

CAPITOLO 7

Controllo in tempo reale della fornitura di banda in presenza di traffico eterogeneo in una rete di telecomunicazioni

Mario Marchese, Dipartimento, Università di Genova

Maurizio Mongelli, Università di Genova

Vincenzo Gesmundo, Selex Communications S.p.A

ABSTRACT

Nel contesto delle reti di telecomunicazioni, questo lavoro propone un modello innovativo per il calcolo della banda da allocare ad un buffer di traffico che serve traffico eterogeneo all'interno di un nodo di una rete. L'eterogeneità consiste nella presenza di flussi di traffico caratterizzati da proprietà statistiche e richieste di *Qualità del Servizio* (*Quality of Service - QoS*) differenti. La tecnica proposta non richiede l'uso di forme chiuse per la descrizione dei parametri prestazionali di interesse e, per questo, è di interesse anche per apparati di tipo industriale.

INTRODUZIONE

Il contesto generale ove trova applicazione l'algoritmo presentato in questo lavoro è la gestione della *Qualità del Servizio* (*Quality of Service - QoS*) in reti di telecomunicazioni eterogenee. Offrire servizi con una specifica garanzia di qualità su reti che coinvolgono tratte eterogenee integrate, sia per il supporto fisico (porzioni terrestri, satellitari e radio), che per i protocolli e le soluzioni impiegate, che per il tipo di traffico convogliato, è molto impegnativo in quanto implica la soluzione di problemi di ricerca complessi e multidisciplinari. La trasmissione di informazioni attraverso tali sistemi deve apparire come trasparente dal punto di vista degli utenti. In altre parole, l'obiettivo scientifico-tecnologico è quello di annullare le differenze tra le diverse porzioni tecnologiche della rete attraverso la definizione e l'uso di adeguate interfacce e algoritmi, che rendano "seamless" (dall'inglese, senza giunzioni, senza cuciture, ovvero come costituita da un solo pezzo) l'interconnessione delle varie componenti della rete. Ogni porzione della rete può utilizzare un'apposita pila protocollare e ogni protocollo deve essere

pienamente coinvolto nella fornitura di qualità in modo trasparente per l'utente. I nodi della rete devono comunicare tra loro per stabilire norme comuni per il controllo e per lo scambio di informazioni e dati; ogni singolo livello funzionale deve fornire un servizio ai livelli superiori e l'integrazione deve scaturire dalla cooperazione di ciascun livello e porzione di rete.

Tale scenario è di grande interesse in ambito industriale. Per esempio, nel settore dell'industria satellitare, i terminali d'interconnessione con la tratta satellitare adottano il protocollo di comunicazione di *Digital Video Broadcasting - Return Channel Satellite* (DVB-RCS) che prevede l'incapsulamento del ben noto *Internet Protocol* (IP), di ampia diffusione nell'Internet attuale, nel protocollo cosiddetto di *Asynchronous Transfer Mode* (ATM). In pratica, nel passaggio da una rete terrestre IP ad una porzione DVB-RCS, utilizzando ATM come tecnologia di trasporto dell'informazione, si realizza lo scenario di interconnessione modellato nel seguito (Figura 1).

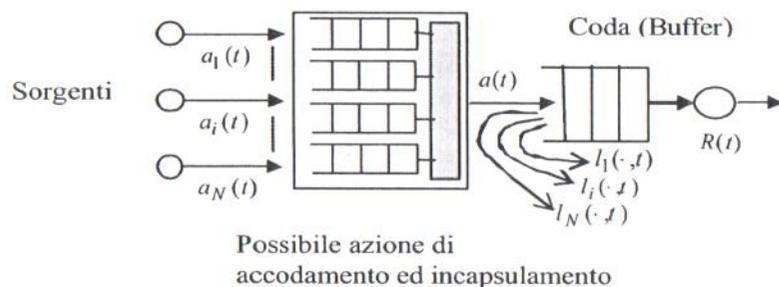


Figura 1: modello del sistema di traffico.

Analoghe considerazioni possono essere svolte per altre tecnologie di reti radio terrestri di imminente immissione nel mercato, come la tecnologia *Worldwide Interoperability for Microwave Access* (WiMAX). Va sottolineato infatti che diverse aziende di reti di telecomunicazioni, nazionali ed internazionali, hanno ormai raggiunto un grado di maturazione nella progettazione di apparati di interconnessioni tra reti eterogenee tale da rendere imminente la presenza sul mercato di prodotti nei quali si rende necessario ottimizzare le prestazioni dello schema di interconnessione attraverso algoritmi in grado di "adattare automaticamente" l'apparato alla condizione di lavoro specifica.

Il concetto di "adattamento" diviene decisivo in presenza di supporto della rete a servizi "tattici" nei quali la QoS va mantenuta anche in situazione di "crisi" della rete stessa, per esempio in applicazioni *mission critical* come in scenari militari o di applicazioni della protezione civile.

Il concetto di interconnessione "seamless" rappresenta quindi una sfida scientifica nell'ambito della tecnologia attualmente in uso. Specificatamente, uno dei principali problemi riguarda l'allocazione della banda. Essa deve essere compiuta in modo tale da garantire un determinato livello di servizio agli utenti, nonostante la presenza di flussi di traffico caratterizzati da diverse proprietà statistiche e da differenti requisiti prestazionali.

In tale ottica, è indispensabile introdurre il concetto di *Banda Equivalente* (*Equivalent Bandwidth*). Essa è definita come la minima capacità di servizio da allocare ad un buffer di traffico in modo da soddisfare un determinato grado (o vincolo) di QoS in termini di parametri oggettivi quali la probabilità di perdita, il ritardo medio od il jitter (varianza del ritardo) dei pacchetti. Le tecniche convenzionali di banda equivalente sono derivate analiticamente per flussi statisticamente omogenei ed in presenza di un unico vincolo di QoS. I moderni apparati di rete spesso invece prevedono l'aggregazione di flussi di traffico con differenti vincoli di QoS e diverse statistiche di generazione dei pacchetti. Questa esigenza comporta la necessità di sviluppare nuove tecniche per il calcolo della banda equivalente in grado di calcolare in tempo reale, con precisione, l'allocazione della banda in presenza della suddetta eterogeneità di traffico.

Un'estensione del lavoro riassunto in questo articolo è attualmente in fase di deposito brevettuale congiuntamente con la Selex Communications S.p.A. Il livello di maturazione del lavoro dal punto di vista operativo è comprovato dal suo utilizzo pratico per controllare un sistema di trasporto per applicazioni audio-video sperimentato con sorgenti di traffici reali (si veda la sezione "Risultati Computazionali" del seguito).

INQUADRAMENTO DEL PROBLEMA

Supponiamo che vi siano N differenti classi di traffico. Sia $a_i(t)$ il processo che regola la *rate* di ingresso dei pacchetti generati dalla i -esima classe di traffico e sia $a(t)$ il processo aggregato di tutti i processi $a_i(t)$, $i=1, \dots, N$. Il traffico è convogliato verso un singolo buffer, modellato attraverso la metodologia di *Stochastic Fluid Model* (Wardi et al., 2002, Cassandras et al., 2003). Si suppone che $a(t)$ sia ergodico in modo tale che una singola realizzazione del processo sia rappresentativa dell'intero processo stocastico. Questa assunzione sarà rilassata nel seguito. Si noti che non si vuole definire e sfruttare alcuna conoscenza a priori dei processi $a_i(t)$, come pure del processo aggregato $a(t)$. Inoltre, è altresì vero che il processo di aggregazione prevede operazioni di accodamento

("bufferizzazione") ed incapsulamento, come tipicamente avviene negli apparati di reti reali. In sintesi, $a(t)$ deriva dalla sovrapposizione dei processi $a_i(t)$ e da una possibile azione di accodamento ed incapsulamento dell'informazione. Ciò rende, di fatto, impraticabile ottenere modelli analitici del processo $a(t)$ anche in presenza di una totale conoscenza sulle proprietà statistiche dei singoli processi $a_i(t)$. L'unica informazione utilizzata nel seguito è ricavata da misurazioni realmente effettuate nel sistema. Sia $R(t)$ la capacità di servizio del buffer in questione e $l_i(R(t), t)$ la frequenza temporale del processo di perdita per l' i -esima classe di traffico, misurata in [bps]. Il relativo valor medio di quest'ultima grandezza è definito nell'espressione (1), $i=1, \dots, N$. Come già detto, non si dispone di alcuna espressione analitica per $l_i(R(t), t)$. Il suo comportamento temporale è valutato anch'esso attraverso misure. Il modello del sistema è riportato in Figura 1.

Il livello di servizio (SLS - *Service Level Specification*) per ogni classe di traffico è definito, in questo caso, da una soglia sulla *Packet Loss Probability (PLP_i^{*})*. Ciò significa che la frequenza ammissibile di perdita dell'informazione è limitata, in ogni istante, dal processo $l_i^*(t) = PLP_i^* \cdot a_i(t)$, misurato in [bps], il cui valor medio è riportato nell'espressione (2), $i=1, \dots, N$.

$$\bar{l}_i = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{\tau} l_i(t) dt \quad (1)$$

$$\bar{l}_i^* = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{\tau} l_i^*(t) dt \quad (2)$$

Lo scopo è dunque quello di calcolare la minima capacità di servizio necessaria per minimizzare la massima distanza quadratica tra \bar{l}_i e \bar{l}_i^* . Matematicamente ciò significa in formule definire il problema di ottimizzazione esposto nella (3), ed identificato nel seguito come problema di *Generalized Equivalent Bandwidth (GEqB)*.

$$R^* = \operatorname{argmin}_R l_{\Delta}(\cdot, R), \quad l_{\Delta}(\cdot, R) = \operatorname{Max}_i \left[\bar{l}_i - \bar{l}_i^* \right]^2 \quad (3)$$

METODO RISOLUTIVO

Essendo i processi stocastici sconosciuti a priori, il problema GEqB è risolto

sulla base di misure prese lungo un opportuno *orizzonte di misurazione* (OH), $T_k = [t_{k-1}, t_k]$, $k=1, 2, \dots$, ed applicando una sequenza di riallocazioni $R(t_k)$, $k=1, 2, \dots$, alla fine di ogni T_k , basata su un opportuno metodo del gradiente delineato nel seguito. La frequenza di perdita $l_i(R(t_k), t)$ e l'ammissibile frequenza di perdita $l_i^*(t)$ sono mediate ogni OH, dando così origine alle quantità \hat{l}_i^k e $\hat{l}_i^{*,k}$, riportate nelle espressioni (4) e (5). Essendo utilizzate per risolvere il problema GEqB, \hat{l}_i^k e $\hat{l}_i^{*,k}$ devono essere rappresentative dei valori medi \bar{l}_i e \bar{l}_i^* , $\forall i=1, \dots, N$ per $\forall k$.

$$\hat{l}_i^k = \frac{1}{T_k} \int_{T_k} l_i(t) dt \quad (4)$$

$$\hat{l}_i^{*,k} = \frac{1}{T_k} \int_{T_k} l_i^*(t) dt \quad (5)$$

L'allocazione di banda al tempo T_k è regolata dall'algoritmo in Figura 2 e chiamato *Gradient-based Generalized Equivalent Bandwidth (G²EqB)*.

a) se $\hat{l}_i^k - \hat{l}_i^{*,k} \geq 0$ per almeno una i

$$\Delta_i(t_k) = \begin{cases} 2 \cdot \left. \frac{\partial \hat{l}_i(R)}{\partial R} \right|_{R=R(t_{k-1})} \cdot [\hat{l}_i^{*,k} - \hat{l}_i^k] & , \text{ if } \hat{l}_i^k - \hat{l}_i^{*,k} \geq 0 \\ 0 & \text{ altrimenti} \end{cases}$$

$$\Delta(t_k) = \max_i |\Delta_i(t_k)|, \quad R(t_k) = R(t_{k-1}) + \operatorname{step}_k \cdot \Delta(t_k)$$

b) se $\hat{l}_i^k - \hat{l}_i^{*,k} < 0, \forall i$; $\Delta_i(t_k) = 2 \cdot \left. \frac{\partial \hat{l}_i(R)}{\partial R} \right|_{R=R(t_{k-1})} \cdot [\hat{l}_i^k - \hat{l}_i^{*,k}]$

$$\Delta(t_k) = \min_i |\Delta_i(t_k)|, \quad R(t_k) = R(t_{k-1}) - \operatorname{step}_k \cdot \Delta(t_k)$$

Figura 2: algoritmo risolutivo proposto chiamato G²EqB

La quantità step_k è il passo del gradiente. La condizione a) afferma che la banda allocata deve essere incrementata. La condizione b) afferma l'opposto. Le derivate $\left. \frac{\partial \hat{l}_i(R)}{\partial R} \right|_{R=R(t_{k-1})}$ rappresentano la sensitività della perdita di

informazione (che è la variabile prestazionale) a variazioni infinitesime della capacità (banda) di servizio del buffer. Intuitivamente, esse dipendono dalla velocità con la quale il sistema passa dallo stato di buffer vuoto a quello di pieno. Possono essere ottenute osservando l'evoluzione dello stato del buffer all'interno di ogni intervallo OH, che è suddiviso in N_{T_k} periodi di attività (in cui il buffer è non-vuoto) identificati dalla variabile bp . Nel caso in cui sia presente un'unica classe di traffico, la forma della derivata in questione è presentata in (Wardi et al., 2002). Essa è tuttora sconosciuta in presenza di più di una classe di traffico, come nel caso qui trattato. Per questa ragione, in questo lavoro si introduce l'approssimazione riportata nell'espressione (6) dove la condizione di eguaglianza vale solo in presenza di una singola classe di traffico, come provato in (Wardi et al., 2002).

$\left[iat_{T_k}^{bp}(R(t_{k-1})) - iIl_{T_k}^{bp}(R(t_{k-1})) \right]$ è il contributo alla perdita di informazione per l' i -esima classe di traffico per il periodo di attività bp all'interno di T_k , $k=1,2,\dots$. $iat_{T_k}^{bp}$ è il tempo di arrivo del primo pacchetto dell' i -esima classe di traffico all'interno del periodo di attività del buffer bp . $iIl_{T_k}^{bp}$ è l'istante temporale relativo all'ultima perdita di un pacchetto appartenente all' i -esima classe durante il periodo di attività bp .

$$\left. \frac{\partial \hat{l}_i(R)}{\partial R} \right|_{R=R(t_{k-1})} \equiv -\frac{1}{T_k} \sum_{bp=1}^{N_{T_k}} \left[iat_{T_k}^{bp}(R(t_{k-1})) - iIl_{T_k}^{bp}(R(t_{k-1})) \right] \quad (6)$$

La G^2EqB , insieme ad ogni altro sistema di controllo per un buffer di traffico basato su una tecnica di "sensitività" come la (6), possono essere interpretati come varianti "ottimizzate" del più tradizionale sistema di controllo cosiddetto *proportional-integral-derivative controller* (nel seguito riportato brevemente come PID), si veda (en.wikipedia.org/wiki/PID_controller) ed i riferimenti bibliografici al suo interno. Il PID è un meccanismo di controllo di tipo closed loop universalmente utilizzato nei sistemi di controllo industriale tradizionali. In sintesi, il PID cerca di correggere l'errore tra la variabile prestazionale di interesse (nel nostro caso il vincolo di QoS) ed il suo valore di target desiderato attraverso il calcolo di un'opportuna azione correttiva tale da minimizzare tale errore. Come nella maggior parte dei casi applicativi reali, non disponendo di un modello analitico dell'errore in funzione dei parametri del sistema (spesso non si dispone neanche di un'equazione differenziale del sistema stesso sufficientemente precisa), il PID esegue il calcolo dell'azione correttiva basandosi sugli andamenti

misurati delle componenti proporzionale, derivativa ed integrativa dell'errore. Naturalmente, possedere un modello analitico più raffinato dell'errore consente di tracciarne la minimizzazione con maggiore precisione. Questa è esattamente l'azione svolta dalla G^2EqB . Come si può infatti apprezzare dal problema di ottimizzazione di G^2EqB espresso nella (3), l'azione correttiva svolta dalla G^2EqB "insegue" direttamente la minimizzazione dell'errore stesso. Si può mostrare, in risultati sperimentali non riportati nel seguito per brevità, come la G^2EqB sia in grado di riprodurre lo stesso risultato del PID per il problema in oggetto, ma con più precisione e riducendone considerevolmente i tempi di convergenza.

Le condizioni tecniche che assicurano la convergenza della G^2EqB ad un'ottimo globale ($R(t_k) \xrightarrow{k \rightarrow \infty} R^*$) sono: 1) ergodicità dei processi stocastici, 2) comportamento decrescente del passo del gradiente $step_k$, 3)

gradiente limitato all'interno del dominio di controllo $R(t) \in \mathfrak{R}^+$, $\forall t$, 4) assenza di ottimi locali. 1) e 2) sono assunzioni ragionevoli. Per quanto riguarda la condizione 3), la lunghezza temporale dei periodi di attività è limitata dalla dimensione stessa dell'orizzonte di osservazione; perciò le misure effettuate riguardanti la perdita di informazione alla fine di ogni orizzonte di misura non possono essere infinite. Per quanto riguarda invece la condizione 4), si noti che la frequenza di perdita di un buffer di traffico può essere assunta come continua, differenziabile con derivata negativa rispetto alla capacità di servizio; per questa ragione anche la funzione di costo non può che mantenere le stesse proprietà di continuità e differenziabilità, nonché la presenza di un unico minimo globale.

La velocità di convergenza dipende dalla lunghezza dell'orizzonte di misura e dal passo del gradiente. La lunghezza dell'orizzonte di osservazione è determinante in quanto, da un lato, deve essere sufficientemente lunga in modo da assicurare che \hat{l}_i^k e $\hat{l}_i^{*,k}$ siano rappresentativi dei valori medi \bar{l}_i e \bar{l}_i^* , $\forall i=1,\dots,N$ e $\forall k$, ma, al contempo, deve essere abbastanza breve da assicurare una rapida convergenza. Nello specifico contesto di questo lavoro, l'assunzione sull'ergodicità può essere rilassata e limitata al tempo necessario alla sequenza di riallocazione $R(t_k)$, $k=1,2,\dots$ per convergere al

valore ottimo R^* . Quando il processo $a(t)$ cambia il suo comportamento statistico (per esempio aumenta o diminuisce il valor medio della *rate* di ingresso dei pacchetti nel buffer), un nuovo problema G^2EqB deve essere risolto supponendo $a(t)$ ergodico almeno all'interno del tempo di convergenza e facendo ripartire l'algoritmo G^2EqB dall'inizio. Di

conseguenza, fissare con precisione la lunghezza dell'intervallo di osservazione è importante anche per reagire velocemente a variazioni del traffico. Per quanto riguarda il dimensionamento del passo del gradiente, si può notare che test non riportati in questo lavoro hanno mostrato la non criticità di tale parametro per la convergenza. In pratica, la condizione 2) può essere rilassata utilizzando un valore costante per il passo del gradiente come mostrato anche in (Cassandras et al., 2003). Per esempio, se $step_k = 1$ nei test riportati nel seguito l'algoritmo converge. Un adattamento accurato del passo del gradiente aiuta ad aumentare la velocità di convergenza e limita eccessive oscillazioni durante il processo di riallocazione.

RISULTATI COMPUTAZIONALI

Come detto in precedenza, data la complessità statistica del processo $a(t)$, le tecniche convenzionali per il calcolo della banda equivalente, che utilizzano descrittori di traffico complessi (come la "burstiness" delle sorgenti), sono di difficilissima applicazione al caso in questione. L'approccio sviluppato in (Guérin et al., 1991), chiamato EqB nel seguito, semplice ed efficace, è invece applicabile in questo contesto e verrà utilizzato come paragone con l'algoritmo G^2EqB .

Siano $m_a(t_k)$ e $\sigma_a(t_k)$ la media e la deviazione standard di $a(t)$ misurate lungo il k -esimo orizzonte di osservazione. La legge di riallocazione della banda assegnata ai tempi $t_k, k=1,2,\dots$ dell'algoritmo EqB è mostrata nell'espressione (7). PLP_{EqB}^* è la massima PLP (Packet Loss Probability) concessa al sistema ed è definita come il vincolo di PLP più stringente tra quelli delle N classi di traffico.

$$R(t_k) = m_a(t_k) + d \cdot \sigma_a(t_k), \quad d = \sqrt{-2 \ln(PLP_{EqB}^*) - \ln(2\pi)} \quad (7)$$

G^2EqB ed EqB a confronto: allocazione e convergenza

Si consideri un livello di servizio (un SLS) per un traffico di tipo VoIP (i.e., voce su IP). Ogni sorgente di traffico è costituita da un processo on-off. Le variabili aleatorie delle durate dei tempi di on ed off sono distribuite esponenzialmente con medie 1.008 s and 1.587 s, rispettivamente. La banda di picco è 16 kbps. Il traffico VoIP entra in un buffer IP la cui capacità di servizio (dimensionata secondo la banda di picco) garantisce una perdita di informazione nulla. Il traffico IP è poi incapsulato in ATM (via AAL5) in modo da generare il processo $a(t)$. L' $a(t)$ a sua volta entra nel buffer ATM

(di 1600 bytes), dove la frequenza di perdita \hat{l}^k misurata in pacchetti IP è misurata ogni T_k . PLP_{VoIP}^* è $2 \cdot 10^{-2}$; l'orizzonte di osservazione (OH) per l'algoritmo G^2EqB è 30 s; il passo del gradiente è 6.0. L'OH dello schema EqB è fissato in due modi, a seconda delle prove effettuate: o a 30 s o al valore calcolato attraverso l'approssimazione nota come principio della *Dominant Time Scale*, il quale calcola l'esatto dimensionamento dell'OH in modo da ottenere delle stime affidabili dei parametri di media e deviazione standard del traffico. Le due alternative sono rispettivamente identificate come "EqB OH 30s" e "EqB OH DTS" nel seguito. La Figura 3 e la Figura 4 mostrano, rispettivamente, la PLP e la corrispondente allocazione di banda per gli algoritmi G^2EqB e EqB. Il numero di sorgenti di traffico VoIP è incrementato di 10, partendo da 70 fino a 110, ogni 3000 secondi.

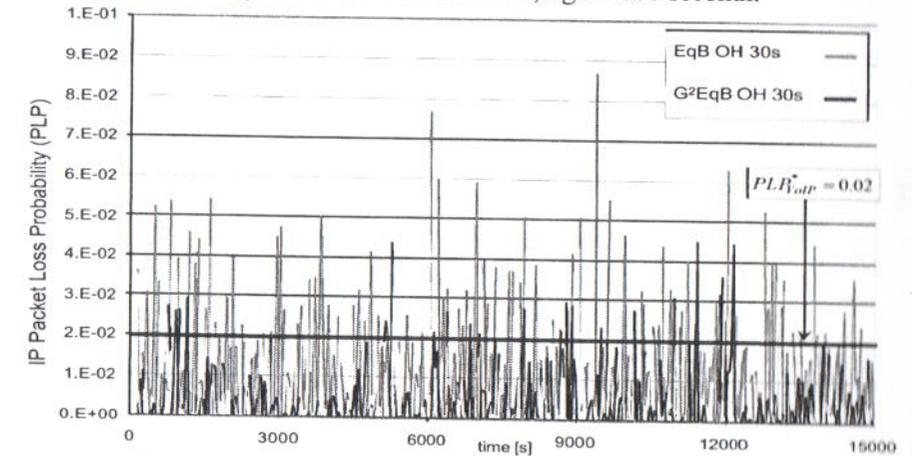


Figura 3: G^2EqB ed EqB a confronto: PLP (Packet Loss Probability) – Probabilità di perdita dei pacchetti.

I risultati medi della PLP sono: $4.40 \cdot 10^{-3}$ per G^2EqB e $1.18 \cdot 10^{-2}$ per EqB OH 30s. EqB OH DTS garantisce una PLP nulla. Le medie della allocazione di banda sono: 0.842 Mbps per G^2EqB , 0.867 Mbps per EqB OH 30s, e 1.31 Mbps per EqB OH DTS. Nonostante i valori medi di PLP sembrano soddisfacenti per tutti gli schemi proposti, una semplice osservazione delle figure 3 e 4 suggerisce che l'algoritmo G^2EqB è in grado di reagire velocemente ai cambi di traffico minimizzando le oscillazioni di banda. EqB OH DTS garantisce sempre la PLP_{VoIP}^* richiesta ma a scapito di un evidente spreco di banda; EqB OH 30s invece fallisce spesso l'obiettivo di qualità

fissato PLP_{VoIP}^* ed introduce forti oscillazioni nella PLP e nell'allocazione di banda. Alcune metriche quantitative aiutano nell'interpretazione di questo comportamento qualitativo. La deviazione standard della PLP è $7.4 \cdot 10^{-3}$ per G^2EqB e $1.33 \cdot 10^{-2}$ per EqB OH 30s. La percentuale dei periodi di osservazione dove la PLP è sopra soglia è il 5% per G^2EqB ed il 18.6% per EqB OH 30s. La differenza media tra la PLP misurata e la PLP_{VoIP}^* , selezionando solo quei periodi di osservazione nei quali la PLP è oltre la soglia, è $4.22 \cdot 10^{-4}$ per G^2EqB e $2.77 \cdot 10^{-3}$ per EqB OH 30s. In sintesi, l'algoritmo G^2EqB permette di minimizzare la distanza tra le misure effettuate e la soglia di QoS richiesta.

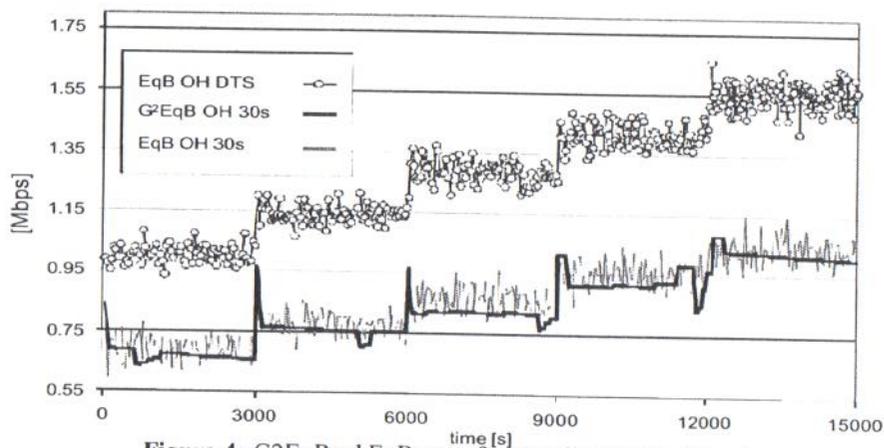


Figura 4: G^2EqB ed EqB a confronto: allocazione di banda.

G^2EqB ed EqB a confronto: traffico eterogeneo

Un tracciato di video reale ("Jurassik park" preso da www-tnk.ee.tu-berlin.de/research/trace/trace.html) è aggiunto allo scenario VoIP usato finora. Le bande di picco e media del video sono 1.418 e 0.280 Mbps, rispettivamente. Il video entra in un buffer IP la cui capacità di servizio (1.418 Mbps) garantisce una perdita di informazione nulla. Sia il traffico VoIP che quello video sono incapsulati in tecnica DVB (pacchetti da 188 bytes) all'uscita del buffer IP in modo da generare il processo $a(t)$.

L'obiettivo di qualità PLP_{video}^* è $5 \cdot 10^{-3}$. La Tabella 1 contiene le misure medie di PLP per video e VoIP insieme alla banda allocata al buffer DVB in [Mbps] per G^2EqB OH 3min (G^2EqB in Tabella 1), EqB OH 3min e EqB OH DTS. La dimensione del buffer DVB è variata insieme al numero di

chiamate VoIP. Ogni singolo campione in tabella simula 107 minuti di traffico reale. La media della PLP per G^2EqB è sempre vicina e sotto la soglia più restrittiva (quella del video). L'algoritmo G^2EqB si adatta alle variazioni della dimensione del buffer perchè il suo comportamento dipende unicamente dalle misure di perdita effettuate. Gli schemi di tipo EqB non si adattano alla lunghezza del buffer: EqB OH 3min sottostima la banda necessaria e fallisce il mantenimento della soglia richiesta; EqB OH DTS si comporta similmente per lunghezze limitate del buffer mentre sovrastima la banda per dimensioni del buffer più elevate.

Tabella 1. G^2EqB ed EqB a confronto: PLP ed allocazione di banda

Numero di utenti VoIP	Buffer [bytes]	Banda Allocata G^2EqB / EqB OH 3min / EqB OH DTS	Video PLP G^2EqB / EqB OH 3min / EqB OH DTS	VoIP PLP G^2EqB / EqB OH 3min / EqB OH DTS
30	9400	0.94 / 0.79 / 0.87	$2.62 \cdot 10^{-3}$ / $8.26 \cdot 10^{-2}$ / $1.054 \cdot 10^{-2}$	$1.20 \cdot 10^{-4}$ / $5.58 \cdot 10^{-3}$ / $5.93 \cdot 10^{-4}$
30	18800	0.84 / 0.79 / 0.87	$3.20 \cdot 10^{-3}$ / $2.964 \cdot 10^{-2}$ / 0.0	$2.22 \cdot 10^{-4}$ / $2.13 \cdot 10^{-3}$ / 0.0
30	28200	0.82 / 0.79 / 0.87	$1.78 \cdot 10^{-3}$ / $1.341 \cdot 10^{-2}$ / 0.0	$1.24 \cdot 10^{-4}$ / $9.14 \cdot 10^{-4}$ / 0.0
60	9400	1.47 / 1.26 / 1.41	$3.94 \cdot 10^{-3}$ / $8.26 \cdot 10^{-2}$ / $1.154 \cdot 10^{-2}$	$2.26 \cdot 10^{-4}$ / $7.774 \cdot 10^{-3}$ / $5.63 \cdot 10^{-4}$
60	18800	1.41 / 1.26 / 1.41	$2.23 \cdot 10^{-3}$ / $5.13 \cdot 10^{-2}$ / $1.544 \cdot 10^{-3}$	$1.55 \cdot 10^{-4}$ / $3.96 \cdot 10^{-3}$ / $8.17 \cdot 10^{-5}$
60	28200	1.39 / 1.26 / 1.41	$1.38 \cdot 10^{-3}$ / $3.04 \cdot 10^{-2}$ / $4.204 \cdot 10^{-4}$	$1.04 \cdot 10^{-4}$ / $2.324 \cdot 10^{-3}$ / $8.17 \cdot 10^{-5}$
90	9400	1.97 / 1.73 / 2.00	$4.13 \cdot 10^{-3}$ / $1.074 \cdot 10^{-1}$ / $1.584 \cdot 10^{-2}$	$2.38 \cdot 10^{-4}$ / $8.11 \cdot 10^{-3}$ / $8.28 \cdot 10^{-4}$
90	18800	1.87 / 1.73 / 2.0	$3.26 \cdot 10^{-3}$ / $5.33 \cdot 10^{-2}$ / $4.354 \cdot 10^{-3}$	$1.94 \cdot 10^{-4}$ / $3.974 \cdot 10^{-3}$ / $2.374 \cdot 10^{-4}$
90	28200	1.84 / 1.73 / 2.0	$1.73 \cdot 10^{-3}$ / $3.074 \cdot 10^{-2}$ / $2.024 \cdot 10^{-3}$	$1.02 \cdot 10^{-4}$ / $2.354 \cdot 10^{-3}$ / $2.374 \cdot 10^{-4}$

CONCLUSIONI

In questo lavoro è stato proposto e testato un metodo innovativo di allocazione della banda per adattare automaticamente la capacità di servizio di un buffer che serve traffico in condizioni di eterogeneità, per quanto riguarda sia le proprietà statistiche, che la qualità del servizio richiesta. L'algoritmo è basato su misure prese in tempo reale sul sistema e non utilizza alcuna forma analitica del funzionale di costo, alcuna informazione a-priori sulle proprietà statistiche delle sorgenti e alcuna assunzione a-priori sulla dimensione del buffer di traffico. L'efficacia dell'algoritmo è stata verificata mediante simulazioni.

BIBLIOGRAFIA

- Wardi, Y., Melamed, B., Cassandras, C.G., Panayiotou, C.G. (2002). Online IPA Gradient Estimators in Stochastic Continuous Fluid Models. *Journal of Optimization Theory and Applications*, 115, 369-405.
- Cassandras, C.G., Sun, G., Panayiotou, C. G., Wardi, Y. (2003). Perturbation Analysis and Control of Two-Class Stochastic Fluid Models for Communication Networks. *IEEE Transactions on Automatic Control*, 48, 770-782.
- http://en.wikipedia.org/wiki/PID_controller.
- Guérin, R., Ahmadi, H., Naghshineh, M. (1991). Equivalent Capacity and its Application to Bandwidth Allocation in High-Speed Networks. *IEEE Journal of Selected Areas on Communications*, 9, 968-981.
- <http://www-tnk.ee.tu-berlin.de/research/trace/trace.html>.