



ISTITUTO NAZIONALE DI FISICA NUCLEARE

CNAF Bologna

INFN/TC-99/18

11 Ottobre 1999

PROGRAMMA QUADIS

STUDIO DI ARCHITETTURE E PROTOCOLLI PER RETI OTTICHE

R. Cucchi, A. Ghiselli, C. Raffaelli, C. Vistoli

INFN-CNAF, V.le Berti Pichat 6/2, I-40127 Bologna, Italy

Abstract

Le reti ottiche basate sulla tecnologia WDM (Wavelength Division Multiplexing) sono molto promettenti per rispondere non solo alle esigenze di capacità trasmissiva ma anche per la realizzazione di reti costituite da apparati esclusivamente ottici dove la lunghezza d'onda rappresenta il circuito di comunicazione fra nodo sorgente e nodo destinatario.

Questo documento prende in esame il “trigger di II livello dell’esperimento ATLAS come case study”, ne analizza le esigenze di trasmissione dei dati e descrive uno studio di fattibilità basato su una rete ottica a stella.

1. L' AMBITO DI RICERCA

In questi ultimi anni si è assistito ad una rapida evoluzione della tecnologia ottica unitamente all'incremento della richiesta di banda da parte di applicazioni distribuite e sistemi client/server. Tutto ciò spinge verso una rapida transizione nel campo delle reti ottiche da sistemi di laboratorio a potenziali sistemi di mercato. Fino ad ora le fibre ottiche sono state inserite nella infrastruttura ad alta capacità della rete in sostituzione dei cavi per raggiungere bande trasmissive dell'ordine dei Gbit/s su distanze dell'ordine di parecchi chilometri, dando origine ad una prima generazione di reti ottiche. Ora stanno emergendo soluzioni basate su un uso più estensivo della tecnologia ottica che coinvolge anche le funzioni di instradamento con la possibilità di ottenere velocità più elevate e soprattutto di trattare l'informazione nel dominio ottico durante l'intero percorso di trasferimento. Le reti di seconda generazione sono caratterizzate dalla trasparenza del trasferimento rispetto alla velocità e alla bit rate dei dati. Tali reti richiedono componenti più sofisticati che attualmente non sono disponibili a livello commerciale a causa dei costi elevati ma che diventeranno di notevole interesse nel medio termine per realizzare reti ad altissima velocità e flessibilità.

In quest'ultimo contesto si vuole sviluppare l'attività di ricerca proposta con l'obiettivo di pervenire a un progetto di rete completo in grado di soddisfare le esigenze di una applicazione che per requisiti di comunicazione si presenta adatta a sfruttare pienamente le nuove tecnologie. Si ritiene che, benchè il contesto scelto sia molto specifico, esso serva tuttavia a individuare le criticità della rete e a individuare criteri progettuali di possibile generalizzazione.

2. L'APPLICAZIONE

L'applicazione che è stata scelta riguarda il contesto degli esperimenti LHC rappresentato in figura 1 [1]. Si prevede di considerare il secondo livello di trigger in cui è necessario interconnettere i buffer (ROB: Read-Out-Buffer) in cui vengono memorizzate le regioni di interesse (ROI: Region-Of-Interest) con le workstation di elaborazione. Senza entrare nel dettaglio della descrizione del sistema di rilevazione ed elaborazione gli elementi che sono stati individuati come utili per una prima fase di studio di fattibilità della rete sono i seguenti:

- I ROB sono 1700 circa
- Le workstation sono 1000 circa
- I ROB vengono scritti con frequenza 100 KHz (ogni 10 microsecondi)
- I ROB interessati da nuove ROI sono dell'ordine della decina
- Il tempo di ciclo a disposizione degli elaboratori è di 10 ms: per tale tempo le ROI permangono nei ROB
- La dimensione di una ROI è 20Kbyte

Nel sistema esiste un supervisore che conosce quali ROB contengono ROI e quali workstation sono interessate alla relativa elaborazione.

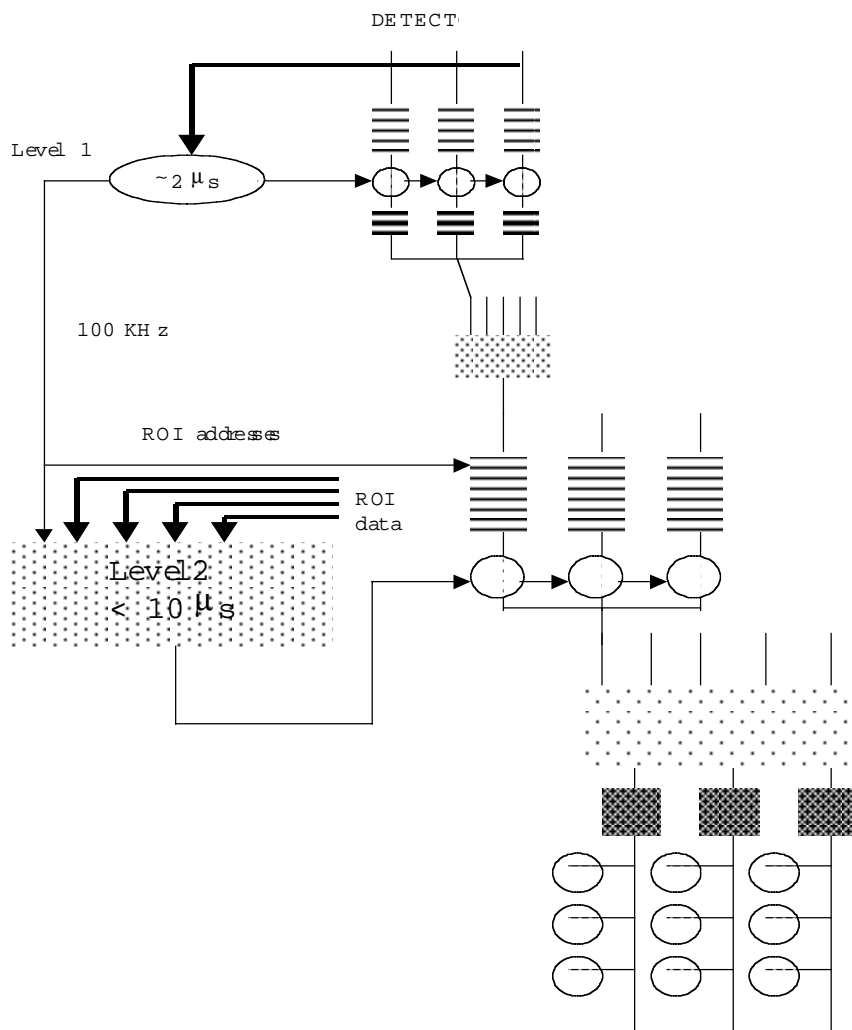


Figura 1 – Schema del sistema di acquisizione

3. IL PROGETTO DELLA RETE OTTICA

Le caratteristiche salienti del contesto applicativo prescelto sono le seguenti:

- ambito locale
- numero di nodi (ROB e stazioni) elevato
- elevata frequenza di riconfigurazione delle connessioni
- requisiti di affidabilità elevati
- necessità di trattare livelli di servizio diversi (supervisione, trasferimento dati)
- requisiti di banda elevati (2 Gbit/s)

In base soprattutto a quest'ultimo requisito si è pensato di scegliere questa applicazione come caso di studio per il progetto della rete ottica. Esistono attualmente soluzioni commerciali in tecnologia elettronica per reti ad altissima velocità che sono già state considerate per il contesto applicativo descritto. Tuttavia la soluzione ottica appare di maggior respiro sia in termini di capacità che di estensione. Nell'ambito ottico si possono prevedere soluzioni diverse basate sulle tecniche WDM e TDM o da combinazioni di entrambe.

Il progetto di una rete ottica si articola in genere nelle seguenti fasi:

- definizione della topologia fisica di rete, ovvero la topologia delle fibre ottiche
- definizione della topologia virtuale, ovvero la topologia dei percorsi ottici (lightpath)
- strategie di controllo

Trattandosi di una rete locale non si hanno vincoli particolari sulla topologia fisica che può essere a stella o ad anello a seconda delle realizzazioni. Tuttavia dato il numero elevato di nodi si rende necessaria la definizione di una rete modulare scalabile tipicamente con organizzazione gerarchica (figura 2).

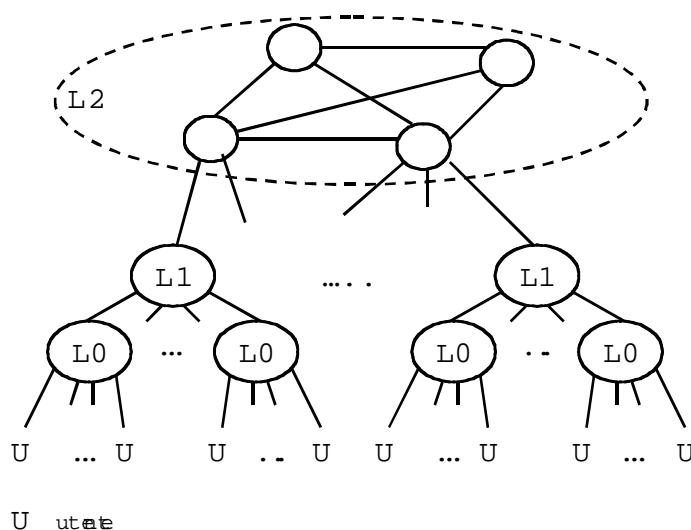


Figura 2 – Architettura gerarchica per reti di grandi dimensioni.

A livello L0 si hanno le singole LAN WDM sulle quali si può effettuare il riutilizzo delle lunghezze d'onda mediante l'adozione a livello 1 di apparati denominati *wavelength router* che selezionano solo le lunghezze d'onda utilizzate per l'interconnessione tra le LAN e confinano le rimanenti nell'ambito delle singole LAN mediante filtraggio (Fig. 3).

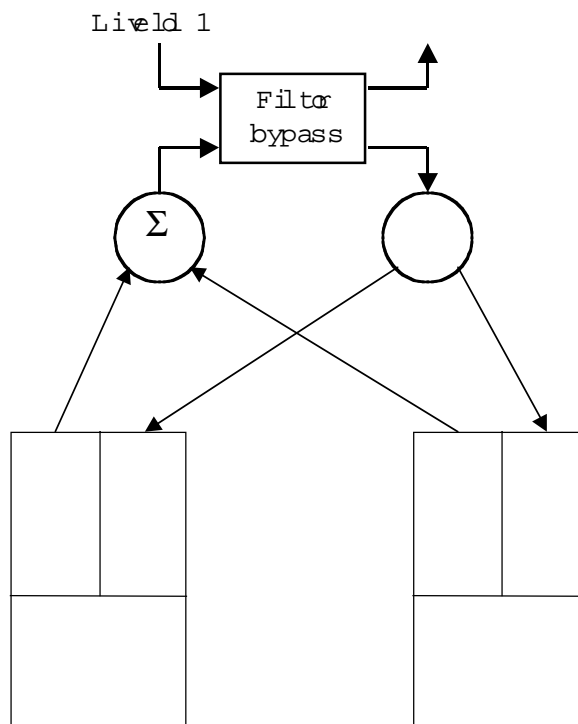


Figura 3 – Sistema di interconnessione a livello L1

A livello L2 si può pensare di avere una rete metropolitana basata su commutatori WDM.

La definizione della topologia virtuale riguarda l'uso del supporto fisico in relazione alle esigenze di connessione. In tale contesto l'elevata frequenza delle riconfigurazioni va considerata attentamente, a seconda della soluzione adottata, in quanto può richiedere l'uso di componentistica sofisticata. Si pensi ad esempio al caso WDM in cui le riconfigurazioni richiedano la sintonizzazione di laser o ricevitori.

Occorre poi analizzare i requisiti di qualità di servizio che devono portare al progetto di una rete in cui sia possibile assicurare le richieste di ciascun tipo di flusso (per esempio in termini di ritardo di trasferimento). Infine la definizione delle strategie di controllo riguarda la modalità e i protocolli che consentono un utilizzo ottimale delle risorse disponibili.

La definizione del progetto è collegata a una conoscenza sufficientemente dettagliata dei flussi di traffico al fine di poter sviluppare studi di simulazione di ausilio alle scelte progettuali. A tale scopo si prevede di interagire con esperti della applicazione scelta.

4. Un primo studio di fattibilità basato su rete a stella

Si riporta a titolo di esempio un primo studio di fattibilità basato sulla rete a stella di tipo broadcast-and-select al fine di rendere più concrete le considerazioni introdotte in precedenza.

La rete a stella broadcast-and-select è stata oggetto di numerosi studi negli ultimi anni [2] che hanno portato alla realizzazione di test-bed e ne hanno mostrato la fattibilità. Non esistono tuttavia al momento prodotti commerciali. Il funzionamento della rete si basa sulla trasmissione da parte delle stazioni su lunghezze d'onda diverse che vengono ridistribuite per mezzo di un accoppiatore ottico a tutte le stazioni. La ricezione da una particolare stazione avviene mediante sintonizzazione sulla corrispondente lunghezza d'onda. Uno schema della rete è riportato in figura 4. Su tale topologia fisica si possono realizzare diverse topologie virtuali a seconda della strategia di controllo delle connessioni adottata. Le soluzioni più classiche sono costituite dalla rete single-hop e dalla rete multi-hop.

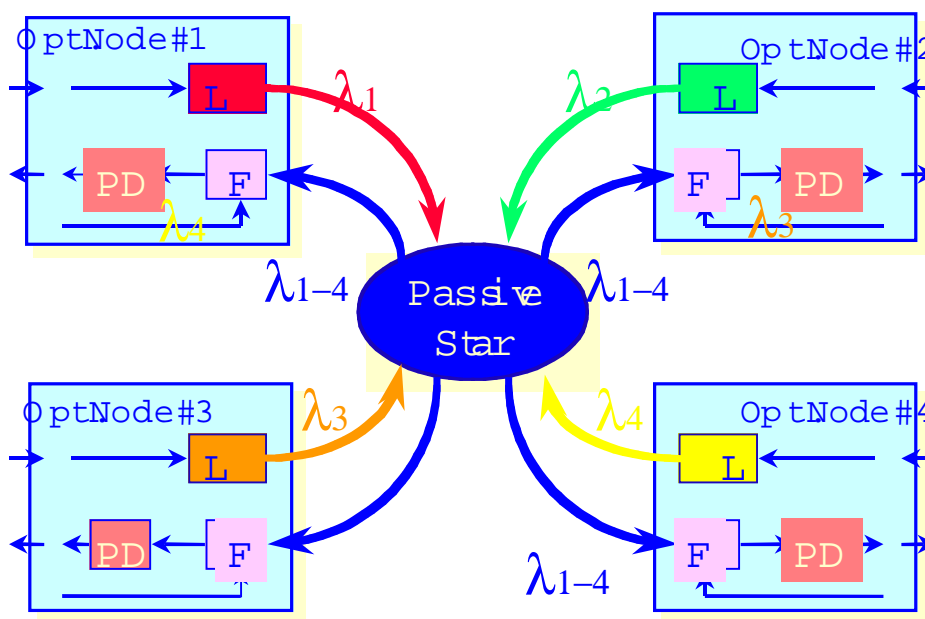


Figura 4 – Schema di funzionamento di una rete Broadcast-and-select.

4.1 Topologia fisica

La topologia fisica proposta si basa su stelle ottiche passive interconnesse per ragioni di modularità e per superare le limitazioni tecnologiche sul numero di lunghezze d'onda attraverso wavelength router che consentono il riutilizzo delle lunghezze d'onda all'interno delle singole stelle. Tale topologia è riportata in figura 5. Il controller opera tipicamente su una lunghezza d'onda dedicata al fine di comunicare alle unità di acquisizione e di elaborazione lo stato delle connessioni e l'esito delle valutazioni. Mediante tale organizzazione è possibile partizionare il sistema al fine di dotarlo delle necessarie

caratteristiche di scalabilità. Il partizionamento può essere pensato al fine di raggiungere una scelta ottimale in relazione al numero di lunghezze d'onda disponibili in ciascuna stella e al numero di stelle interconnesse, che determina il numero di lunghezze d'onda su cui opera il wavelength router. Esso può tenere conto anche di particolari (se esistono) configurazioni dei flussi di traffico attraverso l'intero sistema.

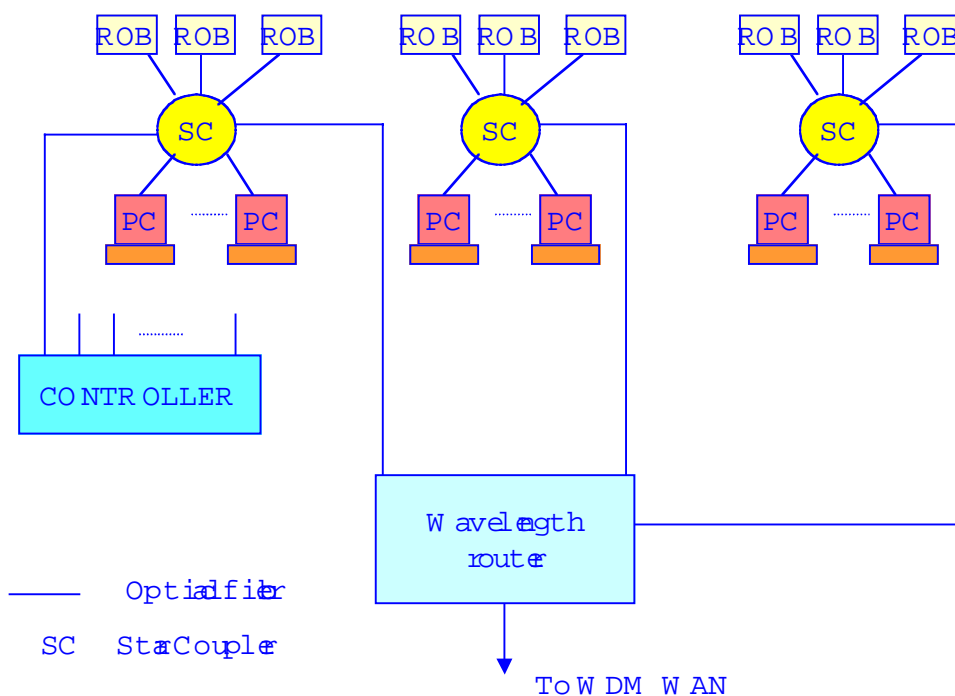


Figura 5 – Schema della rete di stelle interconnesse.

4.2 Topologia virtuale: rete single-hop.

La soluzione single-hop prevede che i nodi tra i quali avviene il trasferimento dell'informazione siano sintonizzati sulla stessa lunghezza d'onda. A tale scopo le connessioni vengono instaurate mediante la sintonizzazione sulla lunghezza d'onda di trasferimento del trasmettitore e del ricevitore interessati. Ciò può avvenire sotto il controllo del supervisore che conosce in ogni istante dove si trovano le regioni di interesse e quali elaboratori devono processarli. Si sceglie come tempo base per le operazioni il tempo di acquisizione pari a 10 microsecondi. Ogni 10 microsecondi il supervisore comunica ai nodi la configurazione delle connessioni. Come conseguenza i nodi coinvolti si devono sintonizzare sulla lunghezza d'onda che consente il trasferimento. Si fa l'ipotesi di mantenere fissa la lunghezza d'onda su cui operano gli elaboratori in modo da minimizzare il numero totale di lunghezze d'onda utilizzate nella singola stella. Il supervisore invia a tutti i nodi (PC e ROB) la configurazione delle connessioni al tempo i -esimo su una lunghezza d'onda dedicata su cui tutti sono sempre in lettura.

Schema pull:

- 1- I ROB interessati sintonizzano al tempo i -esimo i propri trasmettitori sulla lunghezza d'onda del PC corrispondente e aspettano la richiesta da parte del PC
- 2- I PC richiedono ai ROB le informazioni entro 10 microsecondi dall'inizio del tempo i -esimo.
- 3- Le informazioni vengono trasferite nel tempo i -esimo
- 4- I PC inviano il risultato delle elaborazioni al supervisore in modo asincrono ognuno sulla sua lunghezza d'onda tramite buffer ottico localizzato nel supervisore.

L'alternativa è lo schema push:

- 1- I ROB interessati sintonizzano al tempo i -esimo i propri trasmettitori sulla lunghezza d'onda del PC corrispondente
- 2- I ROB trasferiscono l'informazione verso i PC al tempo i -esimo senza necessità di richiesta da parte dei PC
- 3- I PC inviano il risultato delle elaborazioni al supervisore in modo asincrono ognuno sulla sua lunghezza d'onda tramite buffer ottico localizzato nel supervisore.

Entrambi questi schemi consentono di realizzare connessioni punto punto in ciascun intervallo temporale. Per la realizzazione di comunicazioni punto-multipunto occorre utilizzare più intervalli temporali con risintonizzazione dei trasmettitori.

La criticità di questa soluzione sta nei tempi di risintonizzazione in relazione alla tecnologia attuale dei dispositivi ottici.

4.3 Topologia virtuale: rete multi-hop.

Attualmente il costo di trasmettitori o ricevitori sintonizzabili è molto più elevato di quelli a lunghezza d'onda fissa. Inoltre la elevata frequenza con cui è necessario riconfigurare le connessioni può divenire molto critica nella realizzazione di una rete single-hop.

Nella soluzione multihop le connessioni sono fisse e ogni nodo riceve, con un ritardo legato al numero di hop dalla sorgente, le informazioni senza necessità di risintonizzazione. Un esempio di rete multi-hop realizzato su una topologia a stella è mostrato in figura 6.

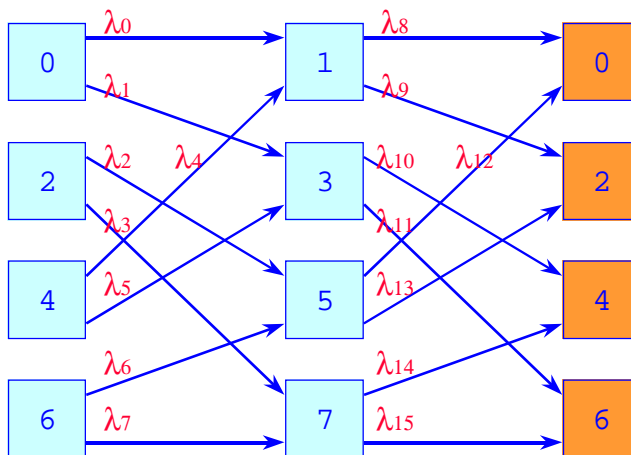


Figura 6 – Rete multihop basata su 8 nodi.

La soluzione multihop risulta quindi più adatta allo scenario in questione in quanto non impone la necessità di riconfigurazione e di instaurazione di sincronismi a livello di tempo di acquisizione.

Si tratta di individuare come assegnare le lunghezze d'onda e quante lunghezze d'onda sono necessarie.

Lo svolgimento delle operazioni di trasferimento avviene sotto il controllo di un supervisore che conosce quali ROB contengono le ROI e lo comunica ai PC sulla lunghezza d'onda di controllo. Quindi questi PC possono interrogare i ROB contenenti le ROI (schema PULL) oppure aspettarsi le informazioni dai ROB (schema PUSH). I risultati vengono comunicati al supervisore utilizzando lo schema di rete virtuale.

Sono possibili comunicazioni punto multipunto sia dai ROB verso i PC sia dai PC verso i ROB. Solo i PC e i ROB "attivi", cioè coinvolti in un trasferimento sono interessati a monitorare l'informazione sulla lunghezza d'onda di ricezione per vedere se è destinata a loro. Rispetto alla soluzione single hop questa soluzione presenta maggiore criticità per quanto riguarda il numero NL di lunghezze d'onda da utilizzarsi che ammonta a $N \cdot \text{numero di trasmettitori per nodo}$. Tuttavia risulta molto meno critico per la gestione delle connessioni. Visto che i tempi non sono troppo stringenti si può anche pensare alla sovrapposizione del TDM che ridurrebbe la criticità sul numero di lunghezze d'onda. Un esempio di utilizzo del dominio temporale in una rete multihop è presentato in figura 7. Si definisce una trama temporale suddivisa in intervalli (slot) in ciascuno dei quali si attribuisce l'uso di ciascuna lunghezza d'onda a un sottoinsieme delle stazioni. Le stesse lunghezze d'onda in diversi slot sono utilizzate da diverse stazioni così che si può ridurre il numero complessivo di lunghezze d'onda utilizzate.

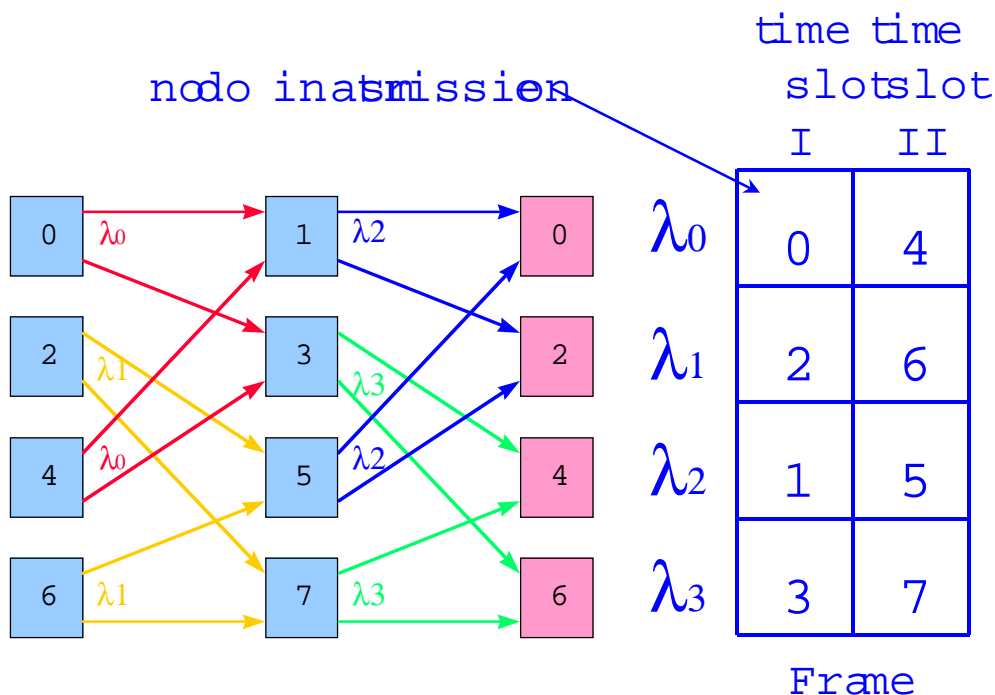


Figura 7 – Esempio di utilizzo del dominio temporale su una rete multihop.

Per l'interconnessione tra stelle multihop si tratta di stabilire come opera il WR in relazione alle richieste di interconnessione. Le richieste vengono inviate dal supervisore e il WR può riconfigurarsi di conseguenza. Tali riconfigurazioni sono meno critiche rispetto alle riconfigurazioni dei singoli nodi in quanto si può pensare avvengano più raramente. Oppure il WR si comporta in modo analogo agli altri nodi della rete ma trasmette su lunghezze d'onda che non sono sentite dalla sottorete ma solo dalle altre. Non necessariamente in questo caso è lo stesso nodo che fa da trasmettitore verso le altre reti o da ricevitore dalle altre reti. Un esempio di schema di interconnessione è presentato in figura 8.

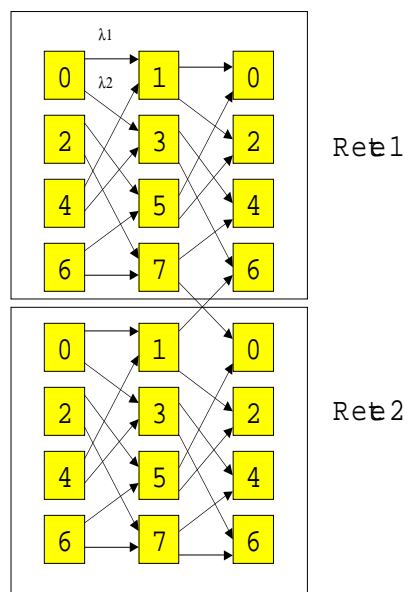


Figura 8 – Esempio di interconnessione tra due stelle multihop.

Un parametro da valutare è il ritardo introdotto dal numero di salti tra la sorgente e la destinazione.

Tale valutazione è riportata a titolo di esempio in figura 9 per una rete a 100 nodi con lunghezza di pacchetto pari a 1000 byte per ritardo medio e ritardo massimo in funzione della velocità della rete. Si può notare che, alle velocità a cui si intende operare, il ritardo è inferiore al tempo di latenza considerato per il trigger di secondo livello che è pari a 10 ms.

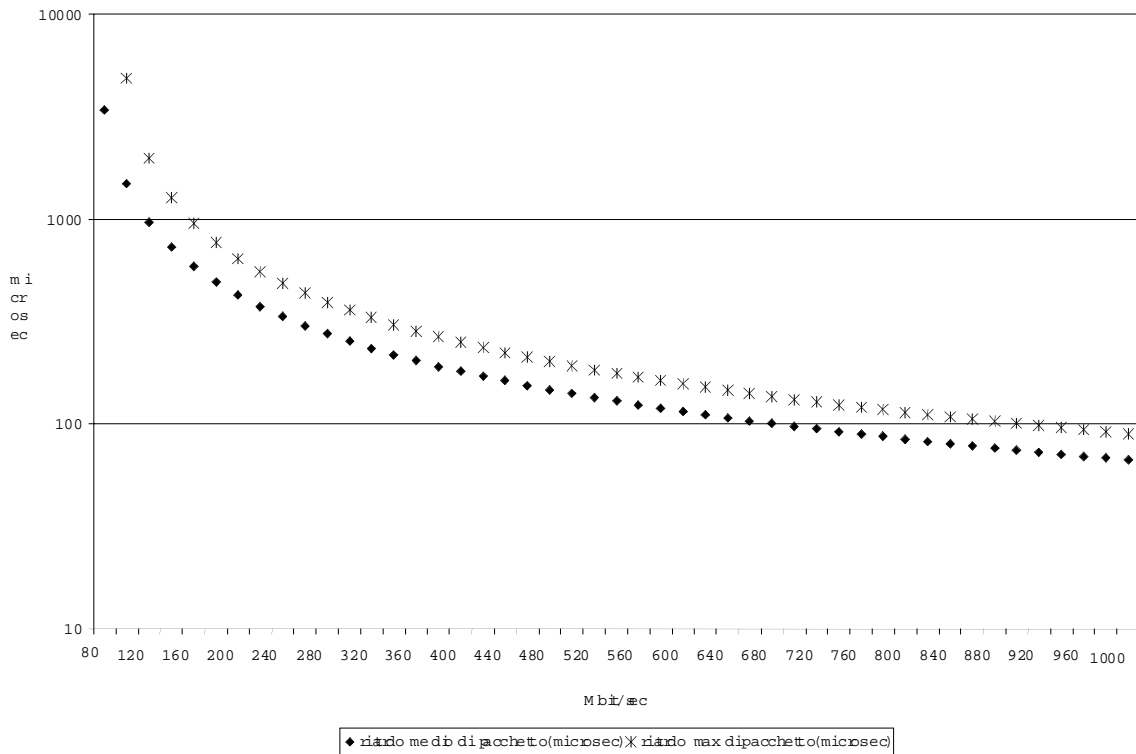


Figura 9 – Ritardo massimo e medio per una rete multihop a 100 nodi con pacchetto di 1000 byte in funzione della velocità.

5. Conclusioni

Con la tecnologia ottica attuale è possibile pensare a soluzioni di interconnessione su area locale a velocità dell'ordine del Gbit/s che però risultano molto costose. Occorre quindi individuare soluzioni che si prestano a realizzazioni con costi accessibili nei prossimi anni. Per quanto riguarda l'applicazione considerata è possibile adottare soluzioni scalabili che consentono di realizzare il collegamento tra un numero molto elevato di nodi. Un aspetto molto importante da definire in tali reti riguarda il controllo delle connessioni in relazione alle esigenze di trasferimento. La soluzione a stella è un esempio ma ne esistono altri come la topologia WDM ad anello che devono essere valutati al fine di individuare la soluzione più idonea sia da un punto di vista tecnico che economico.

References

- [1] ATLAS DAQ/Event Filter Prototype, <http://atddoc.cern.ch/Atlas/>
- [2] R. Ramaswami, K. N. Sivarajan, "Optical Networks: a practical perspective", Morgan Kaufmann Publishers, San Francisco, California, 1998.