

Brevi cenni su Storage Area Network

Andrei Maslennikov
CASPUR

Novembre 2005

Discuteremo di...

- **FC / SAN: terminologia, roadmap, componenti**
- **Esempi di utilizzo della SAN**
- **SAN su WAN**

Fibre Channel (Fiber + Wire)

Fibre Channel e` il nome generale di un insieme di standard, sviluppati dal T11 task group dell'ANSI (dal 1988). Si tratta di un protocollo di trasferimento dati di basso livello.

Privo di un proprio command set, il Fibre Channel costituisce un mezzo che permette ai protocolli di livello piu` alto (come SCSI, IP, Video) di viaggiare, anche assieme, su una stessa linea seriale.

FC e` un misto di tecnologie di channel e di rete:

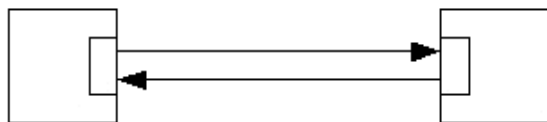
Dal mondo channel (interconnessione chiusa, strutturata, diretta, prevedibile, pochi device):

- capacita` di usare i protocolli preesistenti
- alte prestazioni
- reliable delivery, low error rate

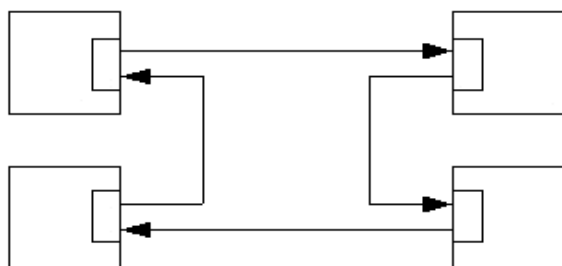
Dal mondo rete (interconnessione non-strutturata, non-prevedibile, ma capace di autoconfigurarsi e di supportare tanti device):

- linea seriale
- i dati viaggiano impacchettati (FC frames)
- e` possibile interconnettere parecchi device usando gli FC switch

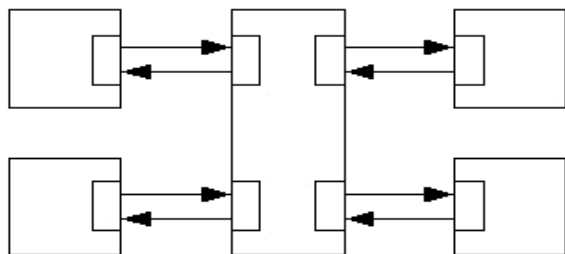
Topologie FC



Point-to-point. Nessuna condivisione, va usata tutta la banda di connessione.



Arbitrated Loop. Media condivisa, fino a 127 device per loop. I device “arbitrano” su Loop con i loro indirizzi fisici (AL_PA). L’Access Fairness Algorithm garantisce che tutti i device abbiano una chance. Vinta l’arbitrazione, la connessione è punto-a-punto.



Fabric (SAN). Fino a 2^{24} device possono essere interconnessi. I device sono indirizzabili su fabric tramite i loro WWNN (World Wide Node Name, simile a MAC address per i NIC). Per fare una SAN serve almeno uno switch.

SAN: vantaggi

- **Enorme flessibilita` nella ricollocazione dei dispositivi storage:**
 - o all'interno di un campus (sale macchine, edifici ecc.)
 - o all'interno di una citta` (MAN)
 - o su LAN/WAN

- **Possibilita` di condivisione di dispositivi tra piu` host:**
 - o Uso migliore di dispositivi costosi (unita` nastro)
 - o SAN-based distributed filesystem (unita` disco)

FC Roadmap

1998/99-2001: 1 Gigabit

2001-2004: 2 Gigabit

dal 2004 → 10 Gigabit

Cavi FC: rame (max 30 m), fibra ottica (max 10 km).

Con WDM (Wave Division Multiplexing) anche decine e centinaia di km.

In commercio oggi: Switch Exchange Port: 10 Gbit, HBA: 4 Gbit

SAN Hardware del momento (prezzi approssimativi)

- Switch FC 4 Gbit

- o Qlogic Sanbox 5600 16p - 5700 USD end user
- o Qlogic Sanbox 5600 20p (4x10Gb uplink) - 7000 USD end user
- o Brocade 200 16p - 5800 USD end user
- o McData 4400 16p - 6300 USD end user

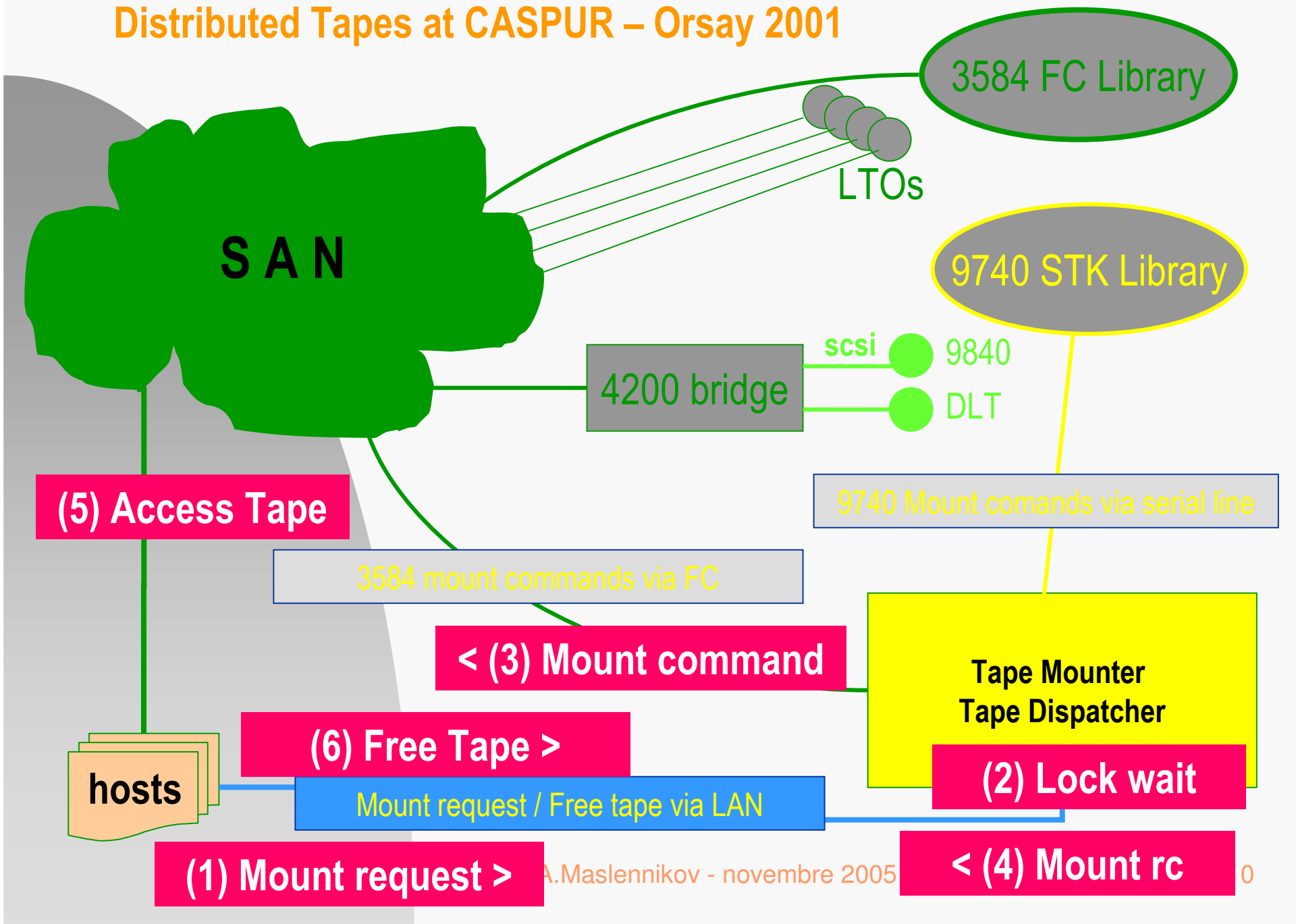
- HBA FC 4 Gbit

- o Emulex LP11000 - 1000 USD end user
- o Qlogic QLA2460 - 900 USD end user
- o Atto 41XS/42XS(2p) N/A (~~ 1000 USD)

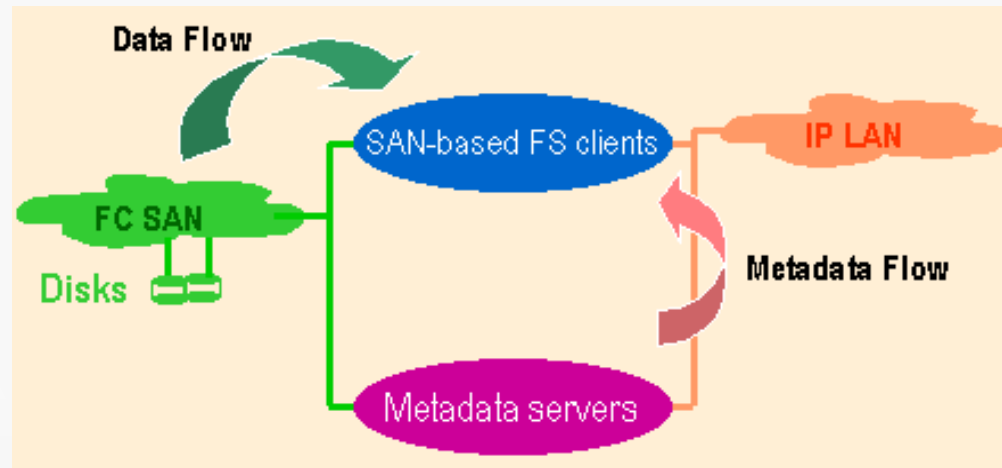
SAN use case 1 – condivisione unita` nastri

- I tape device sono abbastanza costosi
- Condividendo uno o piu` drive tra piu` applicazioni si riesce a ottimizzare l'uso delle risorse
- La condivisione richiede un servizio aggiuntivo di gestione delle risorse
- Una possibile soluzione: **CASPUR Tape Dispatcher**, in produzione dal 2000. Disponibile per download.

Distributed Tapes at CASPUR – Orsay 2001



SAN use case 2 – SAN-based file system



Tipica architettura SAN-based. Dati e metadati usano canali diversi: tutto il traffico dati passa per la Storage Area Network (FC o iSCSI). La gestione dei metadati è invece affidata a un processo diverso che comunica con i client su IP.

- GFS (Linux)
- SGI CXFS (client su IRIX, Solaris, W2K, AIX, HP/UX, Linux). Metadata: IRIX.
- IBM Storage Tank (AIX, Solaris, HP/UX, Linux, W2K),
- IBM GPFS-SAN (AIX, Linux)
- ADIC StorNext (Linux, Windows, AIX, Solaris, HP/UX)
- Polyserve Matrix (Linux)

Prestazioni elevate grazie a FC, ottima scalabilità.

SAN via WAN

- Applicazioni principali:

- o disaster recovery: backup e replicazione dati**
- o condivisione unita` nastro (spesso con virtual tape drive)**
- o video broadcast**

Accesso a SAN via IP

- Protocolli:

- o **iSCSI** (Internet Small Computer System Interface):
incapsulamento dei comandi SCSI nel TCP, trasporto via IP
- o **FCIP** (Fibre Channel Over IP): “tunnel Fibre Channel” via IP
- o **iFCP** (Internet Fibre Channel Protocol): FC layer 4 su TCP/IP,
protocollo gateway-gateway. Rende piu` affidabile la connessione.

- Hardware: SAN Router

- | | | |
|-----------------------------|--------------------|---------------------------------|
| o McData 1620,2640 | iFCP, iSCSI | 35+ KUSD (tante opzioni) |
| o San Valley SL2100 | FCIP | 20 KUSD list price |
| o CISCO 5428-2, 9216 | FCIP, iSCSI | N/A |

Prove al CASPUR Storage Lab - 2003

- In collaborazione con GARR, INFN, CERN
- Provati tre SAN router (Nishan, Cisco, San Valley)
- Obiettivo: verificare le prestazioni delle periferiche FC su distanza (soprattutto dei nastri) e l'affidabilità della connessione
- Queste prove valgono anche oggi perché gli attuali SAN router usano ancora la connessione IP a 1Gbit (seguono alcune slide d'epoca)

2003

San Valley Systems IP-SAN Gateway SL-700 / SL-1000



- 1 or 4 wirespeed Fibre Channel -to- Gigabit Ethernet channels
- Uses UDP and hence delegates to the application the handling of a network outage
- Easy in configuration
- Allows for the fine-grained traffic shaping (step size 200 Kbit, 1Gb/s to 1Mb/s) and QoS
- Connecting two SANs over IP with a pair of SL1000 units is in all aspects equivalent
 - to the case when these two SANs are connected with a simple fibre cable
- Approximate cost: 20 KUSD/unit (SL-700, 1 channel)
 - 30 KUSD/unit (SL-1000, 4 channels)
- Recommended number of units per site: 1

2003

Nishan IPS 3300/4300 multiprotocol IP Storage Switch



- 2 or 4 wirespeed iFCP ports for SAN interconnection over IP
- Uses TCP and is capable to seamlessly handle the network outages
- Allows for traffic shaping at predefined bandwidth (8 steps, 1Gbit- 10Mbit) and QoS
- Implements an intelligent router function: allows to interconnect multiple fabrics from different vendors and makes them look as a single SAN
- When interconnecting two or more separately managed SANs, maintains their independent administration
- Approximate cost: 33 KUSD/unit (6 universal FC/GE ports + 2 iFCP ports - IPS 3300)
48 KUSD/unit (12 universal FC/GE ports + 4 iFCP ports - IPS 4300)
- Recommended number of units per site: 2 (to provide redundant routing)

2003

CISCO IP Storage Services Module (DS-X9308-SMIP)



MDS9216 16 port modular FC switch

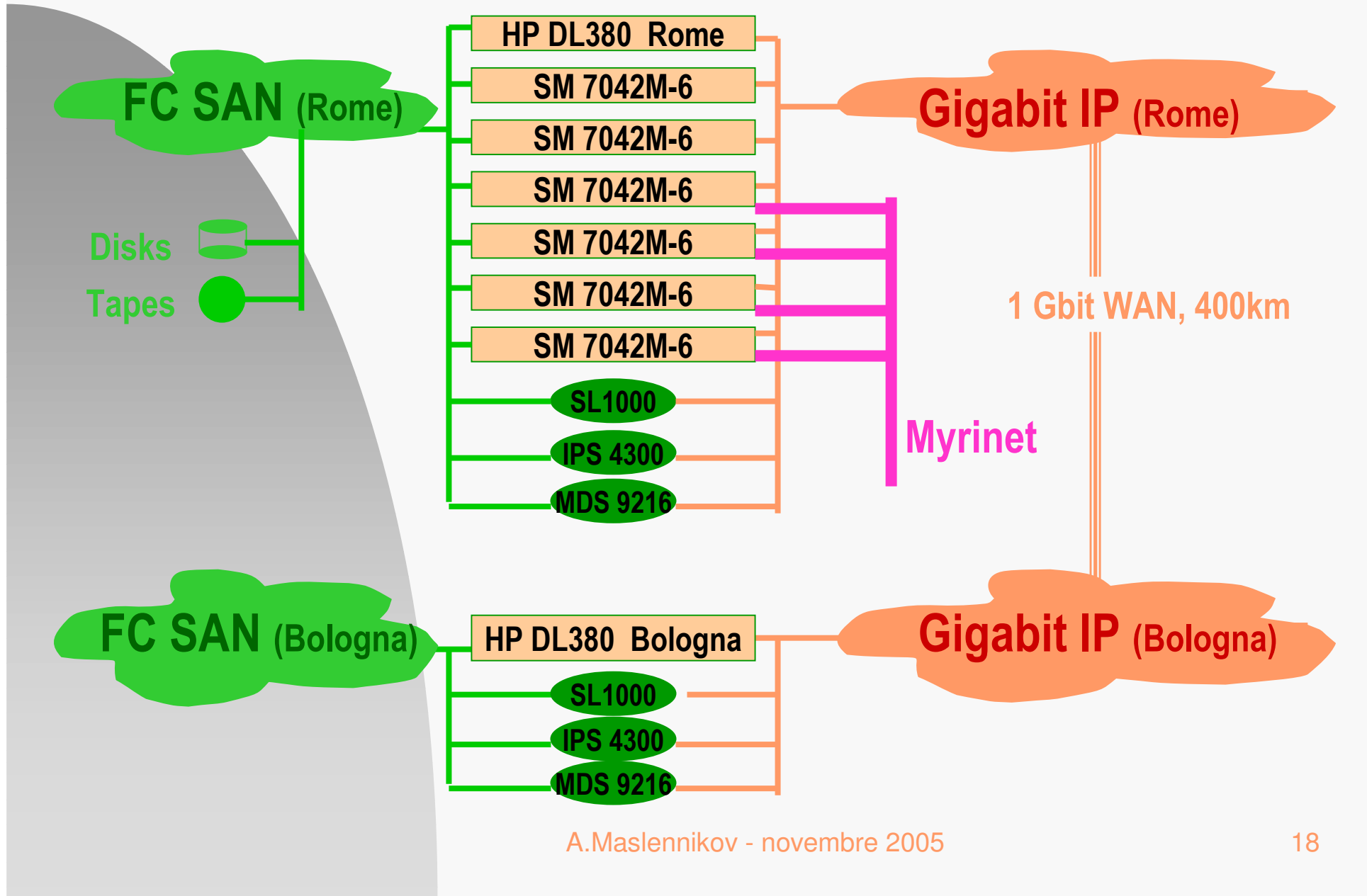


IP Storage Service module

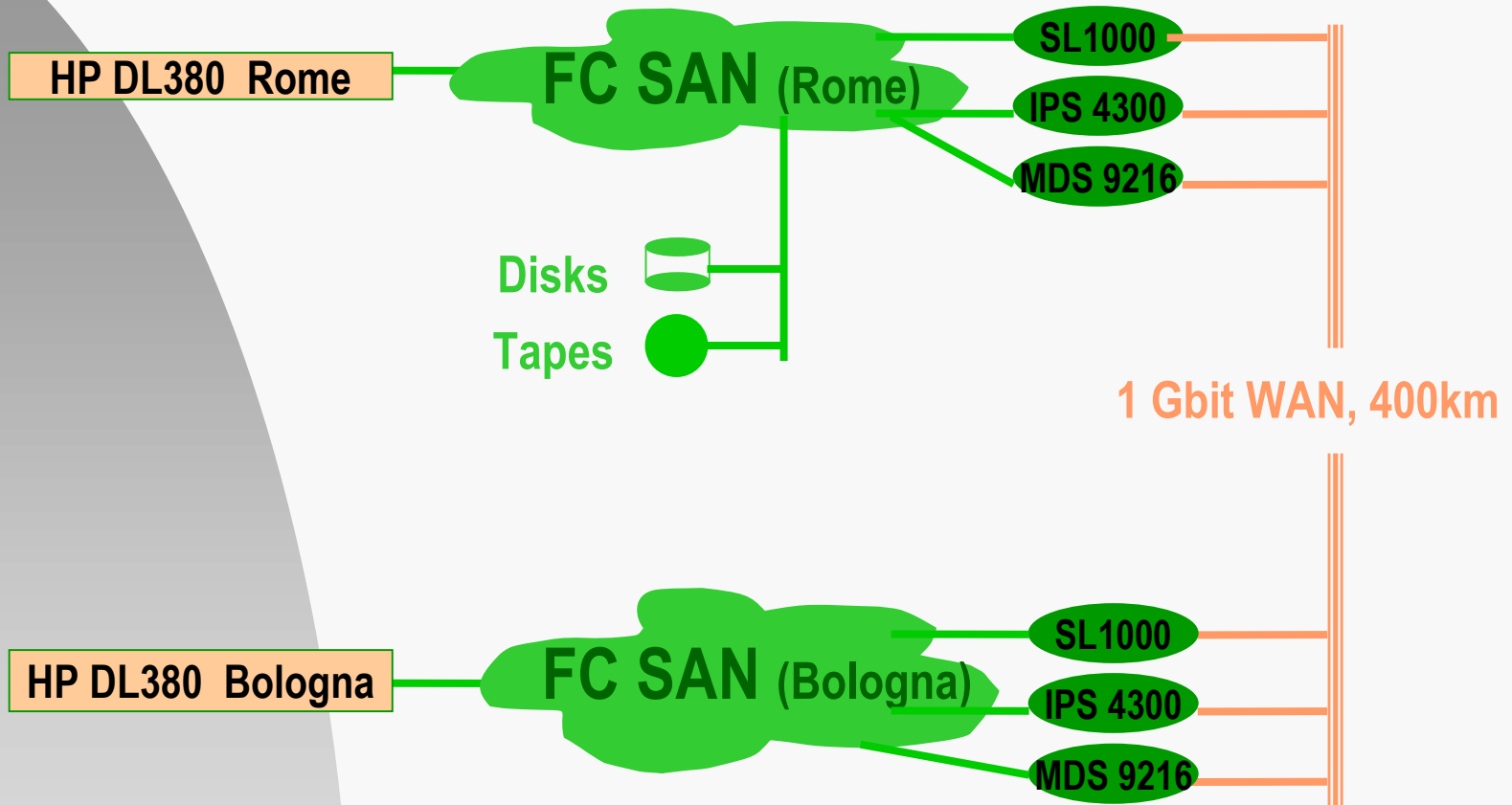
- Provides integration of IP Storage Services into the Cisco MDS9000 FC switches (One MDS9216 switch incorporates 16 FC ports)
- 8 Gigabit Ethernet IP Storage Interfaces:
 - o Wire-rate FCIP on all ports simultaneously
 - o Up to 24 simultaneous FCIP links per module
- Industry standard FCIP protocol uses TCP/IP to provide reliable transport
- VSAN and VLAN services increase stability and security of WAN-connected SAN elements
- Pricing provided through reseller partners (IBM, HP and others)
- Recommended number of units per site: 1

2003

CASPUR Storage Lab - 2003



Series 1: accessing remote SAN devices



Series 1 - current results

- We were able to operate with tape drives at the drive native speed (R and W): 15 MB/sec in case of LTO and 25 MB/sec in case of another, faster, drive
- In case of disk devices we have observed a small (5%) loss of performance on writes and a more visible (up to 12%) loss on reads, on all 3 units. The performance drop on reads increases with distance between the units: 6% at 1 m through 10-12% at 400 km.
- Several powerful devices used simultaneously over the geographical link easily grab the whole available bandwidth of the GigE link between the two appliances (seen so far for Nishan and San Valley units only; CISCO will be tested in the close future).
- in case of Nishan (TCP-based SAN interconnection) we have witnessed a successful job completion after an emulated 1-minute network outage. A similar test will also be performed for the CISCO unit.

Distributed Staging based on a direct tape drive access is POSSIBLE !

Alcuni consigli

- La SAN conviene quando un'organizzazione possiede o sta pianificando un ambiente storage con un numero elevato di periferiche fibre channel e storage server.
- Di tutti gli switch FC i piu` interessanti ad oggi sono quelli che rendono possibile l'espansione della SAN mediante link a 10 Gbit (attualmente solo Qlogic Sanbox).
- L'interconnessione delle isole SAN su WAN o l'accesso remoto alle periferiche FC via WAN va sempre visto nell'ottica dell'applicazione (gestione da parte dell'applicazione della ritrasmissione dei pacchetti nel caso di network outage). Il protocollo iFCP rende la connessione decisamente piu` robusta.
- Non scordare che spesso si puo` raggiungere lo stesso risultato utilizzando mezzi meno costosi. Prima di cominciare con una SAN occorre fare una stima accurata dei costi.