

ARCHITETTURA DEI ROUTER IP

Prof. Ing. Maurizio Casoni



Dipartimento di Ingegneria “Enzo Ferrari”

Università degli Studi di Modena e Reggio Emilia

INTRODUZIONE

- La natura distribuita e scalabile della rete Internet contribuisce alla continua, rapida ed incontrollata crescita dei suoi utenti, host, linee ed applicazioni
- Crescita esponenziale dei volumi di traffico => crescente richiesta di enormi capacità nella core network
- Network providers hanno numerose sfide:
  - Necessità di infrastruttura di capacità sempre maggiore
  - Necessità di architetture negli apparati switch/router sempre più veloci per sostenere le sempre maggiori velocità delle linee di trasmissione
- Evoluzione dei sistemi di trasmissione in fibra ottica ha spostato il collo di bottiglia delle Reti di Telecomunicazioni dalle linee ai nodi
  - Sistemi DWDM (Dense Wavelength Division Multiplexing) consentono di avere su singola fibra 128 canali OC-192 (10 Gbit/s) per complessivi 1.2 Tbit/s
  - Sono previste a breve linee OC-768 (40 Gbit/s)
- ***Obiettivo***: sviluppare soluzioni/tecnologie che consentano di progettare e realizzare core switch/router che operino a velocità maggiori del Terabit/s
  - Controllo della QoS: bassa latenza e perdita delle informazioni
- IP è il protocollo di rete dominante
- Apparati di core di Internet sono router IP e switch ATM



SISTEMI ROUTER IP

- IP: protocollo di rete connection-less, con trasferimento dati “best-effort”
  - Trasferimento dei datagrammi da sorgente a destinazione hop-by-hop
- All’arrivo di un datagramma IP su una delle sue interfacce di ingresso il router deve decidare su quale interfaccia di uscita inviarlo, basandosi sul suo indirizzo IP
- Funzioni dei router IP
  - Relative al percorso dati
    - Eseguite per ogni datagramma arrivato
    - Scelta dell’inoltro (forwarding decision)
    - Commutazione dei dati attarverso l’architettura (backplane)
    - Scheduling sulle linee di uscita
  - di controllo
    - Configurazione e gestione del sistema
    - Scambio delle informazioni di routing
    - Il *route controller* scambia informazioni sulla topologia con altri router mediante i protocolli di routing (RIP, OSPF) e costruisce la tabella di routing; può inoltre creare una *forwarding table* che sarà usata dal *forwarding engine*



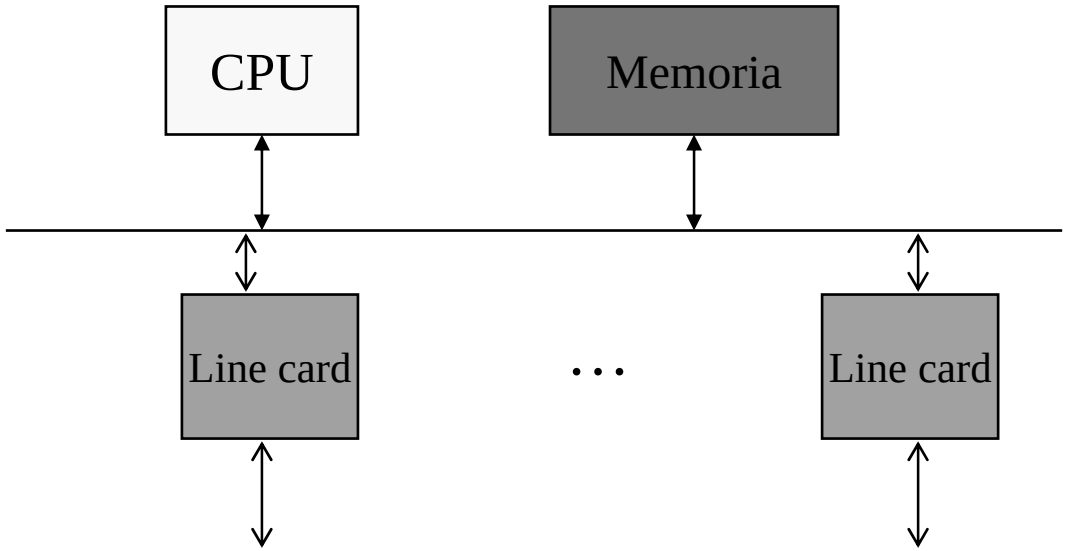
FUNZIONI RELATIVE AL PERCORSO DATI

- Quando un datagramma arriva al *forwarding engine* il suo indirizzo IP viene estratto per essere processato con la *forwarding table*
- Ogni entry della *forwarding table* consiste in un indirizzo IP a 32 bit ed in una *netmask* a 32 bit
- Indirizzo IP del datagramma è processato in AND logico con la netmask di ogni entry alla ricerca della porta di uscita più idonea secondo un metodo chiamato *longest-prefix matching*
- In certe applicazioni i pacchetti sono classificati in base a indirizzi IP di sorgente e destinazione, numeri di porta di sorgente e destinazione e tipo di protocollo
  - Datagrammi possono quindi anche essere eliminati (firewall) o gestiti a diversi livelli di priorità
- TTL è decrementato e nuovo header checksum viene calcolato
- Una volta determinata la porta di uscita, il datagramma vi è trasferito attraverso la rete di commutazione interna (*switch fabric*)
- Sulla porta di uscita può esservi contesa in seguito ad arrivi multipli ed i pacchetti devono essere gestiti opportunamente con eventuali memorie



ROUTER IP DI FASCIA BASSA

- Sono i router delle prime generazioni o quelli attuali più economici
- La decisione dell'inoltro (forwarding) e della commutazione (switching) sono svolte dall'unica CPU centrale con memoria su architettura a bus: tutto svolto via software
- Software => basso costo => apparati prevalentemente di fascia bassa
- ***Prestazioni***: sebbene le CPU migliorino costantemente, permane il collo di bottiglia della gestione di tutti i datagrammi mediante unica CPU centrale





ARCHITETTURA PC: OVERVIEW

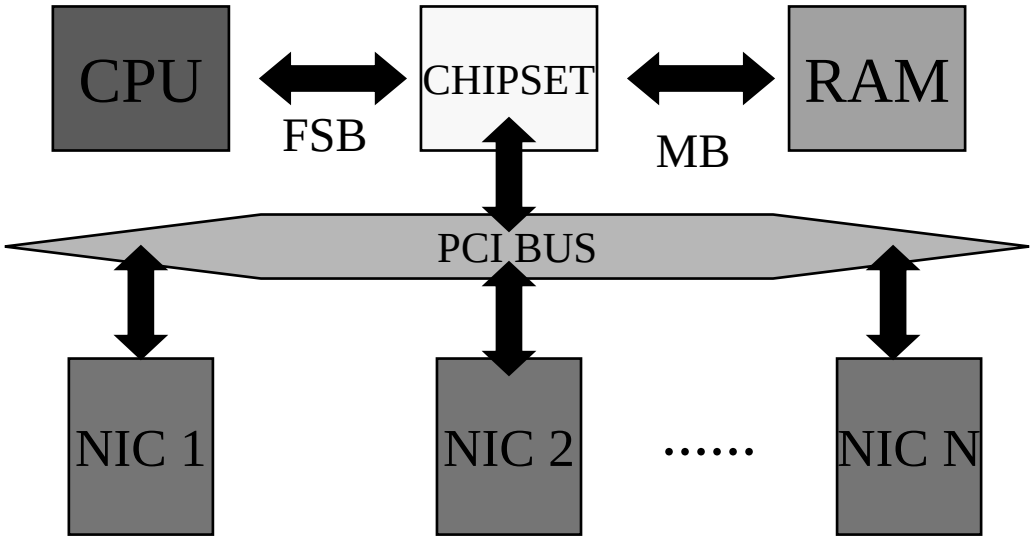
Tre blocchi principali:

1. *CPU*
2. *RAM*
3. *Periferiche*

collegati dal chipset che realizza complesse interconnessioni e funzioni di controllo

- *CPU* comunica col chipset mediante il ***Front Side Bus (FSB),*** secondo terminologia Intel
- *RAM* è accessibile dal controller integrato nel chipset mediante il ***Memory Bus (MB)***
- *Periferiche* sono connesse al chipset mediante il bus condiviso ***peripheral component interconnect (PCI);*** come periferiche vi sono le ***Network Interface Cards (NICs)***

Tutte queste componenti possono essere impiegate pure per implementare un **SOFTWARE ROUTER**





CPU

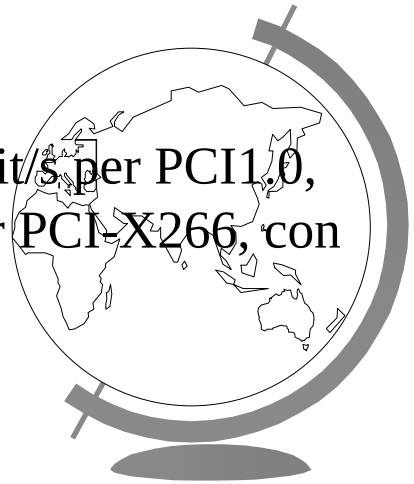
- Frequenze fino a 3.8 GHz e prossime a 4 GHz
- FSB a 64 bit con clock a 100 o 133 MHz con “quad pumped transfer”: dati trasferiti a velocità 4 volte maggiore del clock nominale
  - velocità di picco tra 3.2 e 4.2 Gbyte/s
  - notare che in termini commerciali Intel, i sistemi FSB a 100 e 133 sono presentati come 400 e 533 MHZ, a causa del quad pumping
- High-end PCs possono essere dotati di chipset che supportano molteplici CPU: fino a 16 CPU



RAM + PCI BUS

- **Memory Bus** di solito è 64 bit con clock a 100 o 133 MHZ con “double pumped transfers” (chiamate da Intel PC1600 e PC2100 double data rate DDR RAM)
  - Velocità di picco tra 1.6 e 2.1 Gbyte/s
  - In High-end PCs è possible raddoppiare la banda rendendola uguale a quella del FSB con un raddoppio del MB a 128 bit mediante installazione di banchi di memoria a coppie

- **PCI BUS**: La banda disponibile sul BUS può variare da 1 Gbit/s per PCI1.0, con clock a 33 MHz con parallelismo a 32 bit, a 2 Gbyte/s per PCI-X266, con parallelismo a 64 bit con double pumping a 133 MHz



NETWORK INTERFACE CARD

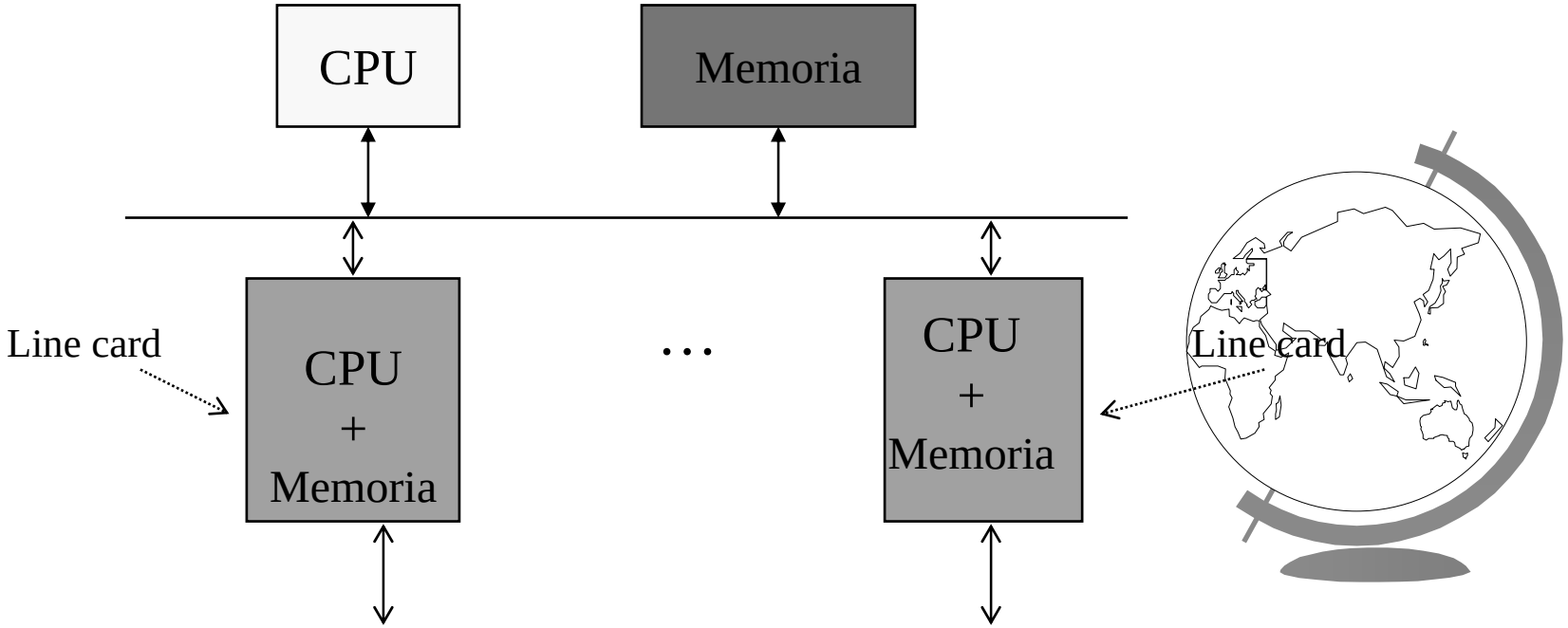
- Le schede GEthernet e Fast Ethernet sono schede PCI ad elevate prestazioni dotate di almeno un dispositivo DMA (direct memory access) che può prendere il controllo del bus per alleggerire la CPU per le operazioni di trasferimento dati tra la RAM e la memoria della scheda

- In linea di principio, l' hardware presente su un PC può consentire di realizzare un router a memoria e bus condivisi, dove le NICs ricevono e memorizzano i pacchetti nella RAM principale, la CPU li instrada verso la desiderata interfaccia di uscita e le NIC di nuovo li prelevano dalla RAM per trasmetterli sulle linee di uscita
  - quindi ogni pacchetto viaggia DUE volte sul bus PCI, dimezzando di fatto la banda disponibile per i flussi di pacchetti da NIC a NIC



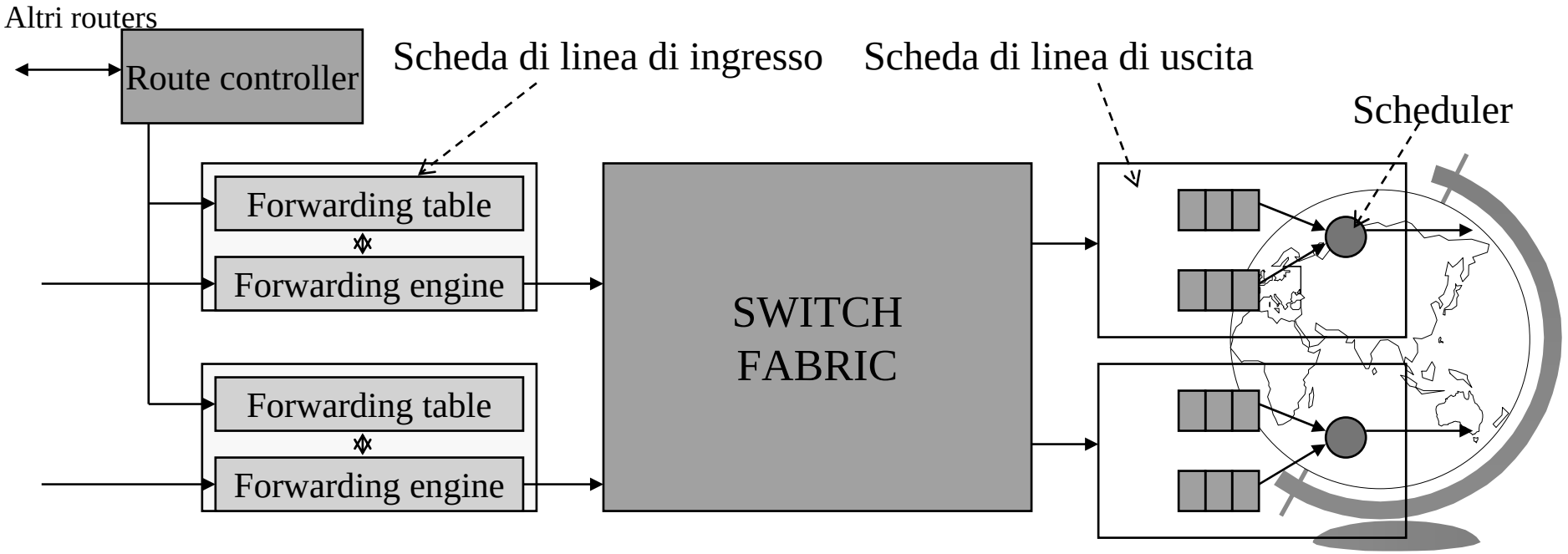
ROUTER IP DI FASCIA MEDIA

- Per superare le limitazioni dell'unica CPU centrale, questi router adottano schede di rete dotate di CPU ed elementi di memoria proprie per l'inoltro dei pacchetti
- Viene ridotto il carico sulla CPU centrale perchè i pacchetti entranti sono processati in parallelo nelle schede di rete prima di essere inviati attraverso il bus alla CPU e poi alla memoria della scheda di rete della linea di uscita desiderata
- ***Prestazioni***: all'aumentare del numero delle linee e della loro velocità, vi è ancora il collo di bottiglia della struttura a bus condiviso e della CPU centrale



ROUTER IP DI FASCIA ALTA

- La funzione di inoltro (*forwarding engine*) risiede in tutte le schede di ingresso
- La funzione di commutazione avviene mediante schede di commutazione
- ***High performance router***: route controller+forwarding engine+switch fabric+scheduler
- Le schede di linea possono comunicare tutte fra loro: throughput complessivo aumentato
- Un pacchetto commutato da ingresso a uscita viene memorizzato nella relativa memoria: risoluzione delle contese
- Scheduler: implementa diverse politiche di servizio, priorità, gestione della memoria



ROUTER IP DI FASCIA ALTA

- ***Domanda fondamentale***: come concepire e progettare il trasferimento dei pacchetti dai *forwarding engine* alle porte di uscita attraverso lo *switch fabric*?
- Tutti i router e switch si distinguono per come rispondono a tale domanda
- Fascia alta: funzioni relative al percorso dati sono implementate in hardware
  - La decisione dell'inoltro deve essere presa dal *forwarding engine* in un intervallo di tempo inferiore alla durata del datagramma di lunghezza minima (*wire speed functioning*)
- Pacchetti possono attraversare lo switch fabric in blocchi dati di dimensioni fisse (celle) o variabili
  - Nel primo caso i datagrammi sono segmentati in celle alle porte di ingresso dello switch e riassemblati nelle porte di uscita
- Per realizzare switch fabric ad alta velocità sono preferibili le celle
  - Trasferimento in modo sincrono con time-slot costanti per le celle: tutto via HW
- In caso di pacchetti di lunghezza variabile vi sono due alternative:
  - Una porta di uscita deve aspettare di ricevere completamente un pacchetto da un certo ingresso prima di poterne ricevere un altro (*event driven*)
  - Un pacchetto variabile viene ancora scomposto in celle che devono essere consegnate all'uscita in ordine senza essere alternate con celle di altri pacchetti (*slot driven*)



CRITERI DI PROGETTO E REQUISITI DI PRESTAZIONI

- *Switch fabric* deve fornire un ***ritardo di transito*** limitato ed una bassa ***probabilità di perdita di cella***, offrendo un ***throughput*** vicino al 100%
- Capacità di supportare linee di ingresso ad alta velocità è pure importante per servizi multimediali come videoconferenza, videotelefonia
- Gestione dei pacchetti FCFS mantiene la sequenza sulle uscite
  - Pacchetti che appartengono alla stessa connessione devono essere serviti, cioè trasferiti attraverso lo *switch fabric*, in sequenza per evitare una consegna sulle uscite fuori ordine

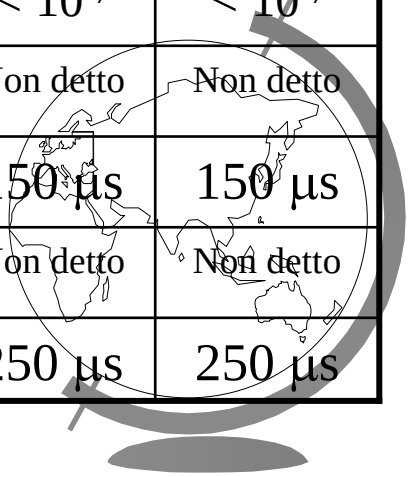
- I Bellcore hanno precisato requisiti di prestazioni per i sistemi di commutazione a larga banda in funzione di tre classi di qualità di servizio (QoS): QoS classe 1, QoS classe 3, QoS classe 4



OBIETTIVI DI PRESTAZIONI

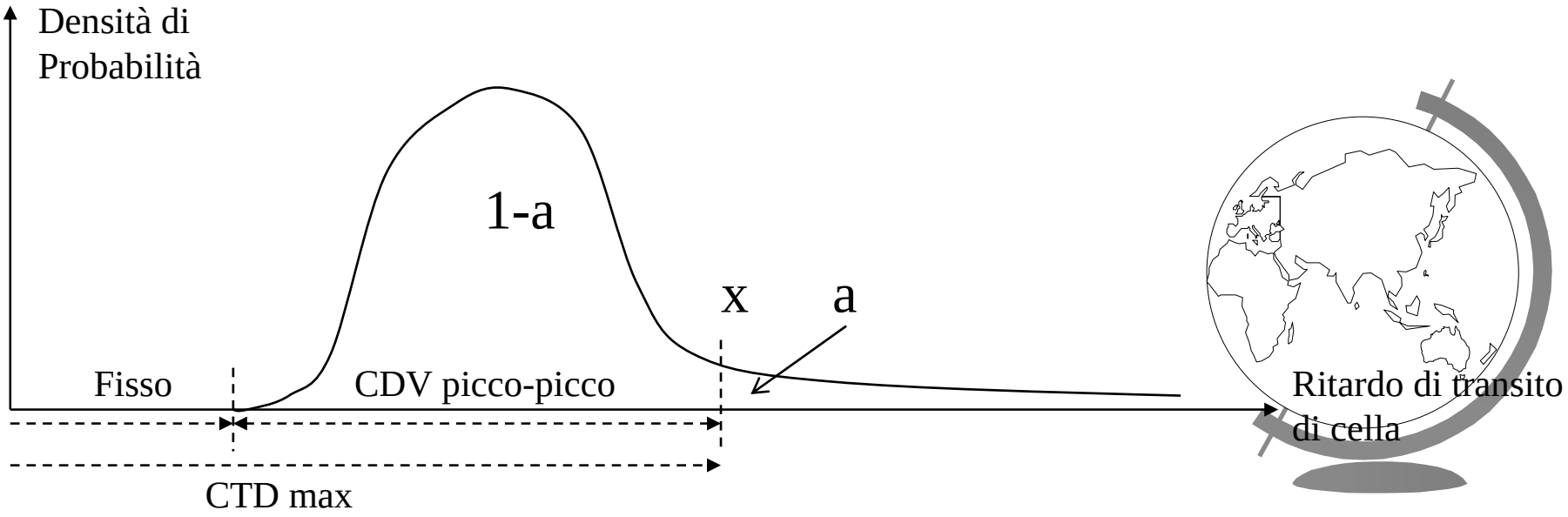
- QoS classe 1: per applicazioni molto sensibili alla perdita e supporto per emulazione di circuito
- QoS classe 3: per applicazioni sensibili al ritardo e connection oriented
- QoS classe 4: per applicazioni sensibili al ritardo e connection less
- Bit di Cell Loss Priority (CLP): CLP=0 (alta priorità), CLP=1 (bassa priorità)
  - Inizialmente ad 1, modificabile dallo switch fabric nel percorso ingresso-uscita

Parametro di prestazione	*CLP*	*QoS 1*	*QoS 3*	*QoS 4*
Probabilità di perdita di cella	0	$< 10^{-10}$	$< 10^{-7}$	$< 10^{-7}$
Probabilità di perdita di cella	1	Non detto	Non detto	Non detto
Ritardo di transito di cella (99 percentile)	1/0	150 μs	150 μs	150 μs
Variazione Ritardo di transito (10^{-10} quantile)	1/0	250 μs	Non detto	Non detto
Variazione Ritardo di transito (10^{-7} quantile)	1/0	Non detto	250 μs	250 μs



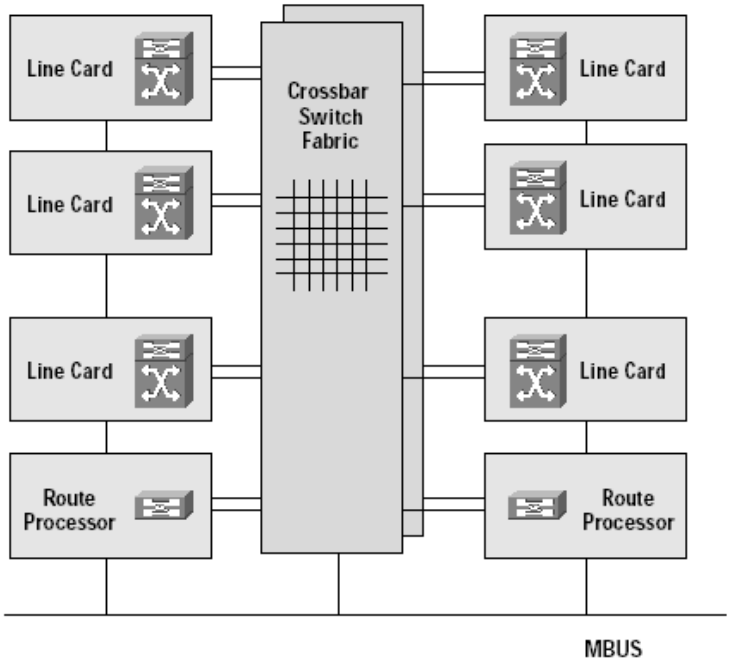
RITARDO DI TRANSITO DI CELLA

- Ritardo fisso: processamento dell'indirizzo IP, dell'header della cella, controllo degli errori, generazione del nuovo header
- Prob(CTD > 150μs) < 1-0.99: a=1%, x= 150μs
- QoS classe 1: Prob(CDV > 250μs) < 10^{-10}
- QoS classe 3 e 4: Prob(CDV > 250μs) < 10^{-7}



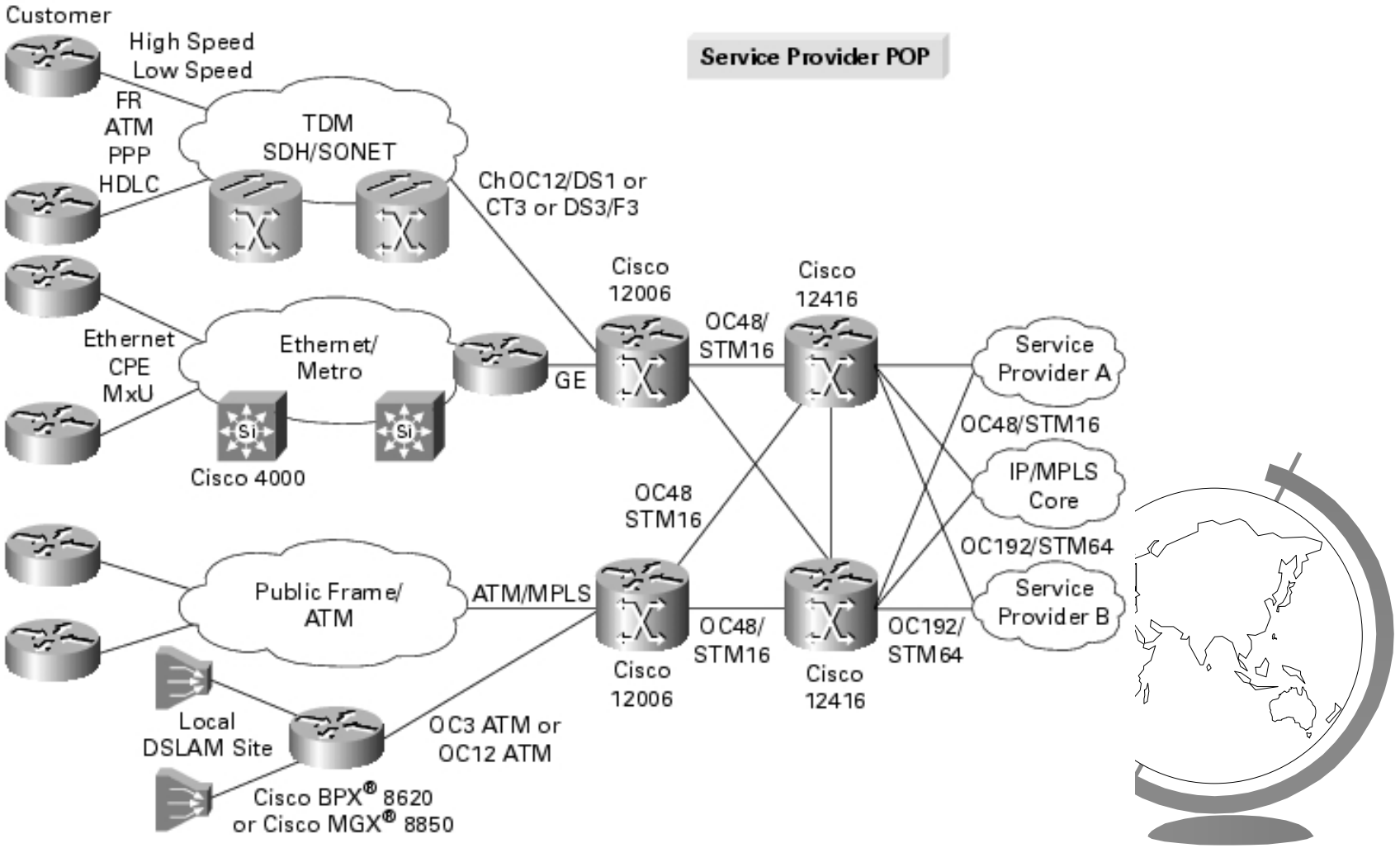
ROUTER CISCO SERIE 12000

12816 Router is the industry's first proven Terabit system available for building high-end IP/MPLS core networks and supporting high-speed edge services in a 16-slot, 40 Gbps (40G)/slot full rack chassis.
It delivers ***1.28 Terabits*** per second switching capacity with wire-speed feature performance
The Cisco 12416 Internet router is a 10 Gigabit, 16-slot chassis member of the Cisco 12000 series that provides a total switching capacity of 320 Gigabits per second (Gbps), with 20 Gbps (10 Gbps full duplex) capacity per slot.



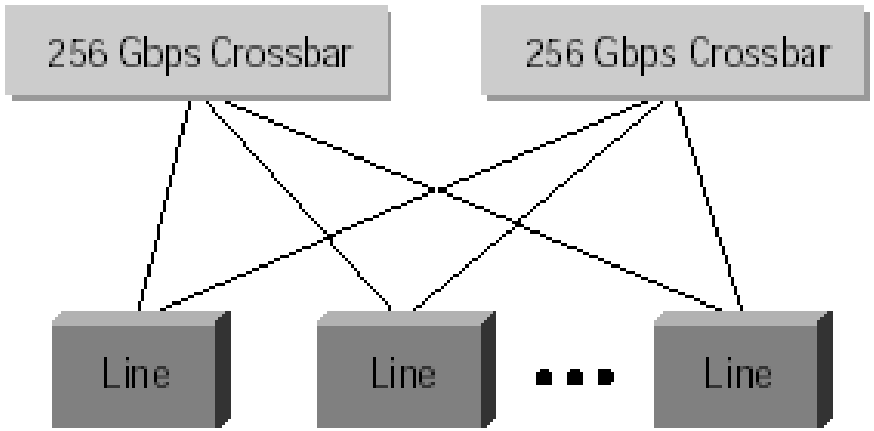


ROUTER CISCO SERIE 12000: ESEMPIO



ROUTER CISCO SERIE 7000

- “The switch fabric modules utilize a crossbar architecture to deliver 256-Gbps total capacity with a high-speed point-to-point connection to each line card. This provides a mechanism to forward packets between all point-to-point connections between the slots simultaneously. Many ports can thus be simultaneously transmitting and receiving data providing much higher aggregate throughput”

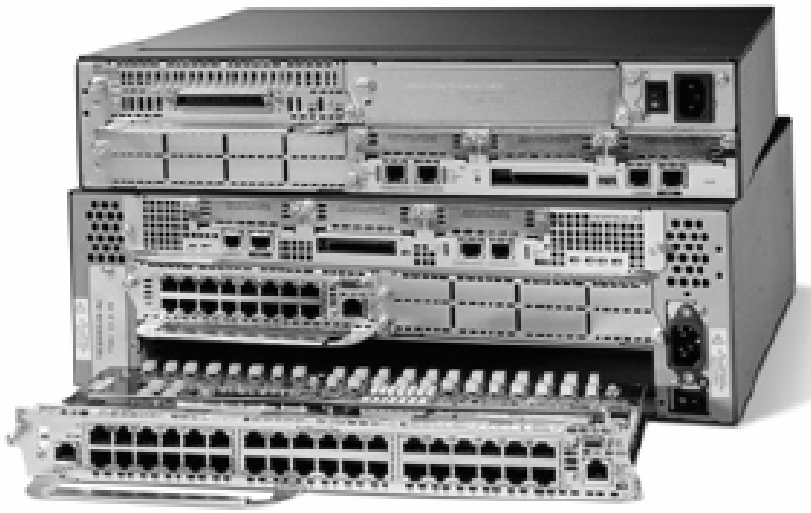




ROUTER CISCO SERIE 3700

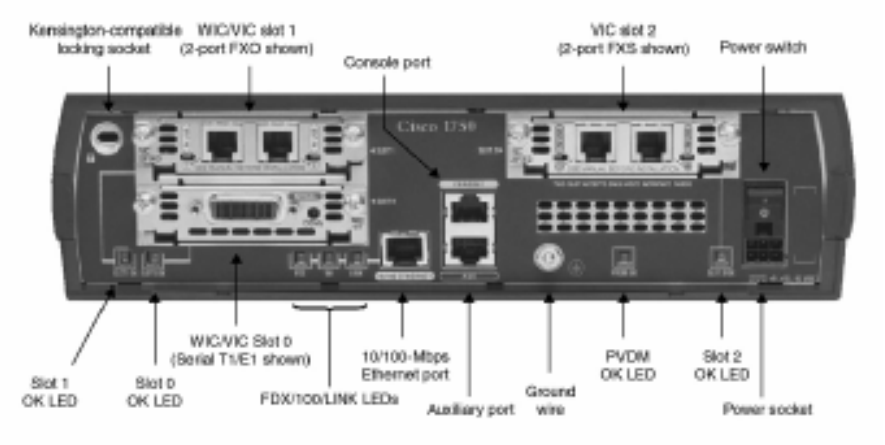
- NM-16ESW-1GIG—One 16-port 10/100 EtherSwitch NM with 1 GE (1000BaseT) port
- NMD-36-ESW-2GIG—One 36-port 10/100 EtherSwitch HDSM with 2 GE (1000BaseT)

Cisco 3725 and 3745 Series Platforms with EtherSwitch Modules



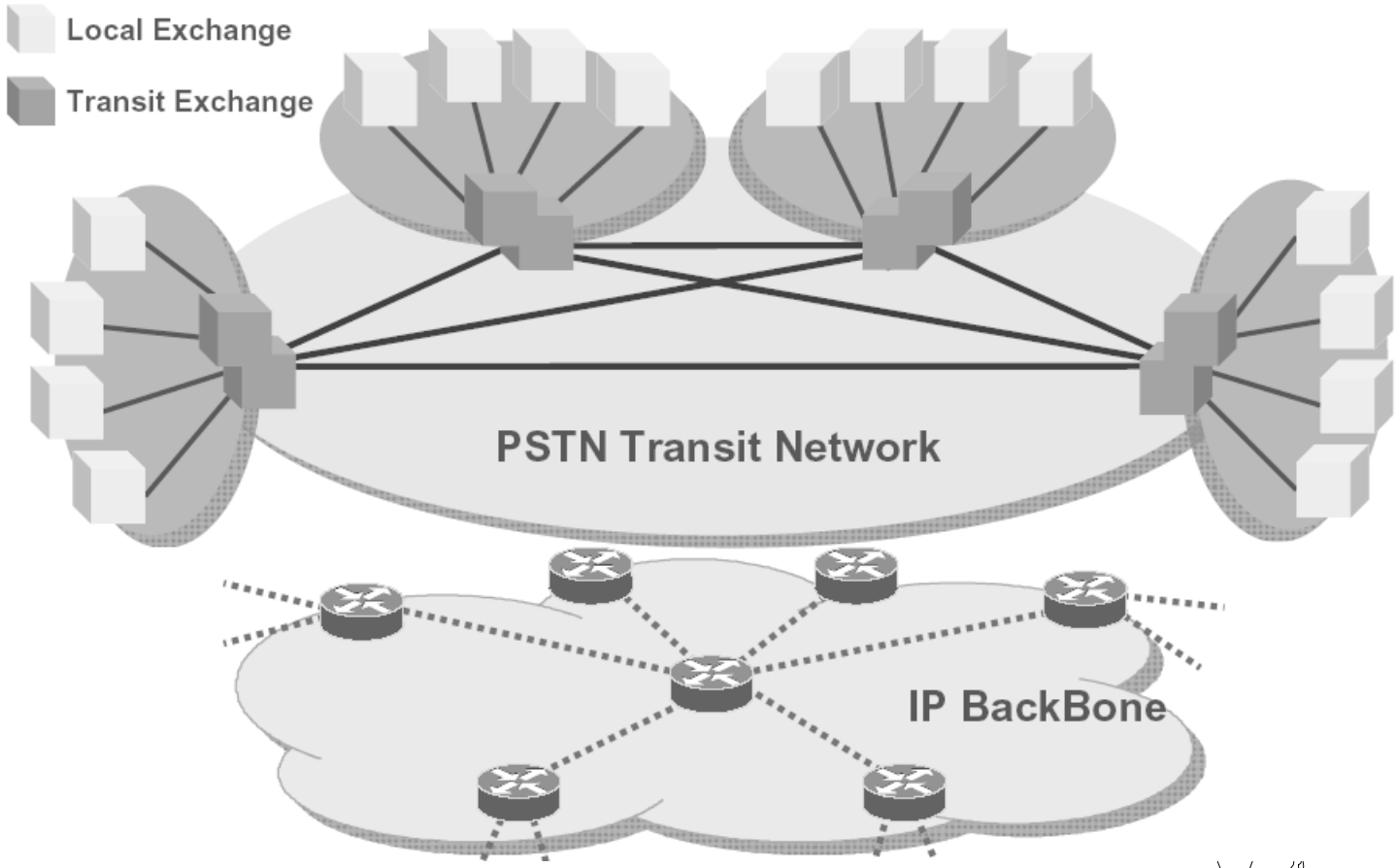
ROUTER CISCO SERIE 1700





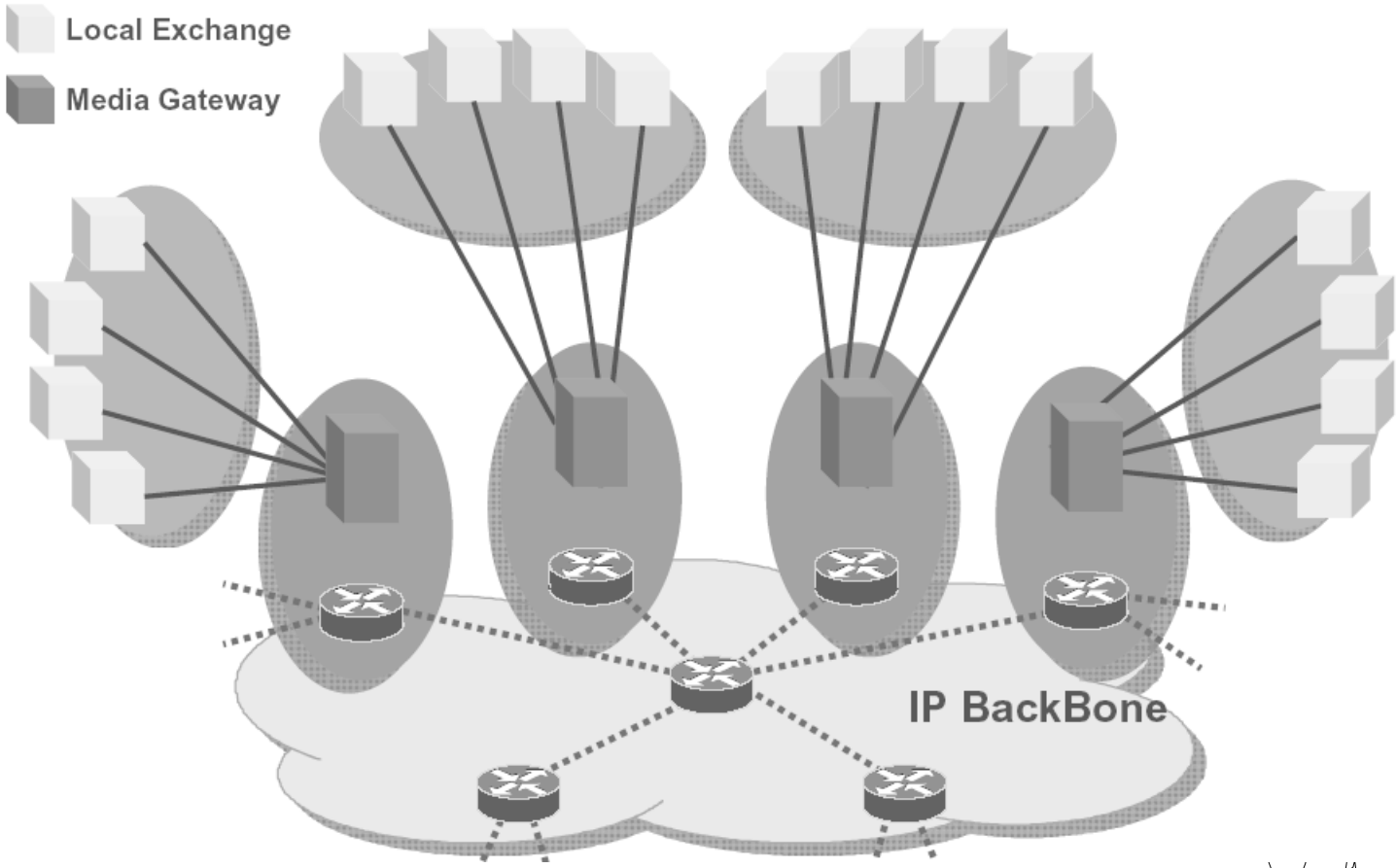


RETE TELEFONICA: EVOLUZIONE (1)



Da presentazione di Maurizio Siviero di TILAB

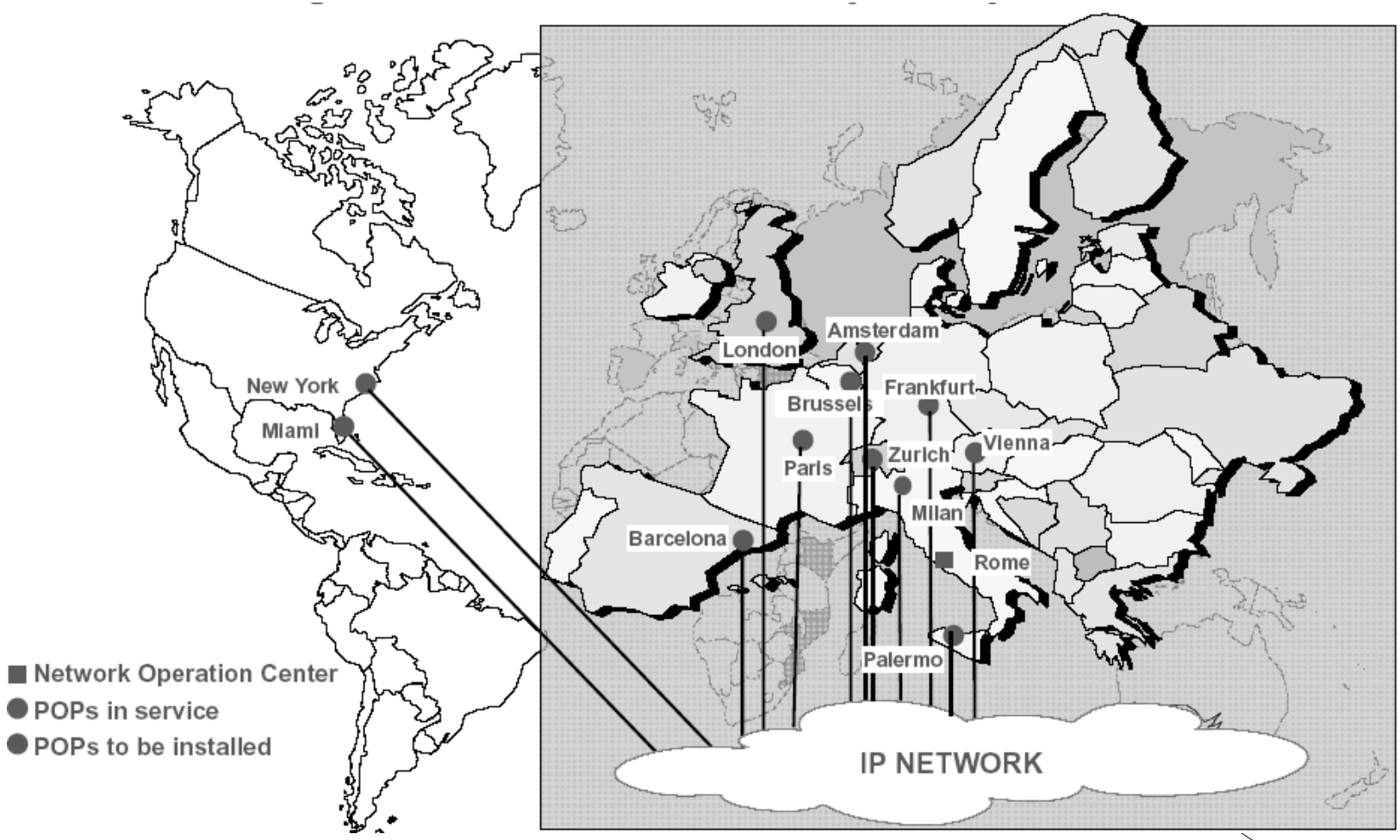
RETE TELEFONICA: EVOLUZIONE (2)



Da presentazione di Maurizio Siviero di TILAB

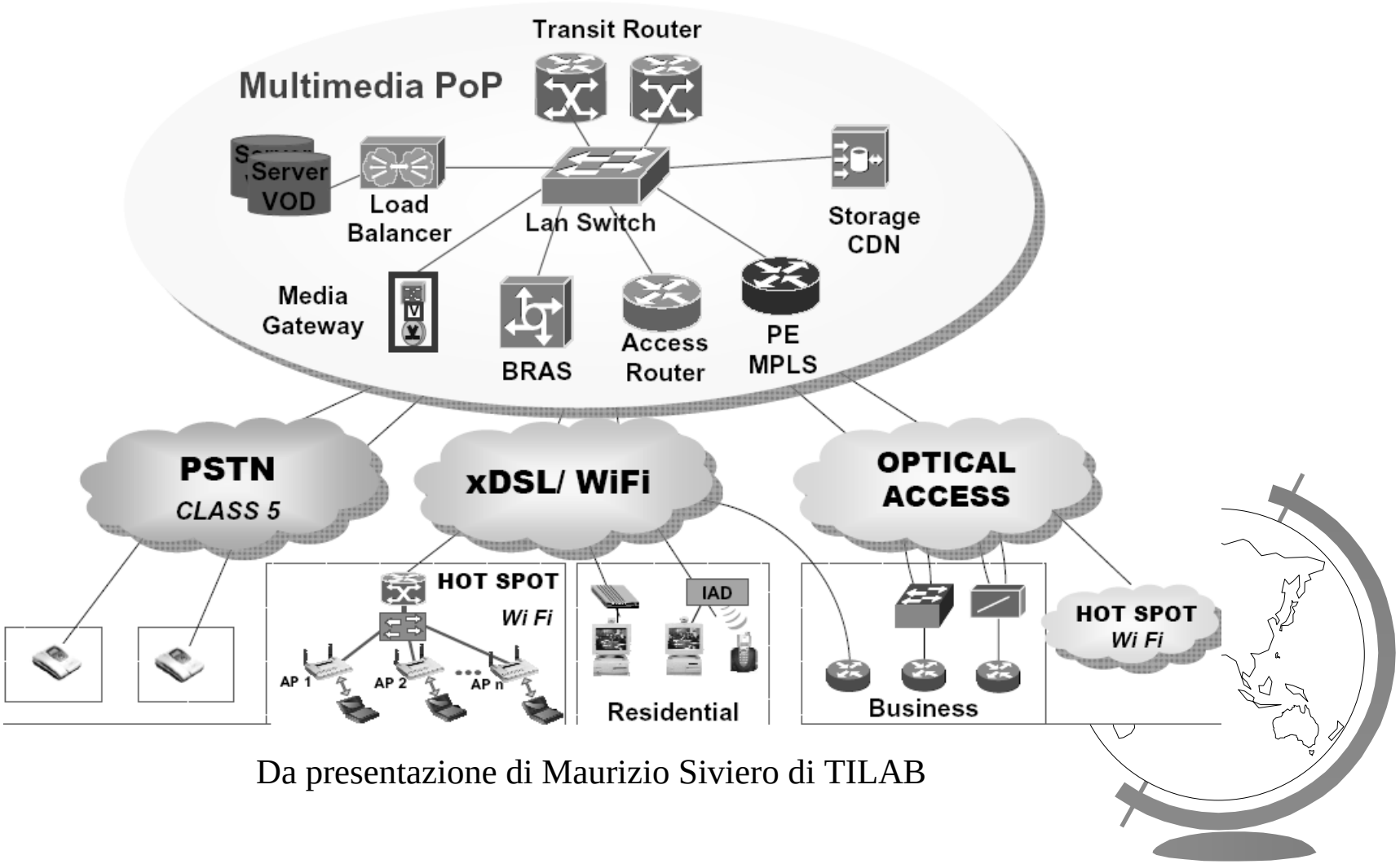


RETE TELEFONICA: TELECOMITALIA



Da presentazione di Maurizio Siviero di TILAB

RETE DI PROSSIMA GENERAZIONE



Da presentazione di Maurizio Siviero di TILAB