

ALMA MATER STUDIORUM · UNIVERSITÀ DI
BOLOGNA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea in Informatica

Modellazione Perdite Pacchetti in uno Scenario VoIP WiFi

Relatore:
Chiar.mo Prof.
Vittorio Ghini

Presentata da:
Alkida Balliu

Sessione II
Anno Accademico 2010-2011

Indice

1	Introduzione	5
1.1	Contesto Applicativo: Always Best Packet Switching (ABPS)	6
1.1.1	Transmission Error Detector - TED	7
1.2	Modellazione Scenario	9
2	Controllo a run-time	15
2.1	Metodo dei Minimi Quadrati	15
2.1.1	Introduzione al Metodo dei Minimi Quadrati	16
2.1.2	Metodo dei Minimi Quadrati: Funzione Lineare	17
2.1.3	Metodo dei Minimi Quadrati: Funzione di Secondo Grado	17
2.1.4	Metodo dei Minimi Quadrati: Funzione Esponenziale	18
2.1.5	Metodo dei Minimi Quadrati: Funzione Logaritmica	19
2.1.6	Minimi Quadrati: Funzione Potenziale	20
2.2	Algoritmo di previsione del prossimo burst	21
2.3	Causa della perdita di un pacchetto	22
3	Modelli Alternativi	23
3.1	Introduzione	23
3.2	Catena di Markov	23
4	Applicazione del Metodo su alcuni esperimenti	27
4.1	Prima Fase	27
4.2	Seconda Fase	32
4.3	Terza Fase	37
5	Conclusioni	43
	Bibliografia	47

Capitolo 1

Introduzione

La creazione e l'elaborazione dei modelli richiede pensiero e inventiva e di solito un approccio puramente meccanico è incompleto e può finire in un fallimento. Modelli diversi possono rivelare diversi aspetti di un problema quindi spesso i problemi non hanno una "miglior risposta". Più volte il teorico e lo sperimentale comunicano tra loro tramite l'analisi matematica fornendo così un terreno comune.

L'obiettivo di questo elaborato è di sviluppare una modellazione della distribuzione di una serie di pacchetti ricevuti e persi, in comunicazioni Voice Over Internet Protocol (VoIP). Le sequenze dei pacchetti persi e quella dei pacchetti ricevuti vengono chiamate rispettivamente burst e gap. Lo strumento utilizzato per la realizzazione del nostro modello è il *Metodo dei Minimi Quadrati*. Inventato dal matematico Gauss¹, questo è un metodo che ha trovato utilizzo in ampie aree della scienza. Nel 1801 l'astronomo italiano Giuseppe Piazzi scoprì l'asteroide Cerere e tracciò il suo percorso per 40 giorni, prima che Cerere fosse perduta nel bagliore del sole. Basandosi sui dati di Piazzi, e grazie al Metodo dei Minimi Quadrati, gli astronomi hanno potuto predire la futura posizione di questo asteroide.²

Una delle prime applicazioni dei computer digitali è stata proprio l'analisi dei dati sperimentali tramite il Metodo dei Minimi Quadrati. Il *Manhattan Project*³ durante la Seconda Guerra Mondiale ha usato questo metodo

¹La prima pubblicazione del Metodo dei Minimi Quadrati risale al 1805, in uno studio del matematico francese Legendre, ma nelle pubblicazioni delle sue memorie del 1809, Gauss afferma di averlo utilizzato per la prima volta nel 1795.

²fonte:Wikipedia

³Il Manhattan Project era il nome dato dalla sezione ingegneristica dell'esercito statunitense allo sviluppo della prima bomba atomica durante la Seconda Guerra Mondiale.

in diversi esperimenti per studiare questioni come gli effetti biologici delle radiazioni, il tempo di dimezzamento degli isotopi radioattivi ect.

In questo studio, basandoci su dati empirici raccolti da vari esperimenti, si vuole trovare la funzione più adeguata ai dati raccolti per rappresentarne al meglio la distribuzione e, di conseguenza, la previsione.

1.1 Contesto Applicativo: Always Best Packet Switching (ABPS)

Il contesto applicativo è quello di un device mobile, equipaggiato con interfacce wireless eterogenee (NIC - Network interface cards), in grado di sfruttare completamente il vantaggio offerto dalle interfacce multiple a disposizione. Le attuali tecnologie infatti permettono al device di usare selettivamente solo una NIC. In alternativa, il modello ABPS estende le capacità di un device mobile con più interfacce di rete, fornendogli la possibilità di instradare ciascun datagramma attraverso l'interfaccia più opportuna, consentendo in questo modo un uso simultaneo di tutte le NIC disponibili.

Un device mobile equipaggiato con NIC wireless eterogenee viene chiamato *multi-homed Mobile Node* (MN).

1.1. CONTESTO APPLICATIVO: ALWAYS BEST PACKET SWITCHING (ABPS)7

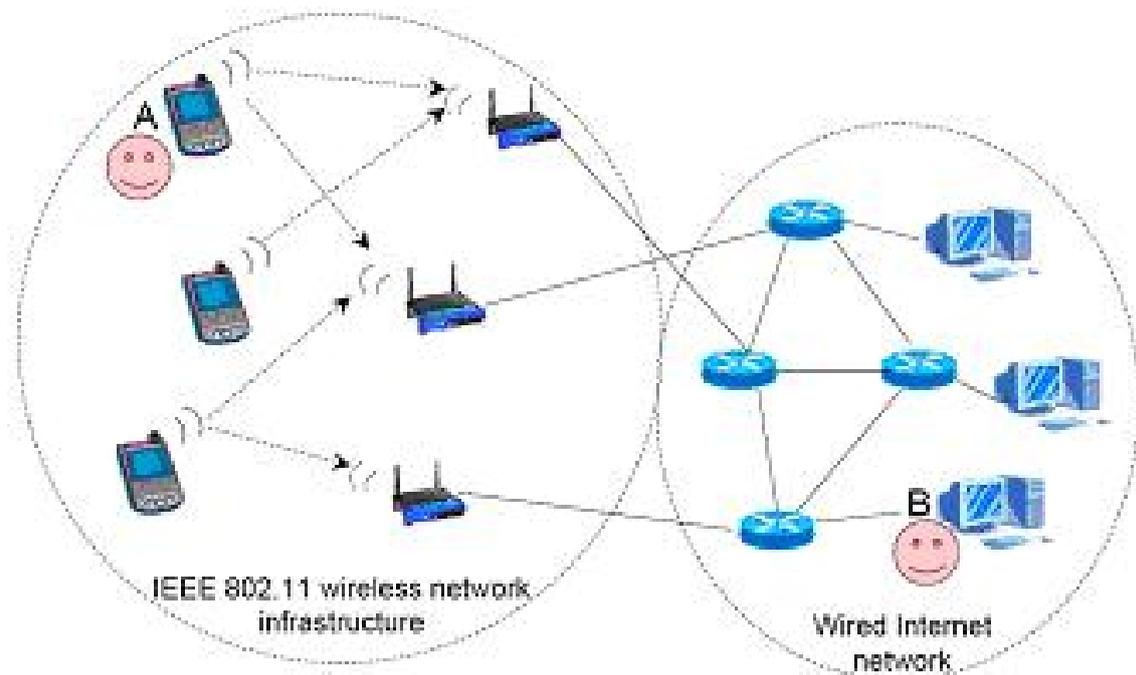


Figura 1.1: Scenario di Multi-homed Mobile Node

1.1.1 Transmission Error Detector - TED

La percentuale di pacchetti persi è un fattore discriminante per il rispetto dei requisiti QoS nei sistemi VoIP. L'utilizzo di un canale di comunicazione wireless, comporta un aumento intrinseco dei pacchetti persi, dovuto all'instabilità di questo tipo di trasmissione rispetto ai classici collegamenti cablati. Le cause principali del degrado della qualità della trasmissione in un canale wireless sono:

- **Interferenze Ambientali:** le onde prodotte da un device wireless si propagano lungo le tre dimensioni dello spazio, seguendo le leggi della fisica. Gli ostacoli che incontrano lungo il percorso, a seconda della forma e del materiale di cui sono costituiti, producono fenomeni di assorbimento o riflessioni che disturbano la trasmissione.
- **Interferenze Trasmissive:** altre sorgenti trasmissive presenti nelle vicinanze del dispositivo mobile possono interferire con la comunicazione (es. forni microonde, dispositivi controllo a distanza come cancelli automatici, allarmi auto, etc).

- **Mezzo Condiviso:** i vari terminali wireless condividono lo stesso mezzo trasmissivo, cioè l'aria. Devono quindi essere applicate politiche di *multiplexing*, per alternare e regolamentare l'accesso a tale mezzo. Una concentrazione di terminali in uno spazio ristretto, comporta una diminuzione della banda disponibile. Un normale Access Point (AP) sopporta circa dieci client che effettuano trasmissioni VoIP contemporaneamente.

TED è un meccanismo che serve a notificare alle applicazioni informazioni sulla perdita di un pacchetto durante il primo *hop*, cioè nel percorso tra l'host mobile e l'AP.

Questo approccio suggerisce di utilizzare un canale trasmissivo alla volta, fino a che le prestazioni di questo non degradino eccessivamente. Quando il canale utilizzato non è più considerato affidabile, si procede alla riassociazione con un altro Access Point, qualora ce ne sia uno più performante disponibile.

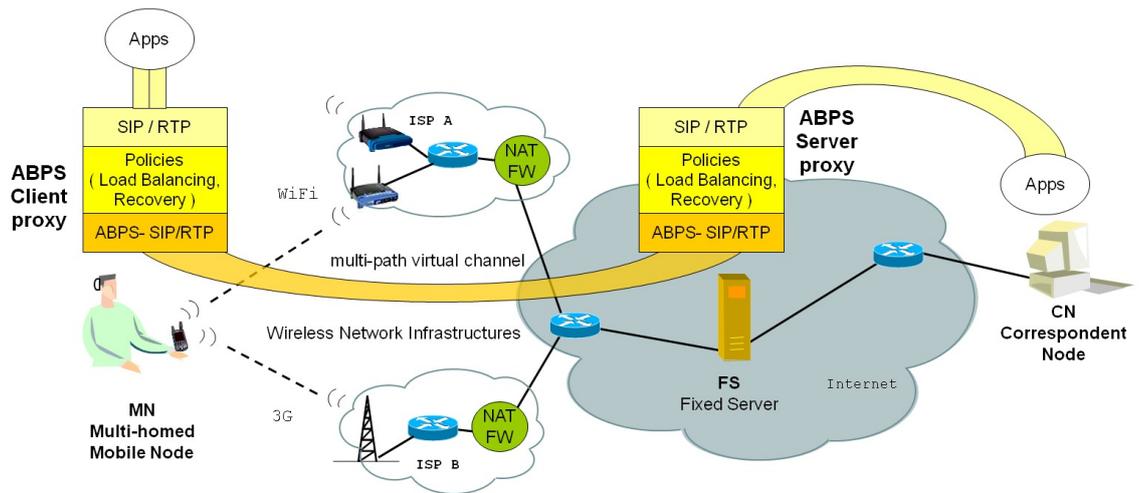


Figura 1.2: Il sistema ABPS

L'ABPS quindi risulta una componente necessaria per la realizzazione della modellazione della perdita dei pacchetti utilizzando il Metodo dei Minimi Quadrati.

1.2 Modellazione Scenario

Nel nostro lavoro prendiamo in considerazione un dispositivo mobile dotato di due schede di rete (Network Interface Cards, NIC). L'ambiente di rete di ciascuna scheda è costituito da un collegamento wireless e un percorso wired (cablato). Dunque il dispositivo mobile preso in considerazione è dotato di due interfacce wireless che sono associate a due Access Point diversi, i quali sono collegati al Nodo Fisso (Fixed Node) tramite una rete cablata.

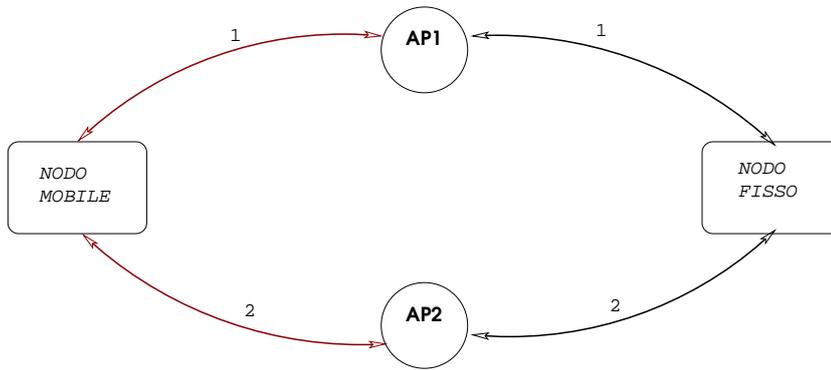


Figura 1.3: *Figura illustrativa dello scenario. Il collegamento rosso indica la wireless network, invece quello nero indica la wired network. Gli indici indicano rispettivamente la prima e la seconda interfaccia.*

I fattori che concorrono alla perdita dei dati sono diversi e possono andare dalla "congestione" della rete, per cui si verificano burst di perdite, a motivazioni di natura casuale, come ad esempio per causa di una scarica elettrica. Tramite il modello ABPS siamo capaci di accorgerci delle ragioni (casuali oppure per via di un burst) della perdita di un pacchetto nel canale wireless. Queste informazioni ci sono molto utili per realizzare la modellazione della perdita di questi pacchetti. Dato questo scenario vogliamo poter essere in grado, tramite alcune tecniche illustrate più avanti, di scegliere in ogni istante l'interfaccia più adatta. Tra le condizioni del nostro lavoro, poniamo che almeno una delle due interfacce sia sempre funzionante e valutiamo come nulla la possibilità che nello stesso momento tutte e due le interfacce si trovino in un burst di perdite.

Consideriamo ora tutti i casi in cui si può verificare la perdita di un pacchetto, quindi i seguenti eventi:

$$\mathcal{E}_1 = \{\text{il pacchetto viene perso nella wireless network della prima interfaccia}\}$$

$\mathcal{D}_1 = \{\text{il pacchetto viene perso nella wired network della prima interfaccia}\}$

$\mathcal{E}_2 = \{\text{il pacchetto viene perso nella wireless network della seconda interfaccia}\}$

$\mathcal{D}_2 = \{\text{il pacchetto viene perso nella wired network della seconda interfaccia}\}$

$\mathcal{A} = \{\text{ci troviamo in un burst di perdite}\}$

$\bar{\mathcal{A}} = \{\text{non ci troviamo in un burst di perdite}\}$

Possiamo esprimere la probabilità che un pacchetto venga perso con la seguente formula:

$$P = P(E_1|\bar{\mathcal{A}}) * P(\bar{\mathcal{A}}) * \left[P(E_2) + (1 - P(E_2)) * \left[P(D_2) + (1 - P(D_2)) * \int_{T-T_R}^{+\infty} f(t)d(t) \right] \right] \\ + P(E_1|\mathcal{A}) * P(\mathcal{A}) * \left[P(E_2|\bar{\mathcal{A}}) + (1 - P(E_2|\bar{\mathcal{A}})) * \left[P(D_2) + (1 - P(D_2)) \right. \right. \\ \left. \left. * \int_{T-T_R}^{+\infty} f(t)d(t) \right] \right] + (1 - P(E_1)) * \left[P(D_1) + (1 - P(D_1)) * \int_T^{+\infty} f(t)d(t) \right]$$

dove:

- $P(E_1)$ è la probabilità che il pacchetto venga perso nella wireless network della prima interfaccia⁴;
- $P(E_2)$ è la probabilità che il pacchetto venga perso nella wireless network della seconda interfaccia⁵;
- $P(D_1)$ indica la probabilità che il pacchetto venga perso nella wired network della prima interfaccia;
- $P(D_2)$ è la probabilità che il pacchetto venga perso nella wired network della seconda interfaccia;

⁴Esprime la probabilità che il pacchetto venga perso nel collegamento wireless della prima interfaccia indiscriminatamente dal motivo, che può essere per via di un burst oppure per una perdita casuale.

⁵Esprime la probabilità che il pacchetto venga perso nel collegamento wireless della seconda interfaccia indiscriminatamente dal motivo, che può essere per via di un burst oppure per una perdita casuale.

- $P(E_1|\bar{A})$ esprime la probabilità subordinata (o condizionata)⁶ che un pacchetto venga perso casualmente nella wireless network della prima interfaccia;
- $P(E_1|A)$ descrive la probabilità subordinata che un pacchetto venga perso in un burst di perdite nella wireless network della prima interfaccia;
- $P(E_2|\bar{A})$ indica la probabilità subordinata che il pacchetto venga perso casualmente nella wireless network della seconda interfaccia;
- $P(\bar{A}) = 1 - P(A)$ è la probabilità che si ha di non trovarsi in un burst di perdite.
- $P(A) = 1 - P(\bar{A})$ indica infine la probabilità di imbatterci in un burst di perdite.
- $\int_T^{+\infty} f(t)d(t)$ è la funzione di densità di probabilità lognormale⁷ che il pacchetto non giunga a destinazione entro il tempo limite T milisecondi. In questo caso, lo consideriamo perso.
- $\int_{T-T_R}^{+\infty} f(t)d(t)$ è la funzione di densità di probabilità lognormale che il pacchetto non giunga a destinazione entro il tempo $T - T_R$ milisecondi, dove T_R è il tempo di ritrasmissione di un pacchetto. In questo caso, consideriamo il pacchetto perso.

Cerchiamo di capire passo per passo la formula estesa sopra. Inizialmente consideriamo il caso in cui un pacchetto venga perso casualmente nella wireless network della prima interfaccia. Questo evento lo rappresentiamo moltiplicando la probabilità subordinata $P(E_1|\bar{A})$ con la probabilità di non trovarci in un burst di perdite, quindi:

$$P(E_1|\bar{A}) * P(\bar{A})$$

In questa occasione, la politica che seguiamo ci suggerisce di cambiare interfaccia e dunque di ritrasmettere il pacchetto nella seconda interfaccia⁸. Consideriamo infatti ora i casi in cui questo pacchetto venga perduto nella seconda interfaccia:

- Può essere perduto nella wireless network (quindi $P(E_2)$).

⁶La probabilità subordinata $P(X|Y)$ è la probabilità che si verifichi l'evento X nell'ipotesi che Y si sia già verificato

⁷La distribuzione di probabilità è una funzione che mette in relazione i valori di una variabile aleatoria con la probabilità che questi valori si presentino. Questa distribuzione di probabilità viene espressa dalla densità di probabilità. Nel nostro caso, la distribuzione dei dati è di tipo log-Normale.

⁸Nel frattempo, facciamo un controllo a run-time per capire le cause della perdita del pacchetto (sezione 2.2).

- Può passare dalla wireless network ($1 - P(E_2)$), ma perdersi nella wired network ($P(D_2)$).
- Può infine essere perduto nel caso in cui giunga a destinazione dopo $T - T_R$ ms.

Sino a questo punto la formula è:

$$P = P(E_1|\bar{A}) * P(\bar{A}) * \left[P(E_2) + (1 - P(E_2)) * \left[P(D_2) + (1 - P(D_2)) * \int_{T-T_R}^{+\infty} f(t) d(t) \right] \right]$$

Ritorniamo ora a considerare i casi rimanenti. Valutiamo il caso in cui il pacchetto venga perso sempre nella wireless network della prima interfaccia ma questa volta per causa di un burst di perdite. Esprimiamo dunque l'evento con la formula subordinata $P(E_1|A)$, e quindi la moltiplichiamo alla probabilità che indica l'evento "ci troviamo in un burst di perdite": $P(A)$:

$$p(E_1|A) * P(A)$$

Quando si verifica questa condizione, abbandoniamo la prima interfaccia per ritrasmettere il pacchetto nella seconda interfaccia e consideriamo quello che può succedere. In questo caso, il pacchetto può essere perso causalmente⁹ nella seconda interfaccia $P(E_2|\bar{A})$. Se così non fosse, e quindi si verificasse la probabilità complementare $1 - P(E_2|\bar{A})$, controlliamo se il pacchetto viene perso nella wired network della seconda interfaccia $P(D_2)$. Altrimenti consideriamo la probabilità che il pacchetto giunga a destinazione dopo il tempo limite $T - T_R$ ms $(1 - P(D_2)) * \int_{T-T_R}^{+\infty} f(t) d(t)$. La nostra formula a questo punto ha preso la forma:

$$P = P(E_1|\bar{A}) * P(\bar{A}) * \left[P(E_2) + (1 - P(E_2)) * \left[P(D_2) + (1 - P(D_2)) * \int_{T-T_R}^{+\infty} f(t) d(t) \right] \right] \\ + P(E_1|A) * P(A) * \left[P(E_2|\bar{A}) + (1 - P(E_2|\bar{A})) * \left[P(D_2) + (1 - P(D_2)) * \int_{T-T_R}^{+\infty} f(t) d(t) \right] \right]$$

Arrivati a questo punto, non ci resta che studiare il caso in cui il pacchetto non venga perso nella wireless network della prima interfaccia, $1 - P(E_1)$, e quindi non effettuiamo alcun cambio. In seguito, osserviamo la probabilità che il pacchetto venga perso nella wired network della prima interfaccia. Se questo evento non avviene, consideriamo la probabilità che il pacchetto venga perso perchè arriva a destinazione dopo il tempo limite T , $(1 - P(D_1)) * \int_T^{+\infty} f(t) d(t)$.

⁹Consideriamo che almeno una delle due interfacce sia funzionante, quindi omettiamo il caso in cui tutte e due le interfacce si trovino contemporaneamente in un burst di perdite.

Dopo questa serie di considerazioni, che passano in rassegna ogni possibile esito negativo, la formula che esprime la probabilità che un pacchetto venga perso è la seguente:

$$\begin{aligned}
P = & P(E_1|\bar{A}) * P(\bar{A}) * \left[P(E_2) + (1 - P(E_2)) * \left[P(D_2) + (1 - P(D_2)) * \int_{T-T_R}^{+\infty} f(t) d(t) \right] \right] \\
& + P(E_1|A) * P(A) * \left[P(E_2|\bar{A}) + (1 - P(E_2|\bar{A})) * \left[P(D_2) + (1 - P(D_2)) \right. \right. \\
& \left. \left. * \int_{T-T_R}^{+\infty} f(t) d(t) \right] \right] + (1 - P(E_1)) * \left[P(D_1) + (1 - P(D_1)) * \int_T^{+\infty} f(t) d(t) \right]
\end{aligned}$$

Nell'ambiente descritto fino ad ora, non si hanno informazioni riguardo all'andamento dei pacchetti mandati dal Nodo Fisso al Nodo Mobile. Cioè, se nell'istante t di trasmissione in una interfaccia i pacchetti giungono a destinazione, non sappiamo se all'istante $t + \text{tempo di trasmissione}$ i pacchetti mandati dal Nodo Fisso verso quello Mobile non vadano a finire in un burst di perdite.

Inoltre non conosciamo, in uno specifico istante, quale interfaccia è più conveniente usare.

Per questi motivi usiamo dei metodi tramite i quali facciamo un controllo a run-time.

Nel controllo a run-time lavoriamo su due casi.

Il primo controllo, consiste nel osservare per ogni perdita se si tratta di una perdita causata da un burst oppure da una perdita casuale. Per fare ciò, utilizziamo il sistema ABPS, le cui politiche e il cui funzionamento abbiamo visto nella sezione 1.1.

Il secondo controllo consiste nel prevedere l'istante in cui avverrà il prossimo burst di perdite.

Intendiamo stimare analiticamente una funzione discreta $b(i)$ che restituisce l'istante d'inizio dell' i -esimo burst di perdite. La stima verrà fatta sulla base di dati empirici ricavati sperimentalmente a run-time. Tali dati empirici sono una sequenza di n coppie $(j, b(j))$ che indicano quando sono iniziati i precedenti n burst di perdite. In particolare, il primo dato della coppia è l'indice j dell'avvenuto burst di perdite di pacchetti, mentre il secondo dato della coppia è l'istante in cui quel burst di perdite ha avuto inizio. A partire da tali dati empirici, sfruttando il Metodo dei Minimi Quadrati, potremo approssimare la funzione discreta $b(i)$ con una funzione parametrica nota $t(i)$. I parametri di questa funzione nota verranno stabiliti proprio grazie alle informazioni sui precedenti n burst di perdite. Ottenuti tali parametri, la funzione nota ci permetterà di prevedere quando inizierà il successivo $(n+1)$ -esimo burst di perdite, sulla interfaccia di rete che stiamo utilizzando. In altre parole, approssimeremo $b(n+1)$ con $t(n+1)$. In tal modo, potremo cambiare l'interfaccia di rete da usare prima dell'istante $t(n+1)$ in cui prevediamo

che comincerà il successivo burst di perdite. La funzione $b(i)$ verrà approssimata nuovamente ogni volta che verrà identificato l'inizio di un nuovo burst di perdite, ovvero ogni volta che otterremo un nuovo dato sperimentale, una nuova coppia $(j, b(j))$. Usando gli ultimi n dati sperimentali stabiliremo i parametri della funzione che meglio interpreta i dati $t()$ e potremo poi stimare l'istante di inizio del prossimo burst di perdite.

La funzione di approssimazione $t()$ è scelta mediante il Metodo dei Minimi Quadrati, tramite il quale basandoci su dati sperimentali, cioè sulla storia recente dell'evento, riusciamo a fare una previsione sul prossimo burst di perdite. La qualità di questa previsione, dipende da quanto sono recenti i dati empirici. Al fine di avere una buona previsione, ci teniamo sempre aggiornati sull'istante in cui la nostra interfaccia subisce una sequenza, più o meno lunga, di pacchetti persi. Per fare questo, ci viene in aiuto il nostro primo controllo a run-time, descritt

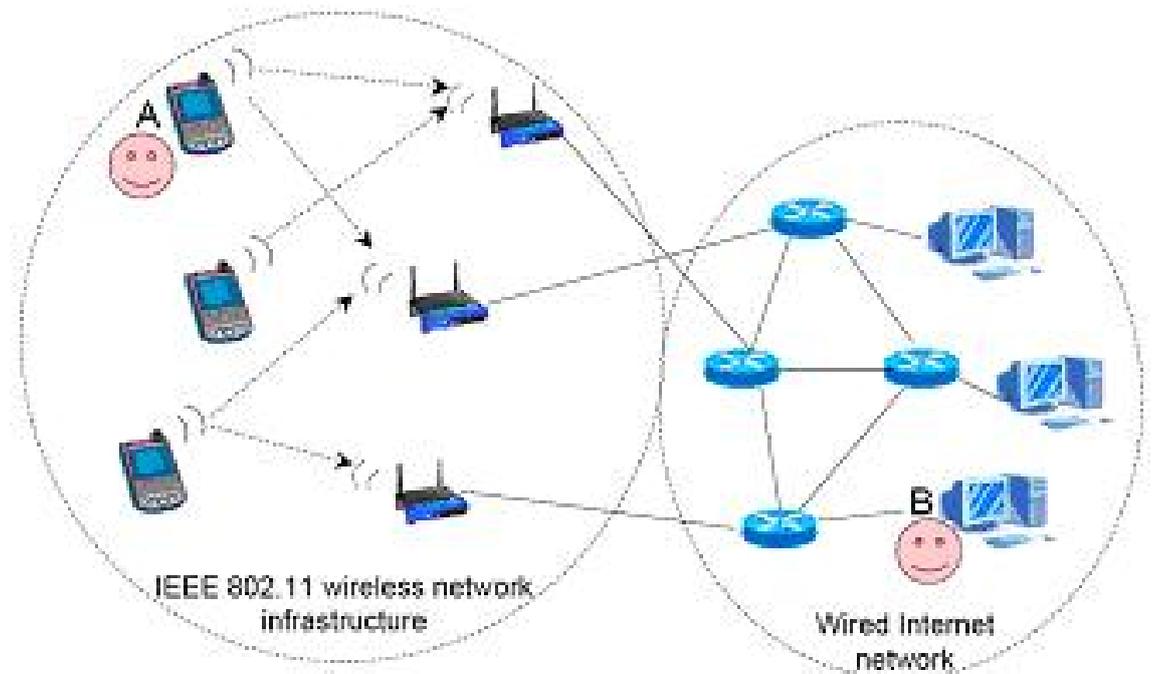


Figura 1.4: Scenario di Multi-homed Mobile Node

o qui sopra. In questo modo, avendo un modello robusto di previsione e facendo i dovuti calcoli, riusciamo ad evitare il burst di perdite cambiando interfaccia al tempo giusto.

Capitolo 2

Controllo a run-time

2.1 Metodo dei Minimi Quadrati

Tramite il Metodo dei Minimi Quadrati noi cerchiamo di ricavare informazioni, il più possibile esatte, riguardo un insieme di dati. Questo insieme di dati descrive un fenomeno che cambia nel tempo e non è detto che cambi sempre nella stessa maniera. Perciò i dati su cui ci basiamo per il nostro modello, devono essere frequentemente aggiornati perchè la modellazione e i risultati da essa ottenuti siano coerenti.

La stima della accuratezza dei dati valutati è parte integrante dell'analisi e quindi per stimare la precisione delle nostre misurazioni, utilizziamo il *Coefficiente di Determinazione*¹, R^2 , il quale misura la proporzione di variabilità totale, spiegata dal modello specifico². Il valore del Coefficiente di Determinazione varia tra 0 ed 1.

$$R^2 = 1 - \frac{\text{Somma Residua}}{\text{Devianza Totale}}$$

dove:

$$\text{Somma Residua} = \sum_{i=1}^n (y_i - \bar{y}_i)^2$$

$$\text{Devianza Totale} = \sum_{i=1}^n (y_i - y_{media})^2$$

dove y_i è l'i-esimo dato osservato, \bar{y}_i è l'i-esimo dato stimato, $y_{media} = \frac{\sum_{i=1}^n y_i}{n}$ è la media complessiva delle variabili y osservati.

¹Il Coefficiente di Determinazione, nelle regressioni lineari è semplicemente il quadrato del Coefficiente di Correlazione. IL Coefficiente di Correlazione è indice di quanto varia mediamente la variabile dipendente (y), al variare di un'unità della variabile indipendente (x).

² R^2 indica la percentuale di quanto il coefficiente di correlazione sia realistico.

2.1.1 Introduzione al Metodo dei Minimi Quadrati

Molte aree della scienza utilizzano gli esperimenti quantitativi per determinare i parametri di interesse. Gli esperimenti quantitativi sono caratterizzati da variabili misurate, un modello matematico e dei parametri sconosciuti.

Nella maggior parte degli esperimenti il Metodo dei Minimi Quadrati viene utilizzato per analizzare i dati al fine di determinare i valori per i parametri sconosciuti. È un metodo basato su modelli parametrici, quindi su modelli che fanno utilizzo delle equazioni matematiche per descrivere i fenomeni osservati. Una volta che un modello parametrico è stato proposto, l'analisi dei dati deve essere eseguita ponendo interesse al calcolo dei valori ignoti, alla stima delle derivazioni standard e all'utilizzo di questi valori ignoti per calcolare la funzione a noi d'interesse. L'obiettivo è quello di prevedere una variabile di output in funzione di una serie di variabili in ingresso.

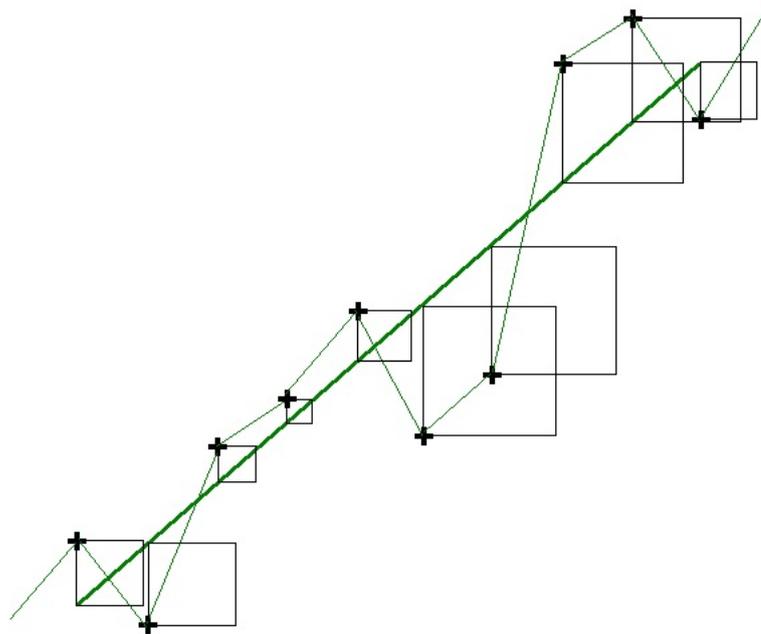


Figura 2.1: Minimi Quadrati

La funzione con maggior grado di accuratezza per un'insieme di dati x,y che possono essere interpretati come punti di un piano, è quella che ha distanza minore da tutti i punti del piano cartesiano. Questa funzione si ottiene costruendo per ogni punto empirico un quadrato, il lato del quale è determinato dalla distanza verticale del punto dalla retta. Sulla base di questo lato, risultante parallelo con

l'asse degli y, si creano gli altri lati che formano i quadrati. Si ripete quindi il procedimento per ogni punto del grafico e si sommano le aree di tutti i quadrati. La funzione che meglio approssima la distribuzione dei punti dati è quella che determina la minore superficie dei quadrati. Da qui il nome Metodo dei Minimi Quadrati.

2.1.2 Metodo dei Minimi Quadrati: Funzione Lineare

In questa sezione vediamo lo sviluppo del Metodo dei Minimi Quadrati per dati che hanno un andamento lineare. La funzione lineare è espressa dalla forma:

$$y = a + b * x$$

Una volta scelto il tipo di funzione di approssimazione, vogliamo ottenere un buon accostamento fra la distribuzione dei valori osservati e quella dei valori teorici calcolati. Perciò la somma dei quadrati delle differenze tra i valori osservati e quelli teorici, deve essere la minore³. Questo concetto viene espresso dalla *Condizione di Accostamento*:

$$f(a, b) = \sum_{i=1}^n (y_i - (a + b * x_i))^2 = \text{minimo}$$

A questo punto, tutto si riduce al trovare i valori a e b tali per cui la funzione appena descritta sia quella minima. Quindi i valori a e b che soddisfano la Condizione di Accostamento sono espressi in questo modo:

$$a = \frac{\sum_{i=1}^n y_i * \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i * \sum_{i=1}^n (x_i * y_i)}{n * \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$b = \frac{n * \sum_{i=1}^n (x_i * y_i) - \sum_{i=1}^n x_i * \sum_{i=1}^n y_i}{n * \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

dove n indica il numero dei dati disponibili.

2.1.3 Metodo dei Minimi Quadrati: Funzione di Secondo Grado

Osserviamo il caso in cui la funzione di approssimazione scelta sia di secondo grado:

$$y_i = a + b * x_i + c * x_i^2$$

³Prendiamo in considerazione la somma dei quadrati delle differenze e non semplicemente la somma delle differenze. Questo perchè essendo alcune differenze positive, altre negative e altre ancora nulle, esse potrebbero compensarsi.

In tal caso, la nostra Condizione di Accostamento sarà:

$$f(a) = \sum_{i=1}^n (y_i - a + b * x_i + c * x_i^2)^2 = \text{minimo}$$

Notiamo che la funzione da risolvere è composta da tre parametri. Per calcolare a , b e c in modo tale che questa funzione sia la minima, dobbiamo risolvere il sistema che si ottiene annullando le derivate parziali prime rispetto ad a, b e c :

$$\begin{cases} \frac{\delta f}{\delta a} = 0 \\ \frac{\delta f}{\delta b} = 0 \\ \frac{\delta f}{\delta c} = 0 \end{cases}$$

Svolgendo le operazioni, il sistema diventa:

$$\begin{cases} -2 * \sum_{i=1}^n (y_i - a - b * x_i - c * x_i^2) = 0 \\ -2 * \sum_{i=1}^n (y_i - a - b * x_i - c * x_i^2) * x_i = 0 \\ -2 * \sum_{i=1}^n (y_i - a - b * x_i - c * x_i^2) * x_i^2 = 0 \end{cases}$$

cioè

$$\begin{cases} -\sum_{i=1}^n y_i + \sum_{i=1}^n a + b * \sum_{i=1}^n x_i + c * \sum_{i=1}^n x_i^2 = 0 \\ -\sum_{i=1}^n x_i y_i + a * \sum_{i=1}^n x_i + b * \sum_{i=1}^n x_i^2 + c * \sum_{i=1}^n x_i^3 = 0 \\ -\sum_{i=1}^n x_i^2 y_i + a * \sum_{i=1}^n x_i^2 + b * \sum_{i=1}^n x_i^3 + c * \sum_{i=1}^n x_i^4 = 0 \end{cases}$$

Portando gli elementi negativi dall'altra parte dell'uguaglianza e svolgendo le sommatorie delle variabili costanti, si ha:

$$\begin{cases} a * n + b * \sum_{i=1}^n x_i + c * \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i \\ a * \sum_{i=1}^n x_i + b * \sum_{i=1}^n x_i^2 + c * \sum_{i=1}^n x_i^3 = \sum_{i=1}^n x_i y_i \\ a * \sum_{i=1}^n x_i^2 + b * \sum_{i=1}^n x_i^3 + c * \sum_{i=1}^n x_i^4 = \sum_{i=1}^n x_i^2 y_i \end{cases}$$

A questo punto il nostro sistema, ormai lineare, si può facilmente risolvere per sostituzione, e una volta ottenuto i valori di

$$\sum_{i=1}^n x_i; \sum_{i=1}^n x_i^2; \sum_{i=1}^n x_i^3; \sum_{i=1}^n x_i^4; \sum_{i=1}^n y_i; \sum_{i=1}^n x_i y_i; \sum_{i=1}^n x_i^2 y_i$$

possiamo calcolare le tre incognite a , b e c .

2.1.4 Metodo dei Minimi Quadrati: Funzione Esponenziale

Nei casi in cui la funzione adatta sia una funzione esponenziale della forma:

$$y_i = a * b^{x_i}$$

con $a, b > 0$ e $b \neq 1$

la condizione di accostamento diviene:

$$\sum_{i=1}^n (y_i - a * b^{x_i})^2 = \text{minimo}$$

Per determinare i due parametri a e b dobbiamo ricorrere ad un artificio che consiste nel prendere il logaritmo decimale di ambo i membri:

$$\log y_i = \log a + x_i * \log b$$

dopodichè facciamo le seguenti trasformazioni:

$\log a = n$; $\log b = m$; $\log y_i = z_i$; alla luce di questo, possiamo riscrivere la funzione nel seguente modo:

$$z_i = n + m * x_i$$

In questo modo abbiamo ridotto il problema al caso della funzione lineare, i cui parametri abbiamo già calcolato nella sezione 2.1.1.

Quindi:

$$n = \frac{\sum_{i=1}^n \log y_i * \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i * \sum_{i=1}^n (x_i * \log y_i)}{n * \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$m = \frac{n * \sum_{i=1}^n (x_i * \log y_i) - \sum_{i=1}^n x_i * \sum_{i=1}^n \log y_i}{n * \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

Trovati n ed m , ricaviamo a e b :

$$\log a = n \implies a = 10^n$$

$$\log b = m \implies b = 10^m$$

2.1.5 Metodo dei Minimi Quadrati: Funzione Logaritmica

Vediamo ora cosa succede quando i dati presentano un andamento logaritmico. La funzione che esprime questo fenomeno è:

$$y = a + b * \ln(x)$$

Per calcolare a e b e trovare la funzione logaritmica che meglio spiega i nostri dati, risolviamo il sistema che otteniamo annullando le derivate parziali prime rispetto ad a e b :

$$\begin{cases} \frac{\delta f}{\delta a} = 0 \\ \frac{\delta f}{\delta b} = 0 \end{cases}$$

Faccendo le dovute operazioni matematiche, il sistema evolve in:

$$\begin{cases} -2 * \sum_{i=1}^n (y_i - a - b * \ln(x_i)) = 0 \\ -2 * \sum_{i=1}^n (y_i - a - b * \ln(x_i)) * \ln(x_i) = 0 \end{cases}$$

Dividendo con -2 entrambe le parti dell'uguaglianza e svolgendo i calcoli:

$$\begin{cases} -\sum_{i=1}^n y_i + \sum_{i=1}^n a + b * \sum_{i=1}^n \ln(x_i) = 0 \\ -\sum_{i=1}^n \ln(x_i) * y_i + a * \sum_{i=1}^n \ln(x_i) + b * \sum_{i=1}^n \ln(x_i)^2 = 0 \end{cases}$$

cioè

$$\begin{cases} a * n + b * \sum_{i=1}^n \ln(x_i) = -\sum_{i=1}^n y_i \\ a * \sum_{i=1}^n \ln(x_i) + b * \sum_{i=1}^n \ln(x_i)^2 = -\sum_{i=1}^n \ln(x_i) * y_i \end{cases}$$

A questo punto è facile calcolare a e b che sono le seguenti:

$$a = \frac{\sum_{i=1}^n y_i * \sum_{i=1}^n \ln(x_i)^2 - \sum_{i=1}^n \ln(x_i) * \sum_{i=1}^n (\ln(x_i) * y_i)}{n * \sum_{i=1}^n \ln(x_i)^2 - (\sum_{i=1}^n \ln(x_i))^2}$$

$$b = \frac{n * \sum_{i=1}^n (\ln(x_i) * y_i) - \sum_{i=1}^n \ln(x_i) * \sum_{i=1}^n y_i}{n * \sum_{i=1}^n \ln(x_i)^2 - (\sum_{i=1}^n \ln(x_i))^2}$$

dove n indica il numero dei dati disponibili.

2.1.6 Minimi Quadrati: Funzione Potenziale

L'ultimo caso da noi considerato studia i dati che presentano un andamento potenziale. La funzione valutata è:

$$y = a * x^b$$

In questo caso, la condizione di accostamento è:

$$\sum_{i=1}^n (y_i - a * x_i^b)^2 = \text{minimo}$$

Per trovare a e b , trasformiamo l'equazione applicando il logaritmo di base 10 a entrambe le parti dell'uguaglianza:

$$\log y_i = \log a + b * \log x$$

In questo modo notiamo che il problema diventa simile al caso della funzione lineare, caso già studiato in questo capitolo. Segnando $\log a = n$ e $\log y_i = z_i$, troviamo a e b :

$$n = \frac{\sum_{i=1}^n \log y_i * \sum_{i=1}^n (\log x_i)^2 - \sum_{i=1}^n \log x_i * \sum_{i=1}^n (\log x_i * \log y_i)}{n * \sum_{i=1}^n (\log x_i)^2 - (\sum_{i=1}^n \log x_i)^2}$$

quindi una volta calcolato n , a diventa:

$$\log a = n \implies a = 10^n$$

Invece il valore di b lo calcoliamo con la formula:

$$b = \frac{n * \sum_{i=1}^n (\log x_i * \log y_i) - \sum_{i=1}^n \log x_i * \sum_{i=1}^n \log y_i}{n * \sum_{i=1}^n (\log x_i)^2 - (\sum_{i=1}^n \log x_i)^2}$$

2.2 Algoritmo di previsione del prossimo burst

La funzione $b(i)$ che restituisce l'istante di inizio dell' i -esimo burst di perdite di pacchetti, dovrà essere ri-calcolata nuovamente ogni volta che verrà identificato l'inizio di un nuovo burst di perdite, ovvero ogni volta che otterremo un nuovo dato sperimentale, una nuova coppia $(j, b(j))$.

Usando gli ultimi n dati sperimentali stabiliremo i parametri della funzione di approssimazione $t()$ e potremo poi stimare l'istante di inizio del prossimo burst di perdite.

Mediante il Metodo dei Minimi Quadrati stabiliamo i parametri che caratterizzano le funzioni che approssimeranno i nostri dati sperimentali, ovvero Lineare, Polinomiale di secondo grado, Esponenziale, Logaritmica, Potenziale.

Stabiliti i parametri per le cinque funzioni, calcoleremo il coefficiente di determinazione R^2 per ciascuna di queste.

La funzione scelta per approssimare la funzione $b(i)$ sarà la funzione che fornisce il coefficiente di determinazione maggiore.

Stabilito quale funzione $t()$ usare come approssimazione della $b()$, se l'indice dell'ultimo burst individuato era n , allora il valore $t(n + 1)$ indicherà l'istante in cui presumibilmente avverrà il prossimo burst di perdite.

L'algoritmo da seguire per prevedere l'istante di inizio del prossimo burst di perdite può essere descritto perciò come nello pseudo-codice qui di seguito:

NIC = Network Interface Card

AP = Access Point

per ogni pacchetto da trasmettere mediante la NIC attuale

```

{
  tenta trasmissione del pacchetto
  se il pacchetto viene perso
  {
    fai ping verso l'AP attuale
    se il ping fallisce
    {
      classifica l'istante attuale come inizio del burst di perdite n-esimo
      per ciascuna funzione in Lineare, Di secondo grado, Esponenziale, Logaritmica, Potenziale
      {
        calcola parametri per la funzione che approssima gli n burst di perdite
        calcola Coefficiente di Determinazione
      }
      scegli la funzione  $t()$  con indice di correlazione maggiore
      il prossimo burst di perdite inizia all'istante  $t(n + 1)$ 
      poco prima dell'istante  $t(n+1)$  cambiare NIC
    }
  }
}

```

```
}
}
```

2.3 Causa della perdita di un pacchetto

La previsione di un burst, ottenuta tramite il Metodo dei Minimi Quadrati, ci da un'indicazione su quando cambiare interfaccia per non cadere in un burst di perdite. Nel nostro modello, il Nodo Fisso manda i pacchetti attraverso l'interfaccia dalla quale li riceve. Quindi sapendo il *tempo di trasmissione*, è facile calcolare il momento più conveniente per cambiare interfaccia in modo da non perdere i pacchetti che vengono mandati dal Nodo Fisso.

Nel momento in cui ci accorgiamo della perdita di un pacchetto, effettuiamo un cambiamento di interfaccia e trasmettiamo da essa i prossimi pacchetti fino a quando non avremo informazioni sulle cause di questa perdita, per poi reagire di conseguenza. A questo punto è nostro interesse sapere se il pacchetto è stato perso casualmente oppure si tratta di un burst di perdite. Se fosse una perdita casuale, e prima della perdita di questo pacchetto l'interfaccia andava bene, ritorniamo ad usarla e a sfruttare il suo buon andamento. Se fossimo nel caso di un burst, allora abbandoniamo questa interfaccia e utilizziamo la seconda che sicuramente avrà un andamento migliore. Come facciamo ad ottenere queste informazioni?

Per avere informazioni sulle cause della perdita di un pacchetto, per esempio nella prima interfaccia, l'idea è stata quella di fare un test di Ping, mandare un pacchetto prova all'Access Point 1 ed aspettare una sua risposta sulla avvenuta ricezione o meno. Se anche questo pacchetto non viene ricevuto, capiamo allora di trovarci in un burst di perdite e quindi continueremo ad usare la seconda interfaccia. Se il pacchetto invece viene ricevuto potrebbe essere che la perdita sia stata casuale oppure può essere che ci troviamo all'inizio di un burst dove la probabilità di perdita è ancora relativamente bassa. In questo caso abbiamo deciso di accrescere la vigilanza e fare un'ulteriore test di Ping. Nel caso in cui anche questo pacchetto venga ricevuto, allora siamo certi che la perdita è stata casuale e ritorniamo alla prima interfaccia, nel caso contrario saremmo in un burst di perdite.

La previsione dei burst di perdite si basa sui dati empirici. Perché le informazioni che vogliamo ottenere siano il più possibile precise, i dati su cui ci basiamo devono essere ripetutamente aggiornati. Ciò è possibile farlo grazie alle informazioni ottenute a run-time. In questo modo la modellazione attraverso il Metodo dei Minimi Quadrati può essere sempre aggiornata per essere costantemente attuale e più affidabile.

Capitolo 3

Modelli Alternativi

3.1 Introduzione

Le Catene di Markov trovano un ampio utilizzo come metodo di modellazione in svariati campi. Il loro impiego ci da informazioni sulla probabilità che abbiamo di trovarci al tempo t , nello stato s_t , sapendo tutti gli stati precedenti. Questa probabilità però dipende solo dallo stato immediatamente precedente. Cioè, se ci troviamo nello stato *buon andamento dell'interfaccia(in un gap)*, le catene di Markov ci forniscono un'informazione sulla probabilità che abbiamo di passare nello stato *burst di perdite*. Le Catene di Markov sono quindi caratterizzate da una certa indipendenza rispetto al passato, cosa che da robustezza al metodo.

Un modello che usa le catene di Markov ci fornisce un approccio diverso rispetto alla modellazione con il Metodo dei Minimi Quadrati. L'informazione raccolta tramite il Metodo dei Minimi Quadrati si basa sulla storia passata e, grazie a questo, possiamo estrapolare informazioni sul futuro. È proprio il Metodo dei Minimi Quadrati che ci da l'opportunità di fare una previsione sull'istante in cui può avvenire il prossimo burst di perdite e di poter reagire di conseguenza.

È comunque di rilevante importanza prendere visione dell'applicazione delle catene di Markov riguardo la modellazione della perdita dei pacchetti.

3.2 Catena di Markov

Diversi studiosi[2] in questo campo hanno proposto modellazioni di perdita dei pacchetti utilizzando le catene di Markov discrete e finite, a due e quattro stati. Si è notato che nei termini di una scala di tempo di pochi secondi o di pochi minuti, i burst di perdite avvengono approssimativamente con la stessa distribuzione. Un sistema di catene di Markov con due stati è in grado di rappresentare bene questo fenomeno.

Riguardo però ad intervalli di tempo più ampi la catena di Markov a due stati non è più sufficiente. Serve quindi un modello più generale che è stato implemen-

tato usando le catene di Markov a quattro stati.

Le catene di Markov usate sono omogenee nel tempo e irriducibili, il che significa che la probabilità di transizione fra stati è costante nel tempo e che tutti gli stati comunicano gli uni con gli altri.

Iniziamo a studiare il problema usando le catene di Markov a due stati.

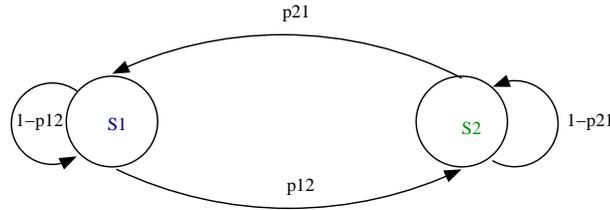


Figura 3.1: *Catena di Markov a due stati. Gli stati errati e corretti sono indicati rispettivamente dai colori blu e verde. p_{xy} rappresenta la probabilità di passare dallo stato x allo stato y .*

Siano $S = S_1, S_2, \dots, S_m$ gli m stati di una catena di Markov e sia p_{ij} la probabilità della catena di passare dallo stato i allo stato j . Possiamo esprimere questa probabilità con la formula:

$$p_{ij} = P(X_i = x_j | X_{i-1} = x_i)^1$$

Indicata così la probabilità, potremo utilizzarla per trovare le formule che esprimono la distribuzione della lunghezza dei burst e dei gap:

$$f_{burst}(k) = p_{12}(1 - p_{12})^{k-1}$$

$$f_{gap}(k) = p_{21}(1 - p_{21})^{k-1}$$

I valori p_{12} e p_{21} sono stati stimati come segue:

$$p_{12} = t_{c \rightarrow e} / n_p$$

$$p_{21} = t_{e \rightarrow c} / n_r$$

dove $t_{c \rightarrow e}$ e $t_{e \rightarrow c}$ sono rispettivamente il numero di transizioni da stati corretti a stati errati e viceversa. Invece n_r e n_p rappresentano rispettivamente il numero di pacchetti ricevuti e persi.²

¹ p_{ij} è la probabilità subordinata di passare dallo stato i allo stato j in un passo. X_i si può interpretare come un numero aleatorio che si evolve nel tempo.

²La differenza assoluta $|t_{c \rightarrow e} - t_{e \rightarrow c}|$ è al massimo uno e può essere considerata zero quando n tende all'infinito.

Nel modello della catena di Markov a quattro stati, la probabilità di transizione fra stati corretti, così come la probabilità di transizione fra stati errati, è stata assegnata a zero.

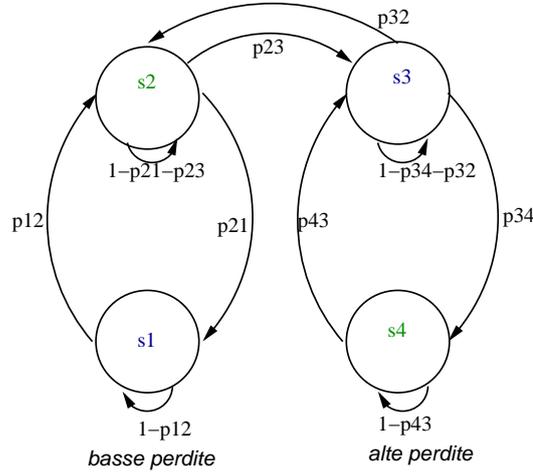


Figura 3.2: Catena di Markov a quattro stati. Gli stati errati e corretti sono indicati rispettivamente dai colori blu e verde. p_{xy} rappresenta la probabilità di passare dallo stato x allo stato y .

La transizione da uno stato corretto ad uno erroneo e la transizione inversa hanno la stessa probabilità, che è data dalla formula:

$$s_2 * (p_{21} + p_{23}) + s_4 * p_{43} = s_1 * p_{12} + s_3 * (p_{34} + p_{32})$$

In questo sistema, la media della lunghezza dei burst e quella dei gap risulta:

$$L_{burst} = (s_1 + s_3) / (s_2(p_{21} + p_{23}) + s_4 * p_{43})$$

$$L_{gap} = (s_2 + s_4) / (s_2(p_{21} + p_{23}) + s_4 * p_{43})$$

dove s_1, s_2, s_3, s_4 sono rispettivamente la probabilità di trovarsi negli stati 1, 2, 3 e 4.

Dai risultati dei vari esperimenti fatti da diversi studiosi, si è constatato che, nel rappresentare l'andamento della distribuzione della lunghezza dei burst, il modello a due stati si è rivelato adeguato e sufficiente. Questo perchè la distribuzione della lunghezza dei burst di perdite si avvicina rapidamente a valori molto bassi (prossimi a zero).

Invece nel caso della distribuzione della lunghezza dei gap, caratterizzata da un decadimento più lento, il modello a due stati non è sufficiente a causa della scarsa

flessibilità della sua formula che, come si è visto sopra, è composta da un solo parametro. In questa eventualità si utilizza il modello che adotta le catene di Markov a quattro stati.

Capitolo 4

Applicazione del Metodo su alcuni esperimenti

Per la maggior parte degli esperimenti quantitativi, il metodo dei minimi quadrati risulta essere la migliore tecnica analitica per estrarre informazioni da un insieme di dati. Questa affermazione è basata sul presupposto che le incertezze, cioè gli errori nei dati, non sono correlate e sono normalmente distribuite. Ciò non toglie il fatto che, anche quando quest'ultima assunzione non sia rispettata, spesso il modello risulta robusto ed efficiente.

Quando la curva che meglio interpreta i dati è lineare, (o polinomiale, esponenziale ect) per descrivere l'andamento dei dati usiamo il termine *regressione lineare (o polinomiale, esponenziale, ect)*

Gli esperimenti consistono nel testare un software VoIP sul nostro dispositivo mobile in un'area coperta da due Access Point, muovendoci casualmente tra vari ostacoli materiali e architettonici. In questo modo abbiamo raccolto i dati empirici sui quali ci siamo basati per fare le nostre previsioni.

Abbiamo così svolto tre esperimenti in condizioni diverse l'uno dall'altro, dividendo il lavoro in tre fasi.

4.1 Prima Fase

In questa prima fase abbiamo utilizzato una connessione VoIP per circa 28 minuti, passando in prossimità di ostacoli fisici. I dati raccolti, vengono rappresentati tramite una tabella, la quale ci indica l'istante effettivo nel quale è iniziato il burst di perdite e la sua previsione.¹

¹Per avere una previsione con il Metodo dei Minimi Quadrati, abbiamo bisogno di almeno tre punti (cioè dati raccolti), dato che con uno o anche due punti, nello scegliere la funzione migliore, abbiamo più di una scelta.

Indice del Burst	Istante Inizio Burst	Previsto Inizio Burst
1	80200	
2	161720	
3	227920	
4	348720	306603
5	526600	471680
6	612000	716631
7	729800	791323
8	813800	826503
9	878400	958717
10	1003320	1004251
11	1172520	1100150
12	1275520	1253941
13	1398000	1382545
14	1482920	1511278
15	1553120	1615364
16	1673600	1697040

Tabella 4.1: Primo Esperimento: tabella illustrativa degli istanti in cui c'è stato un burst di perdite, e loro previsione

Una volta che abbiamo completato l'analisi con il Metodo dei Minimi Quadrati, rivolgiamo la nostra attenzione ad una valutazione dei risultati.

La distanza media fra i dati empirici e la loro previsione è di 44660 millisecondi (0,74 minuti) con un massimo di 104631 ms (1,74 minuti) nella previsione del sesto burst di perdite e un minimo di 931 ms (0.02 minuti) ottenuto nella previsione del decimo burst.

Vediamo più in dettaglio qualche passaggio interessante della tabella.

Nella previsione del quarto burst di perdite, la funzione più adatta a interpretare i dati è una funzione potenziale della forma:

$$y = y = 81021.29 * x^{0.96}$$

con

$$R^2 = 1.00$$

Possiamo rappresentare i dati con il seguente grafico

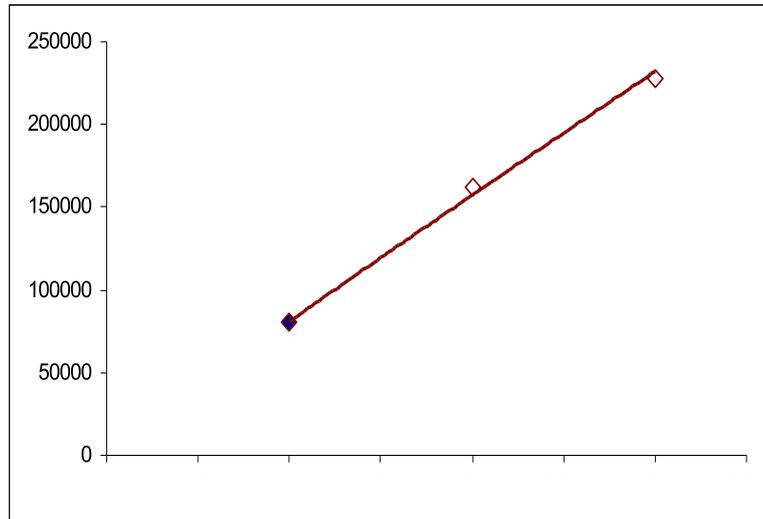


Figura 4.1: Prima Fase: funzione di approssimazione per i primi tre dati

In questo modo abbiamo potuto fare una previsione riguardo all'istante in cui avremo il quarto burst, e tramite il modello ABPS abbiamo potuto rilevare l'effettivo istante di avvenuta di tale burst. A questo punto, nel fare la previsione sul quinto burst di perdite, la funzione che meglio interpreta i dati non è più una funzione potenziale ma è una funzione polinomiale di secondo grado della forma:

$$y = 35800 + 38076 * x + 9820 * x^2$$

con

$$R^2 = 0.99$$

I dati continuano con un andamento polinomiale di secondo grado fino all'ottavo burst di perdite, di cui calcoliamo la previsione con una funzione potenziale della forma:

$$74073.09 * x^{1.16}$$

con

$$R^2 = 0.99$$

Qui di seguito, i nostri dati e la funzione potenziale che li interpreta descritti graficamente:

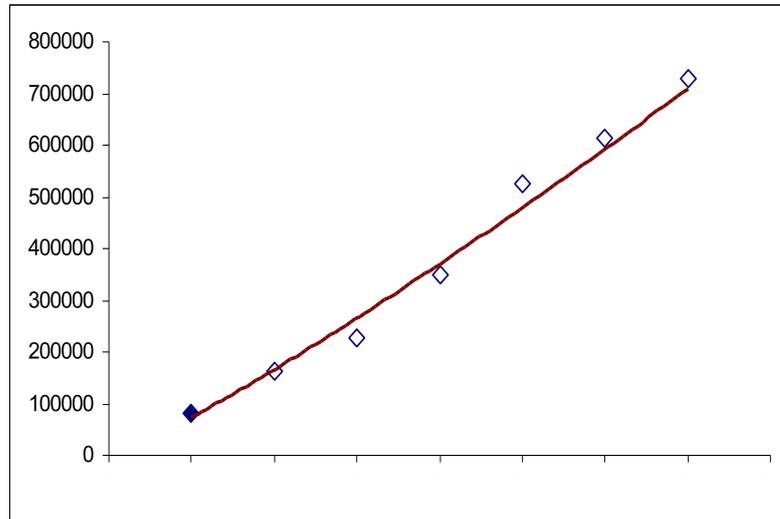


Figura 4.2: Prima Fase: funzione di approssimazione per i primi sette dati

Come possiamo notare, la funzione che interpreta i dati cambia con il cambiare di questi ultimi. In questo primo esperimento, la funzione che meglio si adatta ai dati sarà sempre una funzione di secondo grado. Possiamo così dedurre che i burst di perdite in questo esperimento sono avvenuti secondo una regressione polinomiale di secondo grado.

Facciamo un piccolo salto e prendiamo in considerazione il punto 10 della tabella. La funzione che ci viene in aiuto per calcolare la previsione del decimo burst di perdite è la seguente:

$$y = -65095.24 + 116204.8 * x - 927.01 * x^2$$

In questo caso

$$R^2 = 0.98$$

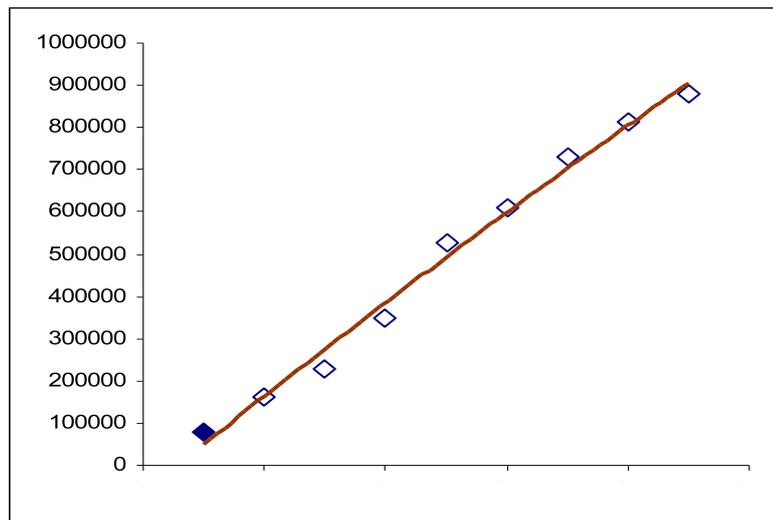


Figura 4.3: Prima Fase: funzione di approssimazione per i primi nove dati

Infine, la funzione che meglio interpreta tutti i 16 dati del nostro primo esperimento è la funzione di secondo grado:

$$y = -55201.76 + 108356.77 * x + 72.4 * x^2$$

con

$$R^2 = 1.00$$

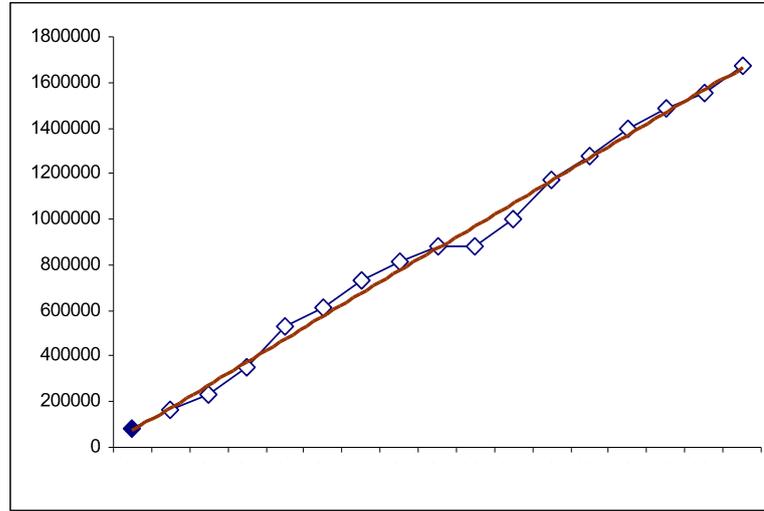


Figura 4.4: Prima Fase: funzione di approssimazione per i dati empirici

4.2 Seconda Fase

Indice del Burst	Istante Inizio Burst	Previsto Inizio Burst
1	35700	
2	39000	
3	123300	
4	125100	192599
5	129000	167025
6	131400	131582
7	141600	151079
8	142200	156074
9	145200	158883
10	145800	161209
11	147600	162702
12	149100	163963
13	151200	165034
14	152100	166100
15	153300	166975
16	156300	167766
17	158700	168766
18	159300	169852

19	160800	170778
20	164100	171691
21	167100	172804
22	168000	174050
23	169800	175191
24	170700	176335

Tabella 4.2: Secondo Esperimento: tabella illustrativa degli istanti in cui c'è stato un burst di perdite e loro previsione

In questo secondo esperimento il tipo di regressione dei burst di perdite è diverso da quello del primo esperimento. Se nel primo i dati avevano un andamento polinomiale di secondo grado, in quest'ultimo hanno una regressione logaritmica. La distanza media fra i dati empirici e la loro previsione è di 13783 millisecondi (0.23 minuti) con un massimo di 67499 ms (1,12 minuti) nella previsione del quarto burst di perdite e un minimo di 182 ms (0.003 minuti) ottenuto nella previsione del sesto burst.

Come possiamo notare dalla tabella, la previsione del 4° e quella del 5° burst di perdite non è molto soddisfacente. Infatti vi è una notevole probabilità che le prime previsioni siano poco riuscite poichè i dati su cui ci basiamo sono pochi per poter determinare un vero e proprio andamento.

La funzione di cui facciamo uso per determinare la previsione del quinto burst è una funzione di secondo grado:

$$y = -9225 + 37125 * x - 375 * x^2$$

con

$$R^2 = 0.83$$

I dati vengono così rappresentati dalla funzione:

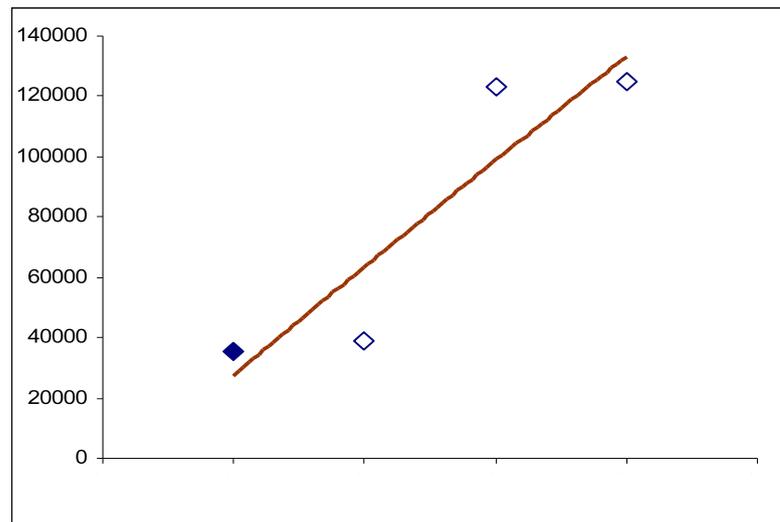


Figura 4.5: Seconda Fase: funzione di approssimazione per i primi quattro dati

Per quanto riguarda invece la previsione del sesto burst, facciamo uso di una funzione di secondo grado, la quale meglio interpreta i nostri dati. La funzione è la seguente:

$$y = -32040.00 + 62112.86 * x + -5807.14 * x^2$$

con

$$R^2 = 0.84$$

Graficamente i dati sono rappresentati in questo modo:

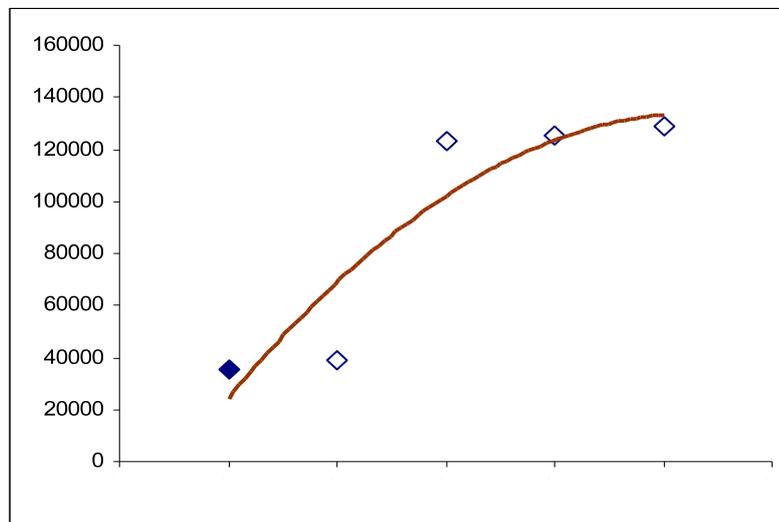


Figura 4.6: Seconda Fase: funzione di approssimazione per i primi cinque dati

Da qui in poi abbiamo notato che la regressione dei dati, quindi quella dei burst di perdite, è di tipo logaritmica. Per esempio, nella previsione del 13° burst di perdite, la funzione logaritmica che meglio si adatta ai dati a disposizione sino a quel momento è:

$$y = 48684.54 * \ln(x) + 40160.98$$

con

$$R^2 = 0.84$$

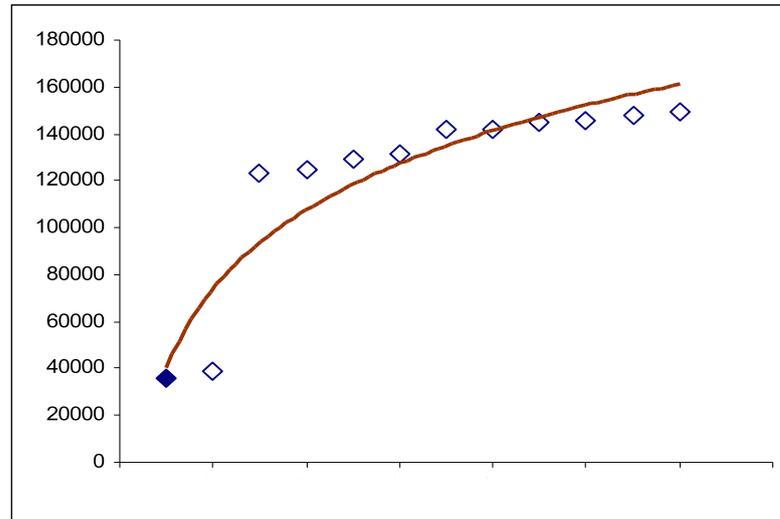


Figura 4.7: Seconda Fase: funzione di approssimazione per i primi dodici dati

Così tutte le altre funzioni che passo passo rappresentano al meglio i dati sono risultate logaritmiche. L'ultima funzione che interpreta tutti i dati raccolti in questo esperimento è la seguente:

$$y = 39091.48 * \ln(x) + 52100.81$$

con

$$R^2 = 0.86$$

ed è graficamente rappresentata in questo modo:

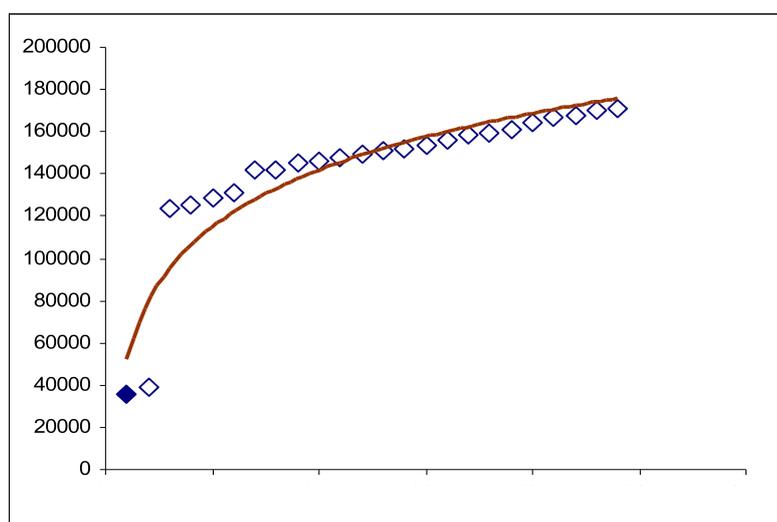


Figura 4.8: Seconda Fase: funzione di approssimazione per i primi dodici dati

4.3 Terza Fase

In questa terza fase riportiamo i risultati del terzo esperimento, svolto in condizioni diverse dai primi due, come si può notare dalla diversità dei dati raccolti:

Indice del Burst	Istante Inizio Burst	Previsto Inizio Burst
1	45270	
2	45630	
3	106410	
4	107550	140693
5	156060	124237
6	156780	195563
7	157620	184067
8	246660	189387
9	380070	309959
10	444120	438860
11	532470	558990
12	533910	721968
13	595470	788445
14	643920	723794
15	737610	768154

16	864510	838016
17	941760	941268
18	944850	996787
19	1033620	1062563
20	1035120	1138959
21	1097310	1182969
22	1098510	1231249
23	1100550	1257744
24	1149960	1269586
25	1240320	1289170

Tabella 4.3: Tabella illustrativa dell'istante in cui c'è stato un burst di perdite e loro previsione

In questo terzo esperimento, i burst di perdite sono avvenuti in maniera alquanto irregolare e per questo motivo è stato difficile descrivere tramite qualche regola l'insieme di questi dati. Infatti i dati alternano frequentemente tipo di regressione da esponenziale a polinomiale di secondo grado e viceversa.

Quindi inizialmente i dati presentano un'andamento potenziale, come nel caso della previsione del 5° burst di perdite. Questa previsione è poco soddisfacente sia per la su citata irregolarità dell'esperimento sia perchè i dati sui quali ci basiamo sono ancora pochi. L'equazione che usiamo per la previsione del quinto burst è:

$$y = 39626.48 * x^{0.71}$$

con

$$R^2 = 0.79$$

I quattro dati sui quali abbiamo fatto la previsione del quinto espressi graficamente:

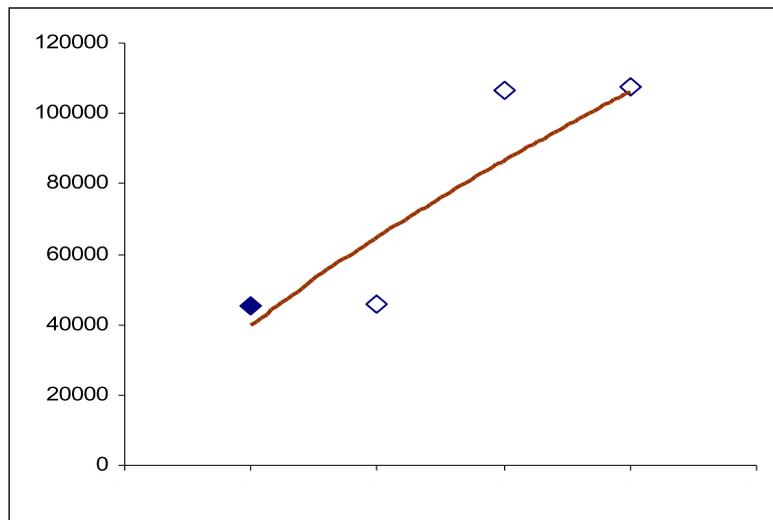


Figura 4.9: Terza Fase: funzione che interpreta i primi quattro dati

L'aggiunta alla tabella dell'effettivo istante in cui è arrivato il quinto burst di perdite, ha cambiato il tipo di regressione dei dati da esponenziale a polinomiale di secondo grado. Infatti per la previsione del 6° burst di perdite usiamo la funzione:

$$y = 25464 + 12638.57 * x + 2618.57 * x^2$$

con

$$R^2 = 0.92$$

Questo è il grafico dei rilevamenti:

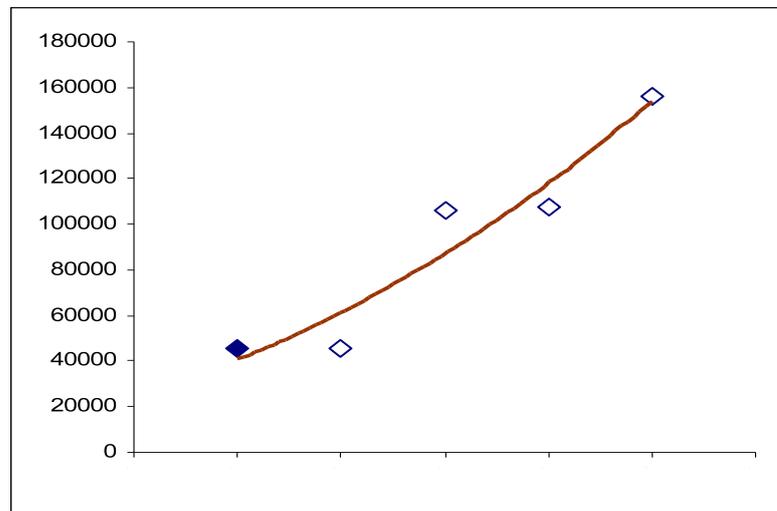


Figura 4.10: Terza Fase: funzione che interpreta i primi cinque dati

Calcoliamo invece l'ottavo burst con una funzione esponenziale, così come tutti i seguenti sino al tredicesimo dato della terza tabella. Riportiamo, per esempio, il calcolo della previsione del 10° burst di perdite tramite la funzione esponenziale:

$$y = 37172.90 * 1.28^x$$

con

$$R^2 = 0.93$$

Dal 14° burst di perdite e sino alla fine di quest'ultimo esperimento l'andamento dei dati cambia ancora una volta. Ritorniamo così a calcolare le nostre previsioni con funzioni polinomiali di secondo grado. La previsione del 14° burst è calcolata attraverso la funzione:

$$y = 31913.08 + 6217.75 * x + 3085.88 * x^2$$

con

$$R^2 = 0.97$$

La rappresentazione grafica è la seguente:

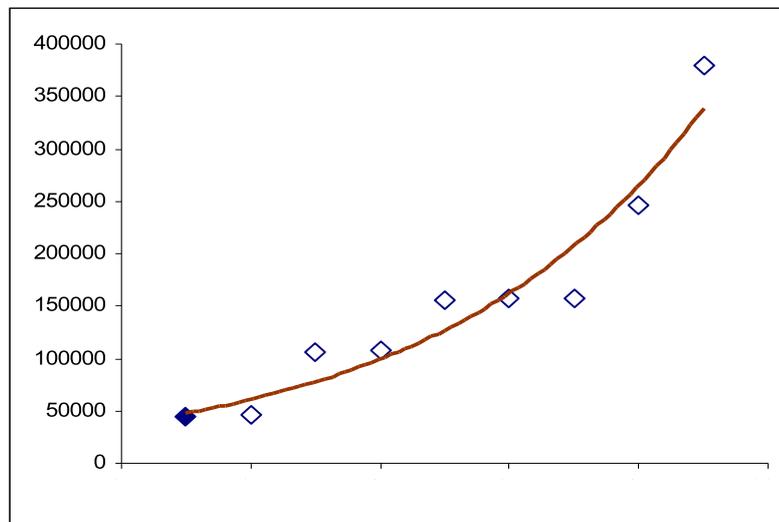


Figura 4.11: Terza Fase: funzione che interpreta i primi nove dati

La miglior previsione in questo terzo esperimento è quella del 17^o burst di perdite, calcolata dalla funzione:

$$y = 8688.33 + 17462.90 * x + 2148.13 * x^2$$

con

$$R^2 = 0.97$$

In fine, la funzione che interpreta al meglio tutto l'insieme dei dati empirici di questo esperimento, è la funzione di secondo grado:

$$y = -106394.58 + 56169.68 * x - 36.06 * x^2$$

con

$$R^2 = 0.98$$

Il grafico dei dati e della funzione che meglio li rappresenta, è il seguente:

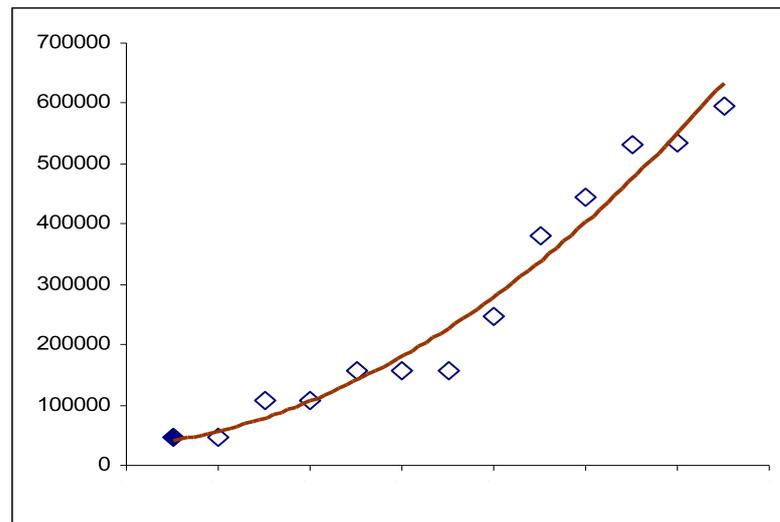


Figura 4.12: Terza Fase: funzione che interpreta i dati empirici

Capitolo 5

Conclusioni

Una volta completata l'analisi dei dati tramite il Metodo dei Minimi Quadrati è naturale rivolgere la nostra attenzione alla valutazione dei risultati. Ci chiediamo se il modello da noi adottato sia una rappresentazione adeguata del problema.

La modellazione dei dati non è sempre basata su un modello matematico corretto, e può essere altresì interessante confrontare modelli teorici alternativi per determinare quali siano i più adatti, i meglio applicabili ai dati sperimentali. Abbiamo voluto qui presentare un modello analitico che può essere usato per descrivere i burst di perdita dei pacchetti e per sfruttare le informazioni raccolte al fine di migliorare le prestazioni ed evitare i burst. Mutando le variabili sperimentali, abbiamo avuto modo di studiare cosa deve essere fatto in modo che gli esperimenti raggiungano gli obiettivi desiderati.

Nonostante il vasto numero dei cosiddetti Statistical Software¹, abbiamo voluto sviluppare dei programmi nel linguaggio C capaci di calcolare, facendo uso del Metodo dei Minimi Quadrati, le funzioni meglio adatte ai dati di tipo Lineare, Polinomiale di secondo grado, Esponenziale, Logaritmico e Potenziale. Ci siamo concentrati su queste funzioni considerandone cinque fra le più essenziali per questo lavoro. È naturale però che più tipi di funzioni abbiamo a disposizione più possiamo essere precisi nel calcolo delle previsioni. Per di più, le previsioni sono state fatte basandoci su tutto l'insieme dei dati ed è stato notato che non tutti i burst avevano lo stesso tipo di regressione. Per esempio infatti, durante il terzo esperimento, per la previsione del quinto burst abbiamo utilizzato una funzione di tipo potenziale, per la previsione del sesto una funzione polinomiale di secondo grado, per quella del decimo invece, una funzione esponenziale. Questo ci suggerisce che il modello si può migliorare con qualche algoritmo che sceglie di interrogare gli ultimi x dati, oppure i dati degli ultimi y minuti, e non tutto l'insieme.

¹Programmi che permettono di applicare e studiare analisi statistiche

Giova inoltre ribadire che nei casi in cui i dati empirici raccolti sono stati eccessivamente irregolari e hanno di conseguenza cambiato frequentemente tipo di regressione, il modello è risultato poco efficiente. Diversamente nei casi in cui i dati sono stati maggiormente regolari, il modello ha dato risultati soddisfacenti.

Abbiamo voluto proporre il Metodo dei Minimi Quadrati come un'interpretazione del problema della perdita dei pacchetti in uno scenario VoIP. In conclusione, vogliamo associarci a quanto affermato incisivamente da un professore d'università:

Non vi è una interpretazione per ogni cosa; l'interpretazione è un veicolo al servizio della comprensione umana. Il valore dell'interpretazione risiede nel consentire ad altri di pensare fruttuosamente riguardo un'idea². Ci auguriamo che questo lavoro possa servire anche a questo.

²Andreas Buja

Elenco delle figure

1.1	Scenario di Multi-homed Mobile Node	7
1.2	Il sistema ABPS	8
1.3	<i>Figura illustrativa dello scenario. Il collegamento rosso indica la wireless network, invece quello nero indica la wired network. Gli indici indicano rispettivamente la prima e la seconda interfaccia.</i>	9
1.4	Scenario di Multi-homed Mobile Node	14
2.1	Minimi Quadrati	16
3.1	<i>Catena di Markov a due stati. Gli stati errati e corretti sono indicati rispettivamente dai colori blu e verde. p_{xy} rappresenta la probabilità di passare dallo stato x allo stato y.</i>	24
3.2	<i>Catena di Markov a quattro stati. Gli stati errati e corretti sono indicati rispettivamente dai colori blu e verde. p_{xy} rappresenta la probabilità di passare dallo stato x allo stato y.</i>	25
4.1	Prima Fase: funzione di approssimazione per i primi tre dati	29
4.2	Prima Fase: funzione di approssimazione per i primi sette dati	30
4.3	Prima Fase: funzione di approssimazione per i primi nove dati	31
4.4	Prima Fase: funzione di approssimazione per i dati empirici	32
4.5	Seconda Fase: funzione di approssimazione per i primi quattro dati	34
4.6	Seconda Fase: funzione di approssimazione per i primi cinque dati	35
4.7	Seconda Fase: funzione di approssimazione per i primi dodici dati	36
4.8	Seconda Fase: funzione di approssimazione per i primi dodici dati	37
4.9	Terza Fase: funzione che interpreta i primi quattro dati	39
4.10	Terza Fase: funzione che interpreta i primi cinque dati	40
4.11	Terza Fase: funzione che interpreta i primi nove dati	41
4.12	Terza Fase: funzione che interpreta i dati empirici	42

Elenco delle tabelle

4.1	Primo Esperimento: tabella illustrativa degli istanti in cui c'è stato un burst di perdite, e loro previsione	28
4.2	Secondo Esperimento: tabella illustrativa degli istanti in cui c'è stato un burst di perdite e loro previsione	33
4.3	Tabella illustrativa dell'istante in cui c'è stato un burst di perdite e loro previsione	38

Bibliografia

- [1] J.Wolberg. *Data Analysis Using the Method of Least Squares*. Springer, 2006.
- [2] L. Estrada, D. Torres, H. Toral. *Characterization and Modeling of Packet Loss of a VoIP Communication*. World Academy of Science, Engineering and Technology 66 2010.
- [3] V.Ghini, G.Lodi, F.Panzieri *Always Best Packet Switching: the Mobile VoIP Case Study*. Journal of Communications, Vol 4, No 9 (2009), 700-713, Oct 2009.
- [4] M.Trovato *Probabilità Statistica Ricerca Operativa: Metodi e Strumenti*. Ghisetti & Corvi Editori.
- [5] T.Hastie, R.Tibshirani, J.Friedman *The Elements of Statistical Learning :Data mining, inference and Prediction*. Springer.
- [6] M. Davidian, D. Giltinan *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall.
- [7] D.M. Bates, D.G. Watts *Nonlinear regression analysis and its applications*. Wiley.