

**CONFRONTO TRA PROTOCOLLI
DI ACCESSO PER RETI
METROPOLITANE DI TIPO FASNET
REALIZZATE CON FIBRE OTTICHE**

Rapporto Interno C88-33

Giugno 1988

M. Conti

CONFRONTO TRA PROTOCOLLI DI ACCESSO PER RETI
METROPOLITANE DI TIPO FASNET REALIZZATE CON FIBRE OTTICHE

M. CONTI .

Rapporto Interno CNUCE C88-33

Consiglio Nazionale delle Ricerche
Istituto CNUCE - Pisa
Copyright Giugno 1988

Questo lavoro si inquadra nell'ambito dello studio di un'architettura per una rete geografica a pacchetto da realizzarsi interconnettendo, tramite nodo di gateway, reti metropolitane, di tipo FASNET, integrate nei servizi (video, voce e dati) e realizzate con fibre ottiche.

In particolare, l'obiettivo di questo lavoro e' lo studio di protocolli di accesso (MAC protocol) per reti metropolitane con configurazione fisica del tipo FASNET; infatti, e' questo livello di protocollo che risulta maggiormente influenzato dalle caratteristiche della configurazione fisica (la lunghezza del mezzo trasmissivo) e dalla tecnologia usata (fibre ottiche).

Ogni protocollo preso in esame viene studiato sia attraverso modelli matematici, sia tramite la simulazione.

1.0 INTRODUZIONE

Questo lavoro si inquadra nell'ambito dello studio di un'architettura per una rete geografica a pacchetto da realizzarsi interconnettendo, tramite nodo di gateway, reti metropolitane di tipo FASNET [LIMB 82] integrate nei servizi (video, voce e dati) e realizzate con fibre ottiche.

La scelta di una architettura di rete di tipo FASNET, per le reti metropolitane, e' motivata dalla esistenza di prototipi di LAN di questo tipo realizzate utilizzando circuiti ottici e dall'intenzione del costruttore di estendere questo protocollo per ottenere un'architettura per reti metropolitane integrate nei servizi [LIMB 84; DIX 86]. Quindi, sebbene altri protocolli (ad esempio, il protocollo QPSX [BUDRIKIS 86; NEWMAN 86; NEWMAN 87]) possano offrire prestazioni migliori, il nostro obiettivo non e' quello di trovare il protocollo che fornisca la soluzione ottima, ma di ricercare una buona soluzione per il problema in esame.

In particolare, l'obiettivo di questo lavoro e' lo studio di protocolli di accesso (MAC protocol) per reti metropolitane con configurazione fisica del tipo FASNET; infatti, e' questo livello di protocollo che risulta maggiormente influenzato dalle caratteristiche della configurazione fisica (ad esempio, la lunghezza delle linee) e dalla tecnologia usata (fibre ottiche).

Per ogni protocollo esaminato viene proposto un modello matematico che consente di evidenziare i parametri rilevanti per definire le *performance* della rete e di stimare i valori medi di alcuni indici di prestazione (numero di pacchetti in attesa di trasmissione e tempi di attesa dei pacchetti prima della trasmissione). Gli indici di prestazione ricavati analiticamente sono confrontati ed estesi con i risultati ottenuti risolvendo, tramite simulazione, un modello del protocollo realizzato con una rete di code.

2.0 PRESENTAZIONE DELL'ARCHITETTURA DI FASNET

Fasnet e' una rete sincrona che utilizza un meccanismo di controllo degli accessi al mezzo trasmissivo del tipo *token passing*. Il mezzo trasmissivo e' formato da una coppia di linee unidirezionali alle quali sono connesse le stazioni.

Per presentare il meccanismo di accesso facciamo riferimento alla rappresentazione della configurazione fisica di una rete di questo tipo presentata in Figura 1.

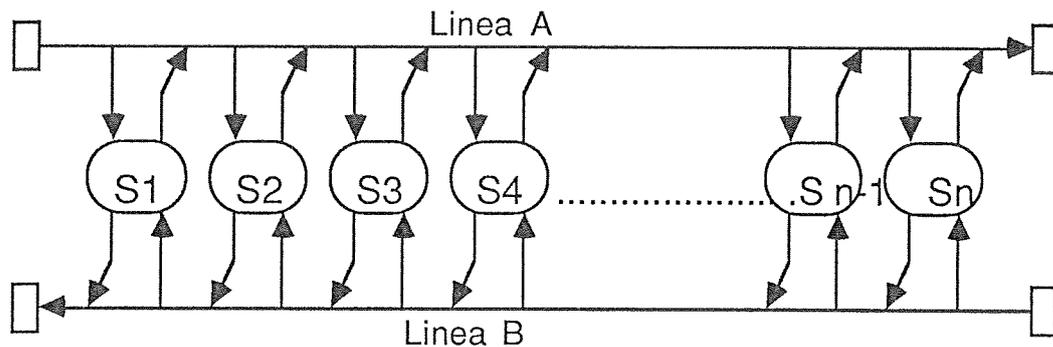


FIGURA 1: Configurazione fisica di FASNET

Data una linea, ad esempio la linea A, la stazione S_1 e' detta *head station* (HS) e la stazione S_n e' detta *end station* (ES); sulla linea opposta le funzioni sono invertite.

La funzione della HS e' quella di trasmettere ad intervalli regolari un codice di sincronizzazione creando cosi' una sequenza di slot che si propagano sulla linea; le altre stazioni si sincronizzano su questo segnale. Uno slot ha una struttura interna formata da un campo di controllo e da un campo informativo. Il campo di controllo viene utilizzato per controllare gli accessi al mezzo trasmissivo, il campo informativo viene utilizzato, invece, per la trasmissione dei pacchetti (quando uno slot contiene un pacchetto si parla di *frame*).

L'accesso al mezzo da parte delle stazioni e' controllato attraverso i bit START (S) BUSY (B) e END (E), contenuti nel campo di controllo degli slot, e attraverso il parametro P_{max} il cui valore puo' assumere valori diversi per le singole stazioni.

La HS settando ad uno il bit S indica l'inizio di un ciclo; una stazione S_i quando legge il bit $S=1$ e' abilitata a trasmettere fino ad un massimo di $P_{max}(i)$ pacchetti consecutivi. La trasmissione effettiva dei pacchetti avviene utilizzando gli slot in cui il bit B e' uguale a zero e settando lo stesso bit ad uno per inibire le stazioni successive dall'utilizzo di tale slot. La posizione delle stazioni lungo la linea determina uno schema a priorita' negli accessi, infatti, la stazione S_i trasmette i suoi pacchetti sulla linea A prima della stazione S_j se $i < j$. In base a questa condizione, quando la ES individua uno slot vuoto ($B=0$) significa che tutte le stazioni hanno gia' trasmesso il numero di pacchetti per il quale erano abilitate in quel ciclo. La ES segnala alla HS la fine del ciclo settando ad uno il bit END del campo di controllo di uno slot della linea opposta; quando lo slot con il bit $END=1$ viene letto dalla HS, questa provvede ad inizializzare un nuovo ciclo settando $START=1$. Durante la fase (detta *sojourn_time*) in cui il bit $END=1$ si propaga, sulla linea B, dalla ES alla HS nessuna stazione e' abilitata a trasmettere, cio' costituisce uno spreco della capacita' trasmissiva del mezzo.

3.0 IL PROTOCOLLO DI ACCESSO IN UNA RETE METROPOLITANA

La scelta del protocollo di accesso (MAC protocol) per una rete metropolitana realizzata con circuiti ottici deve tener conto di alcuni aspetti.

a.1 La capacita' del mezzo trasmissivo (O(Gb/sec))

a.2 Lunghezza delle linee (O(100Km))

la capacita' del mezzo trasmissivo riduce la durata degli slot a misure dell'ordine del microsecondo, mentre, data la lunghezza delle linee, l'informazione di controllo impiega per propagarsi da una estremita' all'altra della rete un tempo pari al tempo necessario a trasmettere alcune centinaia di pacchetti.

b. Costo delle risorse hardware in funzione della semplicita' del protocollo

Le caratteristiche del protocollo di accesso determinano i costi dei componenti elettronici che costituiscono le singole stazioni. In questa analisi centeremo l'attenzione, essenzialmente, sulle dimensioni della memoria veloce nella quale vengono memorizzati i pacchetti prima di poterli trasmettere. La memoria veloce, nel caso delle reti relizzate con fibre ottiche, e' un componente particolarmente costoso, infatti per poter trasmettere correttamente i pacchetti e' necessaria una memoria con una velocita' di lettura-scrittura uguale alla velocita' con la quale l'informazione si propaga sul mezzo.

c. utilizzazione del mezzo trasmissivo

Il costo del mezzo incide pesantemente su i costi di una MAN, pertanto e' possibile offrire servizi a costi competitivi solo garantendo un'elevata utilizzazione del mezzo

L'obiettivo di questo lavoro e' quello di confrontare alcuni MAC protocol per una MAN realizzata utilizzando la configurazione fisica di FASNET.

Il confronto verra' condotto tenendo conto dei seguenti parametri:

* utilizzazione del mezzo trasmissivo

* dimensioni delle aree tampone

* i tempi di attesa che i pacchetti sperimentano prima di poter essere trasmessi (questo parametro assume particolare interesse quando esiste traffico voce o video).

4. EFFICIENZA DEL PROTOCOLLO FASNET IN UNA RETE METROPOLITANA

In questo paragrafo vengono stimati alcuni indici di prestazione relativi al MAC protocol di FASNET quando venga utilizzato in una rete metropolitana.

La stima degli indici di prestazione viene condotta caratterizzando una rete metropolitana con il seguente insieme di parametri:

CAPACITA' DEL MEZZO: 1.2 Gb/sec
VELOCITA' DI PROPAGAZIONE: 250000 Km/sec
NUMERO DI STAZIONI: 50
LUNGHEZZA DELLE LINEE: 100 Km
DISTRIBUZIONE DEGLI ARRIVI: Poisson
LUNGHEZZA FRAME: 1000 bit
STAZIONI EQUIDISTANTI

Nel seguito della trattazione si assume che il CARICO sulla rete, rappresentato dal numero medio di pacchetti/secondo che le stazioni cercano di trasmettere sulla rete, sia formato al 90% da traffico di tipo *intranet* e per il 10% da traffico di tipo *internet*. Si assume, inoltre, che il traffico di internet sia generato uniformemente dalle stazioni, mentre il carico di tipo *intranet* che una stazione trasmette su una data linea e' proporzionale al numero delle stazioni alle quali puo' inviare pacchetti sfruttando tale linea. Pertanto, il carico di tipo *intranet* che la stazione S_i trasmette utilizzando la linea A e' proporzionale a (numero_stazioni - i).

Esempio

Supponiamo che il CARICO sulla rete sia di 600 Mb/sec. Poiche' la dimensione di un frame e' di 1000 bit e poiche' esiste una corrispondenza 1:1 tra frame e pacchetti, possiamo dire che le stazioni della rete trasmettono, in media, complessivamente 600 000 pacchetti/secondo; pertanto, il traffico di tipo *internet* e', in media, di 60 000 pacchetti/secondo, mentre il traffico di *intranet* e', in media, di 540 000 pacchetti/secondo.

Il numero medio di pacchetti/secondo generati da una data stazione S_i e' definito dalla seguente equazione:

$$\lambda_i = [\text{NUMERO_STAZIONI} - i] * k + \alpha$$

α e' il numero medio di pacchetti/secondo di tipo *internet* generati da una stazione. k e' il numero medio di pacchetti/secondo che una stazione S_i trasmette ad un'altra stazione S_j appartenente alla stessa rete metropolitana. α e k sono assunte costanti ed il loro valore viene calcolato risolvendo le seguenti equazioni:

Numero_stazioni

$$\sum_{i=1} (\text{numero_stazioni} - i) * k = \text{traffico di tipo intranet (540 000 pacchetti/secondo)}$$

$$\text{Numero_stazioni} * \alpha = \text{traffico di tipo internet (60 000 pacchetti/secondo)}$$

Assumendo Numero di stazioni = 50 si ottiene: $\alpha=1200$ pacchetti/secondo e $k=440.8$ pacchetti/secondo. Quindi, data la stazione S_1 , il numero medio di pacchetti/secondo che trasmette e' dato da: $\lambda_1 = 50*k + \alpha = 21600 + 1200 = 22800$ pacchetti/secondo.

utilizzo del mezzo

Data una rete di tipo FASNET, l'utilizzazione del mezzo puo' essere definita dal seguente rapporto:

Numero medio di slot utilizzati durante un ciclo

$$\rho = \text{-----}$$

Lunghezza media del ciclo (misurata in slot)

il valore di ρ dipende dal numero di pacchetti/secondo che le stazioni trasmettono; in particolare, mantenendo la distribuzione "proporzionale" del carico tra le stazioni della rete, il valore di ρ cresce all'aumentare del carico fino ad un limite massimo (≤ 1); raggiunto il limite massimo ogni ulteriore aumento del carico lascia inalterato il fattore di utilizzazione.

Il valore massimo del fattore di utilizzo permette di stimare il numero massimo di pacchetti/secondo che possono essere trasmessi e quindi definisce il livello massimo del carico che la rete puo' sopportare. Infatti, livelli di carico superiori a quello corrispondente all'utilizzazione massima portano la rete in saturazione (il numero medio di pacchetti in attesa di trasmissione aumenta nel tempo).

Nel caso delle reti di tipo FASNET, la presenza della fase di sojourn_time mantiene il fattore di utilizzo al di sotto del valore unitario. In particolare, possiamo definire il valore massimo del fattore di utilizzo:

$$\text{MAX}(\rho) = \frac{\text{numero_massimo_di_pacchetti_trasmessi_durante_un_ciclo}}{\text{lunghezza_massima_del_ciclo (misurata in slot)}}$$

Il valore $\text{MAX}(\rho)$ si ottiene assumendo che tutte le stazioni trasmettano durante un ciclo il numero massimo di pacchetti ($P_{\text{max}}(i)$) che sono abilitate ad inviare

$$\text{MAX}(\rho) = \frac{\sum_i P_{\text{max}}(i)}{\sum_i P_{\text{max}}(i) + S_T}$$

4. EFFICIENZA DEL PROTOCOLLO FASNET IN UNA RETE METROPOLITANA

In questo paragrafo vengono stimati alcuni indici di prestazione relativi al MAC protocol di FASNET quando venga utilizzato in una rete metropolitana.

La stima degli indici di prestazione viene condotta caratterizzando una rete metropolitana con il seguente insieme di parametri:

CAPACITA' DEL MEZZO: 1.2 Gb/sec
VELOCITA' DI PROPAGAZIONE: 250000 Km/sec
NUMERO DI STAZIONI: 50
LUNGHEZZA DELLE LINEE: 100 Km
DISTRIBUZIONE DEGLI ARRIVI: Poisson
LUNGHEZZA FRAME: 1000 bit
STAZIONI EQUIDISTANTI

Nel seguito della trattazione si assume che il CARICO sulla rete, rappresentato dal numero medio di pacchetti/secondo che le stazioni cercano di trasmettere sulla rete, sia formato al 90% da traffico di tipo *intranet* e per il 10% da traffico di tipo *internet*. Si assume, inoltre, che il traffico di internet sia generato uniformemente dalle stazioni, mentre il carico di tipo *intranet* che una stazione trasmette su una data linea e' proporzionale al numero delle stazioni alle quali puo' inviare pacchetti sfruttando tale linea. Pertanto, il carico di tipo *intranet* che la stazione S_i trasmette utilizzando la linea A e' proporzionale a (numero_stazioni - i).

Esempio

Supponiamo che il CARICO sulla rete sia di 600 Mb/sec. Poiche' la dimensione di un frame e' di 1000 bit e poiche' esiste una corrispondenza 1:1 tra frame e pacchetti, possiamo dire che le stazioni della rete trasmettono, in media, complessivamente 600 000 pacchetti/secondo; pertanto, il traffico di tipo *internet* e', in media, di 60 000 pacchetti/secondo, mentre il traffico di *intranet* e', in media, di 540 000 pacchetti/secondo.

Il numero medio di pacchetti/secondo generati da una data stazione S_i e' definito dalla seguente equazione:

$$\lambda_i = [\text{NUMERO_STAZIONI} - i] * k + \alpha$$

α e' il numero medio di pacchetti/secondo di tipo *internet* generati da una stazione. k e' il numero medio di pacchetti/secondo che una stazione S_i trasmette ad un'altra stazione S_j appartenente alla stessa rete metropolitana. α e k sono assunte costanti ed il loro valore viene calcolato risolvendo le seguenti equazioni:

Numero_stazioni

$$\sum_{i=1} (\text{numero_stazioni} - i) * k = \text{traffico di tipo intranet (540 000 pacchetti/secondo)}$$

$$\text{Numero_stazioni} * \alpha = \text{traffico di tipo internet (60 000 pacchetti/secondo)}$$

Assumendo Numero di stazioni = 50 si ottiene: $\alpha=1200$ pacchetti/secondo e $k=440.8$ pacchetti/secondo. Quindi, data la stazione S_1 , il numero medio di pacchetti/secondo che trasmette e' dato da: $\lambda_1 = 50*k + \alpha = 21600 + 1200 = 22800$ pacchetti/secondo.

utilizzo del mezzo

Data una rete di tipo FASNET, l'utilizzazione del mezzo puo' essere definita dal seguente rapporto:

Numero medio di slot utilizzati durante un ciclo

$$\rho = \frac{\text{Numero medio di slot utilizzati durante un ciclo}}{\text{Lunghezza media del ciclo (misurata in slot)}}$$

Lunghezza media del ciclo (misurata in slot)

il valore di ρ dipende dal numero di pacchetti/secondo che le stazioni trasmettono; in particolare, mantenendo la distribuzione "proporzionale" del carico tra le stazioni della rete, il valore di ρ cresce all'aumentare del carico fino ad un limite massimo (≤ 1); raggiunto il limite massimo ogni ulteriore aumento del carico lascia inalterato il fattore di utilizzazione.

Il valore massimo del fattore di utilizzo permette di stimare il numero massimo di pacchetti/secondo che possono essere trasmessi e quindi definisce il livello massimo del carico che la rete puo' sopportare. Infatti, livelli di carico superiori a quello corrispondente all'utilizzazione massima portano la rete in saturazione (il numero medio di pacchetti in attesa di trasmissione aumenta nel tempo).

Nel caso delle reti di tipo FASNET, la presenza della fase di sojourn_time mantiene il fattore di utilizzo al di sotto del valore unitario. In particolare, possiamo definire il valore massimo del fattore di utilizzo:

$$\text{MAX}(\rho) = \frac{\text{numero_massimo_di_pacchetti_trasmessi_durante_un_ciclo}}{\text{lunghezza_massima_del_ciclo (misurata in slot)}}$$

Il valore $\text{MAX}(\rho)$ si ottiene assumendo che tutte le stazioni trasmettano durante un ciclo il numero massimo di pacchetti ($P_{\text{max}}(i)$) che sono abilitate ad inviare

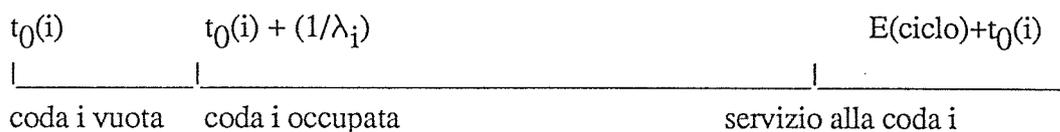
$$\text{MAX}(\rho) = \frac{\sum_i P_{\text{max}}(i)}{\sum_i P_{\text{max}}(i) + S_T}$$

Consideriamo la stazione i -esima, il suo carico in ingresso e' caratterizzato da una sorgente poissoniana di media λ_i (pacchetti al secondo), pertanto il numero medio di pacchetti che arrivano alla stazione i -esima durante un ciclo e' $E(N_i) = E(\text{ciclo}) * \lambda_i$. Se assumiamo che la stazione si trovi in condizioni stazionarie deve essere $E(N_i) \leq P_{\max}(i)$, cioe' il numero di pacchetti in attesa di trasmissione, ad una data stazione, e' in media minore o uguale al numero massimo di pacchetti che la stessa stazione e' abilitata a spedire in un ciclo. Pertanto, possiamo dire che, nel caso medio, quando il servente finisce il suo servizio alla coda i -esima questa e' vuota (cioe' nel caso medio il servizio e' esaustivo); indichiamo con $t_0(i)$ l'istante in cui il servente finisce il suo servizio alla coda i -esima, poiche' la distribuzione dei tempi di interarrivo e' esponenziale (di media $1/\lambda_i$) si ha:

$$\text{Probabilita'}(X > t_0(i) + t | X > t_0(i)) = \text{probabilita'}(X > t)$$

X = Tempo di interarrivo

cioe' la probabilita' che si verifichi un arrivo dopo $t_0(i)$ ha una distribuzione poissoniana di media λ_i , pertanto la coda i -esima rimane vuota in media per un intervallo temporale $1/\lambda_i$ durante ogni ciclo. Graficamente la situazione puo' essere rappresentata nel seguente modo:



durante la fase "coda i occupata" almeno un pacchetto e' in attesa di trasmissione nella coda i , nella realta' fisica il pacchetto non puo' essere tuttavia servito (trasmesso) poiche' il servente "e' vacante", nel modello assumiamo, invece, che durante la fase "coda i occupata" il servente stia erogando un servizio supplementare al primo pacchetto della coda. Con questa assunzione il nostro modello viene ricondotto al classico modello $M/G/1$ [KLEINROCK 75]. Facciamo notare che questa assunzione non altera la stima degli indici di prestazione in esame, infatti i tempi di attesa dei pacchetti e le lunghezze delle code sono le stesse sia assumendo che il servente sia vacante per un certo periodo di tempo, sia assumendo che il servente sia sempre disponibile e fornisca ad un utente un servizio supplementare pari al tempo di "vacation".

In particolare nell'appendice si dimostra che il nostro modello si comporta come una coda $M/D/1$.

Utilizzando le ipotesi fatte possiamo stimare il fattore di utilizzo relativo "all'ipotetico" servente $M/D/1$ associato permanentemente alla coda i -esima¹:

¹Il seguente algoritmo non puo' essere applicato se $\lambda_i * E(\text{ciclo}) \leq 1$, cioe' se il numero medio di utenti che arrivano in un ciclo alla stazione S_i e' minore di uno. Quando si verifica la condizione $\lambda_i * E(\text{ciclo}) < 1$, il valore $\rho^*(i) = [n * E(\text{ciclo}) - 1/\lambda_i] / [n * E(\text{ciclo})]$, dove n e' il piu' piccolo intero che garantisce $n * E(\text{ciclo}) \lambda_i > 1$.

$$\rho^*(i) = \frac{E(\text{ciclo}) - (1/\lambda_i)}{E(\text{ciclo})}$$

$\rho^*(i)$ = utilizzazione del servente associato alla coda i

Applicano il teorema fondamentale della teoria del canale singolo sviluppata da Khintchine e Pollaczek si ha:

$$E(W) = \frac{\rho^2}{2 \cdot (1-\rho)}$$

ρ = utilizzazione del canale

$E(W)$ = numero medio di utenti in attesa di servizio

sostituendo nella formula il valore $\rho^*(i)$ si ottiene la stima del valore $E(W_i)$ e quindi anche $E(Q_i)$ (numero medio di utenti nella stazione S_i) attraverso l'equazione: $E(Q_i) = E(W_i) + \rho^*(i)$. Utilizzando la legge di Little si ricava quindi $E(T_q)$ (tempo medio di attesa di un pacchetto).

Esempio

Con questo esempio vogliamo mostrare come può essere utilizzato il modello presentato per stimare gli indici di prestazione del MAC protocol FASNET. In questa analisi si assume che una MAN sia definita dall'insieme di parametri presentati nel paragrafo 4.0.

Supponiamo che la rete sia sottoposta ad un carico in ingresso di 600 Mb/sec; questo carico garantisce una utilizzazione $\rho = .5$ del mezzo trasmissivo.

Sostituendo i parametri nel modello si ottiene:

$$E(\text{ciclo}) = 1601.66 \mu\text{sec}$$

Vediamo come sia possibile stimare alcuni indici di prestazione relativi alla HS della linea A (stazione S_1). Utilizzando le assunzioni fatte sulla distribuzione del carico si ha:

$$\lambda_1 = 22800 \text{ pacchetti/sec}$$

$1/\lambda_1 = 43.86 \mu\text{sec}$ quindi $\rho^*(1) = .9726$. Applicando la formula di Khintchine e Pollaczek

si ottiene: $E(W_1)=17.26$ e quindi $E(Q_1)= E(W_1) + \rho*(1) = 18.23$. Infine applicando la legge di Little si ottiene la stima dei tempi di attesa: $E(T_q)= 799 \mu\text{sec}$.

A questo punto dell'analisi non abbiamo un termine di paragone per stabilire il grado di efficienza garantito dal protocollo in esame; tuttavia e' possibile notare che, sebbene il carico non sia elevato (infatti, il carico garantisce un fattore di utilizzazione = 0.5) il tempo di attesa dei pacchetti e' circa 1000 volte superiore rispetto al tempo di trasmissione ($0.83\mu\text{sec}$)

Prima di passare ad analizzare MAC protocol alternativi rispetto a quello di FASNET, presentiamo la stima di alcuni indici di prestazione relativi al protocollo di FASNET ottenuti risolvendo tramite la simulazione un modello realizzato con reti di code del MAC protocol di FASNET. La presentazione di questi risultati ha una duplice funzione:

a. permette di verificare, nei casi esaminati, quale precisione garantisce il modello analitico presentato

b. fornisce ulteriori informazioni sul MAC protocol di FASNET; in particolare informazioni sulla distribuzione dei tempi di attesa e sulla distribuzione della lunghezza delle code.

4.1.2 Indici di prestazione relativi al MAC protocol FASNET ottenuti attraverso la simulazione

Utilizzando il package RESQ2 [SAUER 82] e' stato costruito e risolto attraverso la simulazione un modello a reti di code del MAC protocol di FASNET [CONTI 88]. Gli indici di prestazione sono stati stimati al 90% del livello di confidenza; i risultati ottenuti sono riassunti nella tabella 1

Dai risultati presentati in tabella 1 risulta evidente che utilizzando valori elevati per il parametro P_{max} e' possibile garantire un'utilizzazione elevata del mezzo trasmissivo, ma al crescere dell'utilizzazione i tempi medi di attesa crescono in maniera "esponenziale", come risulta dalle figure 4.1 e 4.2.

Nella prima figura viene presentato l'andamento della distribuzione dei tempi di attesa dei pacchetti ("*access delay* ") definita attraverso il suo valor medio, il 95-th, 99-th percentile e dal massimo tempo di attesa sperimentato da un pacchetto. Nella figura 4.2 si confronta l'andamento del tempo medio di attesa ($E(T_q)$) con il grado di utilizzazione del mezzo (dal primo grafo e' possibile notare che gli altri parametri che caratterizzano la distribuzione del tempo di attesa hanno lo stesso andamento del valor medio al variare del fattore di utilizzazione).

Indici di prestazione	CARICO	%	CARICO	%	CARICO	%	CARICO	%
	600	10.5	830	8.1	980	17	1010	17
E(Tq)	808	3.1	1303	6.3	2151	9	2328	
E(Q)	485	2.3	1110	8.7	2276	6.5	2552	
MAX(Tq)	1701		2872		5200		6227	
MAX(Q)	55		102		208		249	
99-th Tq	1600	1.5	2800	0.3	4900	2.0	5500	4.2
95-th Tq	1500	5.4	2500	4.1	4400	5.0	4900	10.7

i tempi presenti nella tabella sono espressi in microsecondi

% e' l'ampiezza percentuale dell'intervallo di confidenza

E(Q) = numero complessivo di pacchetti in attesa

MAX(Q) = lunghezza massima di una coda

CARICO e' espresso in Mb/sec

TABELLA 1

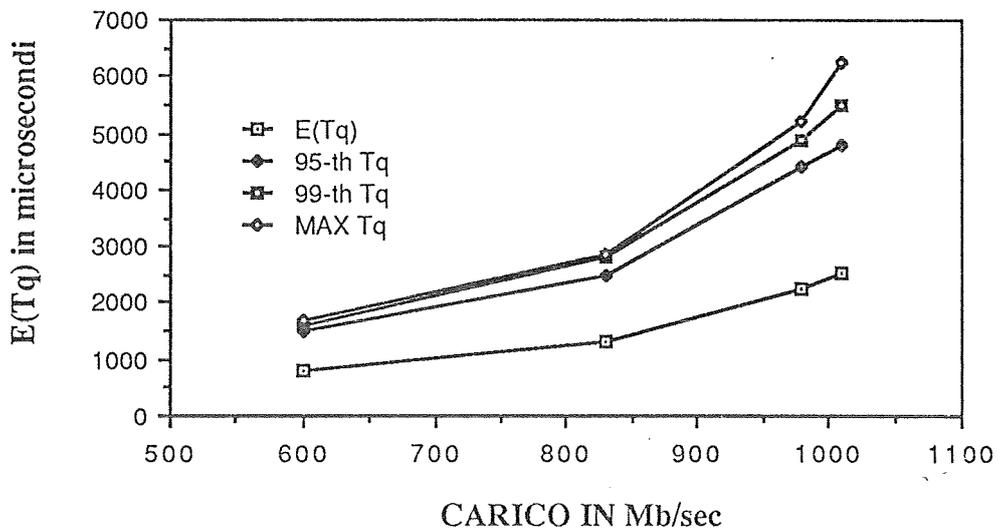


Figura 4.1: andamento della distribuzione dei tempi di attesa in funzione del carico

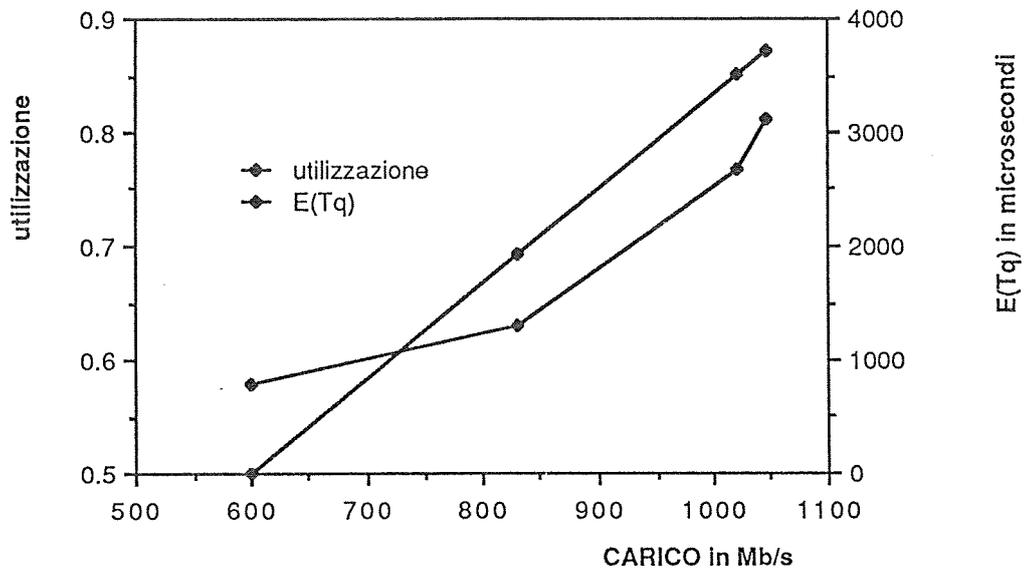


Figura 4.2: Confronto tra l'andamento del fattore di utilizzazione del mezzo trasmissivo e l'andamento del tempo medio di attesa ($E(Tq)$) al variare del carico

Nelle tabelle 2, 3 e 4 sono confrontate le stime di alcuni indici di prestazione ottenute utilizzando il modello analitico presentato con le misure degli stessi indici ottenute tramite simulazione. I tempi riportati nelle tabelle sono espressi in microsecondi.

Dai risultati presentati nelle tabelle risulta evidente che, nei casi esaminati, il modello analitico proposto fornisce stime assai accurate.

CARICO 597 Mb/sec		CONFRONTO ANALISI SIMULAZIONE									
STAZIONI	S(1)	S(10)	S(20)	S(25)	S(30)	S(40)	S(50)				
ANALISI	22686	18738	14352	12159	9966	5580	1194				
ρ*	.9723	.9665	.9562	.9484	.9370	.8875	.4744				
E(Q)	18.06	14.90	11.41	9.66	7.91	4.39	0.68				
E(Tq)	796	795.9	795.3	794.7	793.7	786.8	576.8				
Inf E(Q)	13.61	12.46	9.41	8.28	7.57	3.95	0.19				
Med E(Q)	16.48	13.36	11.03	9.30	7.87	4.52	0.58				
Max E(Q)	19.35	14.79	12.63	10.32	8.17	5.08	0.97				
Inf E(Tq)	596	660	692	688	728	709	201				
Med E(Tq)	696	727	755	758	770	815	435				
Max E(Tq)	796	793	818	827	827	922	663				

TABELLA 2

S: RISULTATI OTTENUTI DALLA SIMULAZIONE

Inf : estremo inferiore dell'intervallo di confidenza

Med : valore centrale dell'intervallo di confidenza

Sup : estremo superiore dell'intervallo di confidenza

CARICO 830 Mb/sec		CONFRONTO ANALISI SIMULAZIONE									
STAZIONI	S(1)	S(10)	S(20)	S(25)	S(30)	S(40)	S(50)				
A	31576	26083	19978	16925	13872	7767	1662				
N	.988	.9856	.9812	.9778	.9723	.95	.77				
A	42.14	34.80	26.65	22.58	18.50	10.34	2.1				
L	1334.6	1334.5	1334.2	1334	1333	1331	1263				
I	41.25	31.79	24.97	21.61	16.61	8.54	1.85				
S	42.87	34.84	26.76	22.47	18.97	9.99	2.46				
I	44.49	37.89	28.56	23.32	21.33	11.43	3.07				
S	1261	1267	1277	1269	1274	1248	1041				
S	1303	1321	1317	1324	1336	1297	1206				
S	1346	1375	1356	1380	1397	1347	1371				

TABELLA 3

S: RISULTATI OTTENUTI DALLA SIMULAZIONE

Inf : estremo inferiore dell'intervallo di confidenza

Med : valore centrale dell'intervallo di confidenza

Sup : estremo superiore dell'intervallo di confidenza

CARICO 930 Mb/sec		CONFRONTO ANALISI SIMULAZIONE									
STAZIONI	S(1)	S(10)	S(20)	S(25)	S(30)	S(40)	S(50)				
A	35340	29190	22356	21810	15525	8702	1860				
N											
A	.9920	.9903	.9874	.9871	.9819	.9677	.8489				
L											
I	62.88	51.90	39.78	38.80	27.62	15.47	3.23				
S											
I	1779	1779	1779	1779	1779	1777	1739				
Inf E(Q)	42.77	39.51	32.69	34.65	23.94	16.36	2.36				
Med E(Q)	54.31	46.28	39.16	36.51	29.71	17.43	3.23				
Max E(Q)	65.85	53.05	45.64	38.37	35.48	18.49	4.11				
Inf E(Tq)	1596	1660	1692	1688	1728	1709	1201				
Med E(Tq)	1696	1727	1755	1758	1770	1815	1435				
Max E(Tq)	1796	1793	1818	1827	1827	1922	1663				

TABELLA 4

S: RISULTATI OTTENUTI DALLA SIMULAZIONE

Inf : estremo inferiore dell'intervallo di confidenza

Med : valore centrale dell'intervallo di confidenza

Sup : estremo superiore dell'intervallo di confidenza

5.0 ANALISI DELLE PRESTAZIONI DI MAC PROTOCOL PER MAN DI TIPO FASNET

Abbiamo visto che l'estensione del MAC protocol della LAN FASNET quale protocollo di controllo dell'accesso al mezzo per una MAN non permette un'efficiente utilizzazione del mezzo a causa dell'overhead rappresentato dalla propagazione dell'informazione di controllo. Nel paragrafo precedente abbiamo visto come, incrementando il valore del parametro P_{max} , sia possibile ridurre il costo % dovuto al sojourn time. Questo approccio richiede tuttavia che la lunghezza media del ciclo venga incrementata notevolmente, cio' comporta un aumento almeno proporzionale dei tempi medi di attesa che i pacchetti sperimentano nelle stazioni e del numero medio di pacchetti in attesa di trasmissione (cio' si ripercuote negativamente sul dimensionamento della memoria veloce).

In questo paragrafo vengono presentati ed analizzati due MAC protocol che cercano di garantire un'utilizzazione massima del mezzo trasmissivo senza penalizzare gli altri indici di prestazione: metodo **I** e metodo **distribuito**.

Il metodo **I** utilizza il MAC di FASNET e cerca di migliorarne l'efficienza utilizzando anche la fase di sojourn_time per la trasmissione dei pacchetti.

Il metodo **distribuito** [LIMB 82] rappresenta un superamento della logica del MAC protocol di FASNET; infatti, mentre nel protocollo di FASNET e' la HS che decide, utilizzando informazioni sullo stato della rete (END=1), l'inizio di un nuovo ciclo, in questo metodo ogni stazione, quando ha pacchetti da trasmettere, analizzando lo stato della rete decide se e' iniziato un nuovo ciclo ed inizia quindi a trasmettere. Pertanto nel metodo distribuito non esiste una fase di sojourn_time.

5.1 Metodo **I**: fine ciclo segnalata da una stazione intermedia

L'idea su cui si basa questo metodo e' l'invio anticipato del segnale END=1. Una stazione S_i se conosce l'esatto ammontare del numero di slot necessari, nel ciclo attualmente in corso, per la trasmissione dei pacchetti da parte delle stazioni che devono ancora trasmettere (relativamente alla linea A: $S_{i+1}, S_{i+2}, \dots, S_{ES}$) puo' stimare se, inviando lei il segnale END=1, il numero di slot generati dalla HS prima che le arrivi il segnale di fine ciclo e' sufficiente a soddisfare le richieste delle stazioni che devono ancora trasmettere. Pertanto, supponendo che i requisiti delle singole stazioni possano essere conosciuti, o almeno stimati, e' possibile che il segnale END=1 possa essere inviato anticipatamente senza che cio' alteri il comportamento equo del MAC protocol. Per chiarire il funzionamento di questo metodo utilizziamo le seguenti definizioni:

$$ES \\ MAX_REQUEST(I) = \sum_{I+1} S_{req}(i)$$

$S_{req}(i)$ = numero slot richiesti dalla stazione S_i

allora \underline{I} e' il piu' piccolo indice tale che:

$$\Delta_{\underline{I}1} = > \text{MAX_REQUEST}(\underline{I})$$

$\Delta_{\underline{I}1}$: e' il numero di slot a disposizione delle stazioni che seguono $S_{\underline{I}}$ generati dalla HS prima di vedere il segnale END=1 lanciato dalla stazione $S_{\underline{I}}$.

Rimane adesso da stabilire come stimare $S_{req}(i)$. La soluzione ottima, ma non realizzabile in pratica, e' rappresentata dalla conoscenza delle richieste delle stazioni in ogni ciclo. Una soluzione facilmente realizzabile sfrutta invece la conoscenza dei valori dei parametri P_{max} associati alle stazioni, e ipotizza $S_{req}(i) = P_{max}(i)$ (cioe' si stimano le richieste di ogni stazione con il limite superiore). In questo caso il valore di \underline{I} e' fissato staticamente e non viene piu' modificato a meno che non vengano modificati i valori di P_{max} . Nel seguito dell'analisi utilizzeremo questa ipotesi, facciamo tuttavia notare che utilizzando il metodo \underline{I} con politiche dove la conoscenza delle richieste viene aggiornata quando il numero degli slot richiesti differisce dalle ipotesi formulate ($\sum S_{req}(i)$) oltre una certa soglia, il valore di \underline{I} varia dinamicamente ed e' possibile ottenere prestazioni migliori rispetto al caso statico.

5.1.1 Un modello analitico per valutare gli indici di prestazioni del MAC protocol del metodo \underline{I}

Il modello che prendiamo in esame e' un modello a *server vacation* uguale al modello utilizzato per il MAC protocol di FASNET. I due modelli si differenziano quando si passa ad analizzare il calcolo del tempo di ciclo, nel caso del metodo \underline{I} il tempo di ciclo e' calcolato considerando una rete ridotta che contiene solo le prime \underline{I} stazioni. Infatti, la fase di trasmissione delle stazioni $S_{\underline{I}+1}, \dots, S_{ES}$ avviene parallelamente alla propagazione del segnale END=1 lanciato da $S_{\underline{I}}$ e la propagazione del segnale termina dopo che tutte le stazioni ($S_{\underline{I}+1}, \dots, S_{ES}$) sono gia' state servite (purché \underline{I} sia stato stimato correttamente).

Il tempo medio di ciclo per questo modello puo' essere stimato:

$E(\text{ciclo}) =$ Trasmissione dei pacchetti da parte delle prime \underline{I} stazioni +
+ Soujourn_time di una rete che contenga solo le prime \underline{I} stazioni

$$E(\text{ciclo}) = \sum_1^{\underline{I}} P_{max}(i) * T_s * \rho^{\wedge} + \Delta_{\underline{I}1} * T_s$$

$$\Delta I_1 = 2 * \frac{I-1}{NUM_ST-1} * \frac{LUNGHEZZA}{VEL} * \frac{1}{Ts}$$

VEL: velocita' di propagazione del segnale sul mezzo (in Km/sec)

LUNGHEZZA: e' la lunghezza del mezzo (in Km)

NUM_ST: e' il numero delle stazioni

Ts: e' la durata di uno slot in secondi

NOTA: Nel calcolo di ΔI_1 si assume che le stazioni siano equidistanti, allora

$$\frac{I-1}{NUM_ST-1} * \frac{LUNGHEZZA}{VEL} = \text{distanza (in secondi) tra due stazioni}$$

Il termine ρ^{\wedge} rappresenta il fattore di utilizzazione della banda massima disponibile, in un ciclo, da parte di una singola stazione. Questo fattore viene assunto uguale per ogni singola stazione in quanto negli esempi che prendiamo in considerazione la banda massima a disposizione di una stazione (cioe' il valore del suo Pmax) e' proporzionale al carico assegnato alla stazione. Il valore di ρ^{\wedge} viene ricavato dalla seguente equazione:

$$\frac{1}{E(\text{ciclo})} * \left[\sum_1 P_{\max} * \rho^{\wedge} \right] = \text{CARICO pacchetti/sec}$$

il primo termine rappresenta il numero di cicli effettuati nell'unita' di tempo (un secondo); il secondo termine rappresenta il numero medio di pacchetti spediti in un ciclo.

Noto il valore di ρ^{\wedge} si ricava il valore di E(ciclo). Disponendo del valore di E(ciclo) possiamo utilizzare la stessa tecnica utilizzata per stimare gli indici di prestazione del MAC protocol di FASNET.

5.1.2 Stima degli indici di prestazione relativi al MAC protocol del metodo I

La stima degli indici di prestazione viene condotta supponendo che la rete sia sottoposta ad un carico di 600 Mb/sec ripartito "proporzionalmente" (si veda il paragrafo 4.1.1) tra le stazioni della rete (si tratta delle stesse ipotesi formulate per stimare gli indici relativi al MAC protocol FASNET). Assumiamo inoltre che il numero massimo di pacchetti che possono essere spediti in un ciclo ($\sum P_{\max}$) sia 2500. Il valore del P_{\max} relativo ad una specifica stazione viene calcolato assumendo che sia proporzionale al carico assegnato alla stazione. Da queste ipotesi si

ricava che la stazione con la funzione di inviare END=1 anticipatamente e' S₂₉ (cioe' I=29).
 Applicando il modello presentato si ottiene:

$$E(\text{ciclo}) = (\sum_{i=1}^I P_{\max}) * T_s * \rho^{\wedge} + \Delta_{I1} * T_s$$

Ts: tempo di trasmissione di un pacchetto = .83μsec

$$\Delta_{I1} * T_s = 457\mu\text{sec}$$

$$\sum_{i=1}^I P_{\max(i)} = 2009$$

Introducendo i valori numerici nell'espressione del tempo medio di ciclo, si ottiene E(ciclo) in funzione del solo valore di ρ[^]. Risolvendo l'equazione che lega il numero di cicli fatti nell'unita' di tempo e il numero di pacchetti trasmessi in media in un ciclo al CARICO si ottiene ρ[^]=.1826.

Consideriamo, adesso la stazione S₁, e stimiamo gli indici relativi a questa stazione. Dalle ipotesi fatte si ricava:

$$\lambda_1 = 22800 \text{ pacchetti/sec}$$

$$1/\lambda_1 = 43.28\mu\text{sec}$$

$$E(\text{ciclo}) = 761.9\mu\text{sec}$$

$$\rho^*(1) = \frac{E(\text{ciclo}) - (1/\lambda_1)}{E(\text{ciclo})} = .9424$$

Applicando prima la formula di Khintchine e Pollaczek e quindi la legge di Little si ottengono gli indici di prestazione cercati:

$$E(W_1) = 7.72 \text{ pacchetti}$$

$$E(Q_1) = E(W_1) + \rho^*(1) = 8.66 \text{ pacchetti}$$

$$E(T_q) = 379.8\mu\text{sec}$$

Osservazioni

Confrontando gli indici di prestazione del MAC protocol di FASNET con quelli relativi al MAC protocol del metodo \mathbf{I} si ottiene che:

$$E(Q_1)_F \cong 2 * E(Q_1)_I$$

$$E(Tq_1)_F \cong 2 * E(Tq_1)_I$$

F = FASNET; \mathbf{I} = metodo \mathbf{I}

Risulta pertanto evidente che il MAC protocol relativo al metodo \mathbf{I} fornisce prestazioni migliori rispetto al MAC protocol di FASNET quando venga utilizzato su una MAN.

Vediamo, adesso una generalizzazione della relazione esistente tra gli indici di prestazione relativi al MAC protocol FASNET e quelli relativi al MAC protocol \mathbf{I} .

Abbiamo visto che data la distribuzione scelta dei valori dei P_{\max} il valore $\mathbf{I} = 29$. Assumendo che le stazioni siano alla stessa distanza si ottiene che il Sojourn-time per il MAC protocol \mathbf{I} e' \leq .57 volte il Sojourn_time del MAC protocol FASNET (ricordiamo che durante l'invio del segnale END=1 esistono delle stazioni abilitate a trasmettere). Indicando con S_T il valore del Sojourn_Time (espresso come: "numero di slot vuoti in un ciclo") relativo al MAC protocol FASNET ed assumendo che il fattore di utilizzazione del mezzo sia ρ , possiamo scrivere nel caso del MAC protocol \mathbf{I} :

$$\rho = \frac{\sum_1 P_{\max}(i) * \rho^{\wedge}}{(\sum_1 P_{\max}) * \rho^{\wedge} + \Delta_{\mathbf{I}} 1} = \frac{2500 * \rho^{\wedge}}{2009 * \rho^{\wedge} + .57 * S_T} = \frac{2500 * \rho^{\wedge}}{2500 * \rho^{\wedge} + X * S_T} = \rho_{\mathbf{I}}$$

dove: $2500 * \rho^{\wedge}$ sono gli slot occupati all'interno del ciclo

$X * S_T$ ($X \leq .57$) e' il numero di slot vuoti all'interno di un ciclo

$\rho_{\mathbf{I}}$ = fattore di utilizzo relativo al MAC \mathbf{I}

Nel caso del MAC protocol FASNET, assumendo lo stesso fattore di utilizzo si ha:

$$\rho = \frac{\sum_1 P_{\max}(i) * \rho^{\prime\prime}}{(\sum_1 P_{\max}) * \rho^{\prime\prime} + \Delta_{\mathbf{I}} 1} = \frac{2500 * \rho^{\prime\prime}}{2500 * \rho^{\prime\prime} + S_T} = \rho_{\mathbf{F}}$$

ρ_F = fattore di utilizzo del MAC protocol FASNET

Allora possiamo concludere che deve essere: $\rho^{\wedge} = X * \rho''$.

Infatti, se $\rho^{\wedge} = X * \rho''$ entrambi i protocolli garantiscono lo stesso utilizzo del mezzo:

$$\rho_I = \frac{2500 * \rho^{\wedge}}{2500 * \rho^{\wedge} + X * S_T} = \frac{2500 * X * \rho''}{2500 * X * \rho'' + X * S_T} = \frac{2500 * \rho''}{2500 * \rho'' + S_T} = \rho_F$$

Supponiamo, invece, che esista $Y \neq X$ tale che $\rho^{\wedge} = Y * \rho''$, allora:

$$\rho_I = \frac{2500 * \rho^{\wedge}}{2500 * \rho^{\wedge} + X * S_T} = \frac{2500 * Y * \rho''}{2500 * Y * \rho'' + X * S_T} = \frac{2500 * \rho''}{2500 * \rho'' + (X/Y) * S_T}$$

poiche':

$$\rho_F = \frac{2500 * \rho''}{2500 * \rho'' + S_T}$$

si ottiene: $\rho_I \neq \rho_F$. Poiche' il risultato ottenuto viola l'ipotesi iniziale che l'utilizzazione del mezzo sia la stessa (e cioe' sia uguale il carico sulla rete) concludiamo che **non esiste** $Y \neq X$ tale che $(\rho^{\wedge} = Y * \rho'')$ \Rightarrow $\rho_I = \rho_F$.

Utilizzando il fatto che $\rho^{\wedge} = X * \rho''$ e $X \leq 0.57 * S_T$ si ottiene facilmente che:

$$\begin{aligned} E(\text{ciclo})_I &= X * E(\text{ciclo})_F \\ E(\text{ciclo})_I &\leq 0.57 * E(\text{ciclo})_F \end{aligned}$$

dove $F = \text{FASNET}$ e $I = \text{metodo } I$;

Analizziamo, adesso, una generica stazione S_i della rete alla quale e' associato un carico caratterizzato dal parametro λ_i . Applicando il modello analitico presentato si ottiene:

$$\rho^*(i) = 1 - \frac{1}{\lambda_i * E(\text{ciclo})}$$

$$E(Q_i) = \frac{1}{2} * \{ \lambda_i * E(\text{ciclo}) - \frac{1}{\lambda_i * E(\text{ciclo})} \}$$

Indichiamo con E(ciclo) la lunghezza media del ciclo relativa al MAC FASNET, allora possiamo indicare la lunghezza media del ciclo relativa al MAC I: E(ciclo)*X. Si ottiene così:

$$E(Q_i)_F = \frac{1}{2} * \{ \lambda_i * E(\text{ciclo}) - \frac{1}{\lambda_i * E(\text{ciclo})} \}$$

$$E(Q_i)_I = \frac{1}{2} * \{ \lambda_i * E(\text{ciclo}) * X - \frac{1}{\lambda_i * E(\text{ciclo}) * X} \}$$

indichiamo $a = \lambda_i * E(\text{ciclo})$; si ottiene quindi:

$$2 * E(Q_i)_F = a - 1/a \qquad 2 * E(Q_i)_I = a * X - 1/(a * X)$$

$$2 * E(Q_i)_F * X = a * X - X/a$$

poiché $X/a < 1/(a * X)$ (ricordiamo, infatti, che $X < 1$) si ottiene:

$$E(Q_i)_I < X * E(Q_i)_F \qquad [1]$$

$$E(Q_i)_I < .57 * E(Q_i)_F$$

Vediamo, adesso, come sia possibile definire il valore di X in funzione di ρ . Infatti,

$$\rho = \frac{2500 * \rho^{\wedge}}{2009 * \rho^{\wedge} + .57 * S_T} = \frac{2500 * \rho^{\wedge}}{2500 * \rho^{\wedge} + (.57 * S_T - 491 * \rho^{\wedge})}$$

dopo alcuni passaggi algebrici si ottiene:

$$\rho^{\wedge} = \frac{\rho * .57 * S_T}{2500 * (1 - \rho) + 491 * \rho}$$

assumendo $\rho = .5$ si ottiene $\rho^{\wedge} = .1838$ e quindi: $X = (.57 * S_T - 491 * \rho^{\wedge}) / S_T = .4765$.

Sostituendo il valore di X nelle formule che definiscono $E(Q_i)$ si ottiene:

$$E(Q_i)_I \leq 0.4765 * E(Q_i)_F$$

Conoscendo il valore degli indici di prestazione relativi al MAC FASNET e' possibile ricavare, utilizzando la relazione [1], una buona approssimazione dell'indice corrispondente del MAC I.

Ad esempio, con $\rho=0.5$ si ottiene: $E(Q_i)_I \leq 0.4765 * E(Q_i)_F = 8.68$, mentre nel paragrafo precedente, utilizzando il modello analitico proposto, abbiamo ottenuto che per $\rho=0.5$: $E(Q_1)_F = 18.23$ e $E(Q_1)_I = 8.66$.

Dall'analisi presentata possiamo concludere che l'utilizzo del MAC protocol I introduce un miglioramento nei valori medi degli indici di prestazione (lunghezze delle code e tempi di attesa) $> = (1-X)$ (cioe' gli indici di prestazione sono inferiori di un fattore $(1-X)$) dove $X = (0.57 * S_T - 491 * \rho^2) / S_T$.

Sotto condizioni di basso carico in cui si ha $\rho \rightarrow 0$ il miglioramento e' $\geq 43\%$; sotto condizioni di carico elevato in cui si ha $\rho \rightarrow 1$ (e' una situazione raggiungibile solo in teoria) e conseguentemente il miglioramento degli indici di prestazione e' $\geq 94\%$.

Nella tabella sono riportati i risultati ottenuti risolvendo con la simulazione modelli dei due protocolli realizzati con reti di code.

Gli esperimenti sono stati condotti assumendo:

NUM_ST

$$\sum_{i=1} P_{\max}(i) = 2500$$

$i=1$

Carico ripartito "proporzionalmente"

Valore " $P_{\max}(i)$ " proporzionale a λ_i

CARICO	600 Mb/sec		865 Mb/sec	
Indici	$E(Tq)$	$E(Q)$	$E(Tq)$	$E(Q)$
FASNET	806 [799;813]	480 [472; 489]	1563 [1556;1570]	1354 [1335;1374]
I	384 [381; 386]	229 [226; 233]	550 [542; 558]	476 [466; 486]

i tempi riportati nella tabella sono espressi in microsecondi

$E(Q)$ numero medio di pacchetti in attesa di essere spediti

Analizzando la tabella si ricava:

1. dato un Carico di 600 Mb/sec $E(Tq)_I \cong .47 * (E(Tq)_F)$ $E(Q)_I \cong .47 * (E(Q)_F)$
2. dato un Carico di 865 Mb/sec $E(Tq)_I \cong .36 * (E(Tq)_F)$ $E(Q)_I \cong .36 * (E(Q)_F)$

Utilizzando il modello analitico si ricava che:

1. dato un carico = 600Mb/sec ($\rho=.5$) $\Rightarrow (\rho^{\wedge}=.1838) \Rightarrow X=.47$
2. dato un Carico di 865 Mb/sec ($\rho=.72$) $\Rightarrow (\rho^{\wedge}=.3758) \Rightarrow X=.37 \Rightarrow E(Q)_I < X * (E(Q)_F)$

Nel seguito del lavoro il MAC protocol FASNET non verra' piu' preso in considerazione in quanto dall'analisi presentata abbiamo visto che il MAC protocol **I** fornisce prestazioni migliori.

5.2 Metodo distribuito

Questo metodo prevede che nel campo di controllo degli slot venga aggiunto un bit: REQ; in base al valore di questo bit le stazioni determinano l'inizio del ciclo. Per illustrare le modalita' con le quali le stazioni accedono al mezzo prendiamo in esame una generica stazione S_i ; supponiamo che nella coda della stazione S_i arrivi un pacchetto che deve essere trasmesso sulla linea A. Prima di iniziare la trasmissione la stazione S_i deve attendere l'inizio di un nuovo ciclo; un nuovo ciclo, per la stazione S_i inizia quando controllando il bit REQ degli slot che transitano sulla linea B trova il valore del bit =0. Dal momento dell'arrivo del pacchetto e fino a quando non ha completato il suo ciclo, la stazione S_i setta REQ=1 in tutti gli slot che vede passare sulla linea B per inibire l'inizio di nuovi cicli da parte delle stazioni comprese tra S_i e la HS. Il ciclo della stazione S_i termina dopo aver trasmesso $P_{max}(i)$ o non appena la sua coda e' vuota. La trasmissione dei pacchetti avviene (come nel caso del MAC protocol di FASNET) utilizzando gli slot liberi (BUSY=0) presenti sulla linea A

5.2.1 Modelli analitici per la stima degli indici di prestazione relativi al MAC protocol distribuito

In questo paragrafo vengono presentati due modelli analitici (Modello_1 e Modello_2) che permettono di stimare gli indici di prestazione del MAC protocol distribuito. La necessita' di utilizzare due modelli e' dovuta alla esistenza di una **soglia di carico** che modifica l'insieme di parametri rilevanti per stimare gli indici di prestazione. Infatti, per valori del carico al di sotto della soglia la correlazione tra i tempi di attesa dei pacchetti nelle diverse stazioni e' ridotta e puo' essere espressa attraverso due probabilita': probabilita' che una stazione abilitata a trasmettere trovi gli slot occupati dalle stazioni che la precedono, e la probabilita' che una stazione in attesa

dell'inizio di un ciclo sia inibita da stazioni di indice piu' elevato (cioe' comprese tra lei e la ES). Fissate queste probabilita' che definiscono la correlazione esistente tra le stazioni e' possibile studiare isolatamente ogni stazione.

Quando il carico supera invece la soglia i tempi di attesa dei pacchetti delle stazioni sono molto correlati e pertanto e' necessario studiare il sistema (la rete) nella sua globalita'.

L'esistenza della soglia non viene dimostrata in maniera formale, ma e' il risultato di una serie di osservazioni supportate da stime ottenute attraverso la simulazione.

Osservazioni

Consideriamo una stazione S_i , abilitata a trasmettere. Questa stazione puo' non essere in grado di trasmettere subito e consecutivamente tutti i pacchetti in attesa per una delle seguenti cause.

1. Gli slot sono stati occupati dalle stazioni $S_{HS}, S_{HS+1}, \dots, S_{i-1}$, attivate precedentemente rispetto a S_i (cioe' da un bit $REQ=0$ precedente rispetto a quello che ha attivato S_i). Questo fatto dipende dalla casualita' con la quale si interlacciano gli arrivi alle varie stazioni e genera una correlazione piuttosto limitata (come vedremo in seguito) tra i tempi di accesso al mezzo delle varie stazioni.

2. La stazione S_i quando viene abilitata a trasmettere riesce a trasmettere consecutivamente Δ pacchetti² (Δ e' due volte il numero di slot che si trovano sul mezzo fisico tra la stazione S_i e la stazione S_{i-1}), dopo di che si sospende poiche' anche la stazione S_{i-1} e' stata abilitata (dallo stesso bit $REQ=0$ che ha abilitato S_i) ed ha incominciato a trasmettere. Questo fatto si verifica quando il carico sulla rete e' tale che in media esistono stazioni che ogni volta che sono abilitate a trasmettere hanno in coda un numero di pacchetti maggiore di Δ . Quando esiste una stazione S_i che, in media, ha in attesa di trasmissione un numero di pacchetti maggiore di Δ , allora in media tutte le stazioni di indice inferiore ad i verificano questa condizione³. La presenza di piu' stazioni adiacenti che trasmettono in media piu' di Δ pacchetti per ciclo causa la presenza, sul mezzo di trasmissione, di "treni" di slot occupati. La presenza di sequenze di slot occupati stabilisce una forte correlazione tra gli accessi delle stazioni al mezzo, infatti si verifica che:

a. le stazioni vicine alla ES, o comunque le stazioni di indice superiore ad i , quando vengono

² Al fine di semplificare la trattazione si assume che le due cause non si presentino contemporaneamente, in realta', quando si verifica la causa 2. puo' verificarsi anche la 1.

³ Infatti, per il suo particolare funzionamento questo metodo fa si che date due stazioni S_i e S_j si verifica che se $i < j$, in media, la stazione S_i sara' abilitata a trasmettere un numero minore di volte rispetto alla stazione S_j . Infatti, ogni volta che la stazione S_i viene abilitata anche la stazione S_j viene abilitata purché abbia pacchetti da trasmettere. Inoltre, nel caso in esame il carico e' ripartito proporzionalmente tra le stazioni per cui si ha: $\lambda_i > \lambda_j$. Pertanto, in media, ogni volta che S_i e' abilitata avra' un numero di pacchetti da trasmettere \Rightarrow al numero di pacchetti in attesa di trasmissione alla stazione S_j .

abilitate possono rimanere in attesa di trasmettere perche' non trovano slot liberi . In media queste stazioni dovranno attendere un tempo pari alla meta' della durata media di un *treno* di slot occupati.

b. le stazioni di indice inferiore ad i ogni volta che vengono abilitate hanno in coda un numero di pacchetti maggiore di Δ , e pertanto la stazione S_i dopo aver trasmesso Δ pacchetti viene "inibita" dalla stazione S_{i-1} , questa a sua volta dopo aver trasmesso Δ pacchetti viene inibita da S_{i-2} , quando, infine, viene abilitata S_1 questa stazione riesce a trasmettere tutti i pacchetti consecutivamente, dopo di che anche S_2 e' in grado di terminare la trasmissione e quindi procedendo in ordine inverso tutte le stazioni fino ad S_i riescono a completare la trasmissione.

Risulta, pertanto, evidente che tra le prime i stazioni si instaura una forte correlazione negli accessi al mezzo trasmissivo.

c. durante la fase in cui le prime i stazioni stanno trasmettendo i loro pacchetti, le stazioni di indice superiore ad i si abilitano per la trasmissione dei pacchetti. Questa fase di abilitazione puo' riguardare solo le stazioni di indice $> i$, infatti durante questa fase la stazione S_i e' in attesa di trasmissione e non permette che le stazioni che la precedono (cioe' di indice inferiore) possano abilitarsi. Appena termina la fase di trasmissione per le prime i stazioni, incomincia un nuovo ciclo per le stazioni di indice superiore ad i ; la prima stazione a trasmettere sara' S_{i+1} quindi S_{i+2} ed ultima S_{ES}^4 . Fino al momento in cui S_{ES} non termina la trasmissione dei pacchetti nessuna stazione puo' essere abilitata alla trasmissione (infatti S_{ES} pone $REQ=1$ in tutti gli slot). Quando S_{ES} termina la sua trasmissione le stazioni, a partire da S_{ES-1} possono iniziare un nuovo ciclo. In media tutte le stazioni fino ad S_{i-1} riescono a trasmettere tutti i pacchetti prima di trovare gli slot occupati dalle stazioni di indice inferiore. Le stazioni di indice compreso tra i e 1 si bloccheranno l'un l'altra come gia' descritto in precedenza.

Pertanto, quando il carico supera una certa soglia la rete, in condizioni stazionarie, avra' un punto di rigenerazione rappresentato dallo stato in cui viene abilitata la stazione S_i . Lo studio dei valori medi degli indici di prestazione potra', quindi, essere condotto analizzando la rete in un intervallo compreso tra due successivi punti di rigenerazione.

Nel seguito dell'analisi chiameremo ciclo l'intervallo compreso tra due punti di rigenerazione.

Conclusioni

Da quanto evidenziato nei punti precedenti, possiamo dire che nel caso del metodo distribuito esiste una soglia di carico ad di sotto della quale non esiste una stazione S_i che verifichi la condizione: $P_{\max}(i) * \rho^{\wedge} > \Delta$ (dove ρ^{\wedge} e' l'utilizzazione media della banda massima disponibile ad ogni ciclo). In questo caso la correlazione tra gli accessi al mezzo delle stazioni e' determinata solo dalla causa 1.; sotto queste condizioni lo studio degli indici di prestazione del MAC protocol risulta semplificato e puo' essere affrontato attraverso il Modello_1 che

⁴ Vedremo in seguito che in media, data la particolare ripartizione del carico scelta, e' ragionevole assumere che tutte le stazioni di indice maggiore ad i siano abilitate a trasmettere quando termina il ciclo per le prime i stazioni.

Le stazioni di indice maggiore di I possono trasmettere i loro pacchetti sia durante la fase (3) (cioe' non appena la stazione S_I termina di trasmettere), sia durante la fase (1) (cioe' quando, dopo che tutte stazioni hanno trasmesso, viene inviato un bit $REQ=0$ che le abilita tutte nuovamente).

Facciamo notare che nella discussione precedente avevamo indicato il punto di rigenerazione con l'istante di abilitazione della stazione S_I (rappresentato in figura dal simbolo \$); per la nostra trattazione e' piu' semplice utilizzare come punto di rigenerazione l'istante in cui viene inviato lo slot con $REQ=0$ che abilita, per il ciclo in esame, tutte le stazioni (in base alle osservazioni fatte questo bit e' unico all'interno del ciclo), pertanto nella rappresentazione grafica abbiamo fatto coincidere l'inizio del ciclo con l'inizio della fase (1).

Utilizzando la rappresentazione grafica del ciclo e le osservazioni fatte, siamo in grado di determinare il valore della soglia.

Quando il carico raggiunge il livello soglia si ha: $P_{max}(1) * \rho^{\wedge} = \Delta$ ⁶. Possiamo, allora, dire che se il livello di carico \rightarrow soglia non esiste una stazione S_i tale che $P_{max}(i) * \rho^{\wedge} > \Delta$ e quindi le fasi (2) e (3) non esistono⁷.

Per calcolare il livello soglia e' sufficiente osservare la stazione S_1 e determinare il valore ρ^{\wedge} per cui $P_{max}(i) * \rho^{\wedge} = \Delta$. Poiche' in un ciclo tutte le stazioni trasmettono in media $P_{max}(i) * \rho^{\wedge}$ pacchetti e' possibile risalire alla lunghezza media di un ciclo e da questa al livello di carico corrispondente (come rapporto tra parte del ciclo effettivamente utilizzata per trasmettere e la lunghezza complessiva del ciclo).

Esempio

Supponiamo che $\sum P_{max} = 2500$ e assumiamo che $P_{max}(i)$ sia proporzionale al carico assegnato alla stazione. Vogliamo definire il livello di carico relativo alla soglia.

In base alle assunzioni fatte si ha $P_{max}(1) = 95$ e quindi:

$$P_{max}(1) * \rho^{\wedge} \leq \Delta \rightarrow \rho^{\wedge} \leq .207$$

Quando il carico e' al di sotto della soglia il livello del carico e' dato da:

⁶ Infatti, nel nostro studio si assume una distribuzione del carico proporzionale a (numero_stazioni-i) e pertanto la stazione S_1 e' quella con il piu' elevato tasso di pacchetti da trasmettere.

⁷ La fase (2) non esiste perche' non esiste l'indice I . La fase (3) non esiste perche' durante questa fase vengono trasmessi i pacchetti arrivati durante la fase (2).

$$\rho = \frac{\sum P_{\max} * \rho^{\wedge}}{\sum^I P_{\max} * \rho^{\wedge} + \alpha(I)}$$

poiche', in questo caso, non esiste una stazione S_I , poniamo $I=0$.

$$\alpha(0) = (\text{numero_stazioni} - 0) * \Delta = 50 * 19.67$$

$$\rho = \frac{\sum P_{\max} * \rho^{\wedge}}{\alpha(0)} = .526$$

Possiamo, quindi, fissare il valore della soglia al 52% della banda massima (1.2 Gb/sec).

5.2.1.1 Il Modello_1

Questo modello viene utilizzato quando il carico e' inferiore alla soglia.

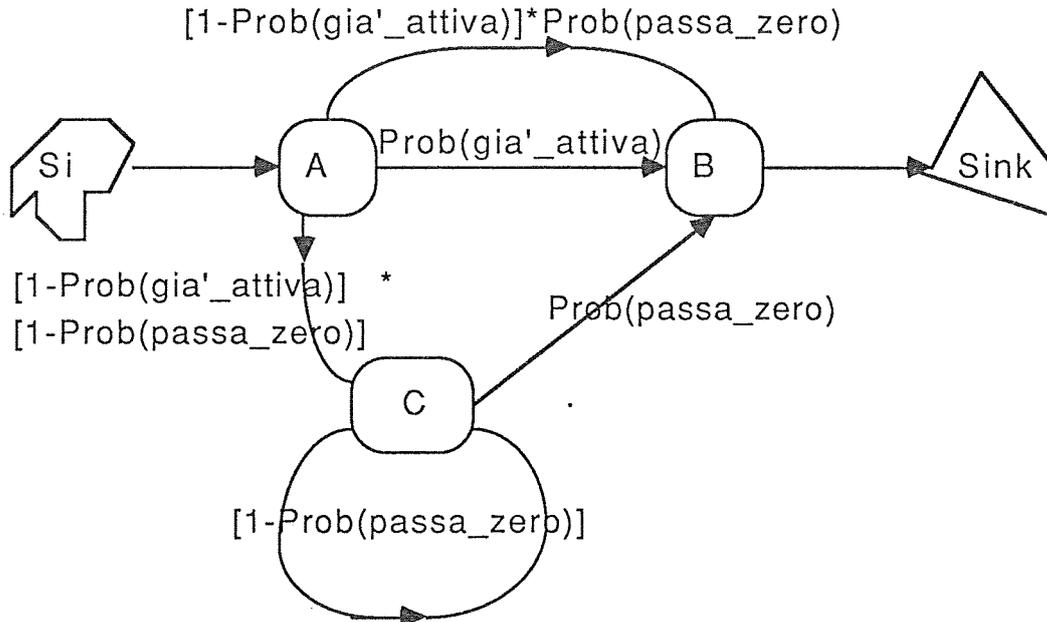
Con questo modello vengono stimati gli indici di prestazione di singole stazioni, rappresentando la correlazione con le altre stazioni attraverso due parametri:

1. La probabilita' che la stazione S_i trovi, nello slot che sta' esaminando, il bit REQ=1. Questo parametro rappresenta la correlazione tra la stazione S_i e le stazioni di indice piu' alto (cioe' le stazioni comprese tra la ES e S_i). Infatti queste stazioni determinano il valore del bit REQ "visto" da S_i .

2. La probabilita' che la stazione S_i , abilitata a trasmettere, trovi libero lo slot che sta' attualmente esaminando sulla linea A. Questo parametro definisce la correlazione esistente tra la stazione S_i e le stazioni di indice piu' basso (le stazioni comprese tra la HS e la stazione in esame). Infatti queste stazioni determinano lo stato degli slot (valore del bit BUSY) che vengono esaminati da S_i .

Attraverso questi parametri vengono rappresentate tutte le dipendenze. Ogni singola stazione viene poi studiata attraverso la risoluzione analitica di un modello realizzato con una rete di code

della classe BCMP [BASKETT 75]. La presentazione del modello viene fatta utilizzando la rappresentazione grafica in Figura.



La sorgente S_i genera i pacchetti in arrivo alla stazione con una distribuzione poissoniana di media λ_i .

Il server A ha un tempo di servizio deterministico $T_{sA}=.42\text{msec}$, cioè metà della durata di uno slot. La presenza di questo server è utilizzata per sincronizzare i pacchetti in arrivo con il passaggio degli slot. La coda associata a questo server è gestita con politica *infinite server* (IS).

Dopo aver ottenuto il servizio al nodo A un pacchetto viene inviato al server B (che gestisce la trasmissione) se la stazione in esame è abilitata a trasmettere. Una stazione è abilitata a trasmettere se si trova già in questo stato quando arriva il pacchetto ($\text{Prob}(\text{gia}'_{\text{attiva}})$), oppure se tra l'istante di arrivo del pacchetto e la fine del servizio al nodo A è passato uno slot con $\text{REQ}=0$ ($\text{Prob}(\text{passa_zero})$).

Quando, invece, la stazione è nello stato "non attiva" il pacchetto, ottenuto il servizio al nodo A, viene accodato al server C in attesa che la stazione S_i venga abilitata dal bit $\text{REQ}=0$. Il tempo di servizio al server C è deterministico ed è uguale alla durata di uno slot ($.83\mu\text{sec}$); dopo aver ottenuto il servizio al server C un pacchetto può essere accodato per la trasmissione con probabilità $\text{Prob}(\text{passa } 0)$, oppure rimane, con probabilità $(1-\text{Prob}(\text{passa } 0))$, al server C dove attende il passaggio dello slot successivo. La coda del server C viene gestita con politica IS.

Il server B gestisce la trasmissione dei pacchetti. Il tempo di servizio al server B ha un valore pari alla durata di un numero_i (dove i è l'indice della stazione) intero di slot; questo numero comprende lo slot utilizzato per la trasmissione del pacchetto, più il numero medio di slot che una stazione trova occupati prima di poter trasmettere un pacchetto. Pertanto numero_i ha una distribuzione geometrica di parametro p_i (dove p_i è la probabilità che uno slot che passa davanti alla stazione S_i sia occupato), quindi si ha: $E(\text{numero}_i)=1/(1-p_i)$. Resta da determinare il tipo di disciplina con la quale gestire la coda del server. La scelta più ovvia prevede l'utilizzo

della disciplina **FIFO**, tuttavia questo implica che per ottenere un modello della classe BCMP i tempi di servizio della stazione B abbiano distribuzione **esponenziale**. Nella realta', invece, i tempi sono deterministici e quindi l'utilizzo di tempi di servizio esponenziali comporta una stima pessimistica degli indici di prestazione. Una possibile scelta alternativa prevede l'utilizzo di una disciplina **IS** e distribuzione geometrica dei tempi di servizio. Utilizzando questa seconda alternativa otteniamo un modello che fornisce stime ottimistiche, in quanto non tiene conto del fatto che piu' pacchetti siano contemporaneamente pronti per la trasmissione. Facciamo, tuttavia, notare che il modello in esame viene usato quando l'utilizzazione del mezzo e' inferiore ad una certa soglia (e il carico e' distribuito tra tutte le stazioni), e quindi la probabilita' che piu' pacchetti siano in contemporanea in attesa di trasmissione ad una data stazione e' ridotta.

Nel modello che presentiamo sono state utilizzate entrambe le ipotesi, al fine di ottenere un limite superiore ed un limite inferiore per gli indici di prestazione in esame.

Vediamo come vengono stimate le probabilita' presenti nel modello esaminando le probabilita' relative ad una generica stazione S_i .

Definizioni

λ_{Tot} = numero totale di pacchetti/sec. che arrivano alle stazioni della rete

λ_i = numero di pacchetti/sec. in arrivo alla stazione S_i

p_i = Probabilita' slot occupato in transito davanti alla stazione S_i

ρ = Livello di carico

$$p_i = \left[\left(\sum_{j=1}^{i-1} \lambda_j \right) / \lambda_{Tot} \right] * \rho$$

t_i = tempo (misurato in slot) in cui la stazione S_i e' occupata in un secondo

$$t_i = \lambda_i * \text{numero}_i$$

infatti numero_i e' il numero medio di slot che e' necessario esaminare prima di poter inviare un pacchetto, mentre λ_i e' il numero di pacchetti che la stazione spedisce in un secondo; da queste definizioni ricaviamo la probabilita' che la stazione S_i sia gia' attiva quando arriva un pacchetto.

$$\text{Prob}(S_i \text{ gia' attiva}) = t_i / (\text{numero di slot in un secondo})$$

Otteniamo, infine, una stima delle probabilita' che una stazione trovi REQ=0 in un dato slot. La probabilita' che una stazione S_i trovi REQ=0 e' definita in funzione dello stato delle stazioni che si trovano tra la stazione in esame e la ES:

$$\text{Probabilita' (ES vede REQ=0)} = 1$$

$$\text{Probabilita' (ES-1 vede REQ=0)} = 1 - \text{Probabilita' (ES gia' attiva)}$$

$$\text{Probabilita' (ES-2 vede REQ=0)} = \text{Probabilita' (ES-1 vede REQ=0)} - \text{Probabilita' (ES-1 gia' attiva)}$$

.

.

$$\text{Probabilita' (S}_1 \text{ vede REQ=0)} = \text{Probabilita' (2 vede REQ=0)} - \text{Probabilita' (2 gia' attiva)}$$

Vediamo, adesso, di confrontare, su uno specifico esempio, i risultati forniti dal modello analitico con le stime ottenute con la simulazione.

Esempio

Il confronto tra i risultati ottenuti analiticamente e quelli forniti dalla simulazione viene effettuato supponendo che il carico in ingresso alla rete sia di 600 Mb/sec e che sia ripartito "proporzionalmente" tra le stazioni.

Nella Figura 5.1 e' rappresentato il confronto tra analisi e simulazione relativamente ai tempi medi di attesa dei pacchetti alle stazioni ($E(Tq)$); I risultati analitici sono stati ottenuti assumendo che il server B abbia disciplina FCFS e distribuzione esponenziale dei tempi di servizio; La curva relativa ai risultati ottenuti tramite simulazione rappresenta per ogni stazione il valore centrale dell'intervallo di confidenza stimato al 90% del livello di confidenza.

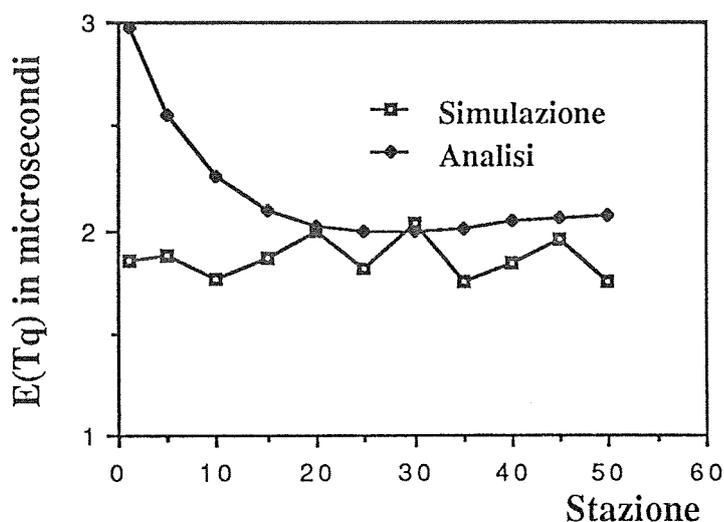


Figura 5.1

Il confronto evidenzia che le stime ottenute analiticamente forniscono un limite superiore di $E(Tq)$. Questo avviene poiché nel modello analitico la probabilità "di trovare slot occupati" e la probabilità "di vedere uno slot con $REQ=0$ " sono costanti (e quindi indipendenti dalla storia passata), mentre nella realtà variano.

Nella Figura 5.2 viene rappresentato un caso limite della dipendenza delle probabilità dalla storia passata: la probabilità "di trovare uno slot occupato" diviene 0 dopo aver trovato uno slot occupato.

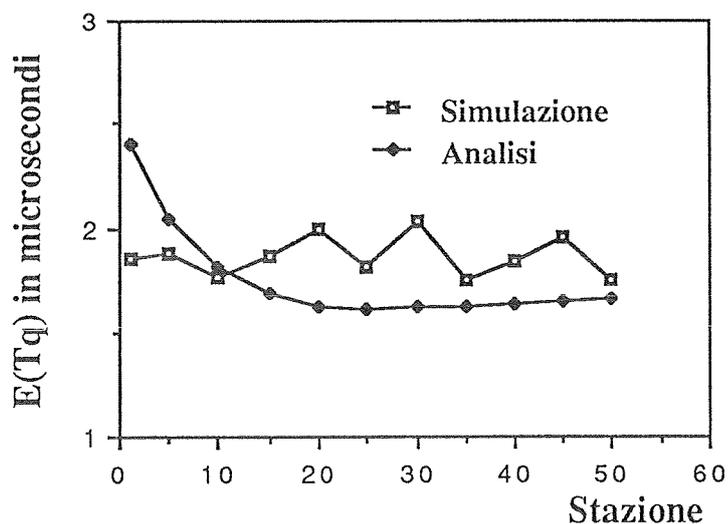


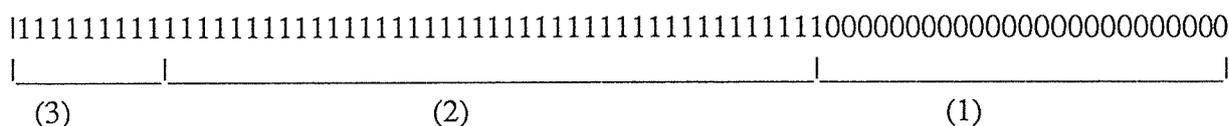
Figura 5.2

Dalla Figura 5.2 risulta che facendo variare dinamicamente la probabilità "slot occupato" si ottiene, nel caso limite, un limite inferiore di $E(Tq_i)$ per le stazioni di indice maggiore di $i=10$, mentre per le stazioni di indice inferiore a 10 i risultati analitici forniscono un limite superiore;

sia durante la fase (1) per un numero complessivo (relativo a ciascuna stazione) di $P_{\max(i)} \cdot \rho^{\wedge}$. La stima del numero di pacchetti trasmessi in ciascuna fase risulta assai complesso in quanto esiste correlazione tra i valori delle variabili ρ^{\wedge} , I , $E(\text{ciclo})$, $\text{lunghezza_fase}(i)$. Per semplificare la trattazione prendiamo, invece, in esame i casi estremi: il **limite superiore** e il **limite inferiore** della lunghezza del ciclo. Utilizzando questi valori saremo poi in grado di stimare un limite inferiore ed uno superiore per ciascuno degli indici di prestazione in esame.

Stima del limite superiore della lunghezza del ciclo

Il limite superiore viene calcolato assumendo che le stazioni di indice superiore ad I trasmettano tutti i loro pacchetti durante la fase (3). In base a questa assunzione la rappresentazione del ciclo diviene quella riportata in Figura



(1) $\Delta \cdot (\text{numero_stazioni} - I)$ slot

(2) Slot occupati dalle stazioni S_1, \dots, S_I

(2) Slot occupati dalle stazioni S_{I+1}, \dots, S_{ES}

Il limite superiore del ciclo può, quindi, essere stimato con la seguente formula:

Limite superiore: $\sum P_{\max} \cdot \rho^{\wedge} + (50 - I) \cdot \Delta$

Rimane, adesso, da calcolare il valore di I in funzione di ρ^{\wedge} . I e ρ^{\wedge} vengono calcolati risolvendo il seguente insieme di equazioni.

$$1. \frac{\sum P_{\max} \cdot \rho^{\wedge}}{\sum P_{\max} \cdot \rho^{\wedge} + \alpha(i)} = \rho$$

ρ : e' la percentuale di CARICO (o anche l'utilizzazione del mezzo)

Il limite inferiore del ciclo puo', quindi, essere stimato con la seguente formula:

$$\text{Limite inferiore: } \frac{I}{\sum P_{\max} * \rho^{\wedge} + (50-I)*\Delta}$$

Rimane, adesso, da calcolare il valore di I in funzione di ρ^{\wedge} . I e ρ^{\wedge} vengono calcolati risolvendo un sistema di equazioni che differisce solo nella equazione 1. da quello utilizzato per calcolare il limite superiore:

$$I' \frac{\sum P_{\max} * \rho^{\wedge}}{\sum P_{\max}(I) * \rho^{\wedge} + \alpha(I)} = \rho$$

Esempio

Supponiamo di avere una rete con un carico tale da garantire una utilizzazione $\rho=.9$. Risolvendo le equazioni presentate si ottiene: $\rho^{\wedge}=.50$, $i = 31$; sostituendo questi valori nell'espressione che definisce il limite inferiore della lunghezza ciclo si ottiene:

$$\text{Limite inferiore: } = 1419 \text{ slot} = 1178 \mu\text{sec}$$

Stima degli indici di prestazione relativi alle singole stazioni

In questa sezione prendiamo in esame la stima degli indici di prestazione relativi alla S_{ES} e S_{HS} quando il MAC protocol e' definito dal metodo **distribuito**. In questa analisi si assume che il carico garantisca un'utilizzazione $\rho=.9$.

Prendiamo in esame per prima la stazione S_{ES} e vediamo di calcolare il limite superiore ed il inferiore dei suoi indici di prestazione: tempo medio di attesa (compreso il tempo di trasmissione) dei pacchetti nella stazione ($E(T_Q)$) e numero medio di pacchetti presenti nella stazione ($E(Q)$).

Calcolo dei limiti superiori degli indici di prestazione

Si esamina il caso in cui la lunghezza del ciclo e' stimata con il suo limite superiore.

Riprendiamo in esame la rappresentazione grafica del ciclo evidenziando le fasi in cui la stazione S_{ES} e' in grado di trasmettere⁸:

CARICO 990Mb/sec CONFRONTO TRA I RISULTATI ANALITICI E LA SIMULAZIONE			
STAZIONE	Stazione_1	Stazione_25	Stazione_50
Analisi	471 <= E(Tq) <= 674	254.6 <= E(Tq) <= 421	40.5 <= E(Tq) <= 287
Simulazione	[471; 645]	[402 ; 533]	[] (*)

TABELLA 5

[a; b] e' l'intervallo di confidenza stimato con la simulazione
 (*) intervallo di confidenza di ampiezza pari al 400% e quindi non significativo; questo si verifica poiche' la stazione S(50) ha un carico ridotto rispetto alle altre e quindi, durante la simulazione, vengono effettuate poche osservazioni relative a questa stazione

CARICO 865 Mb/sec CONFRONTO TRA I RISULTATI ANALITICI E LA SIMULAZIONE			
STAZIONE	Stazione_1	Stazione_25	Stazione_50
Analisi	406 <= E(Tq) <= 579	188.7 <= E(Tq) <= 260	4 <= E(Tq) <= 248
Simulazione	[340; 361]	[298; 300]	[] (*)

TABELLA 6

Dall'analisi delle tabelle risulta che le stime fornite dal modello analitico sono vicine a quelle fornite dalla simulazione, ed inoltre attraverso il modello analitico siamo in grado di fornire le stime degli indici di prestazione relativi alle stazioni meno cariche per le quali la simulazione puo' fornire stime ragionevoli (cioe' intervalli di confidenza "significativi") solo con "runs" particolarmente lunghi e quindi costosi in termini di risorse.

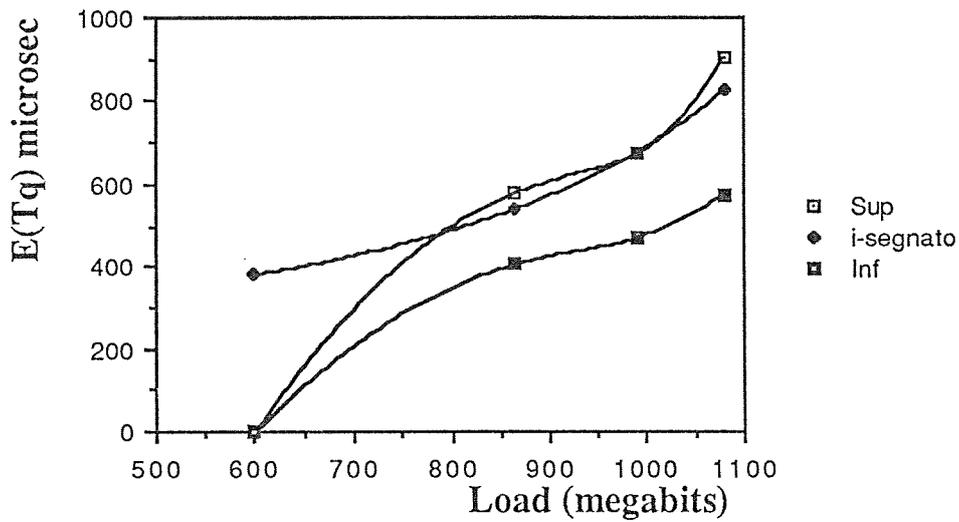
6.0 CONFRONTO DEI MAC PROTOCOL I E DISTRIBUITO

In questo paragrafo vengono messi a confronto gli indici di prestazione relativi al MAC protocol I e al MAC protocol distribuito. Non viene invece preso in esame il MAC protocol di

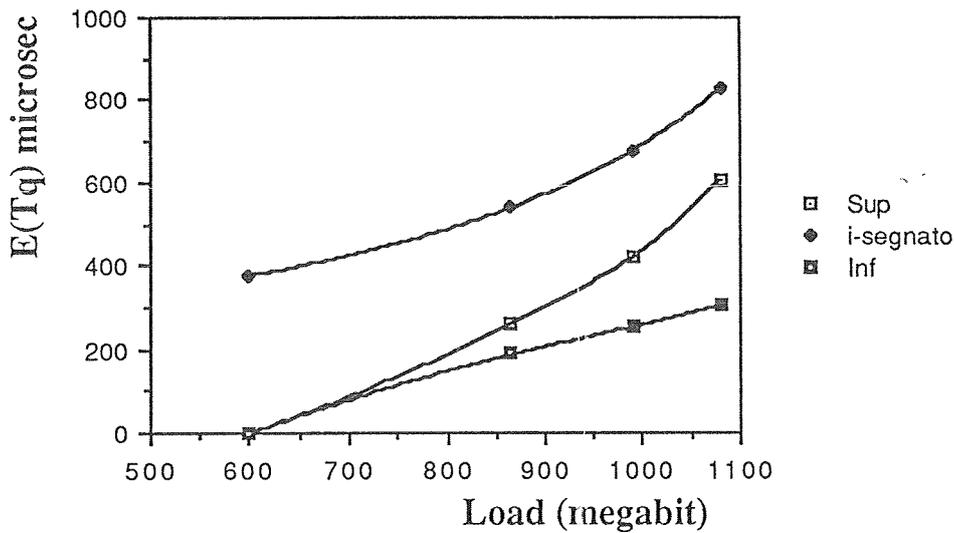
FASNET in quanto, come abbiamo visto nel paragrafo 5.1.2, gli indici di prestazione relativi a questo protocollo sono sempre peggiori degli indici corrispondenti del MAC protocol I.

Il confronto tra i MAC protocol viene effettuato utilizzando le stime degli indici di prestazione, relative alle stazioni S_1 , S_{25} e S_{50} , ottenute con i modelli analitici presentati. I risultati del confronto sono sintetizzati nelle seguenti figure.

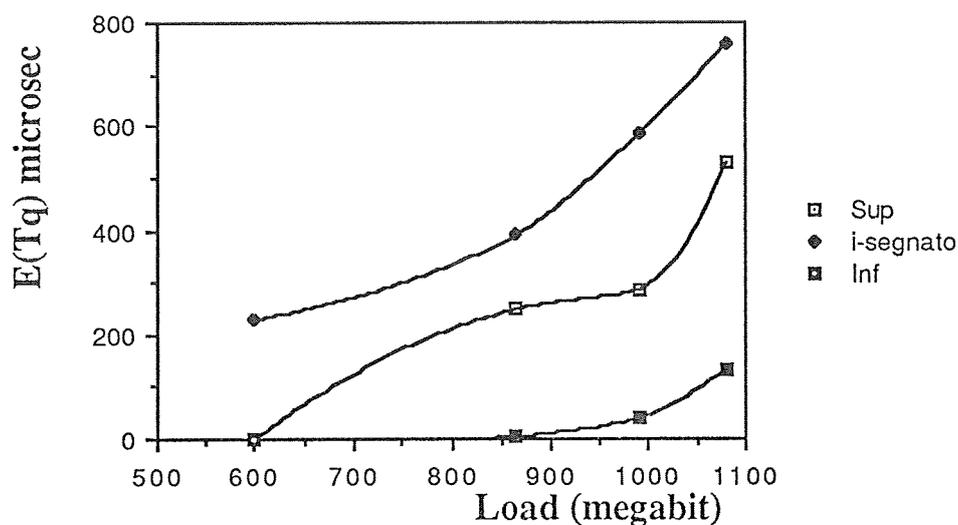
Le tre curve in figura rappresentano rispettivamente l'estremo superiore (Sup) ed inferiore (Inf) del tempo di attesa relativo al MAC distribuito e il tempo medio di attesa (i-segnato) relativo al MAC I.



Stime relative alla stazione S_1



Stime relative alla stazione S_{25}



Stime relative alla stazione S₅₀

Osservazioni

Analizzando le figure possiamo ricavare le seguenti osservazioni.

1. Sotto condizioni di carico "medio-basse" (600 Mb/sec) il MAC protocol **distribuito** fornisce prestazioni molto migliori rispetto al MAC protocol **I**. Utilizzando il MAC protocol **distribuito** otteniamo tempi di attesa e lunghezze delle code ≈ 200 volte più piccole rispetto ai corrispondenti indici di prestazione relativi al MAC protocol **I**.
2. All'aumentare del carico si riduce la distanza tra gli indici di prestazione del metodo **distribuito** e gli indici corrispondenti del metodo **I**.
3. Gli indici di prestazione relativi al MAC protocol **I** sono pressoché uguali per tutte le stazioni; nel caso del MAC protocol **distribuito** il valore degli indici di prestazione è funzione non solo del livello di carico ma anche della posizione della stazione sul mezzo. Possiamo quindi affermare che il protocollo **distribuito** non ha un comportamento equo nel confronto delle stazioni. In particolare il MAC protocol **distribuito** favorisce le stazioni vicine alla ES; in generale le stazioni vicine alla HS sono quelle che richiedono maggiori risorse (hanno in media un numero maggiore di pacchetti da trasmettere) e pertanto questo aspetto del protocollo **distribuito** è opposto a quello desiderato.

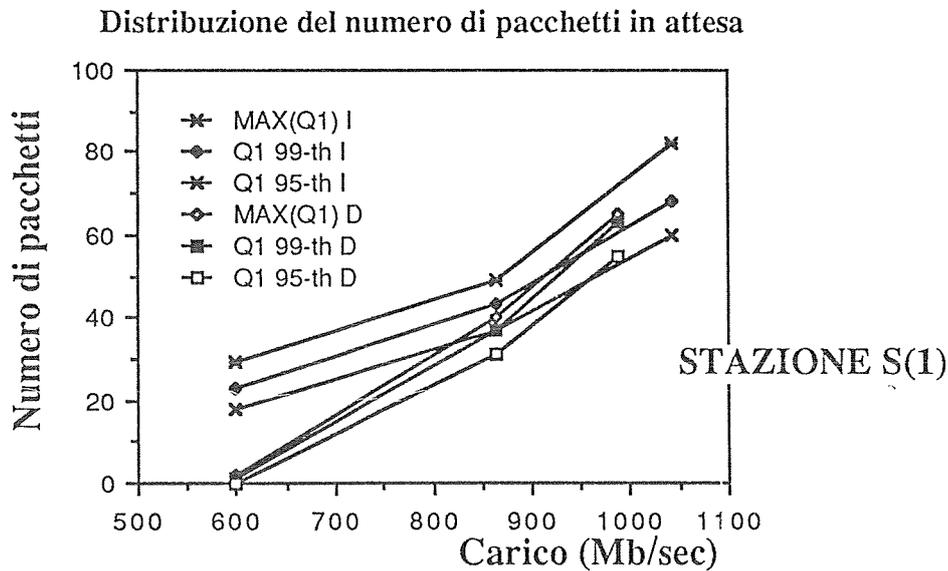
Possiamo quindi affermare che tra i MAC protocol analizzati, il MAC protocol FASNET offre prestazioni inefficienti, mentre il MAC protocol **distribuito** presenta i valori migliori per quanto

riguarda gli indici di prestazione esaminati. Tuttavia, per livelli di carico elevati, le prestazioni del MAC protocol relativo al metodo **I** sono prossime a quelle del MAC protocol distribuito, ed inoltre il MAC protocol **I** garantisce un comportamento equo.

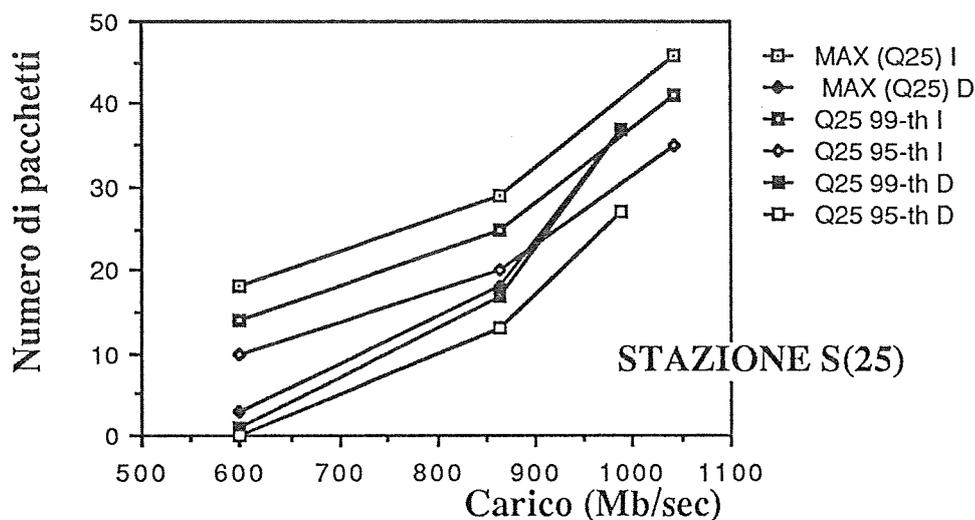
Utilizzando la simulazione e' stato, quindi, verificato che passando ad esaminare le distribuzioni degli indici di prestazione relativi ai due MAC protocol esaminati si mantiene lo stesso andamento osservato confrontando i MAC protocol in base al valore medio del tempo di attesa ($E(Tq)$).

Nelle figure 6.1, 6.2 e 6.3, costruite utilizzando i risultati della simulazione, sono confrontati il massimo (MAX) il 99-th e il 95-th percentile della distribuzione del numero di pacchetti in attesa di trasmissione (Q_i) relativi al MAC protocol **I** con i valori corrispondenti relativi al MAC protocol distribuito.

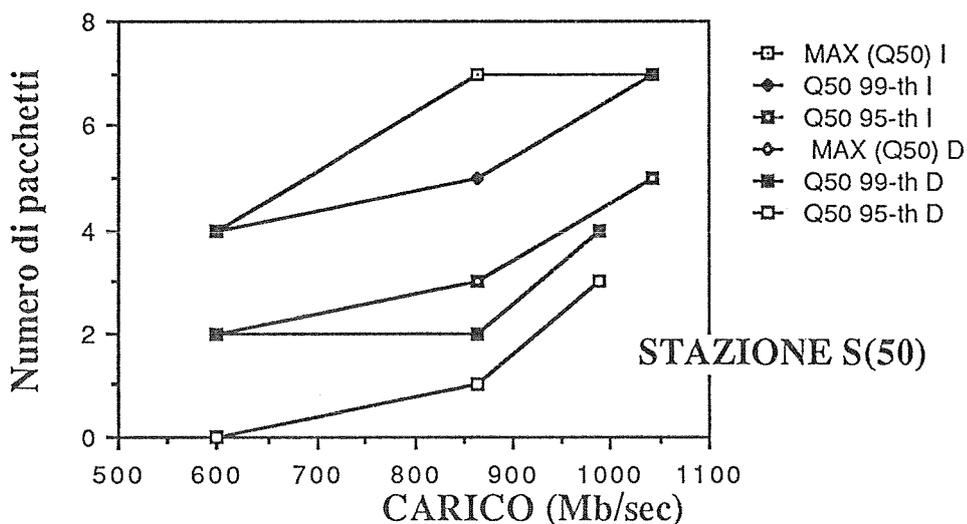
Dall'analisi dei grafici risultano confermate tutte le osservazioni formulate utilizzando i valori medi degli indici di prestazione ottenuti analiticamente.



Distribuzione del numero di pacchetti in attesa



Distribuzione del numero di pacchetti in coda



NOTA

Nel paragrafo 5.2, parlando del MAC protocol relativo al metodo distribuito, abbiamo introdotto il concetto di **soglia di carico**.

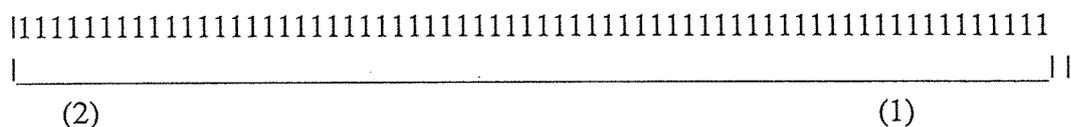
Sotto l'ipotesi di ripartire il carico proporzionalmente tra le stazioni e fissato il valore medio del parametro $P_{max}=50$, abbiamo trovato che il livello soglia $e' \cong 52\%$ della banda nominale.

I risultati ottenuti, sia analiticamente, sia attraverso la simulazione, sembrano confermare le

ipotesi fatte; infatti, per valori di carico inferiori al livello soglia (ad esempio, per un livello di carico del 50%) si ottiene che gli accessi al mezzo delle stazioni sono poco correlati (in media il tempo che una stazione deve attendere prima di poter trasmettere un pacchetto puo' essere espresso solo in funzione del livello di carico¹⁰), mentre per valori del carico maggiori del livello soglia la correlazione tra le stazioni diviene elevata.

Al fine di ottenere un'ulteriore conferma alla nostra ipotesi e' stato effettuato un secondo tipo di esperimento: si e' assunto che il carico sia ripartito uniformemente tra le stazioni e a tutte le stazioni e' stato assegnato lo stesso valore $P_{max}=50$.

Sotto queste ipotesi e' possibile dimostrare che il livello soglia $\rightarrow \rho=1$. Infatti, nelle condizioni fissate, il comportamento delle stazioni e' in media uguale, pertanto se esiste una stazione S_I tale che $P_{max}(I) * \rho^I \geq \Delta$, allora questa condizione e' verificata da tutte le stazioni, pertanto, un ciclo (cioe' l'intervallo compreso tra due punti di rigenerazione) ha una rappresentazione come quella in figura.



$$(1) \Delta * (\text{numero_stazioni} - I) \text{ slot} = 0 \text{ slot}$$

$$(2) \text{ Slot occupati dalle stazioni } S_1, \dots, S_I$$

Questa situazione puo' presentarsi solo se $\rho=1$.

Risolviendo il modello relativo al metodo distribuito, sotto le ipotesi di carico uguale a 1080Mb/sec (cioe' 90% della banda massima) ripartito uniformemente tra le stazioni sono stati ottenuti i seguenti risultati:

$$\text{Tempo medio di attesa } (E(Tq)) = [200\mu\text{sec}; 283\mu\text{sec}] 242\mu\text{sec}$$

$$\begin{aligned} \text{Numero complessivo di pacchetti in attesa di trasmissione} = \\ = [214 \text{ pacchetti}; 302 \text{ pacchetti}] 258 \text{ pacchetti} \end{aligned}$$

dovei valori nell'espressione "[a; b] c" rappresentano rispettivamente l'estremo inferiore ("a"), l'estremo superiore ("b") e il valore centrale ("c") dell'intervallo di confidenza stimato al 90% del livello di confidenza.

Nel caso invece di carico ripartito "proporzionalmente" tra le stazioni e quindi con soglia di carico al 52% della banda massima si ottiene quanto segue.

¹⁰ Infatti possiamo associare all'evento "trasmissione di un pacchetto" una distribuzione geometrica di parametro ρ (e' il livello di carico) che indica la probabilita' che una stazione che ha un pacchetto da trasmettere trovi uno slot occupato. Il tempo medio di trasmissione diviene quindi $[1/(1-\rho)] * .83\mu\text{sec} + .42\mu\text{sec} = 1.98\mu\text{sec}$.

CARICO 1000 Mb/sec (.825 della banda massima)

Tempo medio di attesa ($E(Tq)$) = [475 μ sec; 563 μ sec] 519 μ sec

Numero complessivo di pacchetti in attesa di trasmissione =
= [503 pacchetti; 624 pacchetti] 564 pacchetti

CARICO 865 Mb/sec (.72 della banda massima)

Tempo medio di attesa ($E(Tq)$) = [216 μ sec; 423 μ sec] 319 μ sec

Numero complessivo di pacchetti in attesa di trasmissione =
= [202 pacchetti; 349 pacchetti] 275 pacchetti

Confrontando i risultati risulta evidente che i valori medi degli indici nel caso di carico uniformemente distribuito (nel qual caso, secondo le nostre ipotesi, la soglia $\rightarrow \rho=1$) sono molto migliori dei valori corrispondenti nel caso di distribuzione proporzionale del carico.

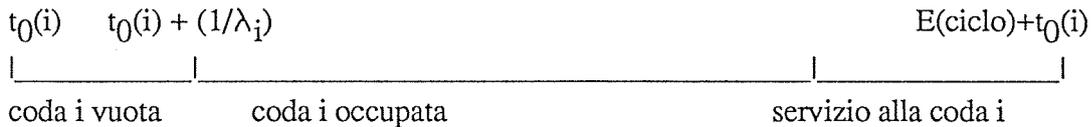
In base alle nostre ipotesi questo risultato era atteso, infatti, secondo le nostre ipotesi, alla distribuzione "proporzionale" del carico e' associata una soglia=.52% (e quindi, per livelli di carico = .72 e .9 si ottiene una forte correlazione tra le stazioni), mentre nel caso della distribuzione uniforme il livello soglia=1.

Quindi, sebbene le prove sperimentali non ci permettano di affermare la correttezza della nostra ipotesi, i risultati ottenuti sperimentalmente sembrano confermarla.

7.0 APPENDICE

Questa appendice viene utilizzata per giustificare la modellizzazione delle stazioni della rete FASNET con code di attesa del tipo M/D/1.

Nel paragrafo 4.1 abbiamo visto che in condizioni stazionarie e con una distribuzione degli arrivi poissoniana (di media λ_i), lo stato della stazione S_i puo' essere rappresentato nel seguente modo:



durante la fase "coda i occupata" almeno un pacchetto e' in attesa di trasmissione nella coda i, nella realta' fisica il pacchetto non puo' essere tuttavia servito (trasmesso) poiche' il servente "e' vacante", nel modello assumiamo, invece, che durante la fase "coda i occupata" il servente stia erogando un servizio supplementare al primo pacchetto della coda. Con questa assunzione il nostro modello viene ricondotto al classico modello M/G/1 [KLEINROCK 75]

Utilizzando queste ipotesi e' possibile definire il fattore di utilizzo relativo "all'ipotetico" servente M/G/1 associato permanentemente alla coda i-esima:

$$\rho^*(i) = \frac{E(\text{ciclo}) - (1/\lambda_i)}{E(\text{ciclo})}$$

$\rho^*(i)$ = utilizzazione del servente associato alla coda i

Vediamo cosa si ottiene se modelliamo con una coda del tipo M/D/1 la stazione S_i . Utilizzando il fattore di utilizzazione $\rho^*(i)$ e applicando la formula K-P si ottiene:

$$E(Q_i) = \frac{[\rho^*(i)]^2}{2*[1 - \rho^*(i)]} + \rho^*(i) = \frac{\rho^*(i)}{2} * \frac{[E(\text{ciclo}) - (1/\lambda_i)]}{E(\text{ciclo})} * \frac{E(\text{ciclo})}{(1/\lambda_i)} + \rho^*(i)$$

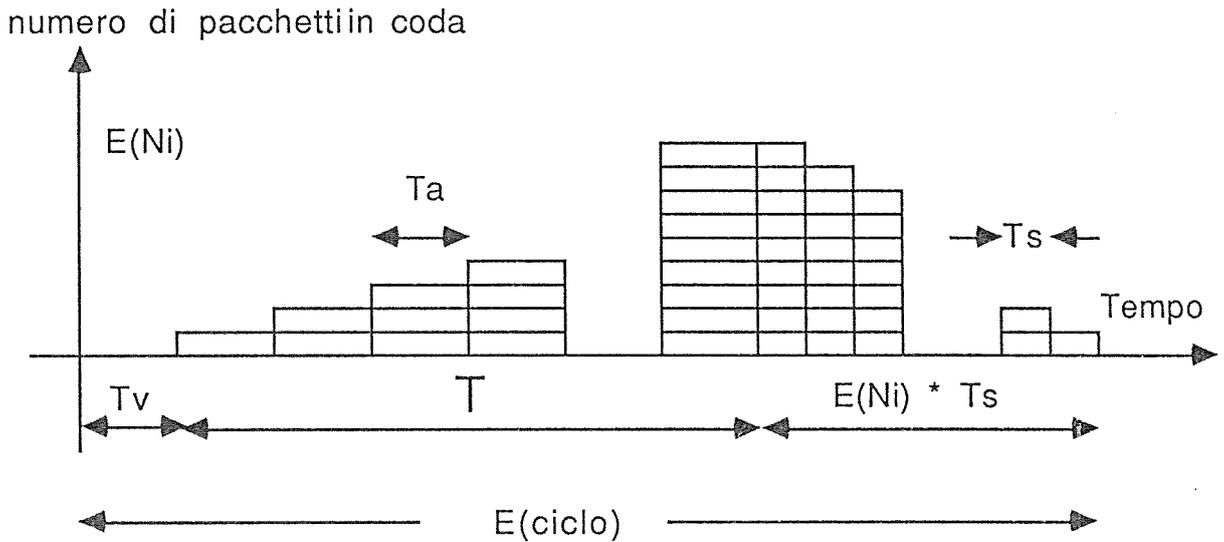
$E(Q_i)$ = numero medio di utenti presenti nella stazione S_i

$E(\text{ciclo})$ = lunghezza media del ciclo

con semplici passaggi algebrici si ottiene:

$$E(Q_i) = \frac{\rho^*(i)}{2} * [\lambda_i * E(\text{ciclo}) + 1]$$

Questo valore di $E(Q_i)$ puo' essere ottenuto se durante la fase "servente occupato" la lunghezza della coda assume, con la stessa probabilita', tutti i valori compresi tra 1 e $[\lambda_i * E(\text{ciclo})]$; questo evento corrisponde alla rappresentazione in figura.



$E(n_i)$ = numero medio di pacchetti che arrivano alla stazione S_i durante un ciclo = $\lambda_i * E(\text{ciclo})$

T_v = durata della fase "servente vuoto"

T_a = tempo medio di interarrivo dei pacchetti

T_s = tempo di servizio (costante)

T = fase di arrivo dei pacchetti

$E(n_i) * T_s$ = fase di trasmissione dei pacchetti

Analizzando la rappresentazione data in figura, si ottiene che la lunghezza media della coda durante la fase $E(\text{ciclo})$ e' data da:

$$E(Q_i) = \rho^*(i) * [1 * P(1) + 2 * P(2) + 3 * P(3) + \lambda_i * E(\text{ciclo}) * P(\lambda_i * E(\text{ciclo}))]$$

dove: $\rho^*(i)$ e' la probabilita' che il servente sia occupato

$P(i)$ e' la probabilita' che durante la fase "servente occupato" la coda sia lunga "i"

poiche', in media, durante la fase "servente occupato" la coda della stazione S_i e' lunga I per un periodo di tempo $(T_s + T_a)$ e questo valore non dipende da I , si ottiene:

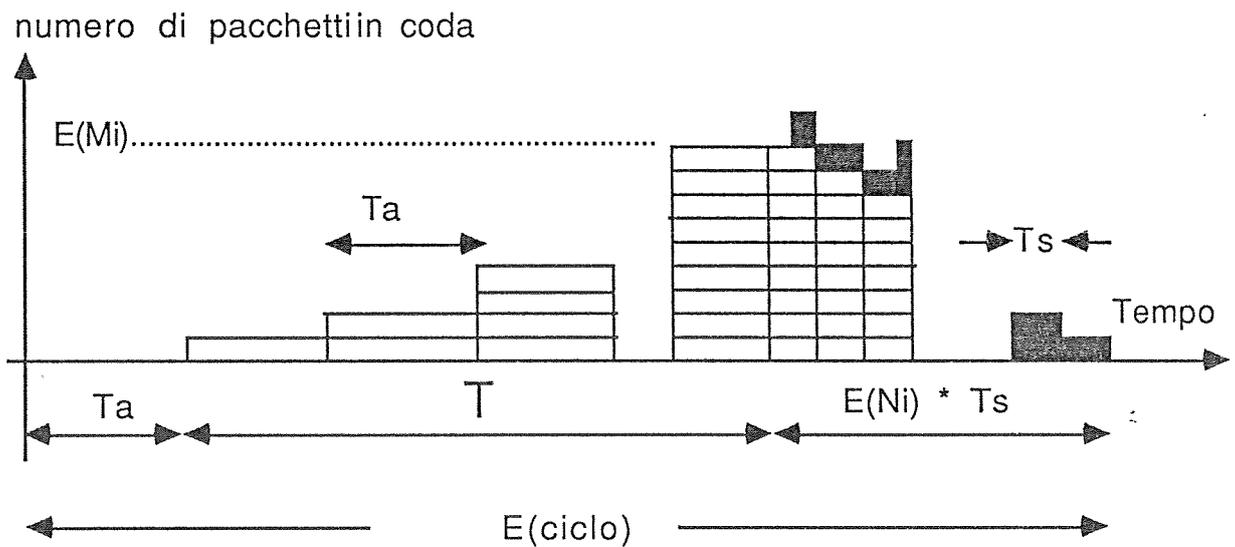
$$P(i) = (T_s + T_a) / [(E(n_i) * T_s) + (E(n_i) * T_a)] = 1/E(n_i) = 1/[\lambda_i * E(\text{ciclo})]$$

$$E(Q_i) = \rho^*(i) * \frac{[\lambda_i * E(\text{ciclo})] * [\lambda_i * E(\text{ciclo}) + 1]}{2} * P(i) = \rho^*(i) * \frac{[\lambda_i * E(\text{ciclo}) + 1]}{2}$$

In questo modo abbiamo dimostrato che la lunghezza media della coda per un servente M/D/1 con fattore di utilizzazione $\rho^*(i)$ e' uguale alla lunghezza media della coda per un servente in cui arrivi e servizi si succedono come rappresentato in figura.

Consideriamo, adesso, come si comporta in media (e sotto ipotesi di servizi esaustivi) la coda della stazione S_i della rete FASNET.

L'andamento della lunghezza della coda della stazione S_i puo' essere rappresentata come in figura.



$E(M_i)$ = numero di pacchetti che arrivano durante la fase T

T_a = tempo medio di interarrivo $= 1/\lambda_i$

T_s = tempo di servizio (costante)

La parte in neretto rappresenta i pacchetti arrivati durante la fase in cui la stazione S_i sta trasmettendo.

Nella realtà che stiamo esaminando esistono, infatti, tre fasi; una prima fase, di durata $1/\lambda_i$ durante la quale la coda della stazione S_i è vuota; una seconda fase che inizia con l'arrivo di un pacchetto nella coda della stazione S_i e termina quando la stazione S_i inizia a trasmettere (questa fase ha una durata media T); ed infine la fase durante la quale la stazione S_i trasmette, in media, tutti i pacchetti arrivati o durante la fase T , o durante la fase di servizio (in condizioni stazionarie, il numero di pacchetti in attesa è in media $\leq P_{\max}(i)$, e quindi in media il servizio è esaustivo).

Passiamo, adesso, a confrontare l'andamento della lunghezza della coda del modello M/G/1 (presentato nel paragrafo 4.1) e di un modello M/D/1 avente lo stesso fattore di utilizzo ρ^* .

Il confronto viene scomposto in tre fasi:

1. coda vuota (la fase di durata T_v nelle figure 1 e 2)
2. coda occupata e *server vacation* (fase di durata T)
3. coda occupata e disponibilità del servente (fase di durata $E(N_i) \cdot T_s$)

Poiché i server hanno lo stesso fattore di utilizzo, la fase 1 è uguale nei due modelli; questo implica che le code di attesa (del modello M/D/1 e del modello M/G/1) assumono con la stessa probabilità lunghezza zero.

Per confrontare l'andamento della lunghezza delle code durante la fase 2, utilizziamo la seguente rappresentazione grafica ottenuta sovrapponendo l'andamento della lunghezza delle code.

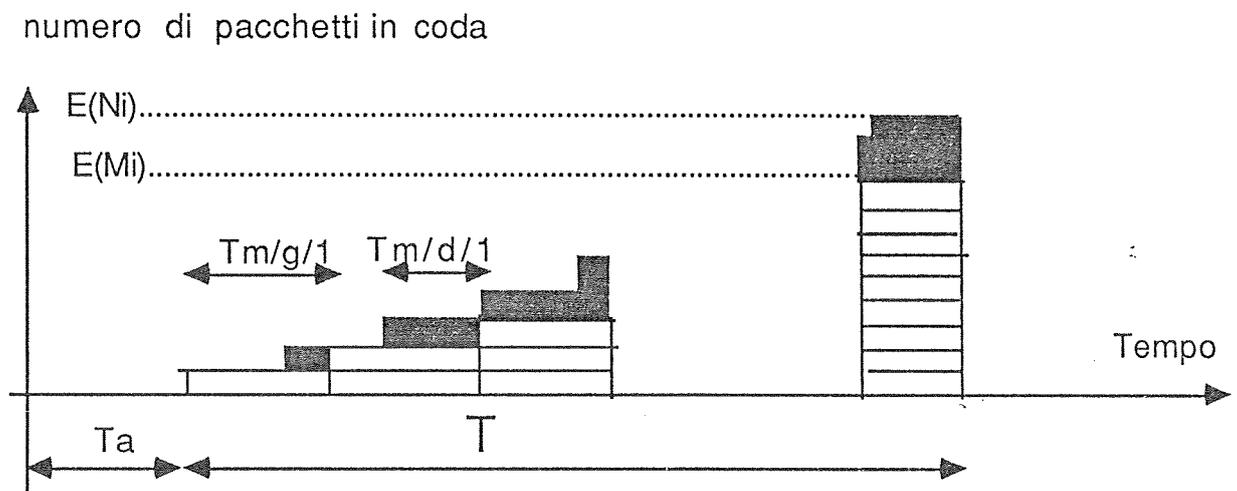


Figura 3

Nella rappresentazione grafica la parte in neretto rappresenta la differenza, nel caso medio, tra la lunghezza del modello M/D/1 e la lunghezza della coda del modello M/G/1. Infatti (per le ipotesi

fatte), nella rappresentazione dell'andamento della lunghezza della coda M/D/1 tutti gli utenti arrivano durante la fase 2, e poiché il numero medio di utenti che arrivano è uguale in entrambi i modelli si ottiene che la lunghezza della coda del modello M/D/1 è maggiore (o al più uguale) alla lunghezza della coda del modello M/G/1.

Passiamo, infine, a confrontare la lunghezza delle code durante la fase 3. All'inizio della fase 3, per quanto visto in precedenza, la lunghezza della coda del modello M/D/1 ($E(N_i)$) è maggiore o uguale alla lunghezza della coda del modello M/G/1 ($E(M_i)$). Durante la fase 3 l'andamento della lunghezza della coda nei due modelli è la seguente:

- la lunghezza della coda relativa al modello M/D/1 diminuisce di una unità ogni T_s secondi
- la lunghezza della coda del modello M/G/1 può avere due tipi di cambiamento di stato:
 - .a l'incremento di una unità dovuto all'arrivo di un pacchetto (questo evento dipende dalla distribuzione dei tempi di interarrivo)
 - .b la diminuzione di una unità ogni T_s secondi

In seguito ai cambiamenti di stato, durante la fase 3 (che inizia all'istante $T+T_v$) possono verificarsi due situazioni che analizziamo con l'ausilio delle figure 4.1 e 4.2.

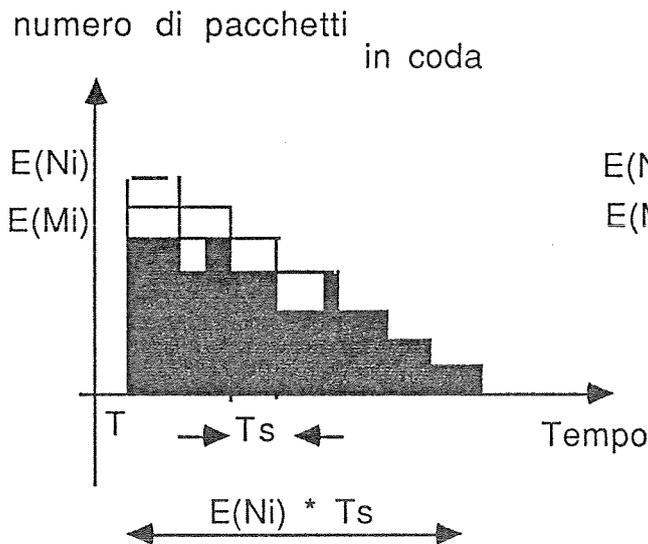


Figura 4.1

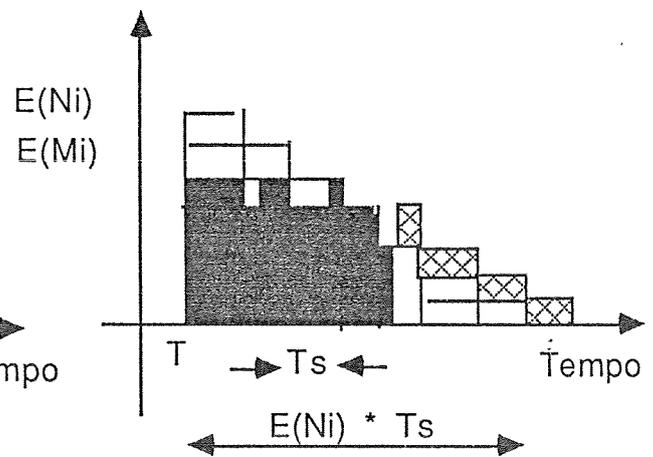


Figura 4.2

Le situazioni che possono verificarsi sono le seguenti:

1. La lunghezza della coda del modello M/G/1 è sempre minore o uguale alla lunghezza della coda del modello M/D/1. Questa possibilità è rappresentata dalla figura 4.1; la parte in neretto rappresenta la lunghezza della coda M/G/1, mentre la parte bianca rappresenta la differenza tra la lunghezza della coda del modello M/D/1 e la lunghezza della coda M/G/1.

2. E' la situazione rappresentata in figura 4.2. La fase 3 viene suddivisa in due sottofasi; nella prima sottofase la lunghezza della coda del modello M/G/1 e' minore o uguale alla lunghezza della coda del modello M/D/1; nella seconda sottofase la lunghezza della coda del modello M/G/1 e' maggiore della lunghezza della coda del modello M/D/1.

Nella rappresentazione in figura 4.2, la parte in neretto rappresenta la lunghezza della coda del modello M/G/1 durante la prima sottofase, mentre la parte tratteggiata rappresenta la differenza esistente nella seconda sottofase tra la lunghezza della coda M/G/1 e la lunghezza della coda M/D/1.

A questo punto e' semplice vedere che la situazione 2. non puo' verificarsi; infatti, se esiste, durante la fase 3, un certo istante di tempo T^* tale che:

$$\text{lunghezza_coda_M/G/1}(T^*) > \text{lunghezza_coda_M/D/1}(T^*)$$

allora in ogni istante di tempo successivo la lunghezza della coda M/G/1 sara' sempre maggiore, almeno di una unita', della lunghezza della coda M/D/1 (ricordiamo infatti che gli eventi "fine_servizio" sono deterministici e si verificano, quindi, in entrambi i modelli allo stesso istante); in particolare, quando l'ultimo pacchetto lascia la coda M/D/1 la coda M/G/1 contiene almeno un pacchetto, e pertanto la fase 3 relativa a questo modello dura almeno T_s secondi in piu' rispetto al caso M/D/1. Poiche', la fase 3 ha, in media, la stessa durata nei due modelli possiamo concludere che la situazione 2 non si verifica mai.

Conclusione

1. Dall'analisi fatta possiamo dire che la lunghezza media della coda di attesa ($E(Q)$) di un modello M/D/1 caratterizzato da un fattore di utilizzo $\rho^*(i)$, costituisce un limite superiore della lunghezza media della coda del modello M/G/1 presentato nel paragrafo 4.1, caratterizzato anch'esso da un fattore di utilizzo $\rho^*(i)$; cioe', $E(Q)_{M/D/1} \geq E(Q)_{M/G/1}$.

2. Dalla teoria delle code sappiamo che *a parita' di fattore di utilizzo* :

$$E(Q)_{M/D/1} \leq E(Q)_{M/G/1}$$

Poiche' dall'analisi presentata abbiamo concluso che la proposizione 1.e' vera, e poiche' la proposizione 2. e' vera in generale possiamo affermare che deve essere:

$$E(Q)_{M/D/1} = E(Q)_{M/G/1}$$

8.0 BIBLIOGRAFIA

[BASKETT 75] Baskett, F., Chandy, K., Muntz, R., Palacios, F., "Open, Closed, and Mixed Networks of Queues with Different Classes of Customers", Journal of the ACM, Vol. 22 No. 2, Aprile 1975, pp. 248-260.

[BOXMA 86] Boxma, O., J., Meister, B., "Waiting-Time Approximations for Cyclic-Service Systems with Switch-over Times", Rapporto Interno IBM, Zurigo 1986.

[BUDRIKIS 86] Budrikis, Z., et alt. "QPSX: A Queue Packet and Synchronous Circuit Exchange", Proceedings of ICCS, 1986.

[CONTI 88] Conti, M., "Modellizzazione di una rete metropolitana di tipo FASNET per servizi integrati voce/dati", Rapporto Interno CNUCE, C88-2, Pisa, Gennaio 1988.

[FUHRMANN 88] Fuhrmann, S., Wang Y., T., "Mean Waiting Time Approximations of Cyclic Service Systems with Limited Service", Proceedings of PERFORMANCE '87, Courtois and Latouche editors, (North-Holland), 1988.

[DIX 87] dix, F., Garrett, M., "Integrated Services Media Access Protocol for FASNET", Rapporto Interno BELL COMMUNICATIONS RESEARCH.

[EISENBERG 70] Eisenberg, M., "Queues with Periodic Service and Changeover Times", Operation Research, Vol 20 (1972), pp. 440-451.

[KLEINROCK 75] Kleinrock, L., "Queueing Systems", Vol I, 1975

[LIMB 82] Limb, J., Flores, C., "Description of FASNET", BSTJ, Vol. 61, No. 7, Settembre 1982

[LIMB 84] Limb, J., Mark, W., "Integrated Voice/Data Services on FASNET", BSTJ, Vol. 63, No. 2, Febbraio 1984.

[NEWMAN 86] Newman, R., Hullet, J., "Distributed Queueing: A Fast Packet Access Protocol for QPSX", Proceedings of ICCS, 1986.

[NEWMAN 87] Newman, R., Budrikis, Z., Hullet, J., "The QPSX MAN".

[SAUER 82] Sauer C., MacNair, E., "THE RESEARCH QUEUEING PACKAGE VERSION 2: CMS Users Guide", Rapporto Interno IBM, Yorktown 1982.

[WATSON 84] Watson, K., "Performance Evaluation Of Cyclic Service Strategies", Proceedings of PERFORMANCE '84, Glembe (editor), (North_Holland), 1984.