

ISTITUTO NAZIONALE DI FISICA NUCLEARE

CNAF - Bologna

INFN/TC-98/02
28 Gennaio 1998

S. Alborghetti, C. Battista, A. Brunengo, M. Campanella, F. D'Isep, T. Ferrari,
L. Gaido, A. Ghiselli, P. Matteuzzi, C. Olivieri, A. Re, C. Salvo, S. Spanu,
E. Valente, C. Vistoli, S. Zani:

**SPERIMENTAZIONI ATM PER L'UTILIZZO DI RETI LAN E WAN AD
ALTA VELOCITÀ**

SPERIMENTAZIONI ATM PER L'UTILIZZO DI RETI LAN E WAN AD ALTA VELOCITÀ

S. Alborghetti³, C. Battista⁴, A. Brunengo², M. Campanella³, F. D'Isep⁵, T. Ferrari^{1,6},
L. Gaido⁵, A. Ghiselli¹, P. Matteuzzi¹, C. Olivieri², A. Re³, C. Salvo², S. Spanu⁴,
E. Valente^{1,4}, C. Vistoli¹, S. Zani¹

¹ Centro Nazionale per la Ricerca e lo Sviluppo nelle Tecnologie Informatiche e Telematiche,
INFN, Bologna

² Istituto Nazionale di Fisica Nucleare, sezione di Genova

³ Istituto Nazionale di Fisica Nucleare, sezione di Milano

⁴ Istituto Nazionale di Fisica Nucleare, sezione di Roma

⁵ Istituto Nazionale di Fisica Nucleare, sezione di Torino

⁶ Dip. di Ingegneria, Università di Bologna

Sommario

Questo documento descrive i risultati della prima fase di sperimentazioni su ATM che l'INFN ha svolto a livello di rete locale e geografica, utilizzando in ambito nazionale la rete SIRIUS e ponti radio, e in ambito europeo l'infrastruttura JAMES. L'obiettivo principale degli studi di questa tecnologia di trasmissione ad alta velocità è stato la verifica della sua validità e dei limiti d'impiego sia a livello locale che geografico per la progettazione di reti di trasmissione dati ad alta banda su area geografica quali la rete GARR-B e TEN-34. I test sono stati effettuati per studiare le prestazioni dei protocolli TCP/IP, la funzionalità, l'interoperabilità e l'efficienza di apparati di diverse case costruttrici (switch, router e workstation) e per analizzare situazioni di congestione in relazione al problema della perdita di celle. È stato inoltre progettato e realizzato un ponte radio a microonde fra i laboratori nazionali di Frascati e l'Università La Sapienza di Roma. Nel complesso le sperimentazioni in ambito locale hanno evidenziato vari punti: la necessità di un'accurata calibrazione dell'intero sistema per massimizzarne l'efficienza; la limitazione della funzionalità dei meccanismi di controllo di flusso al solo ambiente monovendor e l'elevato consumo di CPU nelle workstation per gestire alte velocità di trasmissione. Infine, i test geografici sul servizio PVC di tipo CBR hanno dimostrato l'affidabilità e interoperabilità delle infrastrutture e degli apparati di diverse Telecom europee e degli apparati-utente e quindi la realizzabilità di una rete ad alto carico e ad alta banda su tecnologia ATM.

1 Introduzione

L'introduzione della fibra ottica come mezzo di comunicazione per la trasmissione digitale a velocità fino al gigabit/sec, ha gettato le basi per lo sviluppo di nuovi protocolli di rete ad alta banda. L'Asynchronous Transfer Mode (ATM) è un tipo particolare di tecnologia 'cell switching' sviluppata per la rete ad alta velocità B-ISDN (Broadband-Integrated Service Digital Network) con lo scopo di ottenere l'integrazione di tipi eterogenei di informazioni: dati, video e audio. Le caratteristiche di particolare interesse della tecnologia ATM, basata sulla generazione di segmenti di dati di lunghezza costante (celle), sono: l'isocronia per la voce e il video, la creazione di connessioni virtuali end-to-end di tipo permanente e/o commutato associabili a diversi tipi di qualità di servizio, e la multiplazione statistica di più circuiti virtuali su un'unica linea trasmissiva con possibilità di 'overbooking' e garanzia di servizio.

Nell'ambito dell'INFN ATM può essere impiegato per l'*event building* da apparati sperimentali ad alta velocità di acquisizione e per la realizzazione di reti locali (attraverso il protocollo di LAN emulation) e geografiche ad alta velocità. In particolare l'impiego di ATM a livello geografico è interessante perché l'INFN partecipa alla definizione di un progetto di rete ad alta banda per l'università e la ricerca scientifica italiana, GARR-B, e al progetto europeo TEN-34, per la interconnessione delle reti europee della ricerca a 34 Mbps. Tenuto conto dello stato ancora sperimentale di questa tecnologia, è stato definito un programma di studio e sviluppo per verificare in primo luogo la funzionalità, l'affidabilità e gli eventuali limiti della tecnologia ATM nelle aree sopra citate. Il programma è stato sviluppato utilizzando le seguenti infrastrutture:

- JAMES, la rete ATM sperimentale delle telecom europee
- SIRIUS, la rete ATM sperimentale nazionale di Telecom Italia
- reti locali
- ponti radio.

Scopo di questo documento è la presentazione dei risultati delle sperimentazioni effettuate sulle infrastrutture elencate sopra e la descrizione dell'utilizzo di ATM nel GARR per il trasporto del *traffico di produzione*. Il documento indica infine la linea delle future sperimentazioni.

2 ATM in ambito locale

Le sperimentazioni condotte su ATM in ambito locale erano finalizzate alla verifica delle funzionalità e delle prestazioni di questa nuova tecnologia per la connessione di workstation, e dell'impatto del protocollo ATM sui nodi terminali. Sono stati svolti test di carico variando di volta in volta la configurazione degli apparati (switch e adapter ATM). La rete locale è stata realizzata collegando ai

suddetti switch delle workstation attraverso interfacce ATM a 34 e/o 155 Mbps. In tutti i casi le prove sono state effettuate utilizzando i protocolli TCP/UDP e IP su ATM.

I flussi di dati da memoria a memoria sono stati generati attraverso applicazioni specifiche per la misura del *throughput*, ovvero della quantità di dati (generati dall'applicazione) trasmessi correttamente nell'unità di tempo. Le applicazioni di pubblico dominio utilizzate sono state: *Netperf* e *ttcp*. In entrambi i casi è possibile assegnare il valore a specifici parametri utilizzati dai protocolli TCP e UDP come: la dimensione dei *socket buffer* [4] (che influiscono sull'algoritmo di controllo di flusso di TCP determinando la dimensione massima della *window* TCP), la dimensione del *message* (ovvero della quantità di dati che vengono scritti dall'applicazione nell'area di memoria del kernel del sistema operativo con un'unica system call) e l'opzione *TCP_nodelay*.

Le prove sono servite per studiare le caratteristiche e la configurabilità degli switch e il tipo di configurazione ottimale di TCP e UDP da adottare quando tali protocolli vengono utilizzati su ATM. Tutti i test sono stati svolti utilizzando l'indirizzamento IP.

L'hardware impiegato nelle varie prove è composto da switch: CISCO LS2020, CISCO LS1010, FORE Systems ASX-200 e DEC GIGAswitch; router: CISCO 7000 e 7507; adapter ATM: FORE Systems SBA200 per workstation SUN, DEC ATMworks 350 per bus PCI e ATMworks 750 per bus TURBOchannel e in fine FORE GIA-200 per SGI. Le workstation sono state: Sparc Station SUN 20, DIGITAL Alpha e Silicon Graphics Indy.

Segue la descrizione dei test ripartiti per tipo di switch.

2.1 GIGAswitch/ATM DIGITAL

I test sono stati eseguiti localmente sullo switch GIGAswitch/ATM AN2 della sezione di Roma. Oltre allo switch sono state utilizzate due workstation DEC 3000/600, una stazione DEC 3000/400 con sistema operativo OSF/1 v2.0 e tre adapter ATM DIGITAL TURBOchannel a 155 Mbps. La struttura di rete così ottenuta è illustrata in Figura 1. I tipi di *permanent virtual circuit* (PVC) configurabili erano il CBR (Constant Bit Rate) e l'ABR (Available Bit Rate). Sono stati eseguiti test su entrambi i tipi di PVC e sull'algoritmo non standard di flow control per traffico ABR denominato *FLOWmaster*.

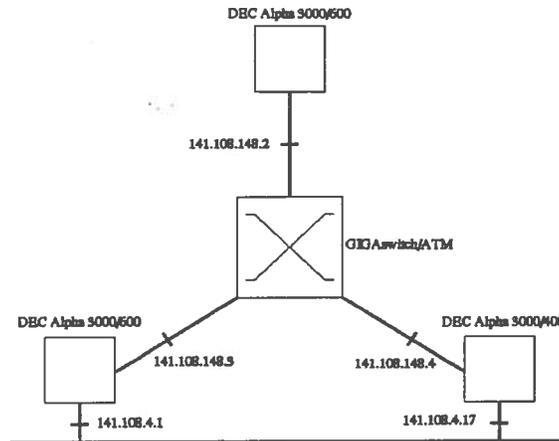


Figura 1: infrastruttura della rete locale basata su GIGAswitch/ATM

Le prove sono state eseguite utilizzando lo stack di protocolli TCP e UDP/IP su ATM e variando la configurazione del protocollo TCP per verificare il valore ottimale dei parametri e delle opzioni di TCP. Attraverso i test di carico è stato misurato l'impiego di CPU delle workstation in ricezione e trasmissione e la capacità di sostenere connessioni multiple e concorrenti.

2.1.1 Studio della configurazione ottimale dei parametri del protocollo TCP

Varie prove sono state condotte per misurare le prestazioni del protocollo al variare della dimensione del messaggio, del *send socket buffer*, del *receive socket buffer* e dell'opzione *TCP_NODELAY*. Attraverso quest'ultimo parametro è possibile forzare l'emissione di uno o più segmenti di dati con dimensione inferiore alla *Maximum Segment Size* (MSS, 9140 byte nel caso di ATM), anche nel caso in cui esistano uno o più pacchetti IP in attesa di *acknowledgment*. Per la maggior parte delle applicazioni non interattive tale opzione è disabilitata per aumentare l'efficienza¹ del protocollo TCP/IP, ovvero per evitare di trasmettere dei *tinygram* [9].

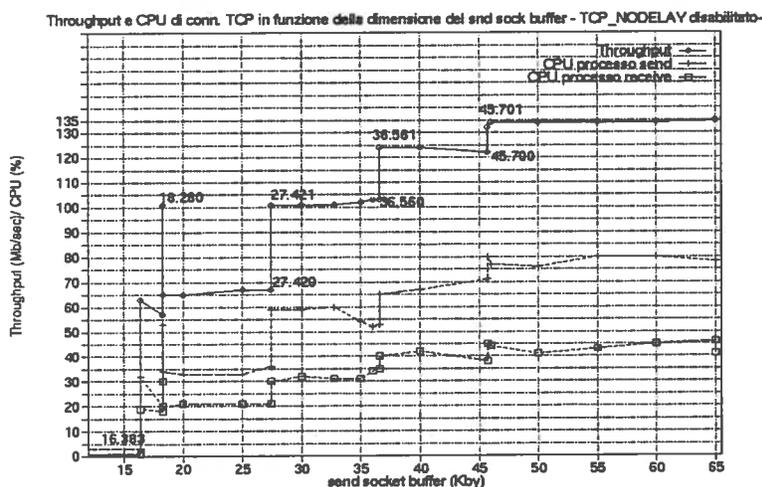


Figura 2: throughput e send socket buffer size

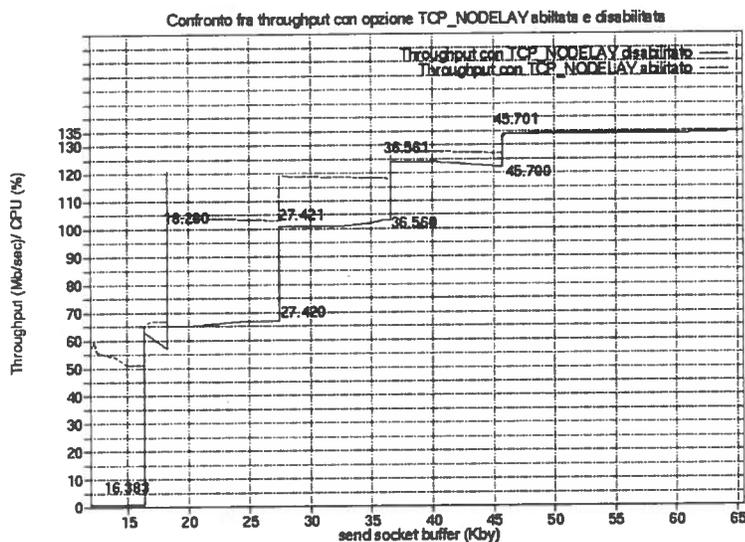
Il grafico in Figura 2 illustra la variazione del throughput in funzione della dimensione della send socket buffer size (le rimanenti variabili vengono mantenute costanti).

L'andamento discontinuo del throughput è dovuto al meccanismo

¹ L'efficienza di TCP o di UDP risultano tanto maggiori quanto minore è l'incidenza dell'*overhead* dei suddetti protocolli rispetto ai dati generati dall'applicazione, in particolare quando la percentuale di byte di overhead è bassa rispetto all'ammontare dei byte della payload area del singolo pacchetto.

interno del sistema operativo OSF/1 utilizzato per assegnare la dimensione ai socket. Dato un valore selezionato dall'applicazione, il sistema operativo sceglie come dimensione finale il minimo multiplo di MSS (9140 byte nel caso di ATM) maggiore del valore dato. Tale meccanismo è introdotto per ottimizzare la trasmissione di pacchetti IP su ATM, ovvero per aumentare il numero di pacchetti trasmessi di dimensione massima. Il throughput più alto ottenuto è stato di circa 135 Mbps, pari al massimo teorico ottenibile se si considera l'overhead di tutti i protocolli utilizzati dal livello di trasporto (TCP o UDP) fino a quello fisico. Esso è pari a circa il 13% della quantità di dati generati dall'applicazione. Il grafico illustra inoltre l'utilizzo di CPU necessario per sostenere una sola connessione TCP a 135 Mbps: esso rasenta l'80%. Inoltre un socket buffer di 64 Kbyte è sufficiente per ottenere il massimo throughput.

Figura 3: throughput e opzione TCP_NODELAY



Il grafico di Figura 3 illustra invece l'effetto dell'abilitazione dell'opzione TCP_NODELAY, ovvero l'incremento delle prestazioni dovuto alla possibilità di inviare segmenti di dimensione inferiore a quella massima. L'opzione TCP_NODELAY ha impatto sulle prestazioni quando si verifica almeno uno dei seguenti casi: 1- l'acknowledgment (ack) è ritardato dal ricevente (il tempo di ricezione dell'ack

è anche funzione del round trip time), 2- l'applicazione produce messaggi piccoli con una velocità non sufficiente per generare un pacchetto IP di dimensione massima prima di ricevere un ack. Nel caso di questo test è la prima condizione a valere. Infatti, quando la dimensione del send socket buffer è di molto inferiore al receive socket (di dimensione costante, pari a 64 kbyte), l'ack viene ritardato dal ricevente secondo il protocollo Delayed Acknowledgment [10]. Se la rimanente area del send socket buffer non è sufficiente per generare un pacchetto di dimensione massima, il processo mittente si pone in attesa passiva dell'ack. L'incidenza dell'opzione TCP_NODELAY è minima in caso di reti locali e di assegnazione bilanciata delle dimensioni del send e receive socket buffer.

Per dimensioni del send socket buffer molto piccole rispetto a quella del receive socket buffer (ovvero nel caso di configurazioni fortemente “asimmetriche”) si verifica una caduta del throughput come illustrato nel grafico di Figura 4, in cui la dimensione del socket varia nell’intervallo [0..16 KB]. L’azzeramento del throughput nell’intervallo [10924..16384] byte è dovuto all’architettura di TCP (ovvero ai protocolli di “delayed acknowledgment”, di Neagle e di “window update” [4]) e alla dimensione elevata dell’MSS standard in ATM (vedi [4,5]).

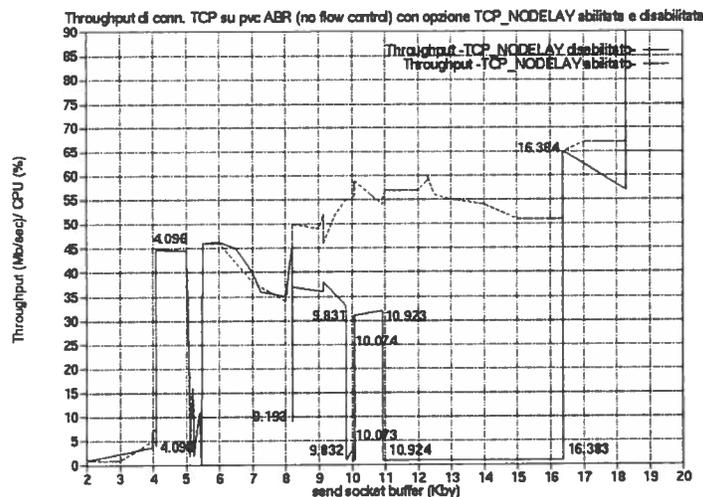


Figura 4: configurazione asimmetrica di socket buffer

2.1.2 Studio delle prestazioni di connessioni TCP multiple

Lo scopo di questo gruppo di prove era quello di verificare la variazione del throughput in funzione del tipo e del numero di connessioni TCP attive. Tutte le connessioni sono state attivate adottando la configurazione TCP ottimale emersa nel precedente gruppo di prove, ovvero utilizzando una configurazione “bilanciata” delle dimensioni dei buffer al lato mittente e ricevente e associando ai buffer la dimensione di 64 Kbyte, essendo essa sufficiente per ottenere il throughput massimo in una rete ATM locale².

Per esempio, data una coppia di nodi e generando tra loro connessioni TCP *full-duplex*, tali che ciascun sistema è impegnato allo stesso tempo sia in trasmissione che in ricezione, si nota che il massimo throughput aggregato teorico ottenibile dalle due connessioni pari a 270 Mbps non è raggiunto a causa della limitata potenza delle workstation impiegate nel test. Il massimo ottenuto è stato di 160 Mbps, ovvero il 56% del massimo teorico, a fronte di un utilizzo di CPU pari a circa il 100%. Inoltre, è stata verificata l’incidenza del numero di connessioni TCP concorrenti sulle prestazioni complessive delle connessioni. Utilizzando un numero crescente di connessioni TCP tra una determinata coppia di nodi collegata da un unico PVC ABR, si osserva che il throughput aggregato non diminuisce. Esso viene

² In una rete ATM locale il valore del RTT è basso, nel nostro caso era inferiore a 1 msec. Per questo in tali condizioni socket buffer di 64 Kbyte sono sufficienti per ottenere una window TCP che permette un trasferimento continuo di dati (ovvero per evitare un comportamento di tipo “stop and wait” nell’applicazione mittente). Nel caso di connessioni ATM geografiche a causa dell’elevato RTT è necessario utilizzare dimensioni di buffer superiori.

suddiviso equamente tra i vari flussi TCP, in modo tale che la somma è sempre pari a 135 Mbps. Lo stesso risultato è confermato quando per ogni flusso TCP si utilizza un PVC ABR distinto.

In conclusione, i test effettuati sullo switch GIGAswitch/ATM DIGITAL hanno rivelato che le prestazioni ottenibili generando flussi di dati TCP e UDP/IP su ATM possono essere ottimali a patto di utilizzare una configurazione bilanciata dei socket buffer al lato mittente e ricevente, generando messaggi sufficientemente ampi e abilitando l'opzione TCP_NODELAY. Attraverso questi strumenti l'applicazione può ottenere un throughput di 135 Mbps al livello dell'applicazione.

2.2 ForeRunner ASX-200/Fore Systems

Sono stati svolti test di interoperabilità e di performance sulla LAN ATM di Milano, la quale comprendeva uno switch ForeRunner ASX-200 e due workstation (una Sparc Station 20 e una Silicon Graphics Indy SGI) dotate di interfaccia ATM. Le prove sono state condotte generando traffico su una singola connessione ATM o utilizzando più connessioni in parallelo come nel caso delle precedenti prove. Il traffico TCP/IP è stato generato su PVC ATM di tipo CBR utilizzando send e receive socket a 64 Kbyte e messaggi di 4 Kbyte. Nella seguente parte del paragrafo vengono illustrati vari gruppi di prove.

2.2.1 PVC singolo (tra SUN e Silicon Graphics Indy)

Tipo di test	Throughput	CPU mitt.	CPU ric.
SUN→SGI	60 Mbps	83%	24%
SGI→SUN	74 Mbps	33%	72%

Tabella 1: prestazioni massime con switch FORE

Le prove sono state eseguite generando traffico in entrambe le direzioni. La tabella indica che è la CPU della SUN a rappresentare il collo di bottiglia che limita il throughput ottenuto rispetto al massimo teorico di 135 Mbps circa. Ripetendo il medesimo test mantenendo il messaggio dell'applicazione pari a 64Kbyte e variando il send/receive buffer in modo tale che send buffer = receive buffer, si registra una condizione di *deadlock*, già osservata anche nei test su GIGAswitch/ATM (vedi Figura 4). Ciò è dovuto alla dimensione ampia della MTU di IP su ATM: 9180 byte che crea situazioni problematiche in presenza di socket buffer piccoli.

Send e Receive socket buffer	Message In Kbyte	CPU mitt. (Indigo)	CPU ric. (SUN)	Throughput in Mbps
65536 byte	64	48%	71%	101
16284 byte	64	16%	51%	52
16383 byte	64	1%	0%	1.4
10000 byte	64	1%	0%	1.3
8110 byte	64	15%	41%	35.21
8000 byte	64	5%	8%	11

Tabella 2: prestazioni con switch FORE al variare della dimensione del messaggio

Il throughput cala al diminuire della dimensione dei buffer e si azzerava quando la dimensione passa sotto la soglia di 16 Kbyte. Aumenta poi quando i buffer si abbassano ulteriormente alla soglia di 8 Kbyte e poi inizia a diminuire progressivamente.

2.2.2 Connessioni multiple su PVC singolo (tra SUN e Silicon Graphics Indy)

Sono state attivate varie connessioni TCP concorrenti (quattro, come in Figura 5) sul medesimo PVC tra SUN e Indy mantenendo costanti le dimensioni dei buffer e del messaggio (ad esse sono stati assegnati i valori ottimali).

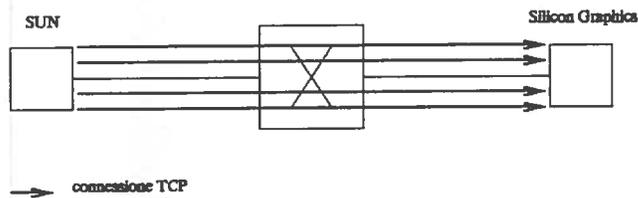


Figura 5: connessioni TCP multiple

Per flussi in direzione SUN → Indy il throughput aggregato scende da 101 Mbps (ottenuto da una connessione singola nel test 2.2.1) a 93 Mbps e risulta ripartito equamente tra le connessioni. Il medesimo valore si ottiene anche nella direzione Indy → SUN.

Attivando quattro connessioni TCP SUN → SGI e altrettante in direzione SGI → SUN per un totale di otto, si ottiene il throughput aggregato seguente:

	Throughput aggregato
Sun→SGI	31.5 Mbps
SGI→Sun	64.3 Mbps

Tabella 3: prestazioni con varie connessioni TCP

In conclusione i test indicano che in presenza di workstation con potenza di CPU limitata e connessioni TCP multiple full o half-duplex, le prestazioni peggiorano.

2.3 LightStream LS2020/CISCO

Le prove di throughput sono state effettuate nella sezione di Torino e al CNAF utilizzando, oltre allo switch, tre workstation a Torino e due al CNAF. In particolare, a Torino sono state utilizzate una AlphaStation 250/4/266 (Digital Unix 3.2G, ATMworks 350), una AlphaStation 600 (Digital Unix 4.0A, ATMworks 350) e una SUN Ultra 1/170E (Solaris 2.5, FORE Runner SBA200). Al CNAF l'hardware utilizzato è stato il seguente: un AlphaServer 1000/4/266 (Digital Unix 3.2C, ATMworks 350), e una AlphaStation 3000/500 (Digital Unix 3.2A, ATMworks 750). Per i test è stato impiegato il programma *Netperf* associando ai send e receive socket buffer la dimensione di 65536 byte (la configurazione ottimale). I parametri utilizzati sono stati la dimensione del messaggio trasmesso e la *max rate*, che sullo switch LS2020 serve per assegnare una data banda ad un PVC.

Con max rate pari a 156250 celle/sec (60 Mbps) si sono ottenute prestazioni piuttosto scadenti: un massimo di 50 Mbps con un messaggio di dimensione ≥ 100 byte. Invece, con max rate pari alla capacità trasmissiva sono stati ottenuti valori ottimali pari a 135 Mbps³.

I seguenti risultati si riferiscono alle prove effettuate utilizzando uno switch LS2020, una AlphaStation 250/4/266, una AlphaStation 600 e una SUN 1/170E.

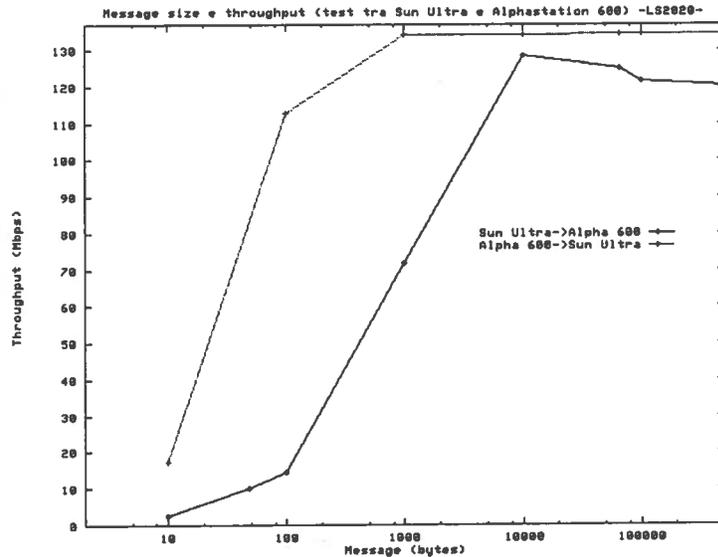


Figura 6: confronto di prestazioni di workstation collegate a un LS2020

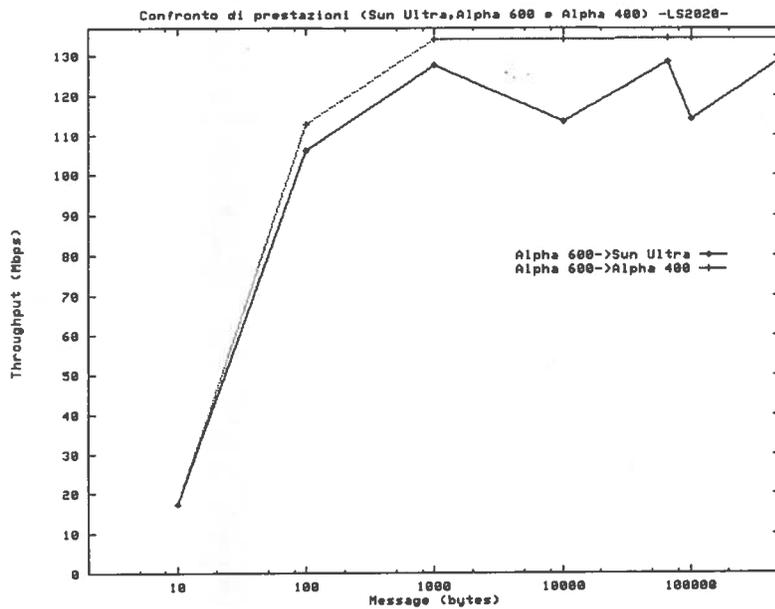
Ogni prova è stata eseguita studiando il throughput al variare della dimensione del messaggio. I risultati mostrano che le prestazioni dipendono strettamente dalla potenza e dal sistema operativo delle workstation.

Per esempio, la Figura 6 mostra il comportamento asimmetrico tra Sun Ultra e Alpha 600 ottenuto al variare della dimensione del messaggio e utilizzando una max rate pari alla capacità della porta. L'Alpha raggiunge un throughput abbastanza elevato utilizzando una message size pari o superiore a 100 byte, mentre quando la Sun Ultra è il nodo mittente le prestazioni crescono più lentamente. Ciò indica che il sistema operativo DIGITAL Unix 4.0a installato sull'Alpha 600 gestisce in modo più efficiente gli *interrupt* software generati dalla system call *write()* ogni volta che un messaggio è spedito dall'applicazione.

Il grafico in Figura 7 mostra invece i risultati comparati ottenuti utilizzando la AlphaStation 600 mittente e variando la stazione ricevente. Il throughput massimo pari a 134,4 Mbps è stato ottenuto nei trasferimenti verso l'AlphaStation 250, che utilizzava sistema operativo DIGITAL Unix 3.2G.

³ Questo tipo di prestazione su connessione singola e con parametri associati al valore ottimale è stata riscontrata su tutti gli switch testati.

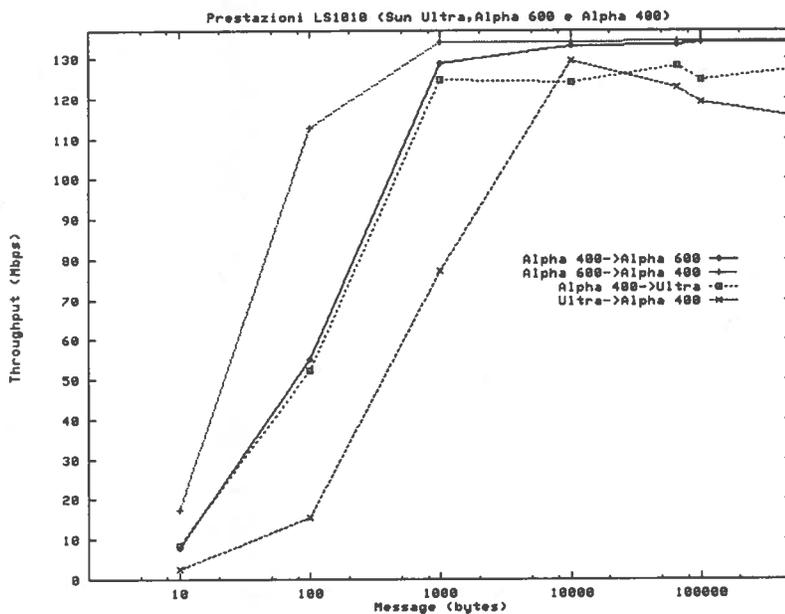
Figura 7: studio comparativo utilizzando l'Alpha 600 come mittente



SIRIUS).

2.4 LightStream LS1010/CISCO

Le prestazioni massime ottenute utilizzando lo switch LS1010 con l'hardware della LAN ATM di Torino (Sun Ultra, AlphaStation 600 e AlphaStation 250) sono le medesime misurate con lo switch LS2020: 134 Mbps scegliendo l'AlphaStation 600 come nodo mittente. Come mostra la Figura 8, anche



in questo caso le prestazioni più basse sono state ottenute utilizzando la Sun Ultra come nodo mittente. In genere l'utilizzo di CPU in trasmissione è superiore a quello in ricezione; ciò giustifica il fatto che il throughput più basso sia ottenuto quando la Sun Ultra è utilizzata come mittente.

Figura 8: risultati dei test effettuati sullo switch LS1010

2.5 Considerazioni

Le prove condotte in ambito locale sono servite per verificare l'interoperabilità degli apparati ATM, la loro funzionalità ed efficienza e la conformità alle loro specifiche. I tipi di servizio sottoposti a test sono stati il CBR (Constant Bit Rate), UBR (Variable Bit Rate) e ABR (Available Bit Rate). Tuttavia non tutti i tipi di switch sono stati in grado di fornire tutte le classi di servizio ATM definite dagli standard. Per

esempio, il GIGAswitch/ATM non implementava il servizio VBR, mentre nell'LS2020 non era disponibile il CBR. L'LS1010 risulta essere il più completo, poiché è possibile configurare connessioni di tipo CBR, VBR, ABR e UBR. Gli switch hanno dimostrato una buona efficienza nell'utilizzo della capacità dei link ATM, tuttavia non è stato possibile effettuare prove di carico sugli switch a causa del limitato numero di stazioni terminali collegate. Con gli switch non sottoposti a carico eccessivo, le prestazioni al livello dell'applicazione non sono funzione del tipo di switch ma piuttosto dalle workstation utilizzate, in particolare del sistema operativo e della potenza della CPU (essendo i processi di trasmissione e ricezione a velocità OC-3c *CPU intensive*). Inoltre, il sistema operativo DIGITAL Unix è stato quello che ha dimostrato l'implementazione più efficiente dello stack TCP/IP, mentre rispetto agli adapter DIGITAL quelli FORE hanno presentato una maggiore flessibilità nella gestione dei parametri dei PVC e nel monitoraggio.

3 ATM in ambito geografico

Il programma di test a livello geografico è stato svolto su una infrastruttura composta da quattro accessi ATM alla rete sperimentale SIRIUS con velocità di linea pari a 34 Mbps e da un collegamento all'infrastruttura ATM europea JAMES. I poli collegati all'infrastruttura SIRIUS sono le sezioni di Milano, Roma, Torino e il CNAF come in Figura 9. Sono state definite diverse configurazioni di VP, ciascuna finalizzata a diversi tipi di prove. Lo scopo di queste sperimentazioni è stato quello di misurare la funzionalità, l'interoperabilità e l'efficienza degli apparati, le prestazioni del protocollo IP su ATM a livello geografico e la prova di nuovi tipi di servizio come l'utilizzo del protocollo di segnalazione.

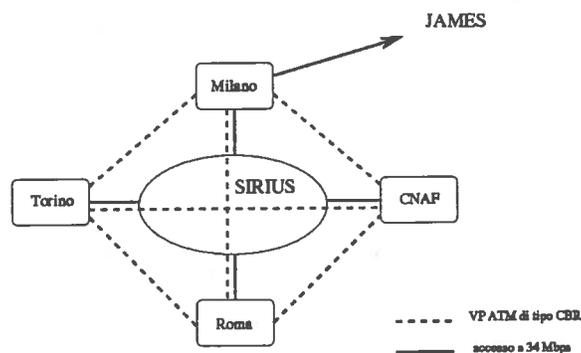


Figura 9: infrastruttura utilizzata per i test in ambito geografico

Segue la descrizione delle prove ripartite in due gruppi: prove su un singolo VP e prove su una topologia di VP magliati.

3.1 Test su singolo VP ATM

3.1.1 Test su VP nazionale

I primi test geografici sono stati condotti tra il CNAF e la sezione di Roma utilizzando un VP ATM su SIRIUS a 5 Mbps. Le prove sono state condotte utilizzando localmente gli switch LS2020 e collegando a SIRIUS due router CISCO 7507, IOS 11.1(2), attraverso una interfaccia a 34 Mbps. Le stazioni terminali utilizzate per la generazione dei dati sono state due Alpha 3000/600 con sistema operativo DIGITAL Unix 3.2. Generando un flusso TCP/IP unidirezionale e utilizzando la configurazione ottimale (64 Kbyte per i socket e per il messaggio) non si è ottenuto un throughput superiore a 3.6 Mbps. Una seconda prova è stata effettuata utilizzando due connessioni TCP in senso opposto in modo da sfruttare l'intera banda in entrambe le direzioni (ovvero 10 Mbps complessivi). Il throughput aggregato massimo è stato pari a 5 Mbps aggregati, cioè il 50% della capacità disponibile a livello utente.

Tale perdita di efficienza è attribuibile alla mancanza di un meccanismo per definire la velocità dei PVC sullo switch LS2020. Infatti, la differenza fra velocità nominale e banda dei PVC a livello geografico può causare la perdita di celle sugli switch. A causa del meccanismo di congestion control di TCP ciò comporta un throughput inferiore a livello di connessione TCP.

3.1.2 Test su VP internazionale

La prima connessione ATM destinata al collegamento del GARR con l'estero per il trasferimento di traffico di produzione è stata un VP CBR a 8 Mbps tra CNAF e il CERN (Svizzera). Sono state effettuate varie prove di carico utilizzando la configurazione illustrata in Figura 10.

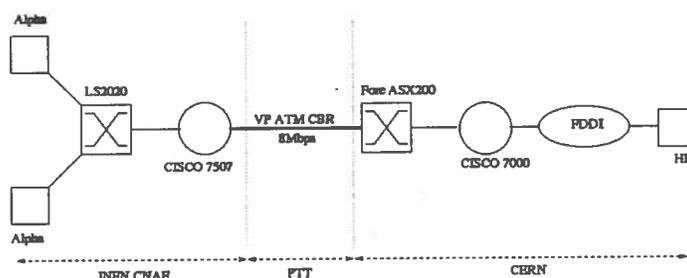


Figura 10: topologia del collegamento ATM tra il CNAF e il CERN

Gli switch sono stati un LS2020 al lato CNAF e un FORE ASX-200 al lato CERN. Nei due router di test sono state definite configurazioni simmetriche dei seguenti parametri: *hold-queue in*, *hold-queue out*, *atm rawq*, *mtu default*, *peak rate* e *average rate*. Poiché al CERN il router era connesso alla workstation attraverso FDDI, nei test non è stata utilizzata la MTU ATM (9180 byte), ma quella FDDI (4352 byte). Ciò implica l'utilizzo di pacchetti IP di dimensione massima inferiore e quindi una maggiore incidenza dell'overhead dei protocolli TCP, UDP e IP.

È stata effettuata una prova per verificare la configurazione ottimale dei parametri *peak* e *average cell rate* da adottare sul router. È emerso che assegnando alla *peak* e *average rate* valori superiori alla banda del VP ATM geografico (8 Mbps), si verifica la perdita di celle sul primo switch Telecom. Ciò significa che sui router la rate non può eccedere il valore della banda assegnata al VP.

I rimanenti test effettuati sul VP verso il CERN hanno permesso di evidenziare i seguenti aspetti.

1. Il massimo throughput misurato generando traffico UDP e rilevato dai contatori del router CISCO del CNAF è stato di 7.0 Mbps.
2. Durante i test in cooperazione con i tecnici di Telecom Italia, non si è mai verificata la presenza di *cell dropping* al lato Telecom. Il massimo throughput registrato generando traffico TCP è stato di 6.0 Mbps. In questa condizione si è verificata la presenza di *cell dropping* sulla porta di ingresso dello switch LS2020. Come spiegato in precedenza, ciò è attribuibile al fatto che l'adapter ATM DIGITAL su workstation non implementa la funzione di *shaping* e che non è possibile assegnare la banda al VP. Ciò è provato anche dal test effettuato collegando la workstation direttamente al router CISCO. In questo modo, a causa della differenza tra le tecniche di gestione dei buffer implementate nello switch e nel router, si è ottenuto un throughput più alto pari a 7.0 Mbps.
3. Il massimo throughput ottenuto nella direzione CERN → CNAF è stato pari a 6.0 Mbps, anche in presenza di una sola connessione TCP.
4. Durante i test l'utilizzo medio della CPU del router CISCO al CNAF è stato registrato pari al 6%.

3.1.3 Test sull'infrastruttura JAMES

I test sono stati condotti su un PVC ATM singolo di tipo CBR a 24 Mbps configurato sull'infrastruttura JAMES tra Milano e la Svezia (KTH). La topologia di JAMES è illustrata in Figura 11, mentre quella utilizzata per il test è illustrata in Figura 12. Lo scopo del test era quello di verificare l'efficienza dei protocolli TCP e UDP/IP su VP ATM a lunga distanza, ovvero l'impatto del round trip time sul massimo throughput ottenibile da una connessione TCP o UDP. Come nel caso di tutti i test in area locale, sono state generate connessioni TCP e UDP attraverso il programma *Netperf*. I test sono stati ripetuti utilizzando diversi tipi di configurazioni degli apparati ATM e anche generando diversi tipi di traffici: connessioni TCP e UDP half-duplex e full-duplex e connessioni multiple concorrenti.



Figura 11: infrastruttura della rete ATM europea JAMES

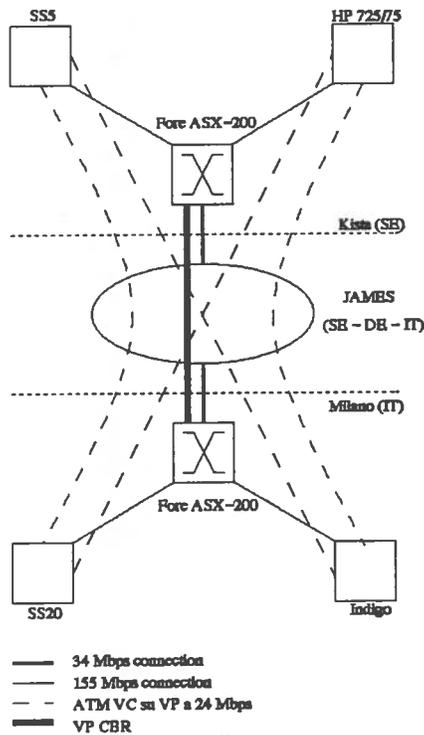


Figura 12: topologia dei collegamenti utilizzati per i test su JAMES tra Milano e la Svezia

Ad entrambi i lati sono stati attestati switch ForeRunner ASX-200. Le workstation utilizzate sono state: a Milano una Sparc Station 20 (Solaris 2.4) e una Silicon Graphics (IRIX 5.3), in Svezia una Sparc Station 5 (Solaris 2.5) e una HP725 (HP-UX9.05). I test hanno dimostrato che il *round trip time* (RTT) può essere un fattore che condiziona fortemente il throughput ottenibile da una o più connessioni TCP/IP attivate su PVC ATM. Il requisito fondamentale per permettere alle applicazioni di saturare la banda disponibile è l'utilizzo di socket buffer di dimensione superiore a 64 Kbyte in un dato sistema operativo, in modo che l'ampiezza della *window* in TCP sia tale da non causare un comportamento "stop and wait" al lato mittente. Esso si verifica quando l'applicazione mittente esaurisce la *window*

TCP prima di ricevere un *acknowledgment* di ritorno, il quale permette alla *window* TCP di avanzare. Se i buffer non hanno dimensione sufficiente, il risultato è che la capacità del VP ATM non può essere completamente saturata da una singola connessione TCP.

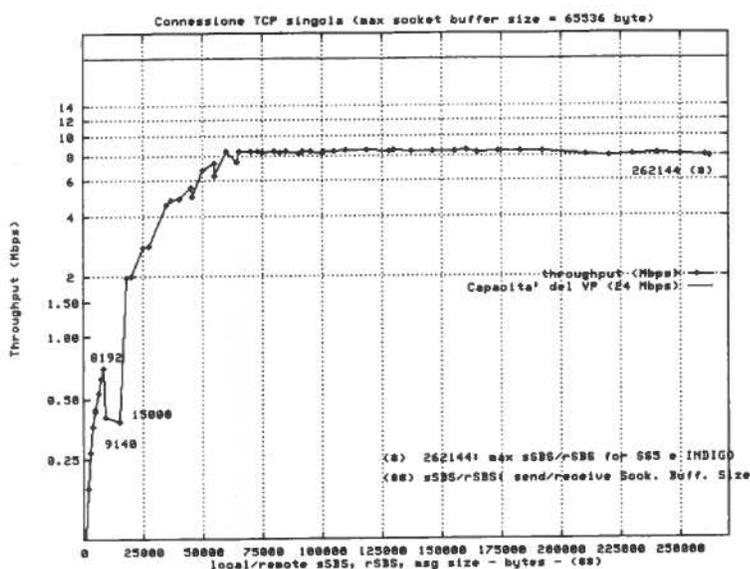


Figura 13: throughput e socket size

La curva di Figura 13 illustra la relazione esistente tra throughput e dimensione dei buffer. L'utilizzo della banda aumenta al crescere della dimensione dei buffer e l'incremento si arresta quando essi raggiungono una dimensione pari a 65536 byte. Infatti questo era il limite reale della window nei sistemi operativi utilizzati nel test. *Netperf* permette di configurare dimensioni anche superiori,

ma tali valori non hanno incidenza sulla dimensione *reale* della window utilizzata nel sistema operativo, poiché essa viene arrotondata per difetto al massimo valore accettabile. In certi sistemi operativi per aumentare il limite massimo della dimensione dei socket buffer (ovvero per adottare la "window scaling option") è necessario applicare specifiche *patch*.

A differenza dei test locali (paragrafo 2), nel caso di connessioni ATM geografiche a banda limitata (24 Mbps invece di 155 Mbps) l'opzione TCP_NODELAY non ha incidenza sulle prestazioni del protocollo. In questo caso la velocità dell'applicazione nel generare dati è più alta di quella con cui essi vengono spediti (il RTT è di due ordini di grandezza superiore), per cui esiste sempre un pacchetto di dimensione massima da spedire.

Se la dimensione dei buffer non è opportunamente configurata sui nodi terminali della connessione, l'utilizzazione della banda disponibile può essere piena solo nel caso in cui vengano utilizzate connessioni TCP multiple e concorrenti, come illustrato nel grafico di Figura 14, in cui il throughput massimo complessivo supera i 20 Mbps. Questo miglioramento è stato ottenuto grazie alla omogeneità dell'hardware utilizzato e alla funzione di traffic shaping implementata negli adapter FORE.

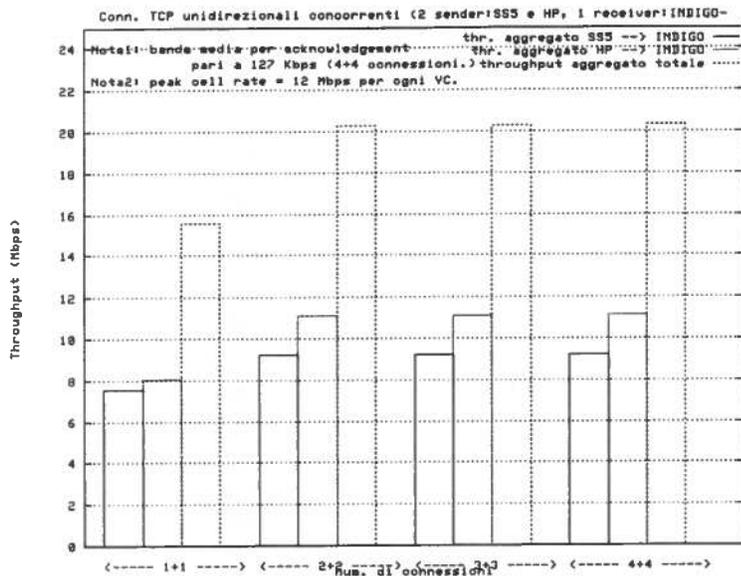


Figura 14: aumento del throughput in presenza di connessioni multiple e concorrenti

Nei test che utilizzano connessioni di tipo bidirezionale o unidirezionale con multipli riceventi si sono ottenuti risultati contrastanti a seconda dei sistemi operativi utilizzati a livello di *end-node*. In alcuni casi, l'attivazione di nuove connessioni TCP manda le connessioni di altri nodi in stato di stallo. Ciò è dovuto alla diversa efficienza delle implementazioni dello

stack TCP/IP nei vari sistemi operativi. Inoltre, sono stati condotte prove per misurare la variazione del RTT di pacchetti al variare della loro dimensione. I valori di RTT sono risultati compresi nell'intervallo [40..60] msec e hanno presentato un incremento lineare, come illustrato in Figura 15.

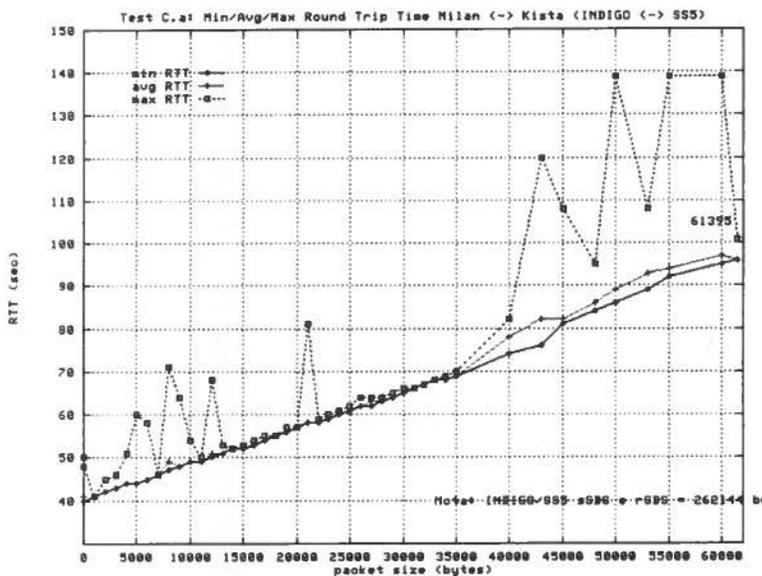


Figura 15: relazione fra RTT e dimensione del pacchetto

Infine, i test del protocollo UDP hanno dimostrato che quando non si utilizza il meccanismo di controllo di flusso di TCP è possibile saturare completamente la banda del PVC ATM, ottenendo così il massimo throughput. Ciò mette in evidenza il fatto che in caso di incompleto utilizzo della capacità, non è la connessione ATM a porre dei limiti di prestazione,

ma è invece il grado di efficienza dell'implementazione di TCP nei sistemi operativi. Tuttavia, nel caso in cui l'opzione "don't fragment" sia abilitata, anche con UDP si possono ottenere prestazioni molto scarse. Generando messaggi che non possono essere inseriti in un unico pacchetto IP (cioè di dimensione superiore a 9152 by) tutti i pacchetti sono rigettati al lato ricevente. Infatti il destinatario non è in grado di effettuare il riassetto dei vari frammenti che costituiscono il *datagram* UDP originario. Il grafico in Figura 16 (ottenuto con l'opzione don't fragment abilitata) illustra che per

dimensioni del messaggio superiori a 9152 byte il throughput della connessione UDP è nullo. Per maggiori dettagli sui test consultare [6].

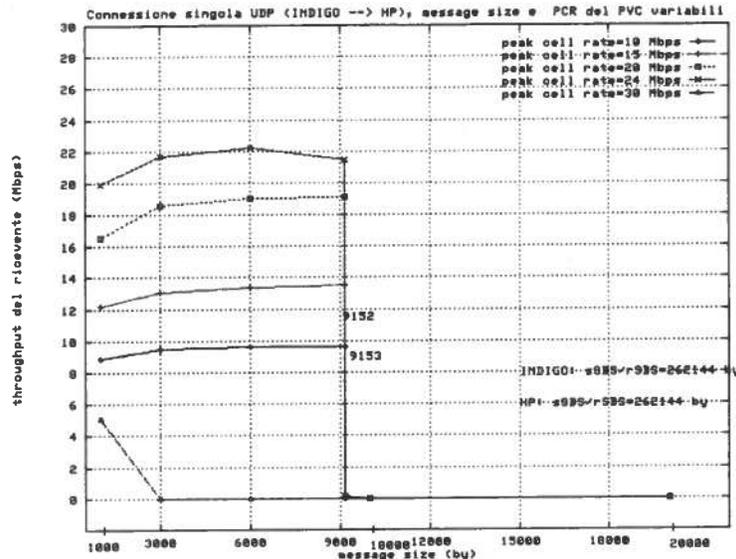


Figura 16: test del protocollo UDP

3.2 Test su topologia magliata di VP ATM

3.2.1 Test di funzionalità tra CNAF, Milano, Roma e Torino (hardware eterogeneo)

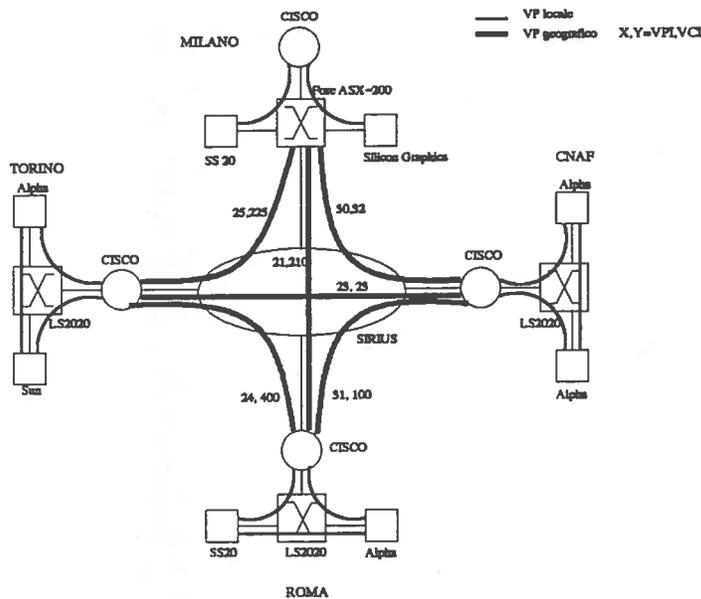


Figura 17: configurazione dei VP su SIRIUS

E' stata effettuata una serie di test su una topologia di VP CBR su SIRIUS completamente magliata tra i quattro poli: Bologna, Milano, Roma e Torino come in Figura 17. Per questo gruppo di test ogni sito era collegato ai rimanenti con 3 VP a *banda payload*⁴ di 8 Mbps ciascuno⁵. Milano era collegato a SIRIUS attraverso lo switch FORE ASX-200, mentre i rimanenti poli hanno utilizzato un

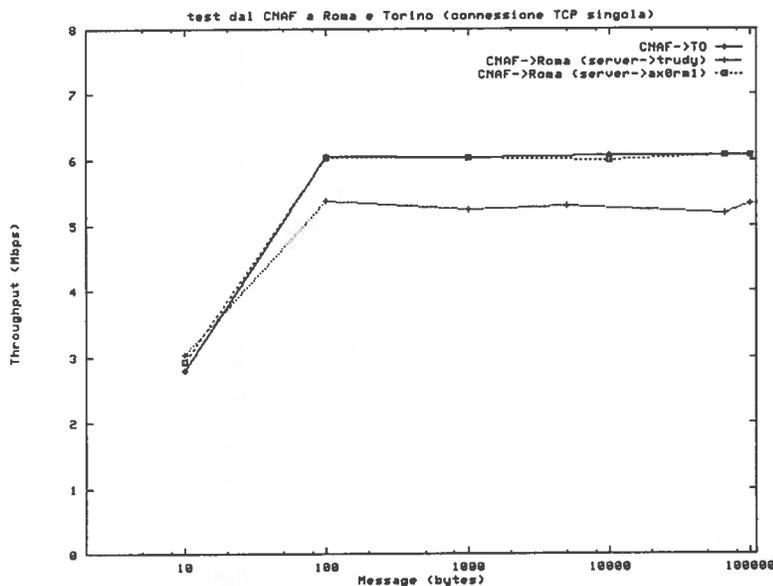
router CISCO e uno switch LS2020 in locale. Lo scopo di queste sperimentazioni era quello di verificare la funzionalità del protocollo TCP in presenza di una configurazione di rete più complessa

⁴ Il termine *banda payload* indica che la banda di un PVC ATM è stata computata considerando solo i byte dell'area payload della cella ATM.

⁵ La capacità assegnata ad un singolo PVP è stata fissata a 8 Mbps tenendo conto del fatto che la massima banda payload allocata da Telecom per il servizio CBR su un accesso a 34 Mbps è pari a 24 Mbps.

rispetto a quelle dei test precedenti e molto simile ad una infrastruttura di produzione basata su protocollo IP. Sono stati impiegati circuiti ATM multipli.

Le prove sono state eseguite variando esclusivamente la dimensione del messaggio e mantenendo le



dimensioni dei send e receive socket buffer costanti e pari a 65536 byte. Tale tipo di configurazione è stato scelto in base ai risultati emersi dalle sessioni di prova in ambito locale. Il grafico in Figura 18 è relativo ai test effettuati generando traffico dal CNAF sui VP verso Roma e Torino.

Figura 18: connessioni TCP dal CNAF verso Roma e Torino

Il massimo throughput ottenuto con connessione singola TCP è stato di 6 Mbps circa, esso è raggiunto utilizzando messaggi di dimensione maggiore o uguale a 100 byte. Questo mette in evidenza che la dimensione del messaggio ha una scarsa incidenza sul throughput. Le prestazioni del protocollo sono inferiori solo nel caso dell'utilizzo di una Sun come stazione ricevente a causa della minore efficienza dell'implementazione dello stack TCP/IP, come già sottolineato. I risultati mettono in evidenza che le prestazioni dipendono strettamente dalle stazioni utilizzate. Ciò emerge anche dalle prove eseguite da Roma, in cui questa volta la stazione SUN è stata impiegata come nodo mittente.

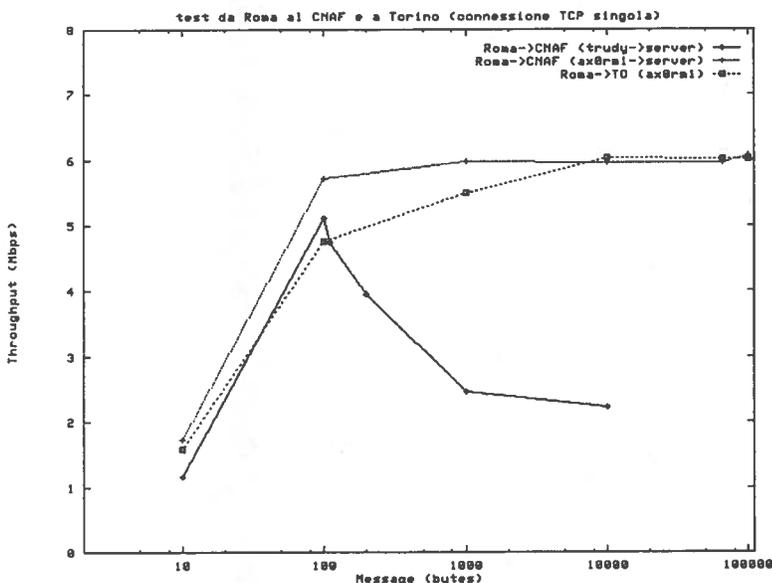


Figura 19: test da Roma al CNAF e a Torino

Il grafico in Figura 19 illustra che il throughput massimo ottenuto sul VP Roma ↔ Torino rimane attestato sui 6 Mbps, come nella prova precedente. Tuttavia scegliendo una Sun come stazione mittente il throughput aumenta oltre i 5 Mbps per messaggi fino a 100 byte, ma poi decresce in modo anomalo.

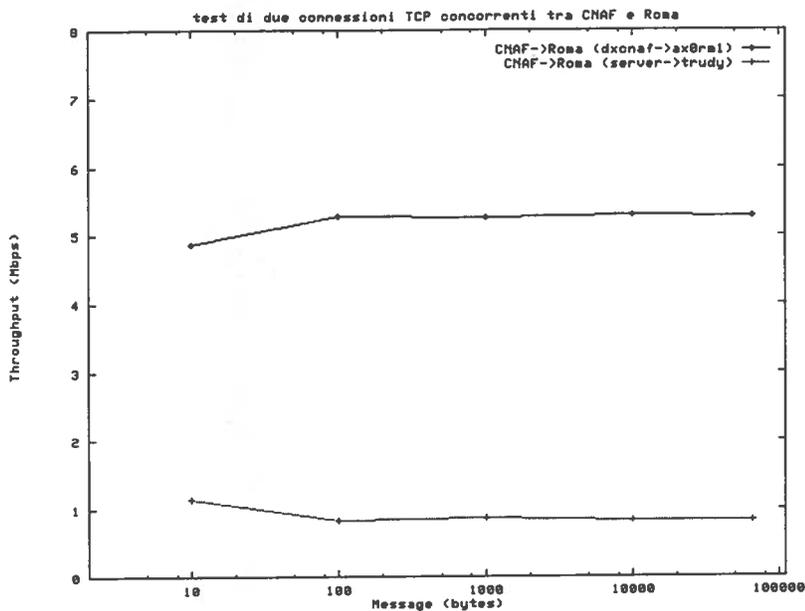


Figura 20: connessioni concorrenti tra CNAF e Roma

Il terzo gruppo di test è stato effettuato tra Roma e il CNAF generando due connessioni concorrenti per verificare la ripartizione del throughput fra i due flussi e misurarne il valore aggregato. Il grafico in Figura 20 illustra le prestazioni di ciascuna connessione. Innanzitutto il throughput aggregato delle due

rimane nel complesso attestato a 6 Mbps (come nel caso di una singola connessione) senza registrare un aumento verso il valore massimo 8 Mbps, come ci si sarebbe invece aspettati. Inoltre, la connessione verso la stazione SUN ottiene un throughput notevolmente più basso, indipendentemente dalla dimensione del messaggio. Anche in questo caso il comportamento della stazione SUN è stato anomalo.

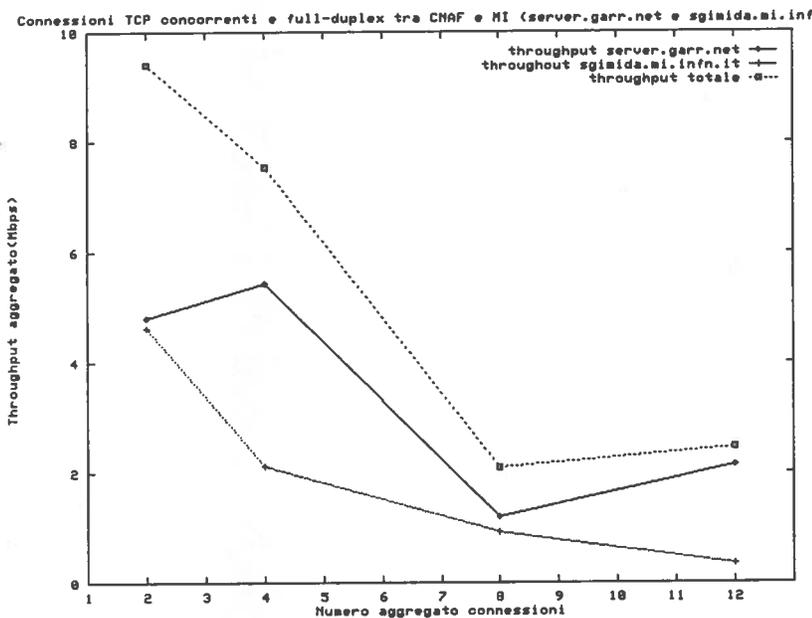


Figura 21: connessioni full-duplex tra CNAF e Milano

Il quarto gruppo di test (tra CNAF e Milano) è stato svolto per verificare l'efficienza delle connessioni TCP in presenza di traffico istantaneo in entrambe le direzioni (CNAF → Milano e Milano → CNAF). Il grafico in

Figura 21 evidenzia la variazione del throughput ottenuto in ciascuna direzione al crescere del numero

complessivo di connessioni, e indica la differenza di comportamento tra la Silicon Graphics e l'Alpha. Per tutti i flussi generati a Milano dalla workstation Silicon Graphics si ottiene un throughput inferiore a quello raggiunto dalla stazione Alpha presente al CNAF. Inoltre, la curva del throughput totale mette in evidenza che le prestazioni complessive calano al crescere del numero delle connessioni. Con dodici

connessioni concorrenti il throughput totale non supera i 3 Mbps a fronte di una capacità disponibile in ciascuna direzione di 8 Mbps (cioè di 16 Mbps massimi teorici).

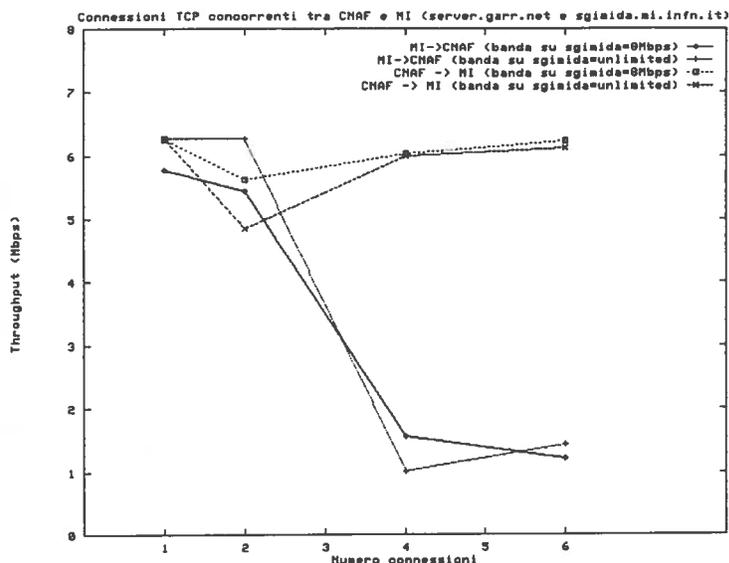


Figura 22: connessioni half-duplex concorrenti tra CNAF e Milano

Altre prove sono state eseguite tra CNAF e Milano per verificare la variazione delle prestazioni in presenza di connessioni concorrenti half-duplex. Il test è stato ripetuto in entrambe le direzioni (CNAF → Milano e Milano → CNAF) e anche con due tipi di configurazioni diverse. Nella stazione di Milano, dotata di adapter FORE, è possibile configurare la banda del PVC. I test sono ripetuti configurando una banda per PVC in un caso pari a 8 Mbps (la banda del VP su SIRIUS) e nell'altro assegnando il valore *unlimited*. Il confronto dei due risultati ha indicato che il valore della banda sull'end-node non è rilevante. In entrambi i casi, il throughput aggregato ottenuto dalla stazione Alpha alla Silicon Graphics è superiore a 6 Mbps, mentre nella direzione opposta (da Silicon ad Alpha) le prestazioni diminuiscono al crescere del numero di connessioni (vedi Figura 22). In conclusione, le prove hanno messo in evidenza che la Silicon Graphics gestisce in modo molto inefficiente la moltiplicazione di più connessioni.

3.2.2 Test di carico tra quattro poli (hardware omogeneo e topologia magliata)

I test sono stati effettuati utilizzando varie topologie nei siti collegati a SIRIUS. Il CNAF accedeva direttamente all'infrastruttura pubblica attraverso lo switch LS1010, mentre le sezioni di Roma e Torino erano collegate via router CISCO. In particolare a Torino era presente lo switch LS1010 (connesso in cascata al CISCO) mentre a Roma era utilizzato lo switch LS2020. La sezione di Milano era l'unica ad avere uno switch FORE attestato a SIRIUS. L'obiettivo era la verifica del massimo throughput aggregato ottenibile su un singolo punto di accesso a 34 Mbps all'infrastruttura pubblica (su cui erano configurati 3 VP a 8 Mbps payload).

La Tabella 4 mostra le prestazioni massime ottenute scegliendo il CNAF come polo ricevente e gli altri tre come mittenti.

Message size	Throughput a Milano(Mbps)	Throughput a Roma (Mbps)	Throughput a Torino (Mbps)	Throughput aggregato (Mbps)
10	2.89	0.70	5.86	9.45
100	6.20	3.07	6.02	15.29
1000	5.92	5.31	6.02	17.25
10000	5.47	6.02	6.00	17.49
65536	5.74	6.03	6.02	17.79
100000	5.99	6.04	6.03	18.06
500000	5.38	5.96	5.91	17.25

Tabella 4: traffico da Milano, Roma e Torino verso il CNAF

Il massimo throughput è prossimo a 18 Mbps ed è inferiore al massimo teorico di 24 Mbps. L'inefficienza nell'utilizzo della capacità su una connessione ATM geografica è stata confermata anche dai risultati ottenuti durante i test su VP singolo.

I risultati del test ottenuto utilizzando tre nodi mittenti (a Milano, Roma e Torino) e due stazioni riceventi a Bologna, sono illustrati in Tabella 5. Il throughput così misurato è all'incirca uguale a quello ottenuto con una sola stazione ricevente (vedi Tabella 4).

Message size	Throughput a Milano (Mbps)	Throughput a Roma (Mbps)	Throughput a Torino (Mbps)	Throughput aggregato(Mbps)
10	1.68	0.72	5.82	8.22
100	5.58	3.19	6.01	14.78
1000	5.89	5.55	6.02	17.46
10000	6.02	6.06	6.05	18.13
65536	5.77	6.06	5.97	17.80
100000	5.78	6.05	5.99	17.82
500000	5.70	6.06	6.00	17.76

Tabella 5: due stazioni riceventi al CNAF e stazioni mittenti a Milano, Roma e Torino

3.3 Considerazioni

Il grado di utilizzo della banda dipende strettamente dalla configurazione delle applicazioni e dall'efficienza degli apparati terminali. Il protocollo UDP rivela che la capacità di un canale ATM può essere saturata. Tuttavia l'efficienza delle connessioni TCP dipende dal tipo di prova.

Nel caso dei test svolti in presenza di router la saturazione della banda può essere ottenuta solo utilizzando dimensioni dei socket buffer appropriate rispetto al valore del RTT. In questo caso mentre in ambito locale la dimensione del messaggio e l'opzione TCP_NODELAY hanno una forte incidenza sulle prestazioni, su una connessione ATM geografica con capacità nell'intervallo [0..24Mbps] essi sono ininfluenti. Il numero di connessioni TCP può incidere sull'utilizzo aggregato della banda solo nel caso in cui il RTT sia maggiore o uguale di 40 msec (vedi il paragrafo 3.1.3). Se nelle workstation non è

possibile assegnare ai buffer dimensioni sufficientemente ampie, il pieno utilizzo della banda è ottenuto solo in presenza di più flussi TCP concorrenti.

Per quanto riguarda invece le prove condotte con router CISCO, è emersa l'incapacità di saturare la banda del VP ATM indipendentemente dal numero di connessioni TCP e dal numero di VP configurate.

Sugli adapter ATM, per minimizzare la percentuale di celle perse è fondamentale applicare la funzione di shaping, attraverso la quale i *burst* di celle generati da una applicazione vengono adeguatamente "spaziati" a seconda della velocità di connessione allo switch. In assenza di tali funzioni, le prestazioni di un flusso TCP dipendono anche dalla topologia di rete impiegata, per esempio dal tipo di apparato attestato sull'accesso alla rete pubblica, e si possono registrare prestazioni diseguali a seconda della direzione del flusso.

4 Test sulla segnalazione ATM su infrastruttura JAMES

Mentre tutti i test descritti fino ad ora sono stati condotti su connessioni ATM permanenti (a causa della mancanza di standard sulla segnalazione ai tempi dei test), per un utilizzo efficiente di ATM è necessario sfruttare una delle caratteristiche portanti di questo protocollo: la capacità di creare dinamicamente su richiesta di una applicazione una connessione (cioè un *Switched Virtual Circuit* – SVC). Questo permette di passare da una rete configurata in modo completamente statico (e quindi non scalabile) a una rete dinamica, ottenibile attraverso l'utilizzo del protocollo di segnalazione detto *signalling*. Test su SVC sono stati condotti sia in ambito locale a Milano che geografico su infrastruttura JAMES [7] in collaborazione con i partner della task force *tf-ten* [6]. Lo scopo principale è stato la misura del tempo impiegato per creare un SVC. Le prove sono state eseguite utilizzando routing ATM statico e utilizzando sempre lo standard UNI 3.0.

A livello locale le prove sono state condotte ripartendole in due sessioni. Nel primo gruppo sono stati effettuati test tra una Sparc Station 20, una Silicon Graphics Indy e uno switch FORE ASX-200, mentre nella seconda sessione lo switch FORE è stato sostituito da un LS1010. A livello geografico è stata utilizzata una *overlay network* costituita da link a 2 Mbps fra le varie nazioni. In particolare, poiché su JAMES non è disponibile il servizio SVC, è stato necessario effettuare il *tunnelling* dell'informazione scambiata dal protocollo di segnalazione. Per questo scopo, il VC dedicato normalmente utilizzato dal protocollo è stato commutato e incanalato utilizzando il VPI assegnato dalle PTT. In questo modo il VP CBR in JAMES è stato utilizzato come semplice strumento di trasporto. I test geografici sono stati condotti sia utilizzando esclusivamente il FORE sia connettendo in cascata il FORE allo switch LS1010.

La metodologia utilizzata per la misura del tempo di *set-up* si è basata sull'applicazione *ping*. Il tempo di Round Trip Time (RTT) del primo pacchetto è all'incirca pari alla somma del tempo di set-up dell'SVC

e del tempo di trasmissione e propagazione. Poiché quando il secondo pacchetto viene trasmesso l'SVC è già stato attivato, il secondo RTT non include più il tempo di set-up e quest'ultimo parametro può essere determinato computando la differenza fra il RTT del primo e del secondo pacchetto generato da ping. Le prove sono state ripetute attendendo il rilascio automatico dell'SVC (*sleep time*) e riattivando poi il ping. Il tempo fra due prove consecutive è un parametro importante nella misura del tempo di set-up.

Test locali: utilizzando lo switch FORE il tempo di RTT del primo pacchetto risulta dipendente dal valore dello sleep-time, infatti nel caso di sleep-time pari a 1 sec la popolazione degli eventi risulta distribuita intorno a due valori di RTT, mentre utilizzando sleep-time pari a 10 sec essa è principalmente distribuita intorno ad un unico valore. Viceversa, utilizzando lo switch LS1010 e aggiornando il driver FORE su SGI alla versione 4.0 si ottiene una popolazione distribuita intorno ad un unico valore di RTT. I risultati dei test sono sintetizzati nelle seguenti tabelle:

tempi medi	Sun-Cisco	Sun-SGI	Sun-FORE	SGI-Cisco	SGI-Sun	SGI-Fore
RTT (primo valore)	23 msec	17 msec	20 msec	21 msec	18 msec	18 msec
RTT (secondo valore)	34 msec	27 msec	31 msec	34 msec	29 msec	29 msec
RTT IP	1 msec	1 msec	1 msec	1 msec	1 msec	1 msec
Set-up (primo val.)	22 msec	16 msec	19 msec	20 msec	17 msec	17 msec
Set-up (secondo val.)	33 msec	26 msec	30 msec	33 msec	28 msec	28 msec

Tabella 6 : test con switch FORE

tempi medi	Sun → SGI	SGI → Sun
RTT	30 msec	21 msec
RTT IP	1 msec	1 msec
Set-up time	29 msec	20 msec

Tabella 7: test con switch LS1010

Test geografici: per la misura del set-up time su infrastruttura geografica sono state effettuate prove con l'Austria, la Svizzera, la Germania e il Belgio. Anche in questo caso la distribuzione dei tempi di RTT risulta concentrata intorno a due valori principali. Inoltre, a seconda dello sleep-time impiegato, le distribuzioni sono diverse visto che all'aumentare dello sleep-time esse si spostano dal primo valore al secondo. Lo stesso set di test è stato ripetuto utilizzando a livello locale sia lo switch FORE, sia lo switch LS1010, ma a differenza delle prove in ambito locale, non è stata notata alcuna differenza. La percentuale di fallimento del ping ⁶ è stata tra il 13% e il 15%, mentre nel caso dei test locali essa è stata pari al 10%. I risultati dei test sono sintetizzati nella seguente tabella.

⁶ Un ping fallisce quando nessun pacchetto ICMP viene spedito correttamente.

tempi medi (msec)	CISCO 7010 (AT)	SUN(CH)	SUN(DE)	CISCO 7010 (DE)	CISC (BE)
RTT (primo valore)	74	42	121	125	207
RTT (sec. valore)	84	51	129	133	216
RTT IP	18	7	47	47	74
Set-up (primo val.)	56	35	74	78	133
Set-up (sec. valore)	66	44	82	86	142

Tabella 8 : test geografici da workstation Sun (IT)

Per il calcolo del set-up time è stato necessario assumere che il tempo di attivazione di un SVC tra il sistema A e il sistema B sia equivalente al caso simmetrico (set-up time tra B e A), assunzione non del tutto corretta perché i test hanno dimostrato che i tempi di set-up non sono simmetrici, essendo essi dipendenti dalla versione dei driver utilizzata. La seconda ipotesi è che in presenza di sistemi intermedi il tempo complessivo di set-up sia la somma dei tempi di set-up introdotti distintamente da ciascun apparato.

Le conclusioni tratte da queste prove sono le seguenti. Innanzitutto il tempo di attivazione di un SVC è un valore abbastanza alto, soprattutto in caso di reti geografiche. Il tempo di set-up medio per uno switch è di circa 5-10 msec, mentre per una workstation esso è tra 10 e 20 msec. Inoltre, la distribuzione dei tempi di set-up è funzione del tipo di hardware e software utilizzati nei test, del livello di carico delle workstation e del tempo trascorso tra il rilascio e la creazione di un nuovo SVC (sleep-time).

5 ATM su ponte radio

Sono stati effettuati lo studio di fattibilità e la progettazione di una connessione ATM a 34 Mbps tramite ponte radio tra i Laboratori Nazionali di Frascati e la sezione INFN di Roma1 [8]. L'analisi teorica delle problematiche dovute alla configurazione geografica ha dimostrato che la connessione di siti remoti via ponte radio è praticabile e risulta economicamente vantaggiosa.

Sono stati condotti test di laboratorio sull'andamento della Bit Error Rate (BER) in funzione della attenuazione da pioggia ed essi hanno dimostrato che la connessione resta affidabile anche nelle peggiori condizioni atmosferiche registrate negli ultimi tempi. In particolare, l'analisi sperimentale è stata condotta utilizzando i protocolli TCP/IP e studiandone le prestazioni in funzione del BER. È stata analizzata la relazione fra BER e S/N (*Signal/Noise*) e tale valore è stato poi confrontato alle previsioni teoriche. A questo scopo in laboratorio sono stati effettuati due tipi di test.

In una prima prova sono stati utilizzati un ponte radio per uso video modificato che trasmetteva a 2 Mbps, VAX Station modelli 4000/90 e 4000/60, un attenuatore, convertitori di segnale di tipo SAT 2001 e un analizzatore di spettro. Durante questi test il protocollo TCP/IP ha presentato la massima efficienza (1.8 Mbps) in caso di BER < 10⁻⁶, mentre per BER pari a 10⁻⁵ TCP/IP mantiene la connettività sia pure con prestazioni scadenti. I medesimi risultati sono stati confermati in una seconda prova realizzata collegando due workstation (una AlphaStation 255/233 e una DECstation 3000/400)

dotate di interfacce ATM OC-3c (155 Mbps) a uno switch LS2020. Due porte a 34 Mbps sono state a loro volta collegate alla stazione trasmittente e ricevente del ponte radio. Inoltre sono stati utilizzati vari attenuatori del segnale generato dal ponte radio, di cui uno di valore variabile, per simulare l'attenuazione dovuta alle condizioni atmosferiche. Il throughput della connessione TCP è risultato massimo per un BER $< 10^{-8}$ (circa 30 Mbps). Con BER= 10^{-6} il throughput si dimezza (circa 15 Mbps) e diminuisce rapidamente fino a raggiungere un valore quasi nullo per valori del BER compresi nell'intervallo $[10^{-4}..10^{-5}]$. Ciò permette di affermare che il protocollo TCP/IP rimane ampiamente funzionale (con un margine di 13 dB sulla zona di degradazione) anche in caso di condizioni atmosferiche sfavorevoli (pioggia intensa). Infatti, in questo caso la potenza in ricezione è all'incirca -57 dBm, il che corrisponde ad un rapporto S/N pari a 40 dB.

A seguito di misure di occupazione spettrale è stata scelta come frequenza di lavoro 14 GHz. La lunghezza d'onda utilizzata corrispondente è $\lambda=2.09$ cm (considerando il valore medio della banda pari a 14,375 GHz). Dati questi parametri, è possibile calcolare l'attenuazione da spazio libero, che ammonta a 139,4 dB, e l'attenuazione atmosferica, che risulta 0.25 dB. Considerando le peggiori condizioni atmosferiche la potenza ricevuta si può quindi stimare pari a -53,5 dBm e il rumore pari a -91 dBm, ottenendo così un rapporto S/N \approx 38 dB. Essendo le funzionalità di TCP/IP garantite per un valore del rapporto S/N > 29 dB, il livello del segnale nelle peggiori condizioni atmosferiche fornisce ugualmente un margine di 8 dB rispetto al limite teorico.

Oltre a risultare fattibile, questa infrastruttura si rivela anche vantaggiosa dal punto di vista economico. Secondo le stime eseguite, su brevi tratte che non richiedono l'utilizzo di ripetitori (circa 30 Km), il costo di una connessione su ponte radio non supera i 150 ML iniziali includendo manutenzione e acquisto delle apparecchiature. Anche su lunghe tratte (500 Km) il costo iniziale della connessione, stimabile in 2149,2 ML, risulta facilmente ammortizzabile se confrontato al costo del servizio ATM offerto da Telecom Italia.

Considerando lo stato attuale di questo esperimento, le successive fasi saranno il completamento dell'installazione degli apparati nei siti predestinati e la verifica delle effettive prestazioni del ponte radio così implementato.

6 Utilizzo dell'infrastruttura ATM per il traffico di produzione

Le prove fatte hanno evidenziato che il servizio ATM VP/CBR a livello geografico ha le caratteristiche di affidabilità ed efficienza necessarie per un utilizzo di produzione. Inoltre è emerso che questo servizio funziona in ambiente di switch ATM *multivendor* e che l'implementazione del traffic shaping sui router CISCO è sufficientemente efficace. Da ciò è scaturita la decisione di realizzare un backbone virtuale di produzione per la rete GARR, chiamato GARR-B pilota. Esso è basato su quattro accessi a 34 Mbps alla rete SIRIUS per il traffico nazionale nei siti INFN/GARR di Milano, Torino, Roma e CNAF, e un quinto accesso a 2*34Mbps nella sede Telecom di Milano per connettere la rete GARR a

TEN-34 ad una velocità di 34Mbps payload. Sugli accessi ATM sono state definite due topologie di VP: una per il traffico nazionale e una per quello internazionale, come illustrato nella Figura 23. In particolare, i VP *nazionali* costituiscono una rete completamente magliata tra Bologna, Milano, Roma e Torino. Le capacità di tali connessioni sono: 6 Mbps sui VP Bologna-Milano e Bologna-Roma, 8 Mbps sui VP Milano-Torino e Roma-Torino.

I VP *internazionali* sono quattro e vengono utilizzati per convogliare il traffico internazionale dai quattro poli verso l'accesso a 34 Mbps di TEN-34. I VP hanno le seguenti capacità: 10 Mbps da Roma, 6 Mbps da Milano e da Torino e 18 Mbps da Bologna. La banda complessiva dei VP internazionali, pari a 40 Mbps, è stata così fissata per saturare al massimo la capacità di accesso a TEN-34 che è di 34 Mbps payload. Per ottenere ciò il router internazionale del GARR è stato collocato in sede Telecom e collegato *back-to-back* al router di TEN-34 con una interfaccia a 155Mbps.

Attraverso il progetto pilota GARR-B sono stati raggiunti i principali obiettivi legati alla progettazione della rete ad alta banda dell'università e della ricerca scientifica, GARR-B, e alla connessione a TEN-34. Tale rete pilota è stata integrata nell'attuale rete GARR-2, anticipando la fase iniziale della migrazione a GARR-B. Inoltre sono state realizzate le prove di accettazione di TEN-34 ed è stato ottenuto un adeguato accesso alla rete europea in attesa della realizzazione di GARR-B.

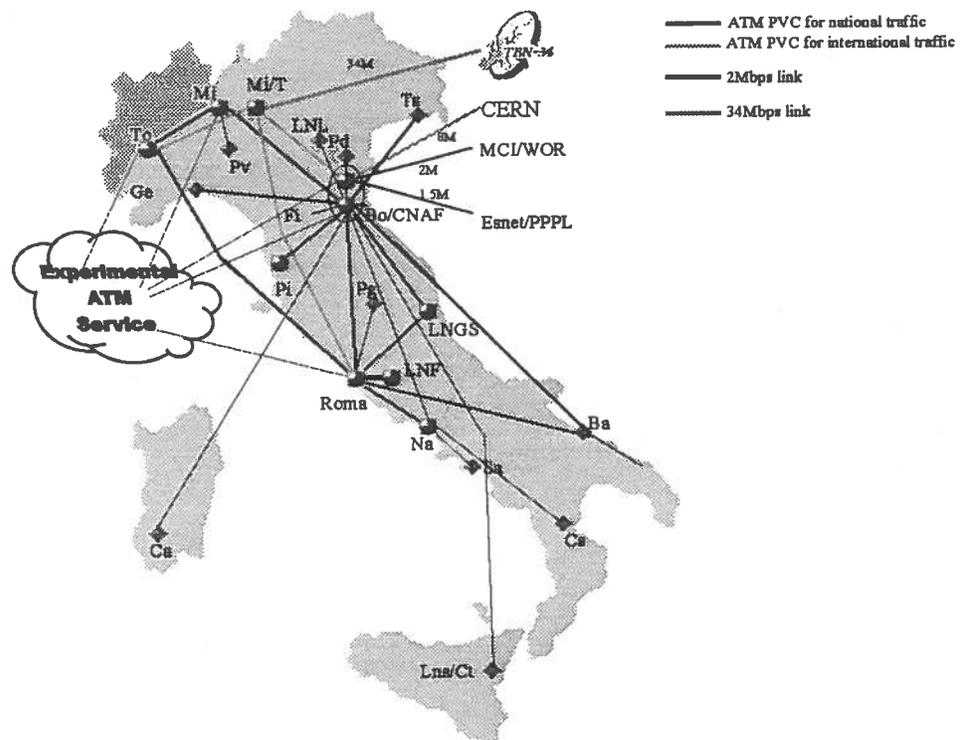


Figura 23: struttura dei VP ATM per il trasporto del traffico di produzione nazionale ed internazionale

7 Evoluzione della sperimentazione

L'attività di sperimentazione fin qui descritta è principalmente focalizzata sulle prestazioni dei protocolli TCP/IP e UDP su reti locali e geografiche, sulla verifica di funzionalità, efficienza e interoperabilità di switch ATM, router e workstation multivendor e sullo studio di situazioni di congestione per analizzare il problema della perdita delle celle. Ciò è stato fatto sui servizi di Permanent Virtual Circuit sia di tipo CBR (a livello geografico) che di tipo VBR e ABR (a livello locale). Il programma delle future prove sperimentali è stato definito per indagare alcuni degli aspetti più avanzati della tecnologia ATM e per porre maggiormente l'accento sulle applicazioni che possono trarre beneficio dalla disponibilità di larga banda. In particolare verranno condotte, anche in collaborazione con gli altri enti della ricerca europei, prove più estese sull'utilizzo del protocollo di segnalazione in ambito geografico, sul protocollo PNNI (Private Network Node Interface, che permette l'utilizzo di connessioni ATM geografiche dinamiche attraverso il protocollo di routing ATM) e di LAN Emulation geografica. Verrà inoltre studiata l'applicazione del protocollo *RSVP* (Resource reSerVation Protocol) per l'introduzione della Quality of Service (QoS) per specifiche applicazioni, del *Tag Switching* e dell' *IP switching* per l'ottimizzazione del routing IP su infrastruttura ATM. Infine, verrà analizzata la prestazione del protocollo ATM nativo. Per quanto riguarda le applicazioni, verrà studiato l'utilizzo di un sistema per il calcolo distribuito denominato *CONDOR*, l'utilizzo di applicazioni per videoconferenza su ATM nativo e di Web su connessioni ATM con QoS.

8 Conclusioni

L'obiettivo principale della sperimentazione su ATM fin qui condotta è stato la verifica della sua validità e dei limiti di impiego sia a livello locale che geografico per la progettazione di reti ad alta banda. I risultati sono stati positivi e ne è una dimostrazione l'attuale funzionamento della rete TEN-34 e della rete pilota GARR-B. Ciò è stato possibile dopo un attento lavoro di calibrazione delle configurazioni degli apparati ATM. Da esso sono emersi vari malfunzionamenti a livello software e hardware. In particolare è stata sperimentata l'inadeguatezza dello switch LS2020 su link geografici, la mancanza di capacità di traffic shaping delle interfacce ATM sulla maggior parte delle workstation e l'impossibilità di definire dei parametri di controllo di traffico sui router CISCO (per consentire un utilizzo dell'intera banda a disposizione). Mentre a livello geografico sono stati utilizzati VP/VC permanenti di tipo CBR (Constant Bit Rate) e commutati di tipo UBR (Unspecified Bit Rate), a livello locale la sperimentazione ha anche riguardato i servizi VBR (Variable Bit Rate) e ABR (Available Bit Rate). I vari test sono stati pianificati via via secondo la continua evoluzione degli standard ATM e secondo la disponibilità di infrastrutture in grado di implementare tali specifiche. Le prestazioni sono state buone a livello locale fra workstation (throughput fino a 135 Mbps su connessione OC-3c) e accettabili a livello geografico (6 Mbps su VP a 8 Mbps in presenza di router e 22 Mbps su VP a 24 Mbps senza router).

Un secondo importante obiettivo è stato l'utilizzo di ATM geografico per le applicazioni specifiche degli esperimenti. L'utilizzo di canali virtuali commutati (SVC) è per ora ancora immaturo, poiché

necessita di ulteriori progressi nelle funzionalità degli switch ATM e nella implementazione delle interfacce verso l'utente (sia sui router che sulle workstation). I test in corso su LAN Emulation geografico e applicazioni native forniranno ulteriori strumenti al fine di valutare l'adeguatezza del servizio SVC per la comunità scientifica dell'INFN. Un importante obiettivo già raggiunto è stato lo studio di progettazione e test per la realizzazione di un ponte radio ATM fra i laboratori nazionali di Frascati e l'Università La Sapienza di Roma.

Finora la maggior parte degli sforzi sono stati impegnati per raggiungere l'efficienza nell'uso di questa tecnologia per reti basate su protocollo TCP/IP. In futuro le sperimentazioni saranno rivolte allo studio e all'applicazione di servizi più avanzati, quali l'SVC o il PVC di tipo ABR, e in generale si orienteranno verso la valutazione di tutte le evoluzioni di questa tecnologia.

9 Ringraziamenti

Si ringraziano per la loro collaborazione nelle sperimentazioni: U.Zanotti (INFN/CNAF), F.Niccolai e S.Pierattini (ENEA), E.Rossetto e S.Lusso (INFN/Torino), D.Colombo e S.Maggi (INFN/Milano).

BIBLIOGRAFIA

- [1] The information superhighway for European R&D, <http://www.dante.net/ten-34.html>.
- [2] JAMES: Joint ATM Experiment on European services, <http://www.labs.bt.com/profsoc/james>.
- [3] GARRB: Progetto di rete a larga banda per le università e la ricerca scientifica italiana, <http://www.garr.net/garr-b>.
- [4] ATM in ambiente scientifico: analisi e valutazione delle tecnologie e dei protocolli "cell switching" nel contesto della rete dell'INFN, Tesi di laurea di Tiziana Ferrari, <http://www.cnaf.infn.it/WWWpages/tesi.html>, marzo 1995.
- [5] Permanent virtual circuits configuration and TCP-UDP/IP performances in a local ATM network, C.Battista, M.Campanella, T.Ferrari, A.Ghiselli, C.Vistoli, Nota interna, <http://www.cnaf.infn.it/ferrari>, luglio 1995.
- [6] Deliverable D11.2, Results of Phase 1 Test Programme (TEN-34), <http://www.dante.net/ten-34/DELIVERABLES/D11.2/>, dicembre 1996.
- [7] Study of SVC set-up times, S.Alborghetti, M.Campanella, D.Colombo, S.Maggi, giugno 1997.
- [8] Progetto e realizzazione di una connessione ATM a 34 Mbps via ponte radio fra i Laboratori Nazionali di Frascati e la sezione INFN di Roma1, A.Brunengo, T.Ferrari, C.Olivieri, C.Salvo, 1996.
- [9] TCP/IP Illustrated, Volume 1 The Protocols, W.Richard Stevens, Addison-Wesley.
- [10] RFC 1122, Requirements for Internet Hosts – Communication Layers -, Ottobre 1989.

1	INTRODUZIONE	2
2	ATM IN AMBITO LOCALE	2
2.1	GIGASWITCH/ATM DIGITAL	3
2.1.1	<i>Studio della configurazione ottimale dei parametri del protocollo TCP</i>	4
2.1.2	<i>Studio delle prestazioni di connessioni TCP multiple</i>	6
2.2	FORERUNNER ASX-200/FORE SYSTEMS	7
2.2.1	<i>PVC singolo (tra SUN e Silicon Graphics Indy)</i>	7
2.2.2	<i>Connessioni multiple su PVC singolo (tra SUN e Silicon Graphics Indy)</i>	8
2.3	LIGHTSTREAM LS2020/CISCO.....	8
2.4	LIGHTSTREAM LS1010/CISCO.....	10
2.5	CONSIDERAZIONI.....	10
3	ATM IN AMBITO GEOGRAFICO	11
3.1	TEST SU SINGOLO VP ATM.....	12
3.1.1	<i>Test su VP nazionale</i>	12
3.1.2	<i>Test su VP internazionale</i>	12
3.1.3	<i>Test sull'infrastruttura JAMES</i>	13
3.2	TEST SU TOPOLOGIA MAGLIATA DI VP ATM.....	17
3.2.1	<i>Test di funzionalità tra CNAF, Milano, Roma e Torino (hardware eterogeneo)</i>	17
3.2.2	<i>Test di carico tra quattro poli (hardware omogeneo e topologia magliata)</i>	20
3.3	CONSIDERAZIONI.....	21
4	TEST SULLA SEGNALAZIONE ATM SU INFRASTRUTTURA JAMES	22
5	ATM SU PONTE RADIO	24
6	UTILIZZO DELL'INFRASTRUTTURA ATM PER IL TRAFFICO DI PRODUZIONE	25
7	EVOLUZIONE DELLA SPERIMENTAZIONE	27
8	CONCLUSIONI	27
9	RINGRAZIAMENTI	28