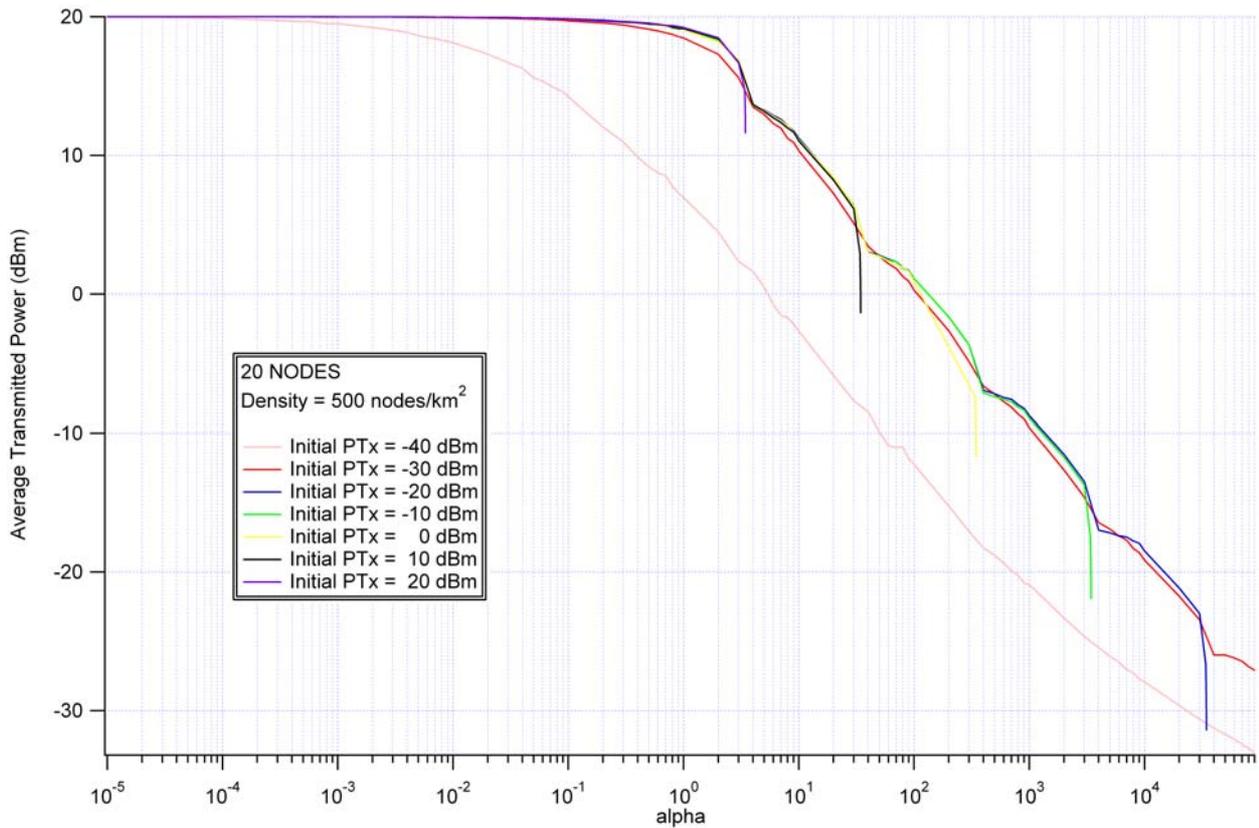


Figura 4.40

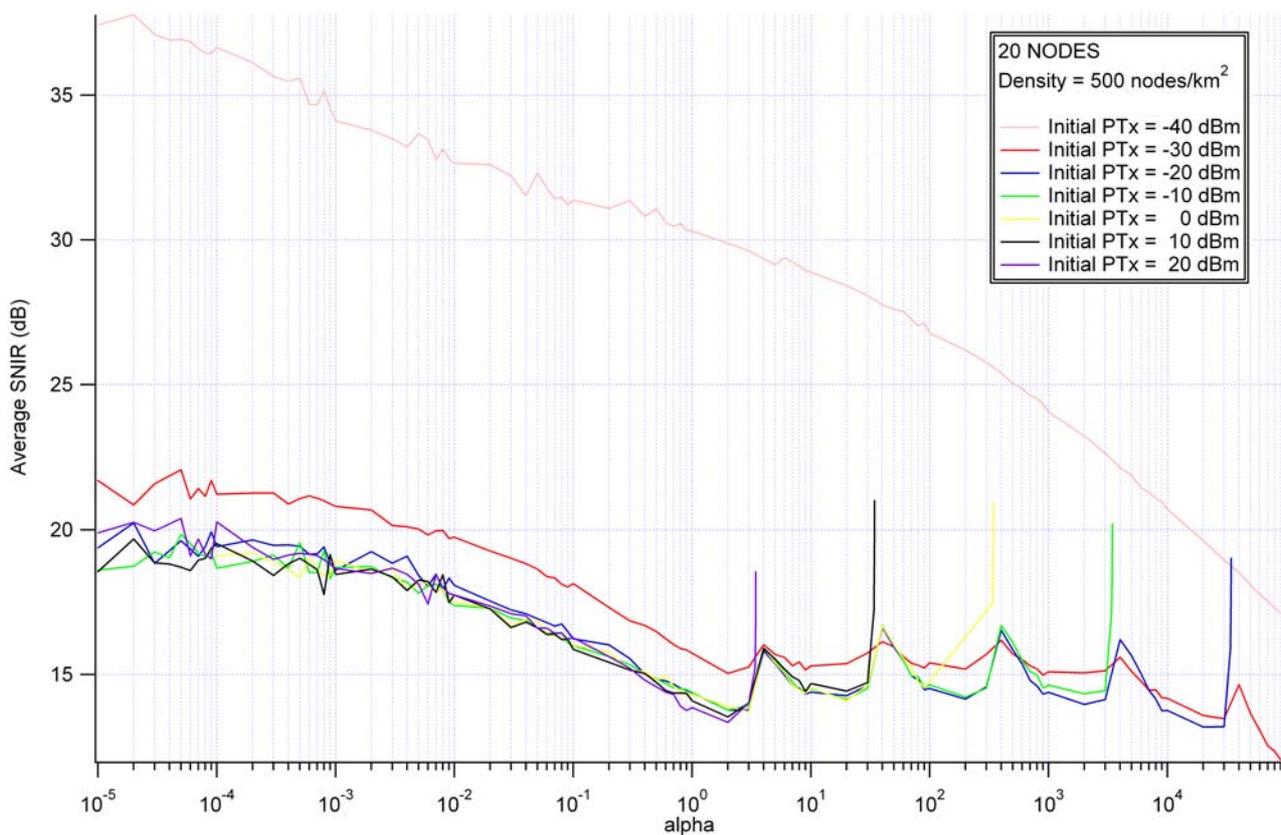


Figura 4.41

Le considerazioni che si possono trarre dall'osservazione di questi grafici sono:

- non esiste alcun valore di α che consenta di ottenere risparmio energetico, qualunque sia la densità dei nodi: d'altra parte, il risparmio di potenza non è un obiettivo di questo criterio;
- per valori di α nell'intorno di 10, si ottiene, per ogni valore di densità, il massimo numero di link attivi, che peraltro è molto vicino al valore ottenuto con l'altro criterio; la potenza media in corrispondenza delle quali tale valore viene raggiunto varia con la densità, come col criterio *power saving*;
- in corrispondenza di questo valore di α (~10), non si ottiene il massimo della capacità di Shannon se non per valori di densità di nodi elevata, però, regolando adeguatamente la potenza iniziale al variare della densità, si ottengono valori di capacità vicini al valore massimo, pagando con circa 1 link attivo in meno: ad esempio, nelle figure 4.26 e 4.27 si nota che, partendo con potenza pari a 0 dBm, con $\alpha = 10$ si ottiene un link attivo in meno rispetto al massimo ottenuto con potenza iniziale 10 o 20 dBm, però la capacità informativa è molto vicina al valore massimo per tale potenza iniziale.

In base ai risultati delle simulazioni, si è pertanto trovato quale deve essere il valore di α al variare della densità dei nodi e della potenza iniziale, come illustrato in figura 4.42.

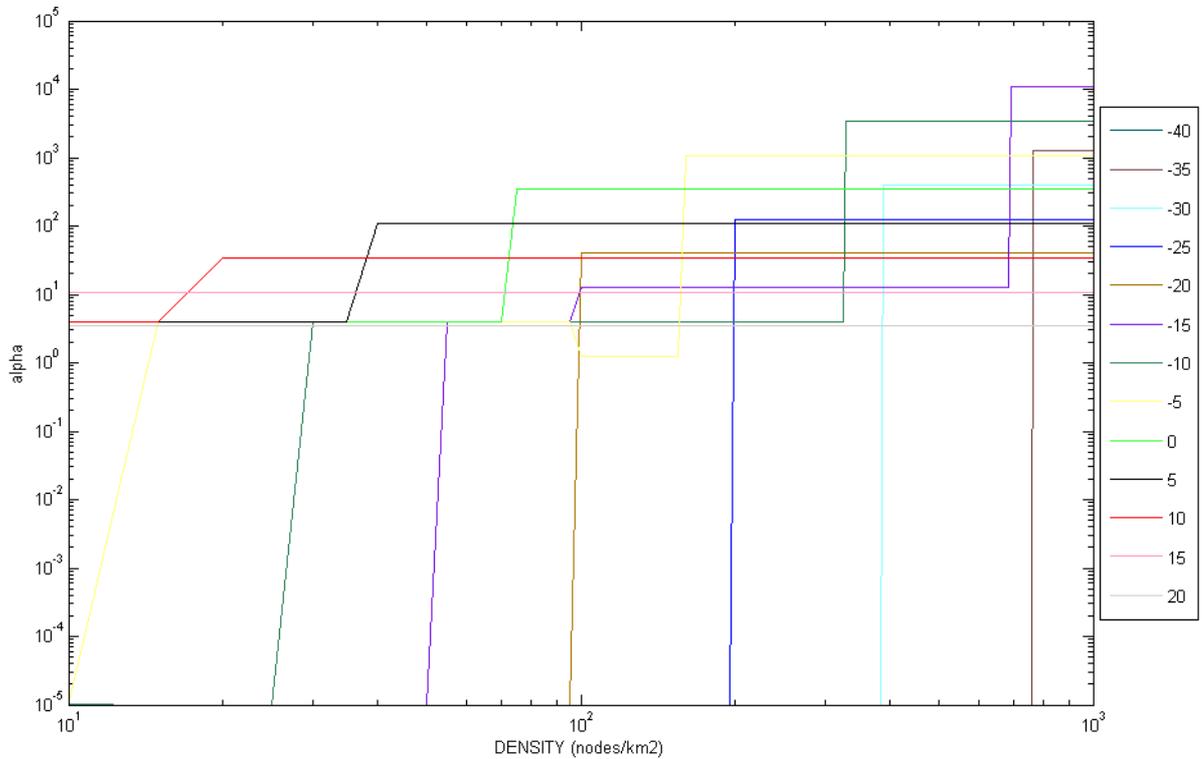


Figura 4.42

Nella figura 4.43, si è applicato il power control di tipo *max capacity* a una cella con 40 nodi, come del resto è stato fatto con l'altro criterio, ed è apprezzabile il guadagno proprio in termini di capacità informativa.

L'andamento spigoloso di molti grafici è legato alla scelta che deve essere fatta per il coefficiente di prezzo: come si vede nel grafico precedente, per ogni valore di potenza iniziale, esistono praticamente 3 range di variazione per il parametro.

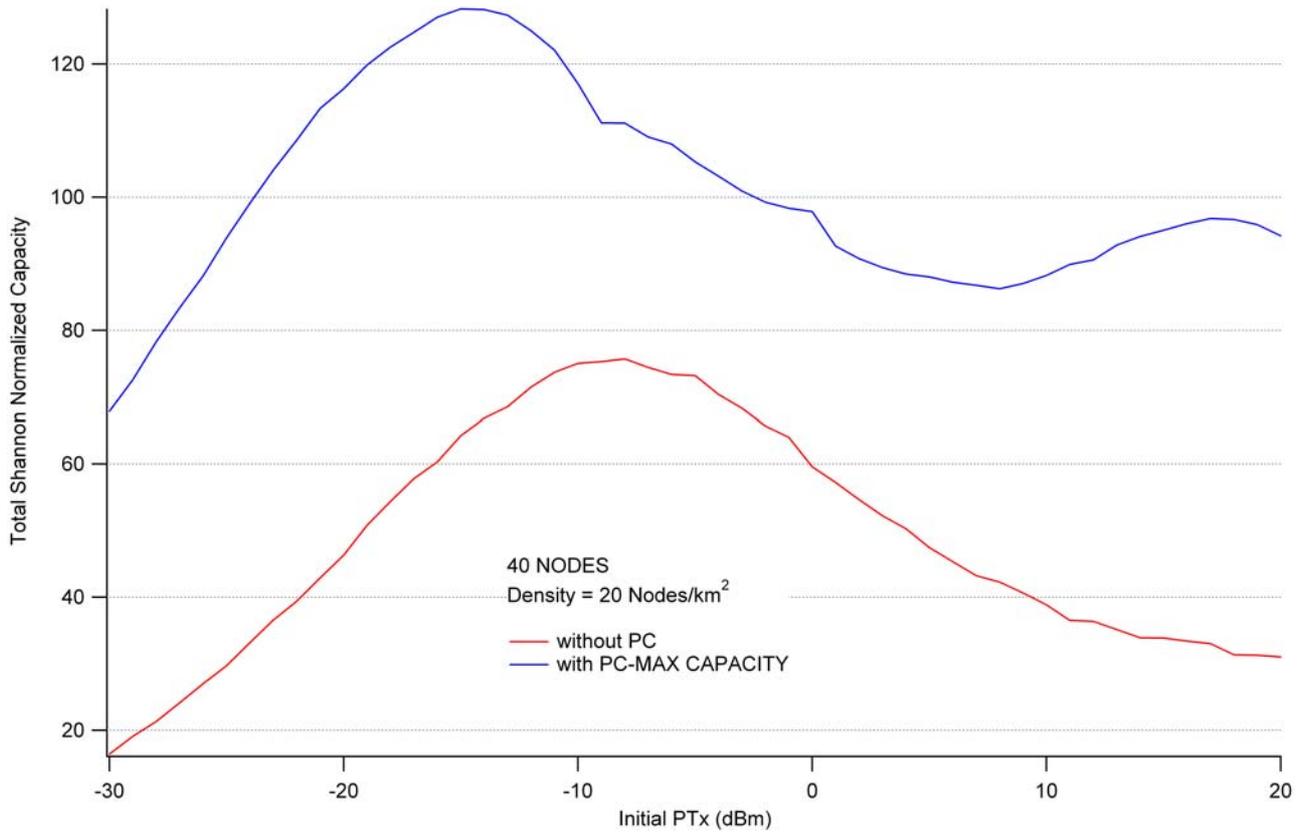


Figura 4.43

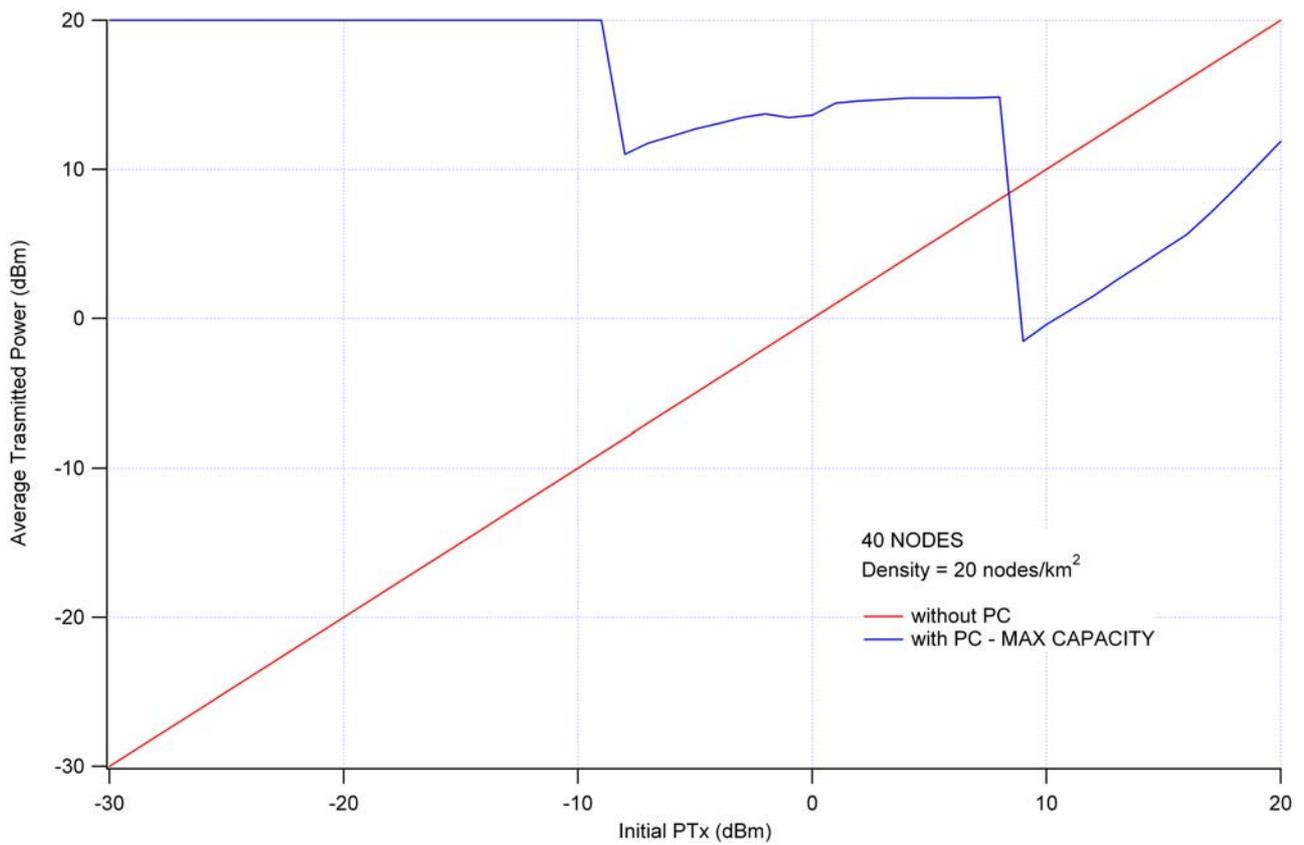


Figura 4.44

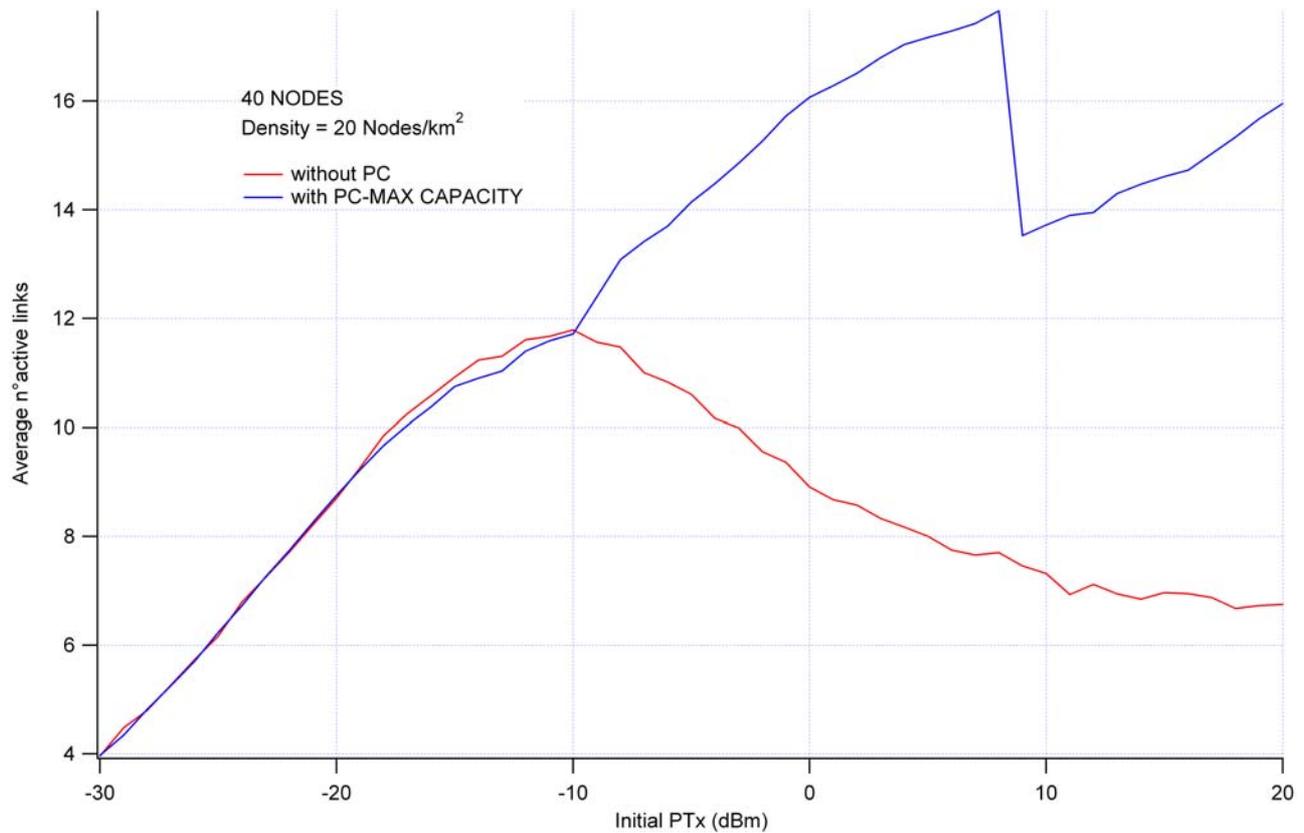


Figura 4.45

I grafici successivi mostrano l'andamento della capacità totale di Shannon al variare della densità: si nota che per raggiungere il massimo livello, ogni valore di densità ha la propria potenza iniziale "adatta"; ad esempio, per densità comprese tra 25 e 70 nodi/km², si deve usare una potenza di partenza di -20 dBm.

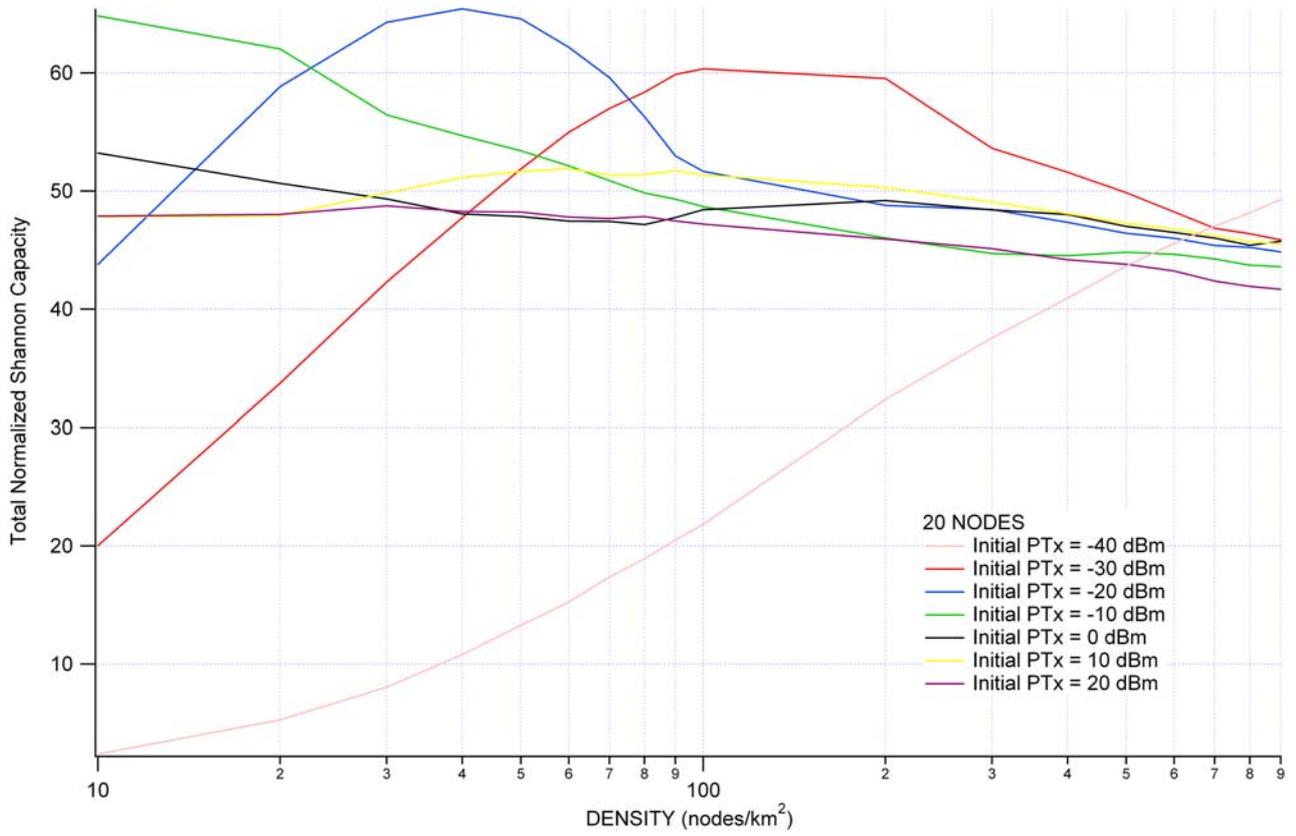


Figura 4.46

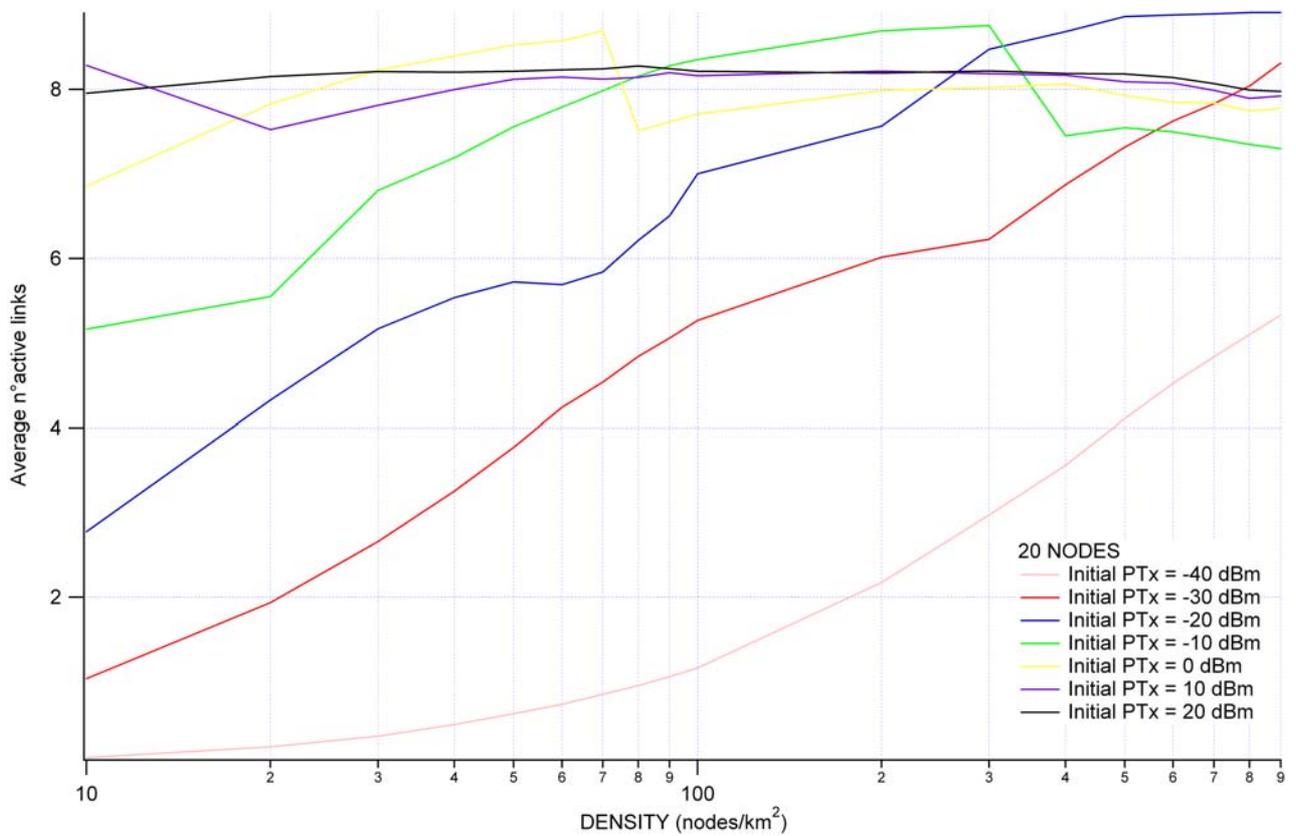


Figura 4.48

4.6 Coefficiente di prezzo “adattativo”

L’idea di utilizzare un coefficiente di prezzo che si auto-adatti alle condizioni ambientali in cui opera un nodo, nasce principalmente dalla volontà di rendere l’algoritmo il più possibile autonomo e indipendente da fattori estrinseci, come lo stesso coefficiente *alpha*. Il tentativo di rendere questo coefficiente adattativo è partito dall’osservazione della figura 3.9, in cui si è mostrato l’andamento decrescente dello SNIR^{tar} al crescere della quantità $A_i = \alpha_i \frac{RI_i}{WG_{ii}}$. Se lo scopo è, ad esempio, far sì

che i nodi ricevitori sperimentino uno SNIR più elevato per aumentare la capacità informativa totale di rete senza risparmiare in potenza, essi devono tendere ad uno SNIR^{tar} più elevato; pertanto, considerato l’andamento dello SNIR^{tar} al variare di A_i , è necessario “spostare” la quantità A_i verso valori più bassi. Una possibile scelta del coefficiente in questo senso è

$$\alpha_i = \beta \frac{I_i}{G_{ii}}$$

dove β costituisce, comunque, una costante estrinseca al problema, il cui effetto va esaminato mediante simulazioni. Con tale scelta, si ha che

$$A_i = \alpha_i \frac{R}{W} \left(\frac{I_i}{G_{ii}} \right)^2$$

Se, allora, le condizioni del link sono buone, cioè il rapporto $\frac{I_i}{G_{ii}}$ è piccolo, ancora più piccolo risulterà il quadrato di quel rapporto, per cui più basso sarà il valore di A_i e, conseguentemente, più alto sarà il corrispondente SNIR^{tar} .

Questa variante dell’algoritmo, come già accennato, può risultare valida quando viene seguito lo scopo della massimizzazione della capacità informativa di rete; essa, infatti, è stata applicata al caso in cui la funzione utilità dipende dalla capacità di Shannon, ossia quando si segue il criterio “max capacity”. I risultati ottenuti sono mostrati nelle figure seguenti, in cui è stata mostrata ancora la dipendenza da un parametro esterno, stavolta chiamato *beta* e in cui ci si è limitati a considerare due valori di densità dei nodi (10 e 50 nodi/km²). Le figure mostrano risultati abbastanza simili a quelli ottenuti con coefficiente di prezzo costante; una differenza importante sta nella maggiore regolarità degli andamenti relativi a potenze iniziali più elevate.

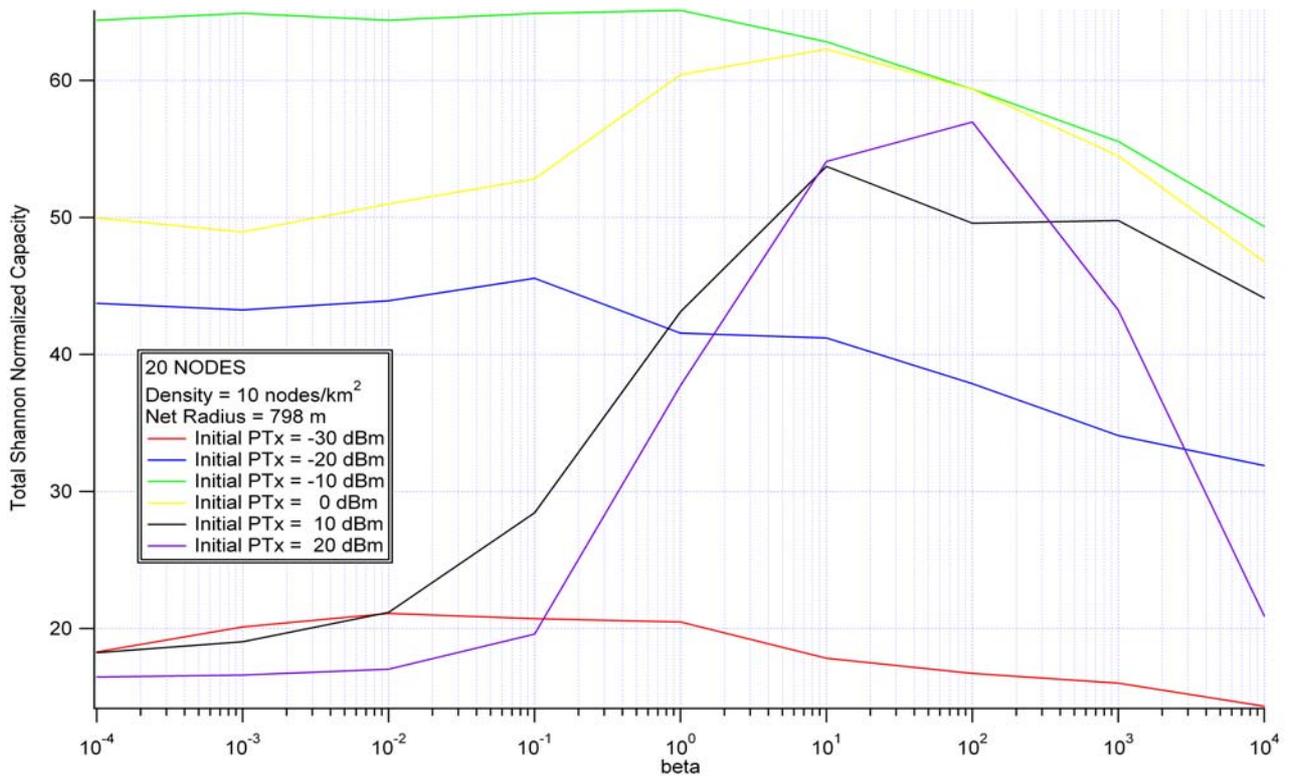


Figura 4.49

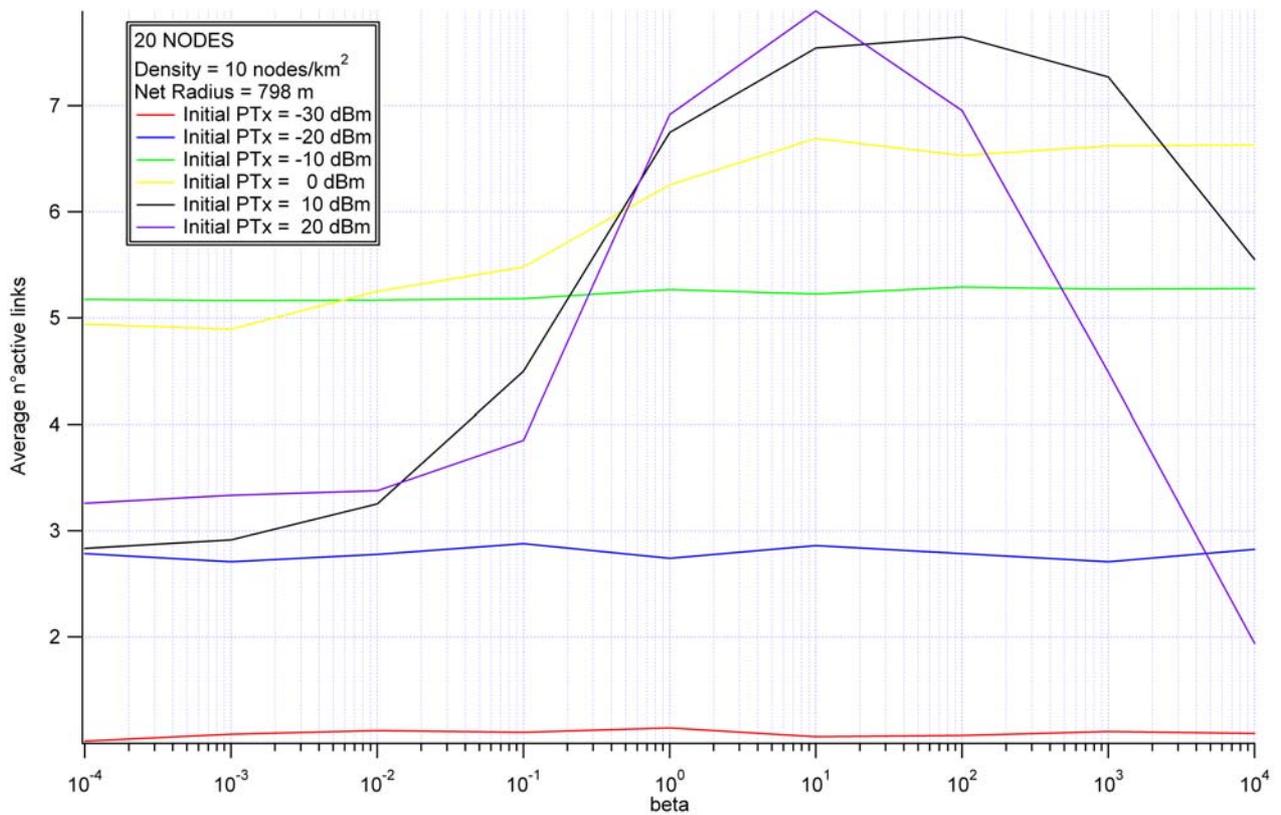


Figura 4.50

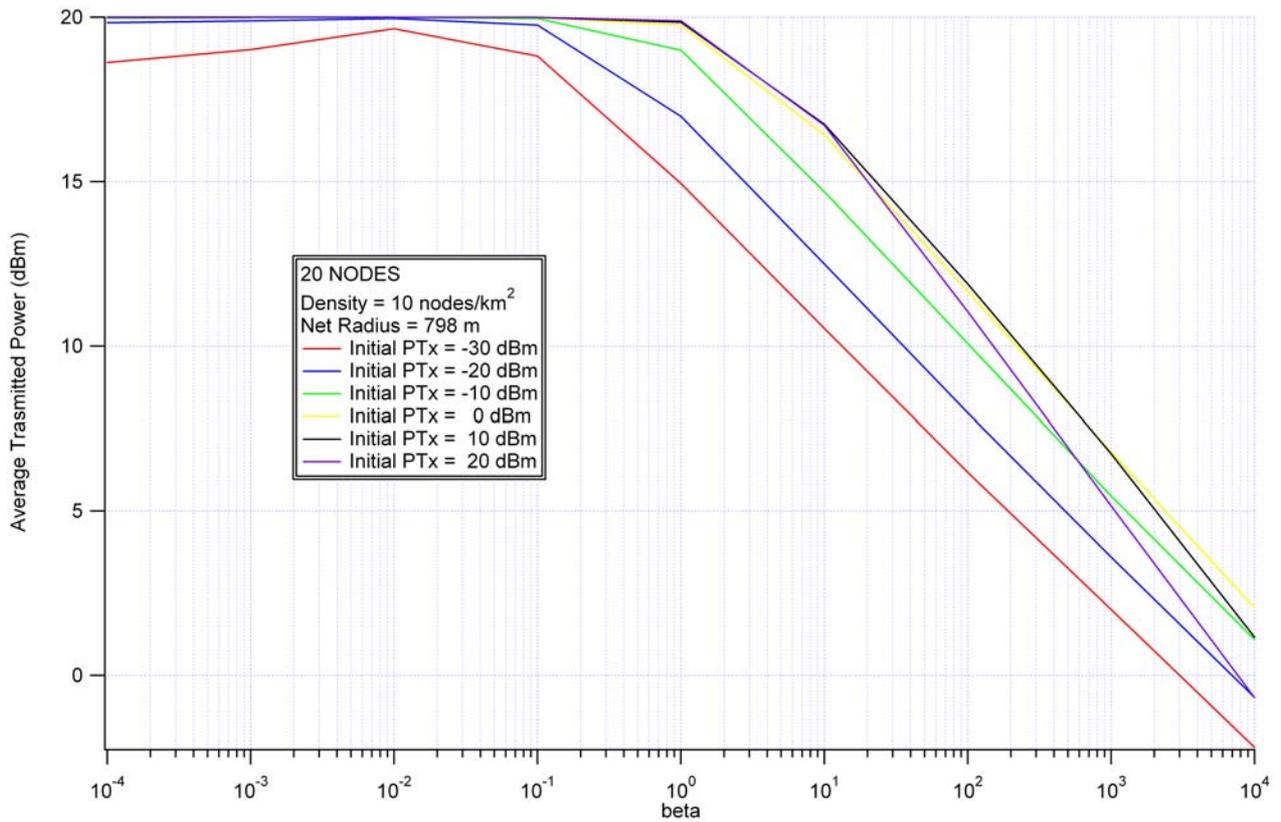


Figura 4.51

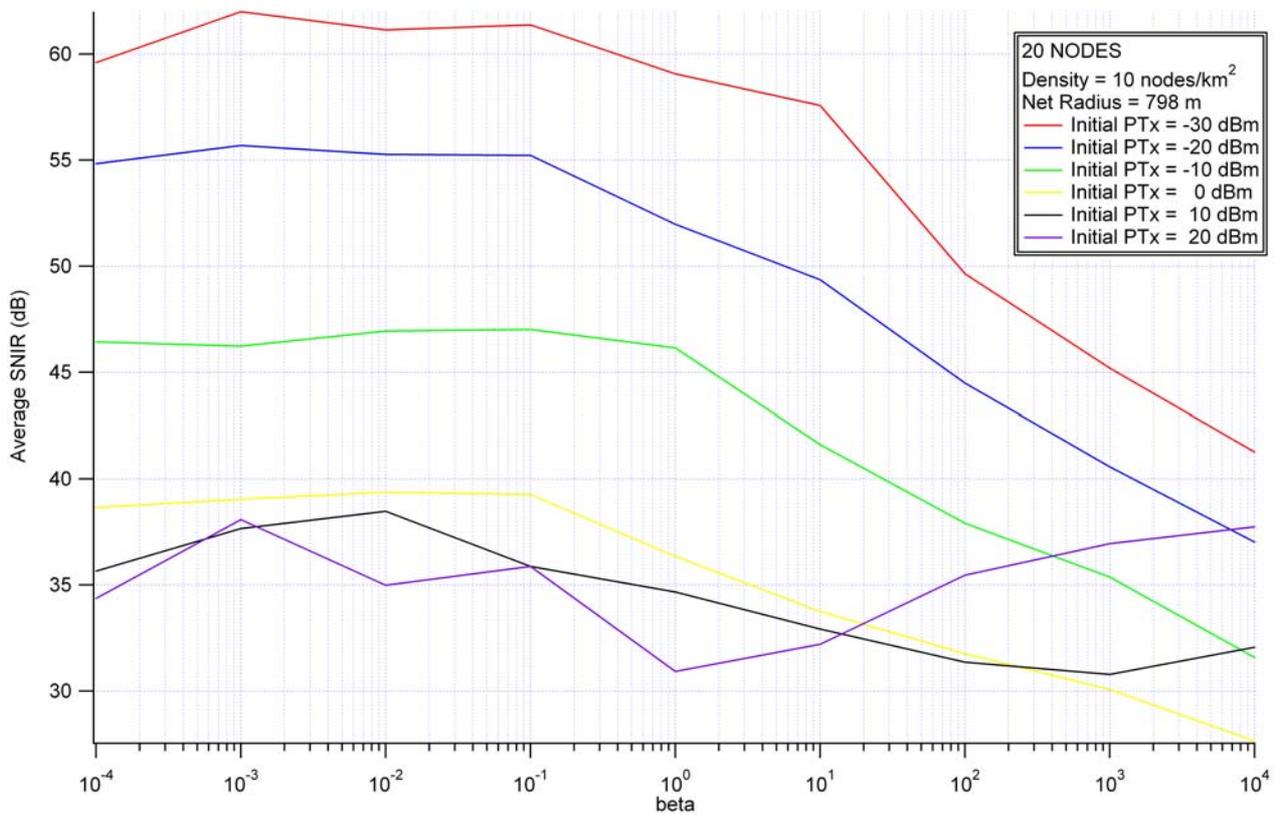


Figura 4.52

CONCLUSIONI

Le ad hoc network costituiscono un paradigma di comunicazione distribuita che può teoricamente essere esteso fino all’*“anytime and anywhere”*. In un’ad hoc network, i dispositivi wireless stessi “sono” la rete, e questo permette l’instaurazione di comunicazioni a basso costo, senza il bisogno di connessioni fisiche vere e proprie, in maniera relativamente rapida, semplice e *self-organized*. L’elevato grado di libertà e le capacità auto-organizzative rendono le MANET completamente differenti da ogni altra soluzione per la realizzazione di una rete. Il prezzo che hanno tutti questi begli aspetti è pagato in termini di complessità tecnologica.

L’analisi condotta in questo elaborato ha evidenziato che l’introduzione di un particolare algoritmo di power control può significativamente migliorare le prestazioni globali della rete, dal punto di vista della potenza utilizzata, del *throughput* complessivo o anche della capacità informativa totale di rete. Questo algoritmo è il frutto della modifica di un altro ben noto algoritmo distribuito di power control [29], la cui correzione si è resa necessaria per problemi di convergenza. Questa modifica ha implicato anche l’introduzione della teoria dei giochi, grazie alla quale gli utenti possono ottenere un livello di qualità del servizio “buono”, ma tale al tempo stesso da non compromettere quella degli altri. Questo livello è espresso quantitativamente in termini di SNIR (quello che è stato chiamato SNIR^{tar}), il quale, immesso nel vecchio algoritmo, elimina i problemi di convergenza, in quanto si “addolciscono” i requisiti individuali sulla qualità del servizio e viene introdotto un certo grado di cooperazione tra gli utenti, nonostante che l’algoritmo rimanga di natura non cooperativa e distribuita: un utente, infatti, automaticamente abbassa il proprio SNIR^{tar} (ed eventualmente interrompe la trasmissione) appena percepisce che il traffico si sta congestionando.

L'introduzione del *pricing* all'interno dell'algoritmo, che consiste in una sorta di "pagamento virtuale" per la potenza usata, non solo migliora l'aspetto della cooperazione implicita tra gli utenti, ma consente di "guidare" il power control verso un obiettivo prefissato, che può essere il risparmio della potenza trasmessa o la massimizzazione della capacità informativa.

Seguendo il primo obiettivo, con l'applicazione dell'algoritmo si ottiene una notevole riduzione della potenza e un corrispondente aumento del *throughput* complessivo di rete rispetto al caso in cui tutte le potenze vengono mantenute fisse; entrambe le variazioni risultano tanto maggiori quanto più grande è la potenza trasmessa iniziale. Si ha, dunque, un miglioramento generale nella gestione delle risorse di rete.

Seguendo il secondo obiettivo, per il quale si è utilizzata un'altra funzione utilità, si può notare che la capacità massima si ottiene partendo con potenze iniziali basse, in modo tale che i link attivi siano pochi e possano aumentare la loro potenza trasmessa fino al limite massimo senza causare danno agli altri, ottenendo così il massimo SNIR.

BIBLIOGRAFIA

- [1] P. Krishna, N. H. Vaidya, M. Chatterjee, D. K. Pradhan, "A Cluster-Based Approach for Routing in Dynamic Networks," *ACM SIGCOMM Computer Communications Review*, 1997.
- [2] M. Jiang, J. Li, Y. C. Tay, "Cluster Based Routing Protocol Functional Specification", *Internet Draft*, Work-in-progress.
- [3] I. Chlamtac, C. Petrioli, and J. Redi, "Energy-Conserving Access Protocols for Identification Networks", *IEEE/ACM Transactions on Networking*, 7, 1, Febbraio 1999.
- [4] C.-K. Toh, "Associativity-Based Routing For Ad-Hoc Mobile Networks", *Journal on Wireless Personal Communications*, 4, First Quarter, 1997.
- [5] J.-H. Chang, L. Tassiulas, "Energy Conserving Routing in Wireless Ad-hoc Networks", *Proceedings of IEEE Infocom 2000*, Tel Aviv, Israel, Marzo 2000.
- [6] P. Gauthier, D. Harada, M. Stemm, "Reducing Power Consumption for the Next Generation of PDAs: It's in the Network Interface," *Proceedings of MoMuC'96*, Settembre 1996.
- [7] I. Chlamtac, C. Petrioli, J. Redi, "Energy-Conserving Selective Repeat ARQ Protocols for Wireless Data Networks", *Proceedings of IEEE 9th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, Boston, Mass, Settembre 1998.
- [8] I. Chlamtac, C. Petrioli, J. Redi, "An Energy-Conserving Access Protocol for Wireless Communications," in *IEEE International Conference on Communications, ICC'97*, Montreal, Quebec, Canada, Giugno 1997.
- [9] J.-C. Chen, K. M. Sivalingam, P. Agrawal, S. Kishore, "A Comparison of MAC Protocols for Wireless Local Networks Based on Battery Power Consumption", *Proceedings of IEEE Infocom 1998*, San Francisco, CA, Marzo 1998.
- [10] L. M. Feeney, "An Energy-consumption Model for Performance Analysis of Routing Protocols for Mobile Ad Hoc Networks", *Journal of Mobile Networks and Applications*, 2001.
- [11] L. M. Feeney, M. Nilsson, "Investigating the Energy Consumption of a Wireless Network Interface in an Ad Hoc Networking Environment", Swedish Institute of Computer Science, Kista, Sweden.
- [12] IEEE. *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications*. LAN MAN Standards Committee, 1999.
- [13] Bluetooth SIG, Inc. <http://www.bluetooth.org> [citato 20 Ottobre 2005].
- [14] J. P. Monks, V. Bharghavan, W. W. Hwu, "Transmission power control for multiple access wireless packet networks", *Local Computer Networks*, 2000. LCN 2000. Proceedings. 25th Annual IEEE Conference on, pp 12-21.
- [15] X. Lin, Y. Kwok, V. Lau, "Power control for IEEE 802.11 ad hoc networks: issues and a new algorithm", *International Conference on Parallel Processing*, 2003.
- [16] S. Agarwal, S. Krishnamurthy, R. Katz, S. Dao "Distributed power control in ad-hoc wireless networks", *International Symposium Personal, Indoor and Mobile Radio Communications*, 2001, pp. F-59-F-66.
- [17] J. P. Monks, J. P. Ebert, A. Wolisz, W. W. Hwu, "A study of the energy saving and capacity improvement potential of power control in multi-hop wireless networks", *LCN 2001*. pp 550-559.

- [18] A. MacKenzie, S. Wicker, "Selfish users in ALOHA: A Game Theoretic Approach", *Vehicular Technology Conference*, 2001, Vol.3, pp. 1354-1357.
- [19] D. Goodman, R. M. Buehrer, J. Neel, "Distributed Dynamic Channel Assignment Schemes", *Vehicular Technology Conference*, 1993, pp. 532-535.
- [20] A. Lozano, D. Cox, "Distributed Dynamic Channel Assignment in TDMA Mobile Communication Systems", *IEEE Transaction on Vehicular Technology*, vol. 51, Novembre 2002, pp. 1397-1406.
- [21] N. Feng, S. Mau, N. B. Mandayam, "Pricing and Power Control for Joint Network-Centric and User-Centric Radio Resource Management", *IEEE Transaction on Communications*, vol. 52, n° 9, Settembre 2004, pp.1574-1557.
- [22] H. Yäliche, R. Mazumdar, C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks", *IEEE/ACM Transactions on Networking*, vol. 8, n° 5, pp. 667-678, 2000.
- [23] Z. Han, "An Optimization Theoretical Framework for Resource Allocation over Wireless Networks", *Ph. D. Dissertation*, 2003.
- [24] D. Goodman, N. Mandayam, "Power control for wireless data", *IEEE Pers. Commun.*, vol. 7, pp. 48-54, Aprile 2000.
- [25] C. Saraydar, N. Mandayam, D. Goodman, "Efficient power control via pricing in wireless data networks", *IEEE Trans. on Comm.*, vol. 50, no. 2, pp. 291-303, Feb. 2002.
- [26] M. Heimbach, "Polar Impact: The Digital Alternative for Multi-Mode Wireless Communications", *Applied Microwave & Wireless*, Agosto 2001.
- [27] S. Srikanteswara, J. Neel, J. H. Reed, P. M. Athanas, "Soft Radio Implementations for 3G and Future High Data Rate Systems", *Globecom*, 2001.
- [28] J. Neel, J. H. Reed, R. P. Gilles, "Game Models for Cognitive Radio Algorithm Analysis", *SDR Forum 2004 Technical Conference*, Novembre 2004.
- [29] G. J. Foschini, Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence", *IEEE Trans. Veh. Technol.*, vol. 42, pp. 641-646, Novembre 1993.
- [30] S. C. Chen, N. Bambos, G. J. Pottie, "Admission control schemes for wireless communication networks with adjustable transmitter powers", in *Proc. IEEE INFOCOM*, Toronto, ON, Canada, Giugno 1994, pp. 21-28.
- [31] D. Mitra, "An asynchronous distributed algorithm for power control in cellular radio systems", in *Proc. 4th Winlab Workshop Third Generation Wireless Information Network*, New Brunswick, NJ, Ottobre 1993, pp. 249-257.
- [32] R. D. Yates, "A framework for uplink power control in cellular radio systems", *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1341-1347, Settembre 1995.
- [33] A. M. Viterbi, A. J. Viterbi, E. Zehavi, "Performance of Power-Controlled Wideband Terrestrial Digital Communications", *IEEE Transaction on Communications*, vol. 41, Aprile 1993, pp. 559-569.
- [34] T. J. Kwon, M. Gerla, "Clustering with Power Control", *Military Communications Conference*, 1999, pp. 1424-1428.
- [35] R. Yates, "Uplink Power Control in Cellular Radio Systems", *IEEE Journal on Selected Areas in Communications*, Vol. 13, No 7, Settembre 1995, pp. 1341-1347.
- [36] R. Kravets and P. Krishnan, "Power Management Techniques for Mobile Communication", *Proceedings of MobiCom '98*, Dallas, TX, Ottobre 1998.
- [37] J. L. Sobrinho, A. S. Krishnakumar, "Quality-of-Service in Ad Hoc Carrier Sense Multiple Access Wireless Networks", *IEEE Journal on Special Areas in Communications*, 17, 8, Agosto 1999.
- [38] M. Rulnick, N. Bambos, "Mobile Power Management for Wireless Communication Networks", *ACM/Baltzer Wireless Networks*, 3, 1, 1996.
- [39] M. Zorzi, R. R. Rao, "Energy Constrained Error Control for Wireless Channels", *Proceeding of IEEE GLOBECOM '96*, London, UK, Novembre, pp. 1411-1416, 1996.
- [40] C. Raghavendra, S. Singh, "PAMAS: Power Aware Multi-Access Protocol with Signaling for ad hoc networks", *ACM Computer Communication Review*, Luglio 1998, pp.5-26.
- [41] L. Feeney, "Energy Efficient Communication in Ad Hoc Wireless Networks", *Mobile Ad Hoc Networking*, S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic (Eds.), IEEE Press-Wiley, 2004.
- [42] H. Woesner, J. P. Ebert, M. Schlaeger, A. Wolisz, "Power-Saving Mechanisms in Emerging Standards for Wireless LAN's: The MAC-Level Perspective", *IEEE Personal Communications (Special Edition on Power Saving)*, 5, 3, 40-48, 1998.

- [43] J. P. Ebert, B. Stremmel, E. Wiederhold, A. Wolisz, "An Energy-efficient Power Control Approach for WLANs", *Journal of Communications and Networks (JCN)*, 2, 3, 197–206, 2000.
- [44] V. Rodoplu, T. H.-Y. Meng, "Minimum Energy Mobile Wireless Networks," *IEEE Journal on Selected Areas in Communications*, 17(8), 1333–1344, Agosto 1999.
- [45] S. Singh, M. Woo, C. S. Raghavendra, "Powerware Routing in Mobile Ad Hoc Networks", in *Proceedings of ACM/IEEE Mobicom*, Ottobre 1998, pp. 181–90.
- [46] Stojmenovic, X. Lin, "Power-Aware Localized Routing in Wireless Networks", *Proceedings of IEEE Symposium on Parallel and Distributed Processing Systems*, May 2000.
- [47] R. Ramanathan, R. Rosales-Hain, "Topology Control of Multi-Hop Wireless Networks Using Transmit Power Adjustment", *Proceedings of IEEE INFOCOM*, Tel Aviv, Israel (Marzo 2000).
- [48] M. W. Subbarao, "Dynamic Power-Conscious Routing for MANETs: An Initial Approach", *Proceedings of IEEE VTC Fall 1999*, Amsterdam, The Netherlands, 1999.
- [49] M. W. Subbarao, "Mobile Ad Hoc Data Networks for Emergency Preparedness Telecommunications-Dynamic Power-Conscious Routing Concepts", 2000.
- [50] S. Khanna, S. Sarkar, I. Shin, "An Energy Measurement Based Collision Resolution Protocol", University of Pennsylvania.
- [51] J. R. Lorch, A. J. Smith, "Scheduling Techniques for Reducing Processor Energy Use in MacOS", *ACM/Baltzer Wireless Networks*, 3, 5, 311–324, 1997.
- [52] M. Weiser, B. Welch, A. Demers, S. Shenker, "Scheduling for Reducing CPU Energy USENIX Association", *First Symposium on Operating System Design and Implementation*, Monterey, CA, pp. 13–23, 1994.
- [53] D. P. Helmbold, D. E. Long, B. Sherrod, "A Dynamic Disk Spin-Down Technique for Mobile Computing", *Proceedings of the Second Annual ACM International Conference on Mobile Computing and Networking*, NY, pp. 130–142, 1996.
- [54] A. J. Viterbi, "CDMA: Principles of Spread Spectrum Communication", Addison-Wesley: New York, 1995.
- [55] J. Chou, D. Petrovic, K. Ramchandran, "A distributed and adaptive signal processing approach to reduce energy consumption in sensor networks", *INFOCOM 2003*.
- [56] A. A. Bertossi, M. A. Bonuccelli, "Code assignment for hidden terminal interference avoidance in multihop packet radio networks", *IEEE Transactions on Communications*, 3(4):441–449, Agosto 1995.
- [57] J. Garcia-Luna-Aceves, J. Raju, "Distributed assignment of codes for multihop packet-radio Networks", *Proceedings of the IEEE MILCOM Conference*, volume 1, pages 450–454, 1997.
- [58] L. Hu, "Distributed code assignments for CDMA packet radio networks", *IEEE/ACM Transactions on Networking*, 1:668–677, Dicembre 1993.
- [59] E. Sousa, J. A. Silvester, "Spreading code protocols for distributed spread-spectrum packet radio networks", *IEEE Transactions on Communications*, 36(3):272–281, Marzo 1988.
- [60] K.-W. Hung, T.-S. Yum, "The coded tone sense protocol for multihop spread-spectrum packet radio networks", *Proceedings of the IEEE GLOBECOM Conference*, volume 2, pp. 712–716, 1989.
- [61] M. Joa-Ng, I.-T. Lu, "Spread spectrum medium access protocol with collision avoidance in mobile ad-hoc wireless network", *Proceedings of the IEEE INFOCOM Conference*, volume 2, pp. 776–783, 1999.
- [62] S. W. Lee, D. H. Cho "Distributed Reservation CDMA for Wireless Lan", *Proceedings of the IEEE GLOBECOM Conference*, volume 1, pp. 360–364, 1995.
- [63] I. Gupta, "Minimal CDMA recoding strategies in power-controlled ad hoc wireless networks", *Proceeding of 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing*, (San Francisco, Aprile 2001).
- [64] A. Penttinen, "Research On Ad Hoc Networking: Current Activity And Future Directions", <http://citeseer.nj.nec.com/533517.html>
- [65] R. Nelson, L. Kleinrock, "Spatial-TDMA: A collision-free multihop channel access control", *IEEE Transactions on Communications*, vol. 33, pp 934-944, 1985.
- [66] C. Sakr, T. Todd, "Carrier-sense protocols for packet-switched smart antenna base-station", *Proceedings of International Conference on Network Protocols (ICNP'97)*, 1997.
- [67] C. A. Pomalaza-Ráez, T. L. Hemminger, "A Unified Approach To Dynamic TDMA Slot Assignment And To Distributed Routing For Multi-Hop Packet Radio Networks".
- [68] Chlamtac et al., "Time-Spread Multiple-Access (TSMA) Protocols for Multihop Mobile Radio Networks", *IEEE/ACM Transactions on Networking*, Vol. 5, no. 6, Dicembre 1997.

- [69] Ji-Her Ju, Victor O. K. Li, "An Optimal Topology Transparent Scheduling Method in Multihop Packet Radio Networks", *IEEE/ACM Transactions on Networking*, Vol. 6, no. 3, Giugno 1998.
- [70] T. Shepard, "A Channel Access Scheme for Large Dense Packet Radio Networks", *SIGCOMM'96 Conference Proc.*, ACM 1996.
- [71] C. Zhu, M. S. Corson, "A Five-Phase Reservation Protocol (FPRP) for Mobile Ad Hoc Networks", *Proc. IEEE INFOCOM'98*.
- [72] Z. Tang, J. J. Garcia-Luna-Aceves, "Hop-Reservation Multiple Access (HRMA) for Multichannel Packet Radio Networks", *Proc. IEEE IC3N'98: Seventh International Conference on Computer Communications and Networks*, Lafayette, Louisiana, Ottobre 12-15, 1998.
- [73] A. Ephremides, T.V. Truong, "Scheduling broadcasts in multihop radio networks," *IEEE Trans. Commun.*, vol. 38, pp. 456-460, Aprile 1990.
- [74] R. Ramaswami and K.K.Parhi, "Distributed scheduling of broadcasts in a radio network", *IEEE INFOCOM '89*, Ottawa, Ont., Canada, Aprile 1989.
- [75] M. H. Ammar, D. S. Stevens, "A Distributed TDMA Rescheduling Procedure for Mobile Packet Radio Networks", *Proc. ICC '91*.
- [76] I.Cidon, M. Sidi, "Distributed Assignment Algorithms for Multi-hop Packet-Radio Networks" *IEEE Transactions on Computers*, 38(10):1353–1361, Ottobre 1989.
- [77] L. Bao, J.J. Garcia-Luna-Aceves, "Collision-Free topology-dependent channel access scheduling", *IEEE MILCOM '00*, Los Angeles, CA, Ottobre 2000.
- [78] P. Gupta, P. R. Kumar, "Capacity of wireless networks", *IEEE Transactions on Information Theory*, vol. 46(2), pp. 338-404, Marzo 2000.
- [79] T. Salonidis, L. Tassiulas, "Asynchronous TDMA ad hoc networks: Scheduling and Performance", *Communication Networks*, 2003.
- [80] R. Jurdak, C. V. Lopes, P. Baldi, "A Survey, Classification and Comparative Analysis of Medium Access Control Protocols for Ad Hoc Networks", *IEEE Communications Surveys & Tutorials*, University of California, 2004.
- [81] C. K. Toh *et al.*, "MARCH: A Medium Access Control Protocol for Multihop Wireless Ad Hoc Networks", *MILCOM 21st Century Military Commun. Conf.*, vol. 1, 2000, pp. 512–16.
- [82] J. Kim, N. Bambos, "Power-Efficient MAC Scheme using Channel Probing in Multi-rate Wireless Ad Hoc Networks", *IEEE Vehic. Tech. Conf.*, vol. 4, 2002, pp. 2380–84.
- [83] L. Bononi, M. Conti, L. Donatiello, "Distributed Contention Control Mechanism for Power Saving in Random-Access Ad-Hoc Wireless Local Area Networks", *Int'l. Workshop Mobile Multimedia Commun.*, IEEE, 1999, pp. 114–23.
- [84] F. A. Tobagi, L. Kleinrock, "Packet Switching in Radio Channels: Part II — The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution", *IEEE Trans. Commun.*, vol. 23, no. 12, 1975, pp. 1417–33.
- [85] Z. J. Haas, J. Deng, "Dual Busy Tone Multiple Access (DBTMA): A Multiple Access Control Scheme for Ad Hoc Networks", *IEEE Trans. Commun.*, vol. 50, no. 6, 2002, pp. 975–85.
- [86] Y. C. Tseng *et al.*, "A Multi-Channel MAC Protocol with Power Control for Multi-Hop Mobile Ad Hoc Networks", *Int'l. Conf. Distributed Comp. Sys. Wksp.*, 2001, pp. 419–24.