

# Qualità di Servizio

## Istruzioni per l'uso in GARR-G

Mauro Campanella

V Incontro del GARR - Roma 24-26 Novembre 2003

Mauro.Campanella@garr.it

# Agenda

- I servizi di QoS che saranno offerti da GARR-G
- Premium IP e LBE e stato su GÉANT
- Non solo QoS a livello 3
- Overprovisioning e CAR

*Per un tutorial su QoS :*

*<http://www.garr.it/ws4/agenda.shtml>*

*Tutorial: QoS*

## Architetture di QoS sviluppate da



- Un gruppo di lavoro su tematiche avanzate

di rete di



e



<http://www.dante.net/tf-ngn>



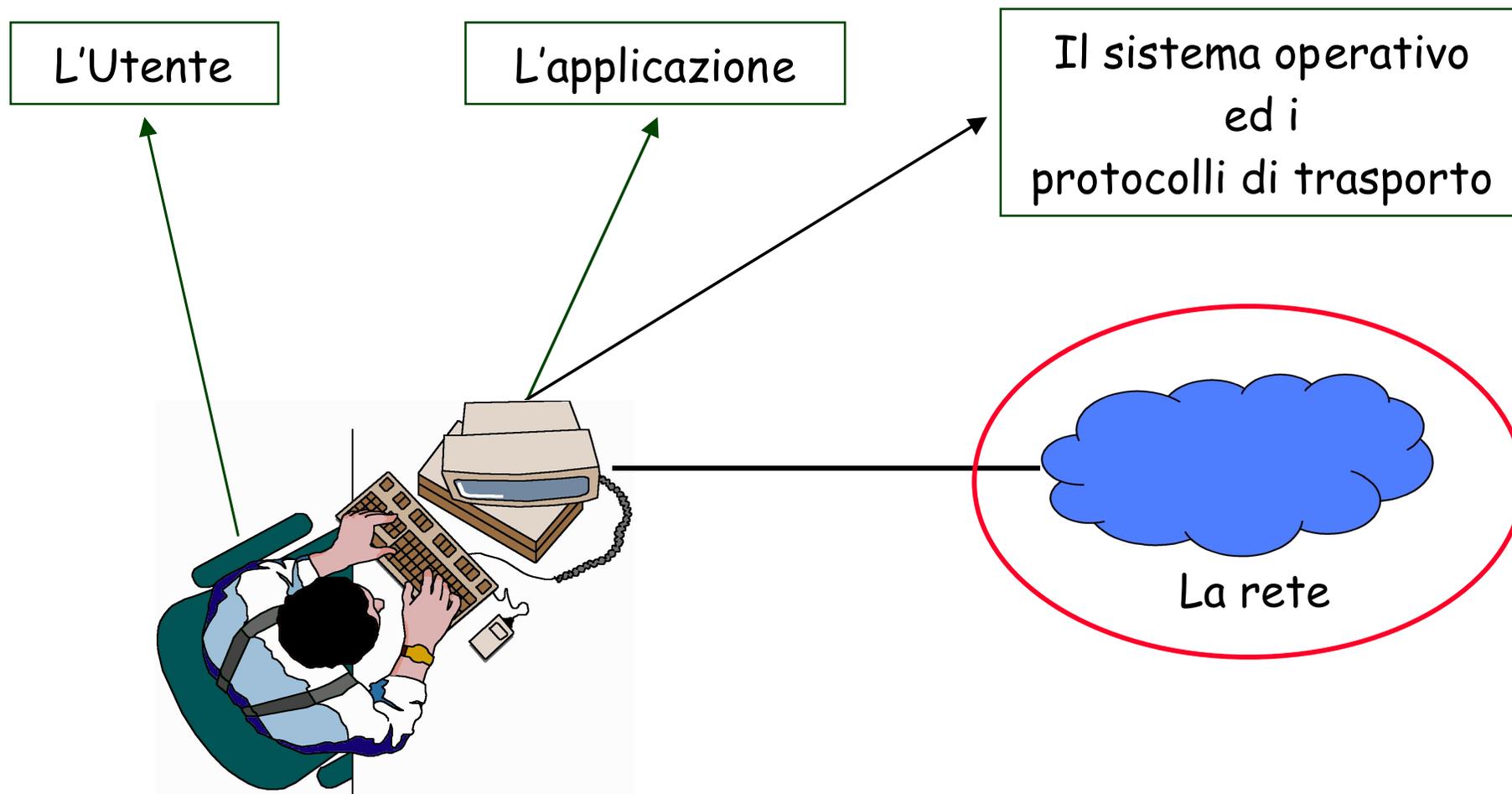
- Un progetto EU RN2 sulla qualità di servizio fra domini interconnessi (GARR ed altri).

<http://www.dante.net/sequin>



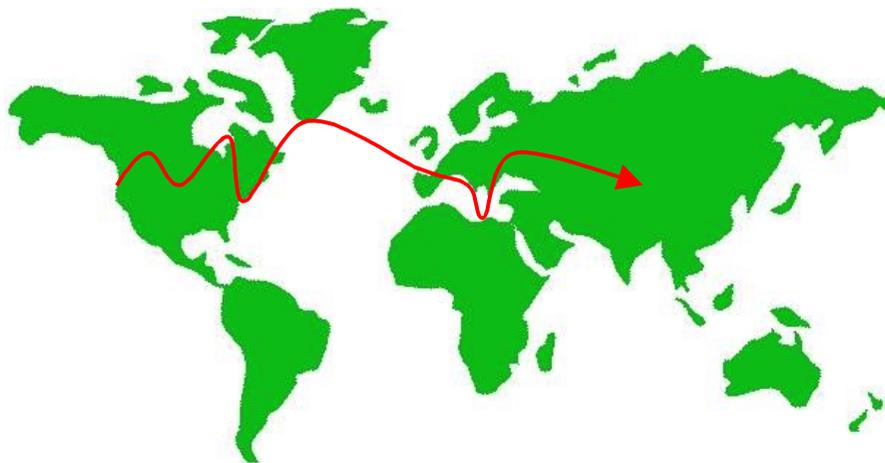
- Internet2 (<http://qos.internet2.edu/wg/>)

## Le componenti della QoS



## End to End

Il servizio di QoS *deve* essere presente in tutte le tratte del percorso per fornire garanzie *e2e*.

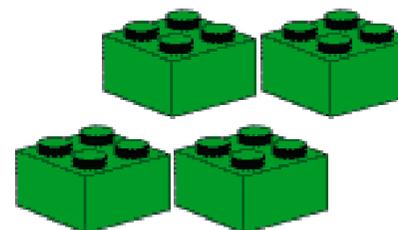


*ma*, anche abilitare la QoS su una singola tratta, può essere molto utile, per esempio su una particolare linea congestionata.

## Quantificazione della QoS a livello 3

Parametri scelti per quantificare un servizio di QoS con misure sul comportamento di pacchetti:

- ✓ - one-way delay (owd, tempo di attraversamento in una direzione);
- ✓ - one way IP packet delay variation (ipdv);
- ✓ - capacità;
- ✓ - perdita di pacchetti in una direzione.

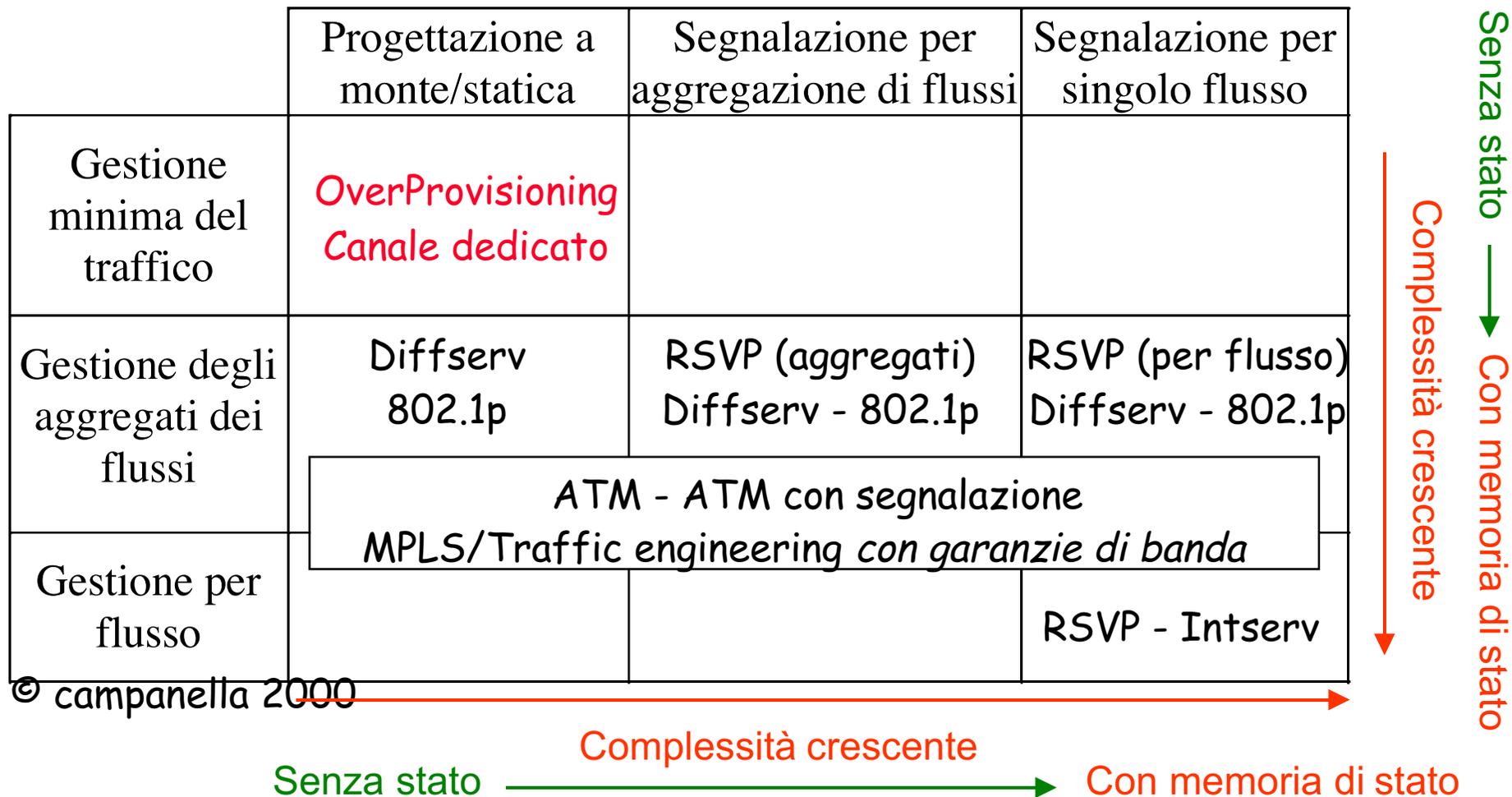


L'insieme è comune a IETF e ITU-T.

Nomenclatura e definizioni seguono la RFC 2330 (Framework for IP Performance metrics) e sono modificate secondo l'evoluzione seguita dal gruppo di lavoro IPPM di IETF.

## Architetture e protocolli di QoS end to end

### Inquadramento generale



## IP QoS in GARR-G

In GARR-G saranno disponibili 3 servizi di QoS a livello di rete per quanto riguarda il trattamento dei pacchetti con protocollo IPv4.

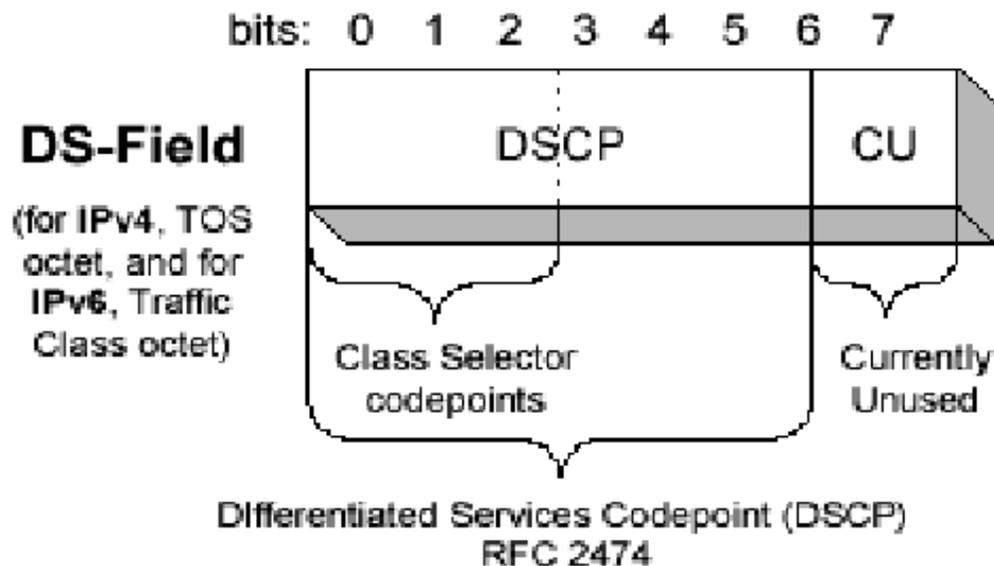
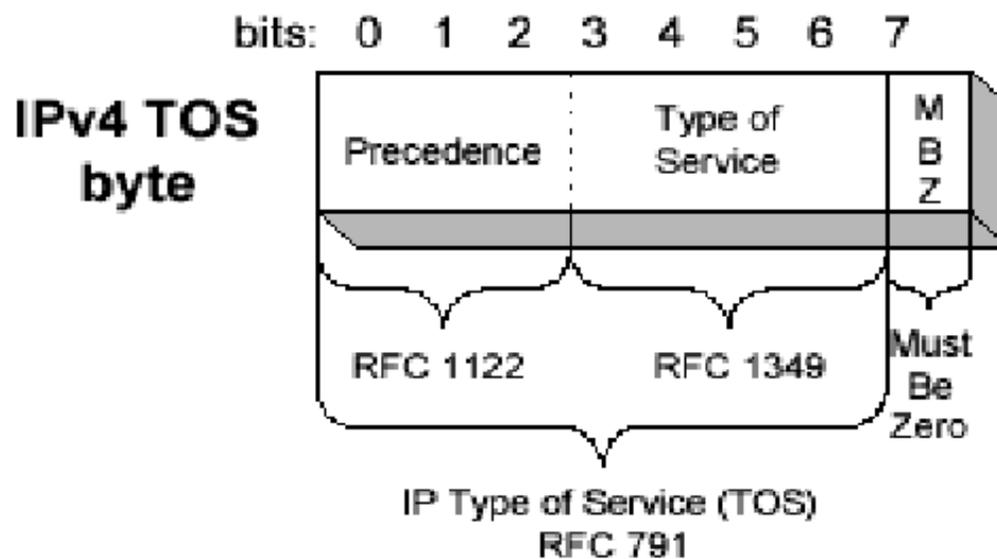
- Premium IP (emulazione di circuito e2e)
- Best effort
- Less than Best Effort (LBE)

Premium IP e LBE saranno introdotti gradualmente, di pari passo con il rinnovo delle apparecchiature e delle linee.

Questi servizi sono introdotti in tutte le reti della ricerca europee, in particolare sono già attivi in GÉANT.

E' intenzione estenderli anche a IPv6.

La ridefinizione  
del  
Byte ToS  
in  
DSCP  
in IPv4



## IP v4

Version == 4	4 bits IHL	8 bits DSCP	16 bits Total Length	
16 bits Identification			4 bits Flags	12 bits Fragment Offset
8 bits Time to Live	8 bits Protocol	16 bits Header Checksum		
32 bits Source Address				
32 bits Destination Address				

Classic IPv4 Header Format

## IP v6

Version == 6	8 bits Traffic Class	20 bits Flow Label		
16 bit Payload Length		8 bits Next Header	8 bits Hop Limit	
128 bits Source Address				
128 bits Destination Address				

IPv6 Header Format

# Agenda

- I servizi di QoS che saranno offerti da GARR-G
- Premium IP e LBE e stato su GÉANT
- Non solo QoS a livello 3
- Overprovisioning e CAR

## Caratteristiche base

Basati sul trattamento del singolo pacchetto attraverso il valore del campo di qualità di servizio.

**Premium IP:** emulazione di un circuito e2e che garantisca

- one-way delay limitato superiormente
- IPDV limitato superiormente e piccolo
- packet loss nullo o molto piccolo
- capacità garantita
- fra due specifici prefissi IP (destination aware)
- <http://archive.dante.net/sequin/> [deliverables e publications]

**Less than Best Effort:** senza alcuna garanzia

- utilizza la banda lasciata libera dalle altre classi.
- <http://qos.internet2.edu/wg/>

## Architettura Premium IP

Pensato per offrire l'equivalente di una linea punto-punto attraverso domini di gestione multipli .

- basato su Differentiated Services, utilizza l'Expedited forwarding Per Hop Behaviour (EF PHB)
- richiede che l'interfaccia fra domini si comporti come EF PHB
- richiede la specifica sia della sorgente che della destinazione (destination aware, non è assegnabile la pura capacità Premium)
- effettua un controllo rigoroso della capacità all'inizio, basato sui prefissi IP, e poi più lasco basato sul valore del DSCP per scalabilità
- pacchetti sono marcati con un valore di DSCP 46 (EF - 101110)
- evitare perdite di pacchetto (permettere una piccola burstiness)

## Architettura Premium IP

- Pacchetti con altri valori di DSCP sono lasciati inalterati.
- La frazione della capacità delle linee assegnata incrementalmente al servizio Premium IP è limitata a non più del 10%
- i pacchetti Premium sono smistati con priorità massima
- minimizzare il numero di azioni per nodo
- approccio modulare che permetta tecniche diverse di realizzazione per ogni collegamento e dominio e che permetta ad ogni dominio di attivare il servizio dove e quando necessario

### Nota:

Premium IP non cerca di risolvere il problema generale, ma piuttosto è un modello che può essere realizzato a livello europeo, usando gli strumenti disponibili oggi.

# Premium IP

Classifica (DSCP)  
Smistamento a massima  
priorità su tutti i nodi

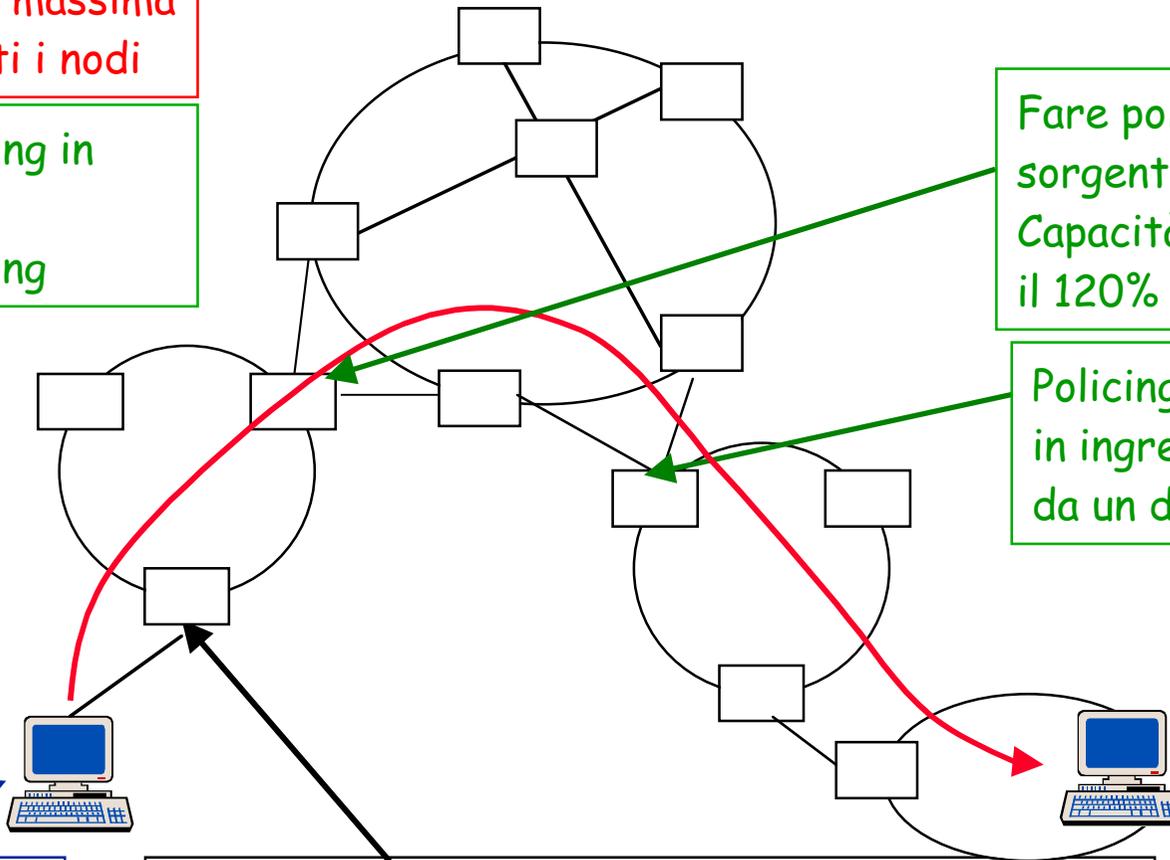
Non fare policing in  
uscita  
Non fare shaping

Fare policing su [coppia AS  
sorgente e destinazione e  
Capacità aggregata] fino a circa  
il 120% del concordato

Policing può essere evitato  
in ingresso quando si riceve  
da un dominio "a monte".

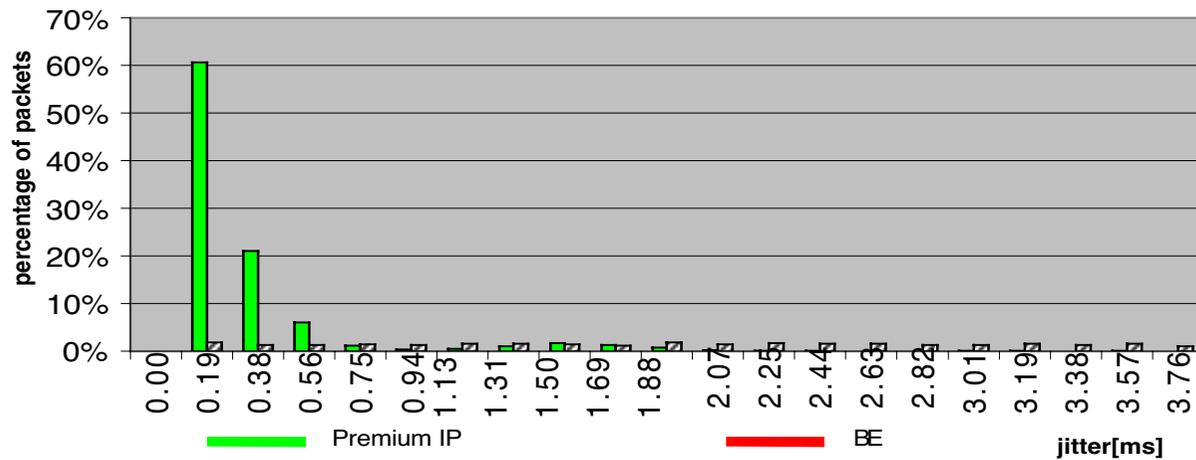
Fare shaping SOLO  
nella sorgente

Classificare (prefissi IP sorgente  
e destinazione)  
Fare policing rigoroso su capacità  
Marcare con DSCP - traffico in eccesso scartato



# Alcuni risultati (end-to-end)

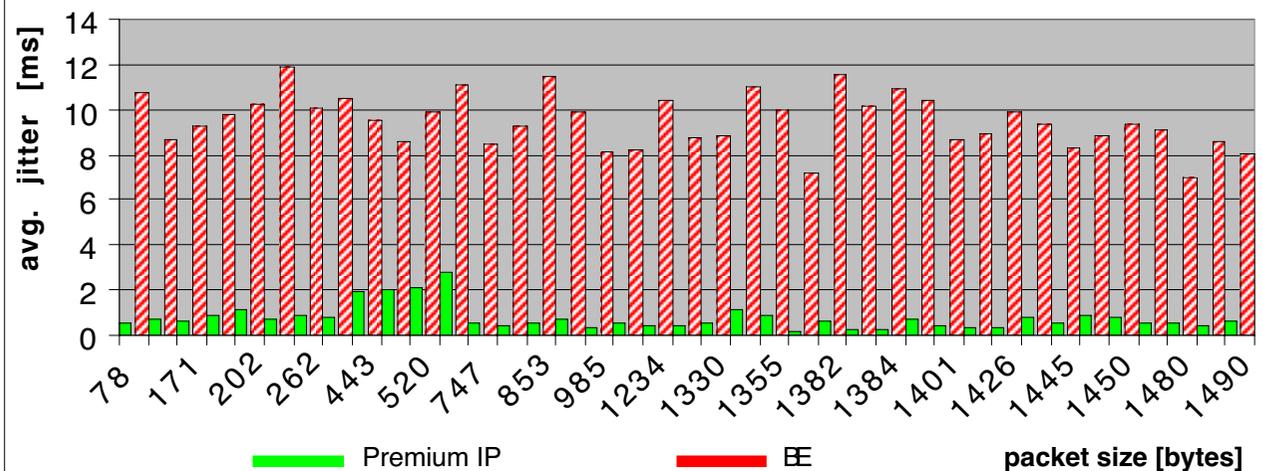
Jitter distribution in VBR traffic - BE & Premium IP



Distribuzione in percentuale dei valori di ipdv per pacchetto

ipdv in funzione della dimensione del pacchetto

Avg. jitter vs. packet size - BE & Premium IP



## Less than Best Effort

LBE è una classe di traffico che utilizza la banda rimasta libera dopo l'uso da parte di ogni altra classe.

- In caso di competizione per qualsiasi risorsa il traffico LBE è il primo ad essere scartato (WFQ a bassissima percentuale)
- Usa il valore DSCP 8 (001000) - compatibile con il servizio scavenger di Internet2 .

Utile per trasferimenti massicci di file a costo basso, accesso dei dormitori degli studenti.

La dorsale si limita a configurare (in ogni router...) le tecniche di accodamento e scarto, l'utente deve marcare da sè il traffico e deve avere un incentivo (economico) per farlo.

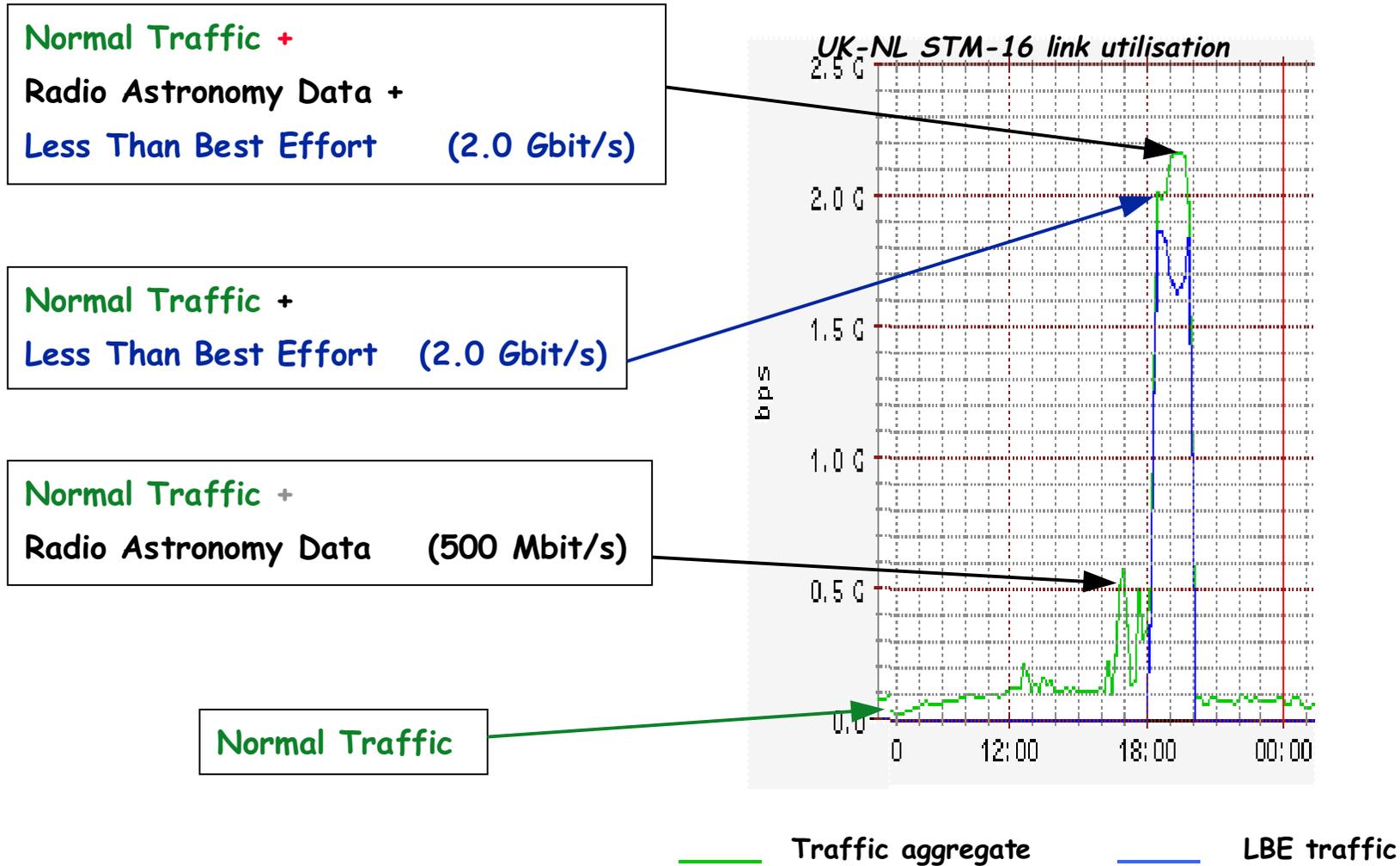
## Less than Best Effort

LBE non richiede la conoscenza della destinazione (destination unaware).

In tale caso è però fondamentale per il corretto funzionamento che le applicazioni che ricevono una richiesta con DSCP 8 rispondano con pacchetti con lo stesso valore di DSCP.

Richiede pertanto la modifica delle applicazioni (per ora quasi nessuna prende in considerazione il campo DSCP che NON è accessibile in alcuni stack attraverso le socket)

# ER2002 Demo - VLBI - dataGRID



## Premium IP su GÉANT

Si veda <http://archive.dante.net/nep/geantqos/>

Nei Juniper vi è un meccanismo di Weighted Round Robin

- 90% della capacità della linea è dedicata a Premium IP su buffer hardware dedicato (per costruzione al massimo il traffico Premium è il 10% del totale !)
- 5% per Best Effort
- 5% per network control.

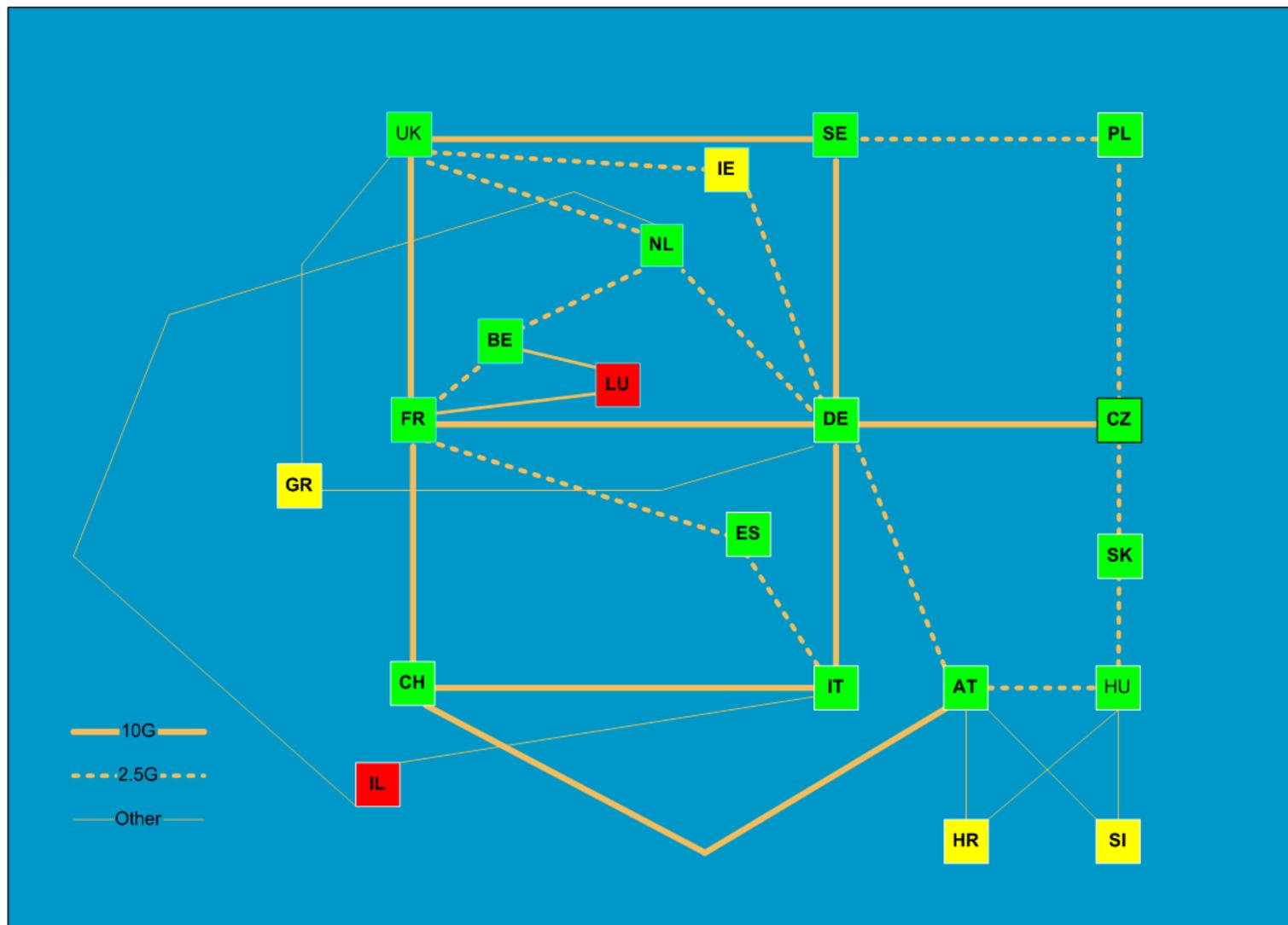
In totale le code hardware utilizzabili su Juniper sono 4 (simile CISCO)

Su un CISCO si può usare, per esempio, Priority Queueing per Premium IP.

Per altri dettagli si veda:

<http://www.switch.ch/network/sequin/conf-samples/>

# Configurazione QoS su GÉANT



## Strumenti e sviluppi in atto

**traceroute** modificato per evidenziare anche il campo DSCP

<http://www.switch.ch/misc/leinen/fasl/traceroute/>

<http://archive.dante.net/nep/geantqos/>

**Interfaccia web di riservazione Premium IP a Livello europeo**  
(DANTE, SA3 GN2).

**Inter ed intra-domain monitoring** (fondamentale per debugging e controllo)

Performance monitoring activity (incluso in JRA1 GN2)

<http://www.dante.net/tf-ngn/perfmonit/>

**PERT** Performance Enhancement Response Team (SA3 GN2)

Un anello di congiunzione fra Applicazioni e trasporto

<http://www.dante.net/tf-ngn/pert/>

## Come si attiverà Premium IP

- Quantificazione delle caratteristiche dell'applicazione (molto spesso capacità massima utilizzata ) in funzione del set-up hardware  
(per esempio ricordarsi che TCP è intrinsecamente a burst ed un PC con FastEthernet invierà burst a 100 Mb/s  $\Rightarrow$  usare scheda a 10Mb/s)
- Identificazione prefissi IP della sorgente e della destinazione
- Accordi fra utente sorgente e destinazione per il reciproco invio e ricezione di banda Premium
- Configurazione del dominio di LAN affinché i flussi arrivino al router attraverso un percorso che ne preservi la QoS (802.1P/Q, cavo di rete dedicato) (richiede attivazione nei nodi locali e almeno nel router utente di tecniche di classificazione ed accodamento preferenziale in e out)
- Richiesta ai relativi NOC della capacità richiesta (via web)
- Debugging e controllo

# Agenda

- I servizi di QoS che saranno offerti da GARR-G
- Premium IP e LBE e stato su GÉANT
- Non solo QoS a livello 3
- Overprovisioning e CAR

## Ma

Quanto abbiamo visto descrive come fornire Qualità di Servizio al puro livello IP.

Esistono inoltre inoltre possibilità diverse di fornire un servizio corrispondente alle qualità richieste, tipicamente di capacità, e2e:

- overprovisioning
- circuiti dedicati a livello 1 (lunghezze d'onda)
- circuiti virtuali dedicati a livello 2 (SDH, ATM CBR, FrameRelay)
- *in futuro*: Capacità a richiesta in tempo "quasi" reale
  
- agire solo sui punti di congestione, utilizzando l'overprovisioning delle altre parti (con CAR, WFQ, WRR, ... )

## Alla base della QoS di rete

La rete stessa deve essere affidabile, cioè soddisfare ad alcuni requisiti di base:

- stabilità fisica (livello 1) del data link (livello2) e del routing (3)
- avere un Bit Error Rate di almeno  $10^{-12}$

*Inoltre*

- la minima dimensione della MTU dovrebbe essere scelta tale da evitare la frammentazione.
- la quantità di pacchetti duplicati e fuori sequenza deve essere al livello fisiologico attuale (che non è nullo, ma molto piccolo)

*e soprattutto:*

- prestazioni adeguate dell'hardware di rete;

## Per chiarezza

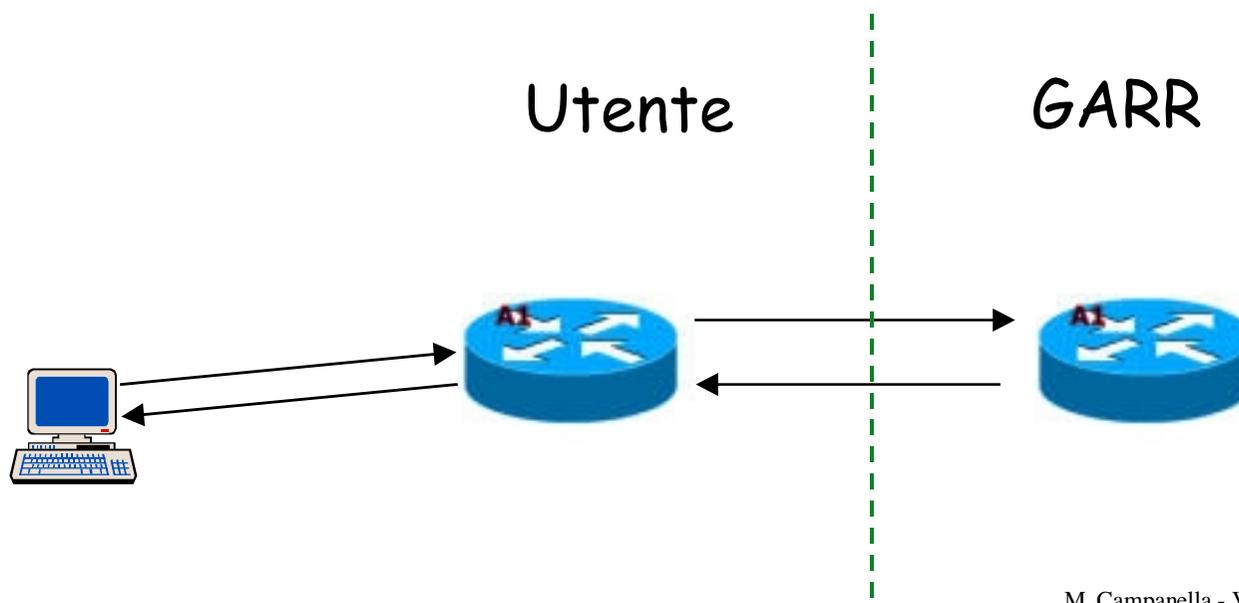
Una VPN a livello 3 ed una VLAN a livello 2 o 3 non forniscono garanzie di QoS, a meno che non le siano applicati algoritmi di prioritizzazione e gestione della capacità (basati o meno su tag DSCP).

Un LSP MPLS analogamente da solo non fornisce alcuna garanzia di QoS (traffic engineering non è sinonimo di QoS).

Un circuito dedicato non garantisce da solo una trasmissione con Qualità di Servizio se usato da più di un flusso

## Full Duplex

Le tecniche base di classificazione ed accodamento per avere qualità di servizio deve avvenire in entrambi i sensi nella trasmissione. Questo implica filtri sia nella parte GARR che utente. I filtri "a monte" impediscono che la capacità delle linee venga utilizzata per pacchetti che poi vengono scartati.



# Agenda

- I servizi di QoS che saranno offerti da GARR-G
- Premium IP e LBE e stato su GÉANT
- Non solo QoS a livello 3
- Overprovisioning e CAR

## Overprovisioning

Due possibili definizioni:

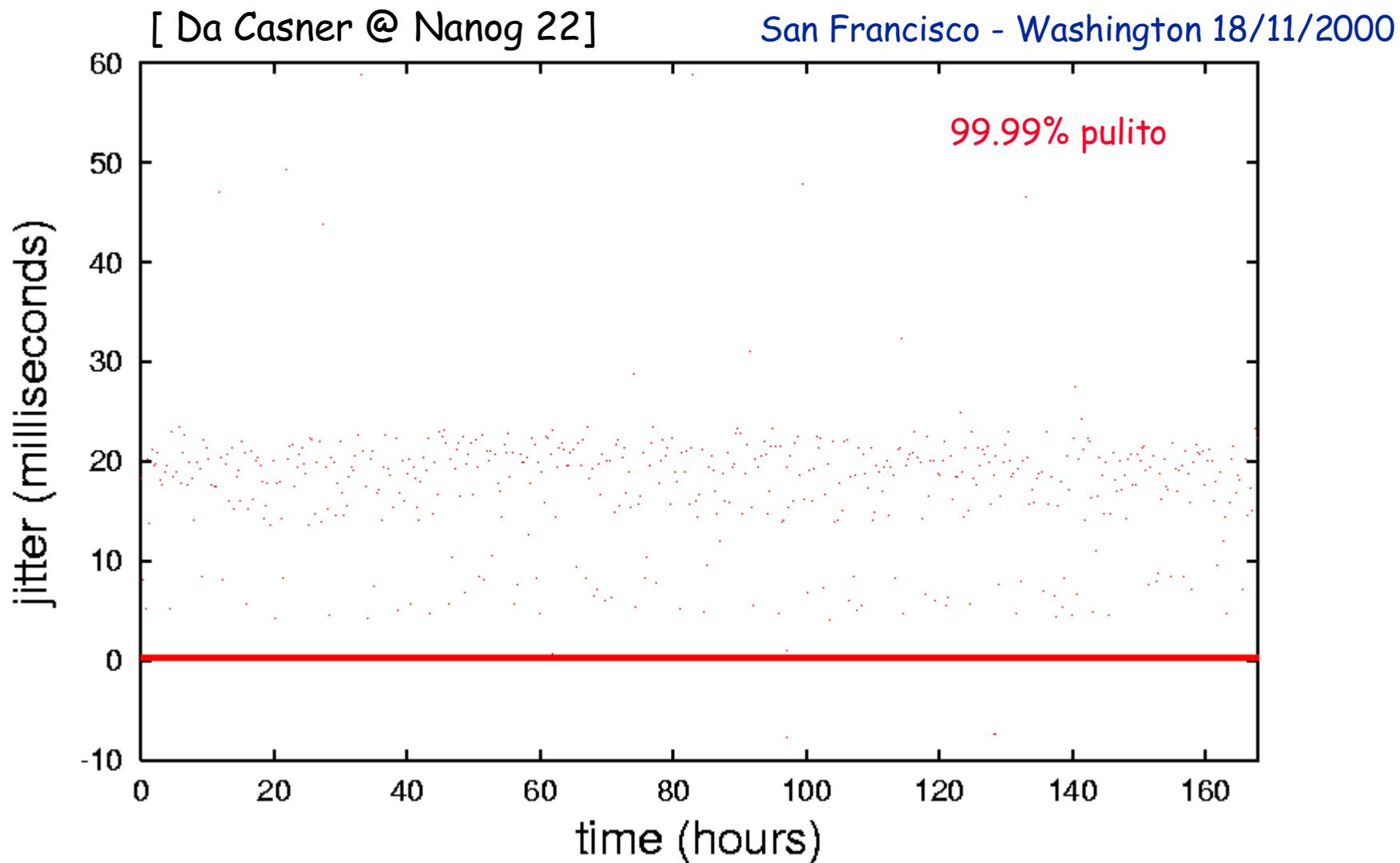
- carico istantaneo della tratta mai superiore al 30%
- nessuna perdita di pacchetti (più debole)

Il sovradimensionamento funziona (fornisce garanzie per i parametri di QoS) nel 99.9 % dei casi, ma la capacità è lontana (per ora) dall'essere abbondante su tutta l'Europa.

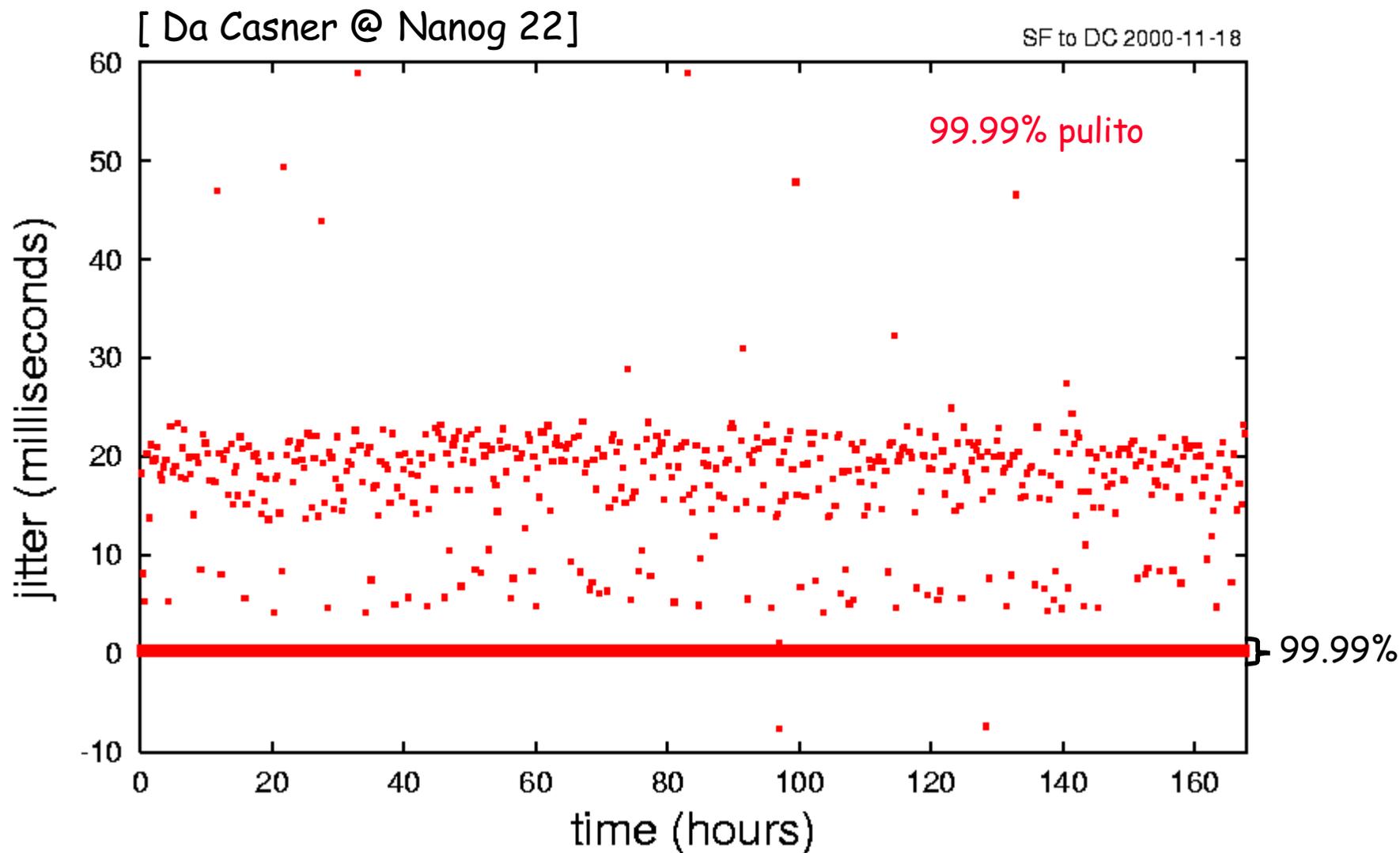
Anche in molte LAN la capacità non è sovradimensionata.

Overprovisioning permette comunque di fornire una qualità di servizio più che accettabile per la maggior parte delle applicazioni.

# Tier 1 USA - dorsale



## Tier 1 US dorsale (continua)



## Overprovisioning (continua)

Lo 0,001% di deviazione del jitter dallo zero è dovuta a:

- problemi di routing ;
- configurazioni dei timers dei protocolli di routing ;
- ARP cache timeouts;
- ...

E' soprattutto instabilità / errori / sviste dello strato software nei router o switch.

## Committed Access Rate

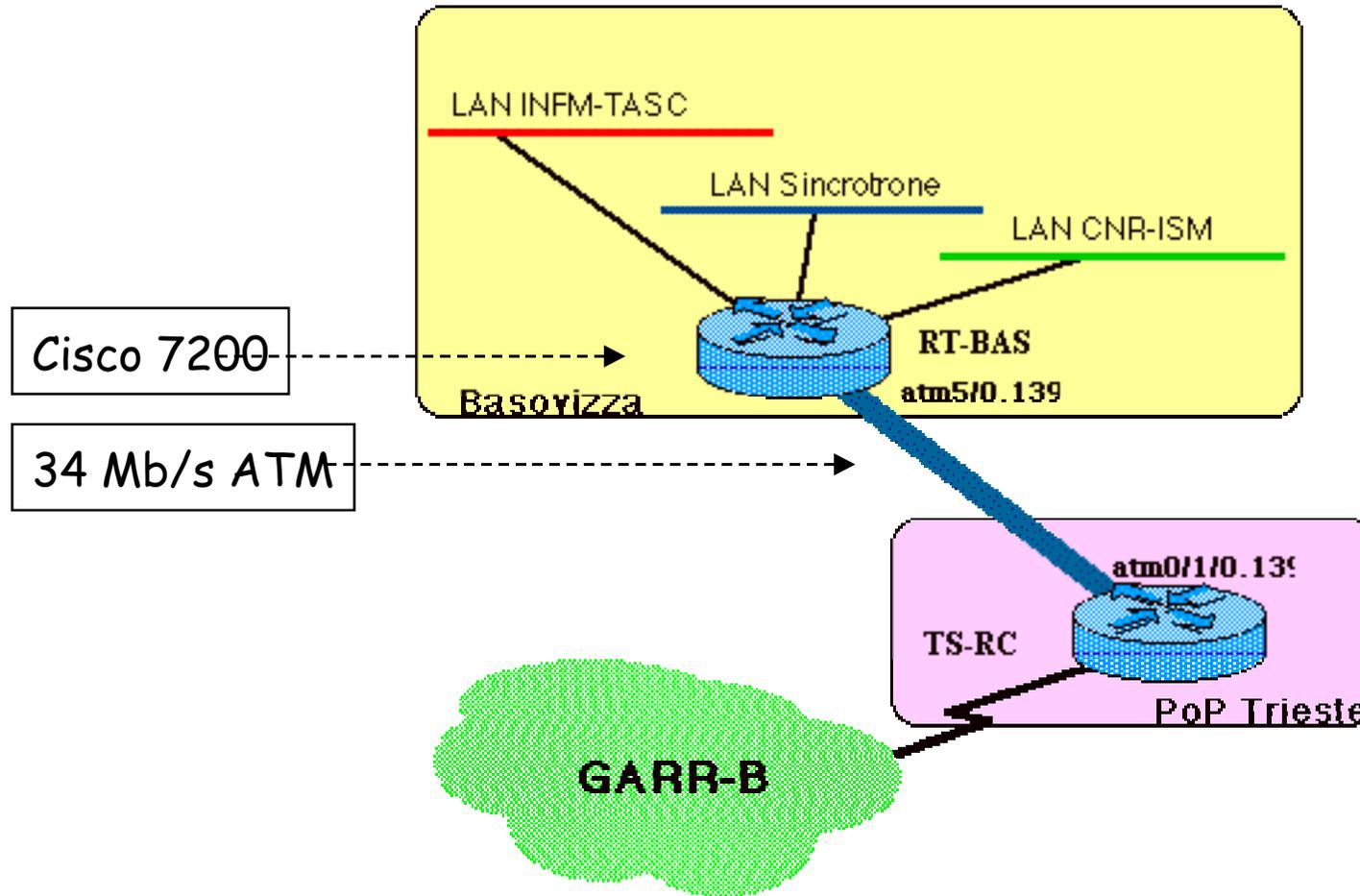
Problema: tre enti diversi condividono la stessa linea verso la dorsale. Configurare QoS sulla linea per garantire che ciascun utente abbia un minimo garantito di capacità e non ecceda la capacità contrattuale, sia per il traffico in uscita che per quello in ingresso.

Soluzione:

- usare CAR sulla dorsale e/o sul router utente
- classificare secondo gli indirizzi IP utente
- scartare il traffico in eccesso (permettendo un margine del 10% in più)
- usare lo stesso peso per ogni utente

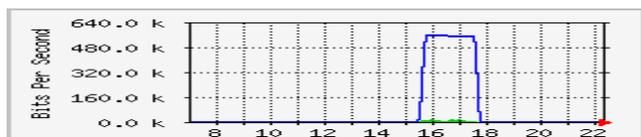
marcatatura, formatura e gestione congestione non sono necessari

# CAR (continua)

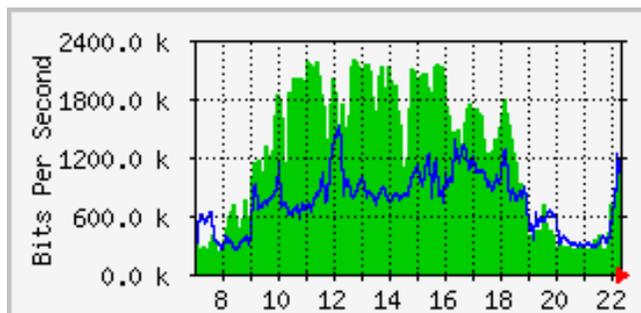


# CAR (continua)

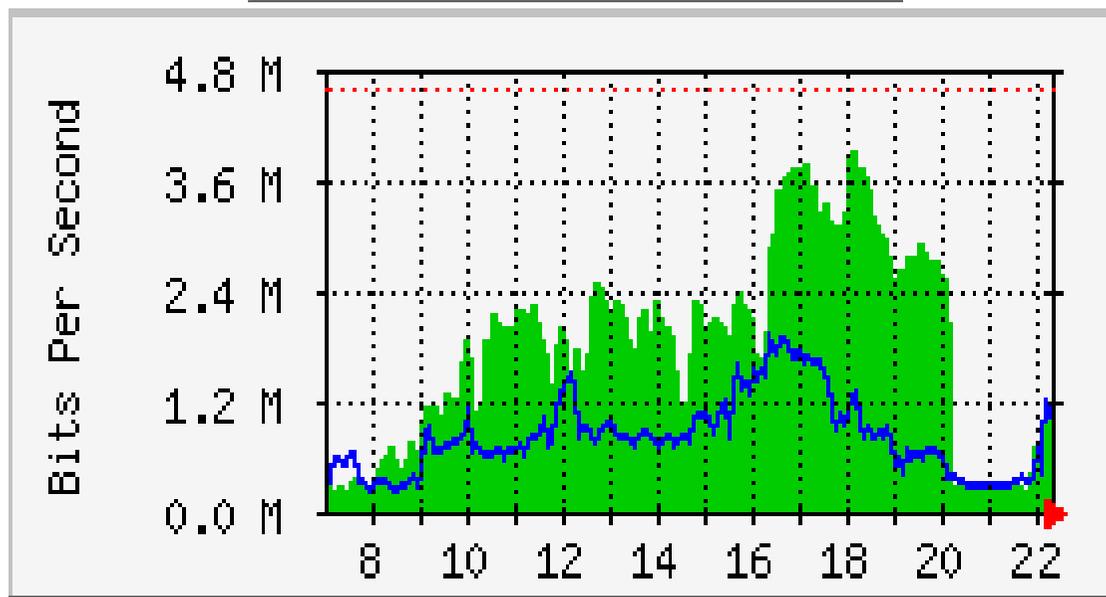
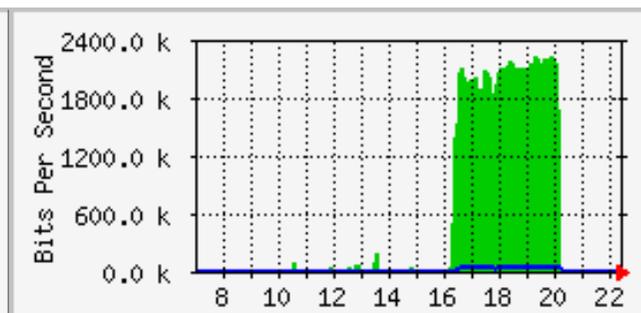
max 640 Kb - CNR



max 2400 Kb - Syncr.



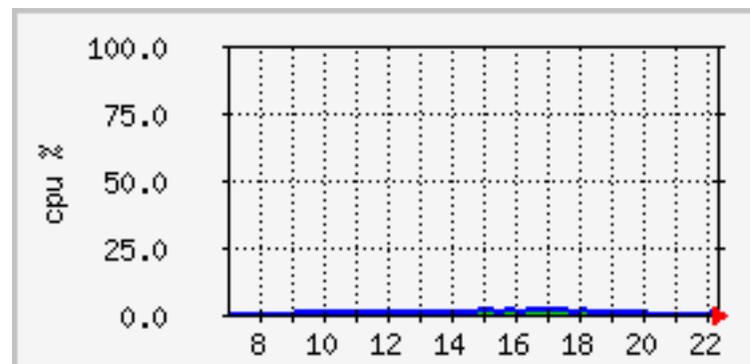
max 2400 Kb - INFM



5000 Kb max - traffico TOTALE

## CAR (continua)

L'utilizzo di CPU è  
qualche per cento



Funziona da circa un anno senza problemi.

Si veda <http://www.garr.it/docs/garr-car-01.shtml>

## Referenze ulteriori

Tool da installare in Linux per avere maggiori informazioni sulle performance di rete (di TCP)

<http://www.web100.org/>

Unn insieme di regole base per vari sistemi operativi per aumentare le performance di trasferimento con TCP

[http://www.psc.edu/networking/perf\\_tune.html](http://www.psc.edu/networking/perf_tune.html)

Modifiche a TCP per permettergli di raggiungere e mantenere trasferimenti a Gigabit/s

<http://www.icir.org/floyd/hstcp.html>