

**Analisi delle prestazioni del collegamento di rete  
tra i laboratori esterni e quelli sotterranei dei  
LNGS**



# **Analisi delle prestazioni del collegamento di rete tra i laboratori esterni e quelli sotterranei dei LNGS**

Piero Spinnato<sup>a</sup>

<sup>a</sup> *INFN - LNGS, S.S. 17/bis km 18+910, 67010 Assergi (AQ), ITALY*

## **Sommario**

Il collegamento di rete tra i laboratori esterni e quelli sotterranei dei LNGS è assicurato da un fascio di fibre monomodo, sulle quali la tecnologia di rete attuale rende possibile una velocità di trasmissione dati nominale pari a 10 gigabit al secondo (Gb/s). Per verificare l'effettiva capacità di tale connessione di supportare queste prestazioni, sono state compiute delle misurazioni, allo scopo di testare anche i dispositivi utilizzati per creare l'infrastruttura di rete. Le prestazioni misurate si sono rivelate più che soddisfacenti, con valori prossimi ai 6 Gb/s rispetto ad un massimo teorico possibile, limitato dalle risorse hardware a disposizione per le misure, pari a 7,6 Gb/s. E' stato riscontrato che il fattore limitante nelle prestazioni misurate non è da associare alla fibra o ai dispositivi di rete, ma ad alcune delle macchine utilizzate come host per le misure.

# 1 Introduzione

In questo documento è descritta l'attività di misura delle prestazioni di un collegamento in fibra ottica monomodo su una distanza di circa 8 km, tra i laboratori esterni e quelli sotterranei dei Laboratori Nazionali del Gran Sasso. La fibra in questione presenta delle peculiarità, in quanto il fascio steso originariamente nella metà degli anni '80, costituito da fibre monomodo 8/125, è stato oggetto di un intervento nel 2006 per cui una porzione di circa 2 km è stata sostituita con fibre 9/125, le nuove fibre essendo state saldate a quelle vecchie mediante opportuni macchinari. In definitiva, la linea di connessione è costituita, andando dai laboratori esterni a quelli sotterranei, da circa 5,5 km di fibra 8/125, quindi da circa 2 km di fibra 9/125, quindi da rimanenti 500 m di fibra 8/125. Le prove svolte avevano lo scopo di appurare la capacità della linea di supportare un traffico dati su un link con larghezza di banda nominale pari a 10 Gigabit/sec (Gb/s), tenendo anche in considerazione la peculiarità delle fibre appena descritta. I dispositivi utilizzati per il test sono due switch Cisco 3560E-24TD-S. Ognuno dei due switch presenta 24 porte in rame Gigabit Ethernet (GigE) e due slot di uplink X2 10 GigE, uno dei quali dotato di un convertitore X2-10GB-LR con connettore per fibra ottica. Tale modulo fornisce una capacità di trasmissione nominale a 10 Gb/s su una distanza fino a 100 km su fibra monomodo, che quindi è adeguato alle caratteristiche della linea utilizzata. Nella serie di misure effettuate si è voluto simulare uno scenario in cui diverse macchine presenti nei laboratori sotterranei devono trasmettere dati con velocità prossime al Gb/s verso i rispettivi host situati presso i laboratori esterni. Ci si trova quindi in una situazione in cui diversi flussi di traffico al Gb/s sono presenti contemporaneamente sulla linea in fibra, e gli switch devono gestire l'aggregazione sull'uplink di più flussi separati. A tale scopo, è stato predisposto un certo numero di host, ognuno fornito di un'interfaccia di rete GigE, e su ognuno di questi è stato installato un generatore di traffico. E' stato scelto **iperf**[1] (versione 2.0.2) per la sua larga diffusione e la semplicità d'uso. Una misura della larghezza di banda di una connessione mediante **iperf** si esegue facendolo girare sugli host alle estremità della connessione, su di un host in modalità server, sull'altro in modalità client. Facendo girare contemporaneamente tale generatore su più macchine opportunamente connesse ai due switch, ci si pone nelle condizioni di generare un traffico aggregato multigigabit attraverso il collegamento in fibra, come descritto di seguito.

## 2 Configurazione dell'apparato di misura

Le misure sono state svolte generando otto flussi di traffico tra otto coppie di host. Tale valore corrisponde al massimo numero di macchine disponibili al

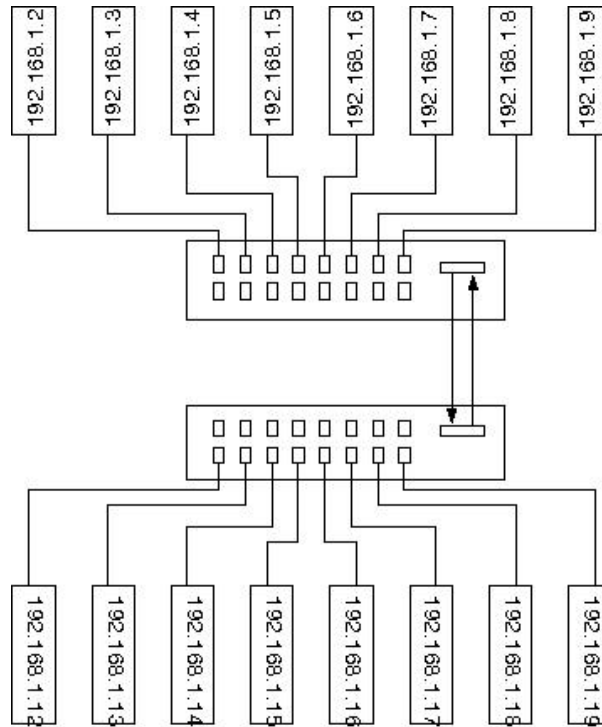


Figura 1: Schema dell'infrastruttura di rete per le misure svolte

momento dei test, aventi scheda di rete GigE. In figura 1 è presentato uno schema della rete locale utilizzata. Come si può vedere in essa, ad ogni switch sono collegate otto macchine, mentre i flussi di traffico avvengono tra coppie di macchine poste non sullo stesso switch, cosicché il traffico tra le due estremità di ogni singola connessione passi attraverso il link in fibra che collega gli switch. A questo scopo, gli host sono accoppiati associando quelli che hanno l'ultima cifra dell'indirizzo IP separata di dieci unità. Come si può vedere dalla figura 1, le due macchine tra le quali passano i singoli flussi non sono mai poste sullo stesso switch. In tabella 1 sono riassunte le caratteristiche hardware e software degli host utilizzati.

L'accesso alle macchine è avvenuto mediante connessioni `ssh` create tramite l'applicazione `clusterssh`,[2] che permette l'apertura contemporanea di terminali su più macchine remote, con la possibilità di esecuzione simultanea di comandi su ognuna di esse (entro le ovvie differenze di tempo di propagazione dei pacchetti verso le singole macchine, comunque stimabile al più nell'ordine del millisecondo, che è trascurabile per le nostre misure). In tal modo, è possibile confrontare le misure prese sulle singole macchine per valutare eventuali effetti di un flusso sull'altro. L'applicazione `clusterssh` era in esecuzione sull'host con ID 1.16 per le connessioni verso gli host presso i laboratori esterni, e sull'host con ID 1.17 per le connessioni verso le macchine situate presso i laboratori sotterranei. Ciò per suddividere sia il carico sulla

	ID	Hardware dell'host	CPU	Cache - RAM	interfaccia di rete	sistema operativo
lab. esterni	1.2	Scheda madre Tyan Thunder K8S (S2880)	2 × AMD Opteron 242, 1,6 GHz	1 MB - 2 GB	Broadcom BCM5704C	Red Hat Linux 8.0.95 - kernel 2.4.20
	1.3	Scheda madre Tyan Thunder K8S (S2880)	2 × AMD Opteron 242, 1,6 GHz	1 MB - 2 GB	Broadcom BCM5704C	Red Hat Linux 8.0.95 - kernel 2.4.20
	1.4	Scheda madre Tyan Thunder K8S (S2880)	2 × AMD Opteron 242, 1,6 GHz	1 MB - 2 GB	Broadcom BCM5704C	Red Hat Linux 8.0.95 - kernel 2.4.20
	1.5	Scheda madre Tyan Thunder K8S (S2880)	2 × AMD Opteron 242, 1,6 GHz	1 MB - 2 GB	Broadcom BCM5704C	Red Hat Linux 8.0.95 - kernel 2.4.20
	1.6	Scheda madre Tyan Thunder K8S (S2880)	2 × AMD Opteron 242, 1,6 GHz	1 MB - 2 GB	Broadcom BCM5704C	Red Hat Linux 8.0.95 - kernel 2.4.20
	1.7	Scheda madre Tyan Thunder K8S (S2880)	2 × AMD Opteron 242, 1,6 GHz	1 MB - 2 GB	Broadcom BCM5704C	Red Hat Linux 8.0.95 - kernel 2.4.20
	1.8	Scheda madre Tyan Thunder K8S (S2880)	2 × AMD Opteron 242, 1,6 GHz	1 MB - 2 GB	Broadcom BCM5704C	Red Hat Linux 8.0.95 - kernel 2.4.20
	1.9	Scheda madre Tyan Thunder K8S (S2880)	2 × AMD Opteron 242, 1,6 GHz	1 MB - 2 GB	Broadcom BCM5704C	Red Hat Linux 8.0.95 - kernel 2.4.20
lab. sotterranei	1.12	Scheda madre Asus K8V-X	2 × AMD Athlon 64 X2 3800+, 2,0 GHz	0.5 MB - 1 GB	Marvell 88E8001	Debian Linux 4.0 - kernel 2.6.18
	1.13	Scheda madre Asus K8V-X	2 × AMD Athlon 64 X2 3800+, 2,0 GHz	0.5 MB - 1 GB	Marvell 88E8001	Debian Linux 4.0 - kernel 2.6.18
	1.14	Scheda madre Asus K8V-X	2 × AMD Athlon 64 X2 3800+, 2,0 GHz	0.5 MB - 1 GB	Marvell 88E8001	Debian Linux 4.0 - kernel 2.6.18
	1.15	Scheda madre Asus K8V-X	2 × AMD Athlon 64 X2 3800+, 2,0 GHz	0.5 MB - 1 GB	Marvell 88E8001	Debian Linux 4.0 - kernel 2.6.18
	1.16	PC HP Pavilion A1679.IT	AMD Athlon 64 X2 4200+, 2,2 GHz	0.5 MB - 1 GB	Realtek RTL8169	Ubuntu Linux 6.06 - kernel 2.6.15
	1.17	Portatile Apple Mac-Book	Intel Core 2 Duo, 2,0 GHz	4 MB - 2 GB	Marvell 88E8053	Mac OS X 10.5.4 - kernel Darwin 9.4.0
	1.18	Portatile HP Compaq nx8220	Intel Pentium M, 1,86 GHz	2 MB - 1 GB	Broadcom BCM5751M	Ubuntu Linux 6.06 - kernel 2.6.15
	1.19	Portatile HP Pavilion dv9000	Intel Core 2 Duo T5600, 1,83 GHz	2 MB - 2 GB	Intel Pro/1000 PL	Ubuntu Linux 6.06 - kernel 2.6.15

Tabella 1: Dati tecnici degli host utilizzati

CPU che l'occupazione di banda su piú host, in modo da limitare l'impatto sulle misure.

Per ogni sessione di misura, sono state lanciate otto istanze di **iperf** in modalità server sulle macchine collegate allo stesso switch. Al fine di porsi nelle condizioni ideali per ottenere un flusso vicino al Gb/s, limite tecnologico delle connessioni in rame utilizzate, si è scelto di far generare ad **iperf** un flusso UDP con dimensione di ogni frame pari a 1470 bytes, tale quindi da entrare nel payload di un singolo pacchetto. Il comando completo lato server è:

```
iperf -s -i 1 -u -l 1470
```

Per quanto riguarda i client, la modalità di trasmissione UDP permette anche di settare la larghezza di banda del flusso,<sup>1</sup>che è stata posta pari a 950 Mb/s, così da generare un flusso prossimo al limite delle possibilità delle macchine, che però non forzasse eccessivamente l'hardware. Il comando completo è (nel caso in cui il client sia 192.168.1.4, che quindi comunica con 192.168.1.14):

<sup>1</sup>Si rammenta che i flussi generati da **iperf** sono diretti per definizione dal client verso il server.

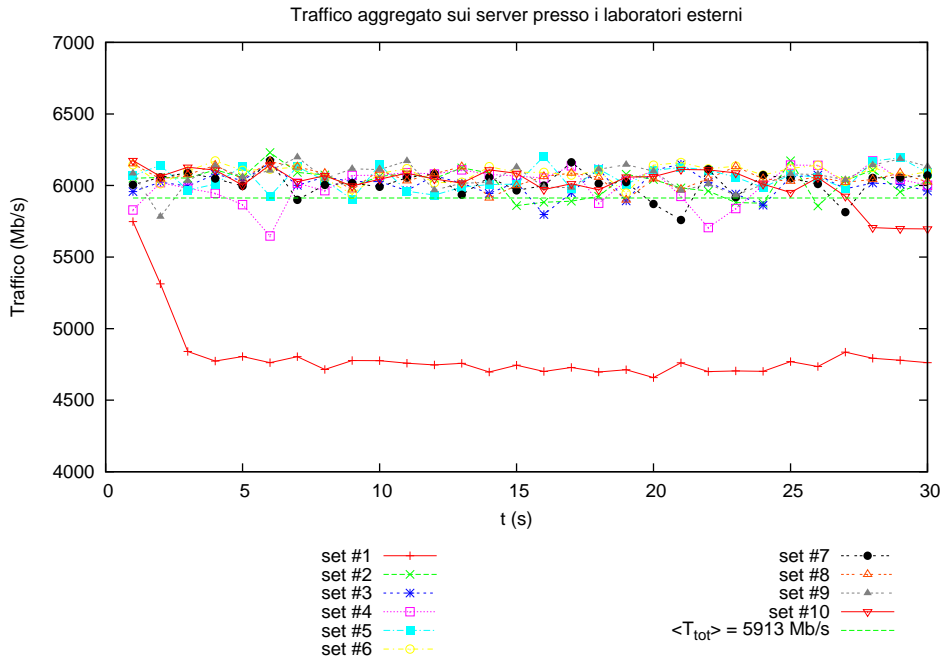


Figura 2: Valori aggregati del traffico misurato sui server posti presso i laboratori esterni.

```
iperf -c 192.168.1.14 -i 1 -t 30 -u -l 1470 -b 950M -r
```

E' stata usata anche l'opzione `-r` tramite la quale alla fine del flusso client  $\rightarrow$  server, viene generato un nuovo flusso in verso opposto. Di seguito sono presentati i risultati delle misure svolte.

### 3 Risultati delle misure

Sono state eseguite cinque sessioni di misurazione con i server attivi sulle macchine situate presso i laboratori esterni (e quindi i client presso i laboratori sotterranei), piú altre cinque con client e server invertiti. Grazie all'utilizzo dell'opzione `-r` che come detto sopra, alla fine di ogni set di misure ne genera un secondo a ruoli invertiti per client e server, si sono cosí ottenute per ogni macchina dieci misure in modalit  client e dieci in modalit  server.

Nei grafici in figura 2 e 3 sono mostrati i valori del traffico aggregato in arrivo sui server istante per istante (i valori misurati da `iperf` hanno una granularit  pari ad un secondo), assieme al valor medio sull'intero tempo di misurazione e su tutti i set di misure. Il singolo valore graficato rappresen-

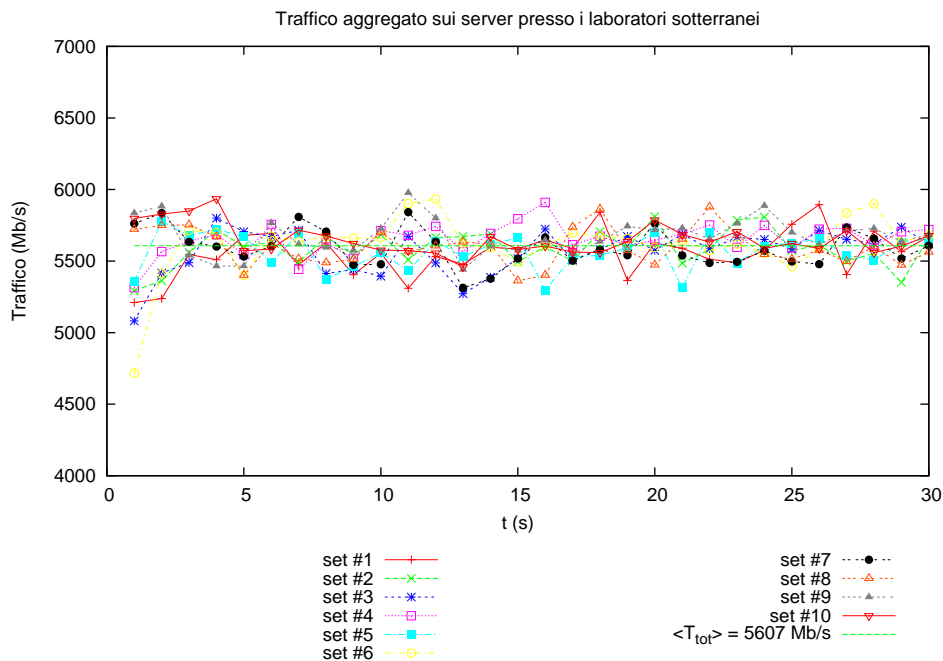


Figura 3: Valori aggregati del traffico misurato sui server posti presso i laboratori sotterranei.

ta quindi il traffico totale ricevuto dai server nell'intervallo di tempo (pari ad un secondo) a cui il valore fa riferimento. Ciò costituisce un limite inferiore per il traffico effettivamente transitato lungo la fibra ottica che collega le due sottoreti. Infatti, per le caratteristiche del protocollo UDP nel quale, contrariamente al TCP, non c'è nessuna verifica dell'effettiva ricezione del pacchetto da parte dell'host di destinazione, e quindi nessun adattamento della velocità di trasmissione da parte dell'host di trasmissione,[3] è possibile che il client trasmetta a velocità maggiore di quanto il relativo server sia in grado di ricevere. In questo caso, una certa percentuale di pacchetti viene persa. Non è stato possibile nel corso delle misure qui descritte verificare l'ammontare di traffico effettivamente transitato su ogni switch, quindi sulla fibra ottica che li connette. Pertanto rimane il valore misurato sui server come stima conservativa del traffico transitato sulla fibra.

La prima osservazione possibile dai grafici in figura 2 e 3 è che i valori medi di traffico (rispettivamente, 5913 Mb/s per il traffico ricevuto dagli host dei laboratori esterni, e 5607 Mb/s per gli host in modalità ricezione presso i laboratori sotterranei) raggiungono un valore pari a circa il 75% del massimo teorico, pari a  $950 \cdot 8 = 7600 \text{ Mb/s}$ . Si nota inoltre una differenza tra le due medie pari a circa il 5%. Infine, gli andamenti dei valori di traffico oscillano di poco attorno ai valori medi, con l'unica eccezione del primo set di misure



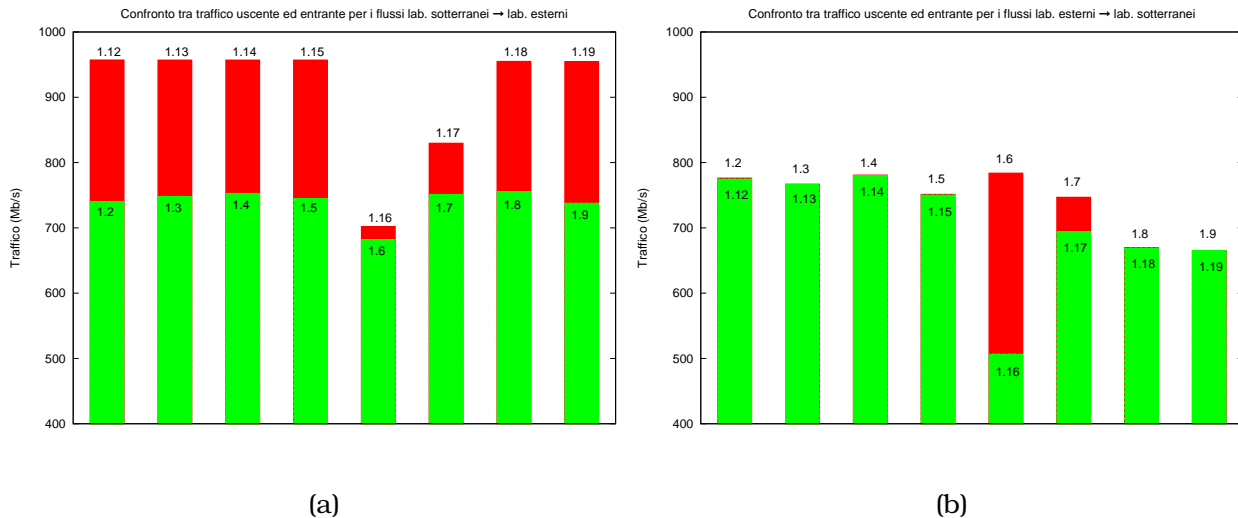


Figura 4: Confronto tra i valori medi di traffico in uscita da ogni singolo client con quelli in ingresso al relativo server per entrambi i sensi di flusso. Le etichette in corrispondenza dei valori identificano gli host (vd. tabella 1).

relativo al traffico verso i server presso i laboratori esterni.

## 4 Analisi dei risultati

Benché i valori di traffico misurati siano più che soddisfacenti, è interessante indagare sulle possibili cause del mancato raggiungimento del picco teorico. L'ipotesi più probabile per spiegare ciò è l'incapacità dell'host che invia o di quello che riceve di gestire un traffico prossimo al Gb/s. Per verificare quest'ipotesi, sono stati confrontati i valori medi del traffico prodotto da ogni client con quello effettivamente ricevuto dal relativo server.

Come è nettamente visibile dalla figura 4(a), gli host situati presso i laboratori sotterranei, funzionanti in modalità client, effettivamente inviano il loro traffico alla velocità attesa di 950 Mb/s, con le due eccezioni degli host 1.16 ed 1.17. La minore velocità di queste due macchine si può giustificare dal fatto che ad esse fanno capo le connessioni `ssh` tramite le quali sono controllati tutti gli altri host. D'altro canto, dalla figura 4(b) si può chiaramente vedere che quando gli host presso i laboratori sotterranei agiscono in modalità server, non c'è nessuna caduta di prestazioni rispetto al traffico inviato dai client,<sup>2</sup> i quali invece mostrano un'incapacità nel raggiungere le prestazioni richieste.

<sup>2</sup>anche qui con le eccezioni di 1.16 ed 1.17, per i quali si possono addurre le stesse argomentazioni avanzate in precedenza.

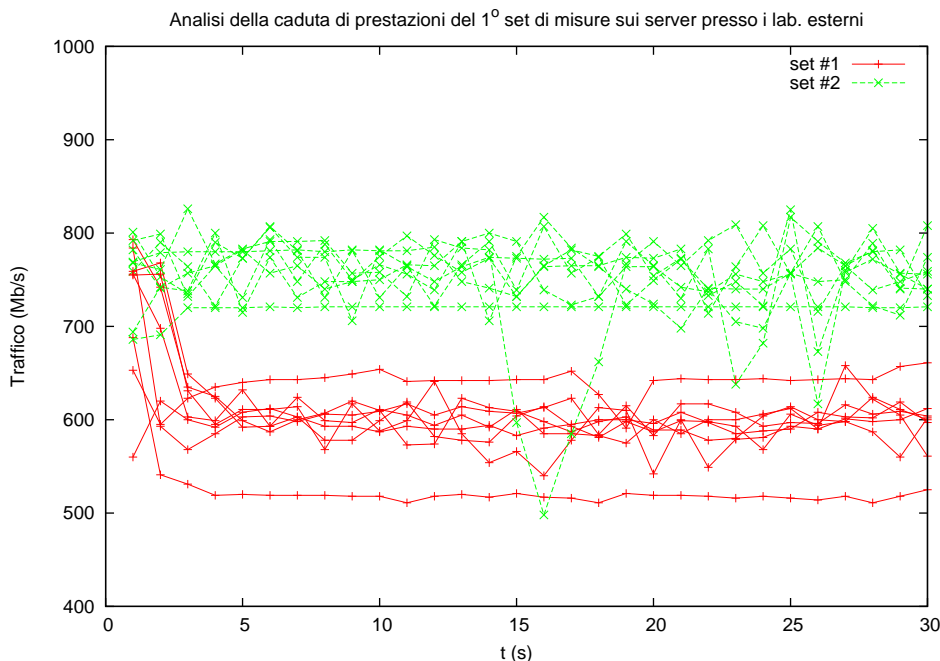


Figura 5: Confronto tra le misure relative al 1° set di misure sui singoli server con quelle di un qualunque altro set (in questo caso il 2°).

Il fattore limitante nel raggiungimento delle prestazioni massime si può pertanto attribuire alle ridotte prestazioni nell'immissione del traffico in rete da parte degli host situati presso i laboratori esterni.

Rimane da discutere l'andamento del primo set di misure di figura 2, per il quale le prestazioni si discostano nettamente da quelle dei rimanenti set. Per identificare la causa della caduta di prestazioni, si può prima di tutto vedere l'andamento delle misure sui singoli host. In figura 5 sono messe a confronto le misure prese sugli otto server posti presso i laboratori esterni nel corso dei primi due set di misure. Tali valori sono, in altre parole, quelli le cui medie sono riportate nei primi due set graficati in figura 2.

Come si può vedere dalla figura, la caduta di prestazioni del primo set di misure coinvolge in modo uniforme tutti gli otto host. Ciò porta ad escludere problemi legati agli host stessi. D'altro canto, una verifica sui dati relativi ai client (dati non mostrati per brevità) mostra che da parte di questi ultimi non c'è nessuna caduta di prestazioni tra il primo set di misure ed i successivi. La causa del fenomeno è quindi da ricercare negli apparati di rete o nella connessione tra essi. Il fatto che la caduta di prestazioni sia stata riscontrata solo su un set di misure, porta ad escludere problemi sulla fibra che connette i due switch, essendo tale elemento un componente della connessione puramente passivo, per cui eventuali problemi causati da esso si sarebbero riscontrati su

tutti i set nelle stesse proporzioni.

Come possibili cause rimangono quindi problemi legati agli switch. Il fatto che la caduta di prestazioni sia omogenea su tutti gli otto server porta ad ipotizzare come spiegazione più plausibile un temporaneo abbassamento di prestazioni da parte delle componenti su cui passa il traffico aggregato, quindi il backplane di uno degli switch, oppure il convertitore a cui è collegata la fibra. Eventuali problemi sulla porta a cui è collegato il singolo host sarebbero stati evidenziati nel grafico in figura 5 con singole cadute di valore, quale quella visibile in corrispondenza di  $t = 16$ .

## 5 Conclusioni

Le principali conclusioni che si possono trarre da questo lavoro sono che, in primo luogo, il collegamento in fibra tra i laboratori esterni e quelli sotterranei si è dimostrato in grado di supportare un traffico multigigabit in entrambe le direzioni di flusso, con prestazioni misurate di 5,6 ~ 5,9 Gb/s, a fronte di un traffico immesso da parte dei client pari a circa 7,25 Gb/s. La qual cosa, considerando sia la lunghezza della connessione pari a circa 8 km, sia il fatto che le fibre presentano delle peculiarità, descritte nell'introduzione, che avrebbero potuto causare delle limitazioni nelle prestazioni massime ottenibili, rende tale risultato del tutto ragguardevole.

E' stata anche evidenziata una buona resa degli apparati di rete, che non hanno quasi mai dato luogo a cadute di prestazioni, benché fossero sottoposti a carichi di lavoro non indifferenti, dovendo aggregare per tempi non brevissimi, dell'ordine del minuto, singoli flussi prossimi al Gb/s.

Ulteriori misure, con un numero sufficiente di macchine in grado di raggiungere le prestazioni nominali sulla propria scheda di rete, cosicché si possa saturare l'interfaccia a 10 Gb/s degli switch, permetteranno di testare la capacità della fibra di supportare moli di traffico al limite dello stato dell'arte della tecnologia attualmente in commercio.

## Ringraziamenti

Si ringrazia Giuseppe di Carlo, Sandra Parlati e Nazzareno Taborgna per aver messo a disposizione le macchine utilizzate come host, ed ancora Sandra Parlati per le chiarificanti discussioni avute durante l'analisi dei dati di misura.

## Riferimenti bibliografici

[1] <http://dast.nlanr.net/projects/Iperf/>

[2] <http://clusterssh.wiki.sourceforge.net/>

[3] A. Tanenbaum, *Computer Networks* 4th ed., Prentice Hall, cap. 6.4