Premium IP service

Mauro Campanella
INFN-GARR
Mauro.Campanella@garr.it

IP Premium goal

Provision QoS for the European research users in the form of an end to end network service offering the equivalent of a leased line.

The service has to be implemented by combining border to border services provided by the NRENs and GEANT networks

The service should be simple, modular, scalable, adapt to network changes easily, based on IP and independent from the transport technology.

The implementation and Service Level Agreements have to match the current status of hardware availability and network topology

Synergy of

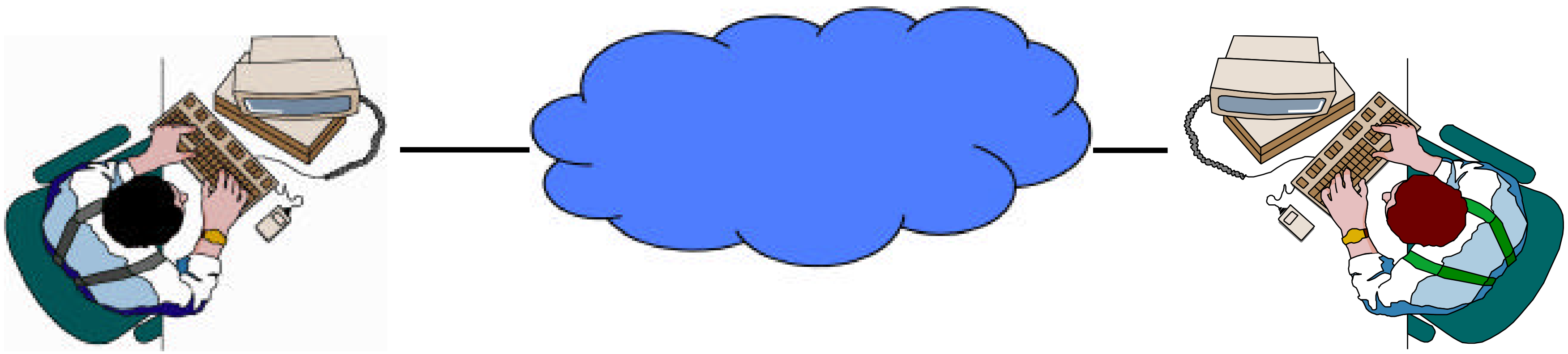
- A joint DANTE and TERENA task force on advanced networking research
  http://www.dante.net/tf-ngn

- A RN2 project on QoS on interconnected domains
  http://www.dante.net/sequin

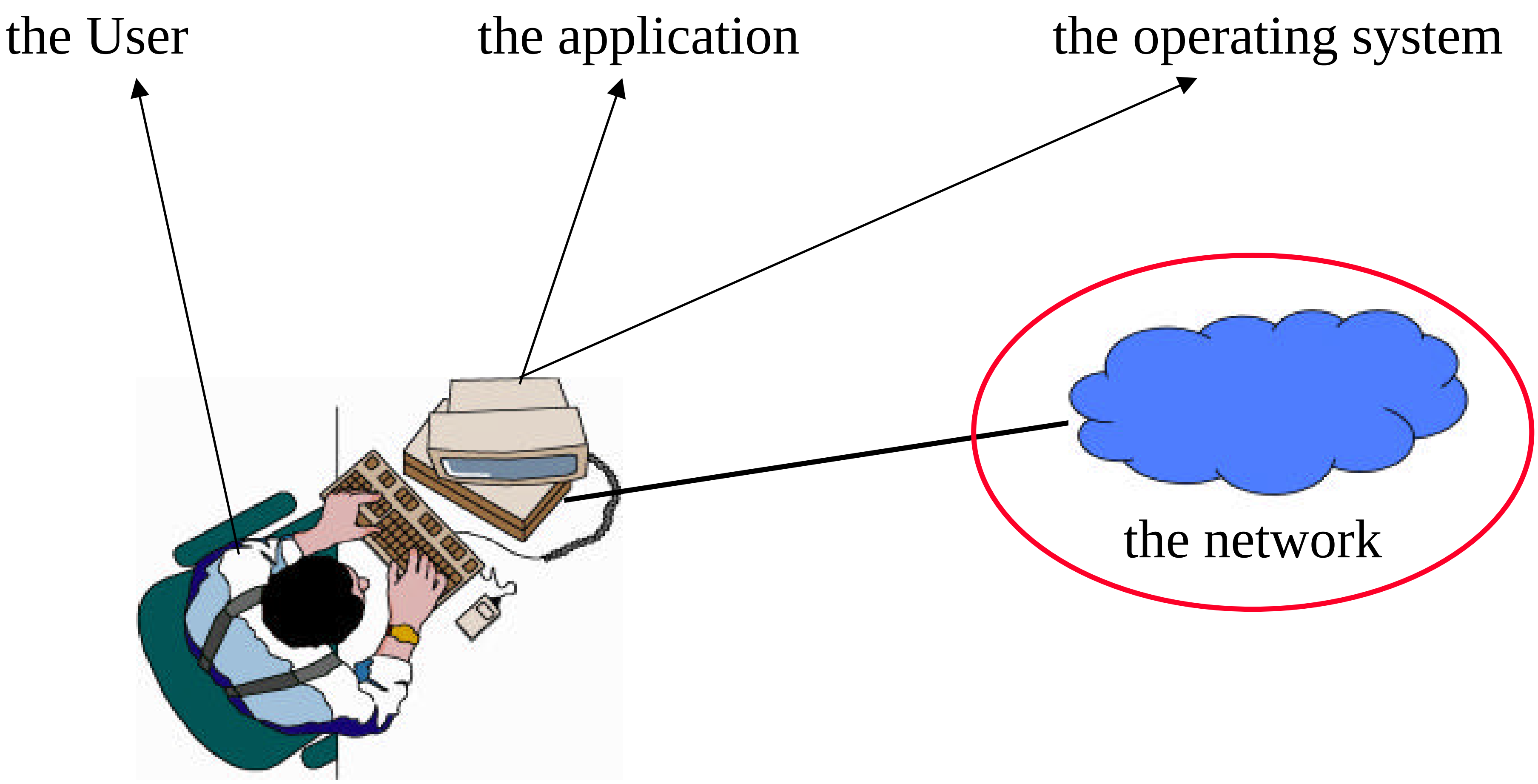
Agenda

- Qualitative definition of Quality of Service
- Users' requirements
- Quantitative definition of QoS
- Which QoS service
- Premium IP model
- Proposed implementation scheme
- Service Level Agreement and Specification

Qualitative definition of Quality of Service using the IP protocol



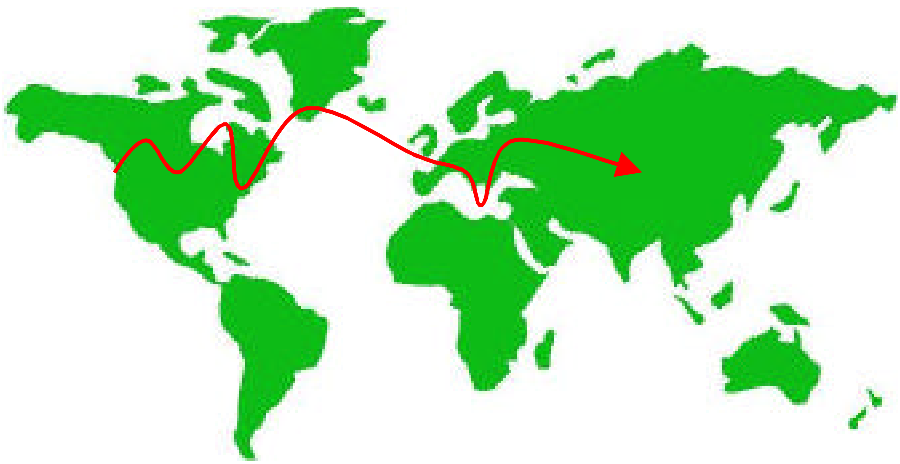
Basic components of QoS



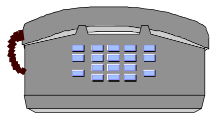
Intuitive definition of QoS

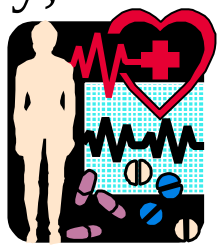
The network offers a QoS service when it's capable of handling selected flows in such a way to fulfill their requirements.

The QoS service must be present at all hops of the path to provide an end to end guarantee.



Why QoS

- Applications sensitive to delay variation or packet losses. 
- Real time control from remote sites (astronomy, medicine) (synchronous, no packet losses)



- Guaranteed capacity.
- Priority channel for monitoring, control signaling of the network itself.
- Bandwidth today (20 Sep 2001) is not infinite.
- ATM is going away (no CBR layer 2 circuits)

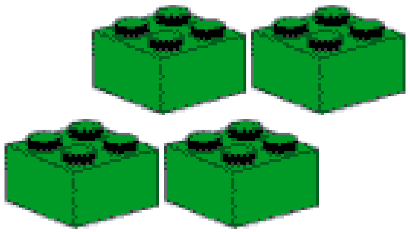
Agenda

✓ Qualitative definition of Quality of Service
✓ Users' requirements
- Quantitative definition of QoS
- Which QoS service
- Premium IP model
- Proposed implementation scheme
- Service Level Agreement and Specification

QoS parameters

From users' requirements and technical considerations :

✓ - one-way delay;
✓ - IP packet delay variation;
✓ - capacity;
✓ - one-way packet loss.

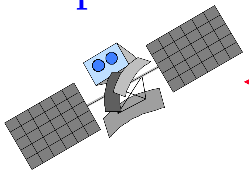


The set is common to IETF and ITU-T.
Naming and definitions are chosen to be comply to RFC 2330 (Framework for IP Performance metrics) and follow the ongoing IPPM IETF working group work.

QoS parameters sample value ranges

	Single value (SV)	Short range (class 0)	Medium (class 1 interactive)	Wide range (class 2 non-interactive)
One-way Delay	Measured value at empty network (baseline)	less than SV + 50 ms (150 ms)	less than SV + 250 ms (400 ms)	less than SV + 10 s (1 s)
ipdv	Between 0 and the time needed to transmit one full MTU at line speed	25 ms (50 ms)	50 ms (50 ms)	none (1 s)
Packet loss (Probability)	null	< 10^-4 (10^-3)	< 10^-3 (10^-3)	< 0.1 (10^-3)
Bandwidth (speed 64Kb/s)	Fixed value, greater than time to transmit one full MTU packet	N/A	N/A	a minimum of one full MTU size packet per second

Between parenthesis are ITU-T Y.1541 draft values, Class 3 (unspecified) is not shown



QoS parameters (continued)

Memento

To build a QoS service based on the previous listed parameters, some basic requirements on the network should be fulfilled:

- physical and data link stability
- exhibit a Bit Error Rate of at least 10^{-12}
- overall network hardware performance;

The minimum MTU size should be chosen large enough to avoid fragmentation.

Duplicate and out-of-order packets at the physiological level (which is not null, but very small)

Examples

	FTP	HTTP	Voce su IP	Video
Capacity	minimal	minimal	minimal	large
Latency	N/A	minimal	medium	medium
Latency variation	N/A	N/A	small	small
Packet loss	medium-high	medium	low	low

The appropriate ranges for each QoS parameter is function of the specific application, for example interactive video and video streaming have quite different requests.

Agenda

✓ Qualitative definition of Quality of Service
✓ Users' requirements
✓ Quantitative definition of QoS
- Which QoS service
- Premium IP model
- Proposed implementation scheme
- Service Level Agreement and Specification

From Questionnaire

QoS service	One-way-delay	ipdv	packet loss	bandwidth
Best effort	wide	wide	medium	wide
Very good (*Premium IP*)	medium	very small	very small	according to SLA
Prioritised Bandwidth (*IP+*)	medium	medium	medium	according to SLA
Guaranteed bandwidth	medium	medium	very small	single value

	One-way-delay	IPDV	Packet loss	bandwidth
Best effort	Unspecified	Unspecified	< 5%	Unspecified
Premium IP	distance delay + 50 ms	< 25 ms	negligible	according to SLA
IP+	distance delay +100 ms	<25-50 ms	< 2%	according to SLA

Which QoS service

Start with the "very good" service and call it "Premium IP":

- it satisfies all the users' requests
- it is "the best" service possible
- it maps to very high priority scheduling techniques available now
- it is similar to a "virtual wire"

Agenda

- ✓ Qualitative definition of Quality of Service
- ✓ Users' requirements
- ✓ Quantitative definition of QoS
- ✓ Which QoS service
- - Premium IP model
- - Proposed implementation scheme
- - Service Level Agreement and Specification

Which QoS framework to use ?

✓ • Differentiated Services - RFC2475 -

• Integrated Services - RFC 1633 -

✓ • Overprovisioning

QoS models

	Initial/static engineering	Signaling for flow aggregates	Signaling for each flow
Minimal traffic handling	Overprovisioning		
Handling of flow aggregates	Diffserv 802.1p	RSVP for aggregates Diffserv / 802.1p	RSVP Diffserv / 802.1p
Handling per flow			RSVP Intserv

Increasing complexity (↓)

Increasing complexity (→)

Differentiated Services

Diffserv achieves scalability by aggregating traffic classification state which is conveyed by means of IP-layer packet marking using the Diffserv Code Point (DSCP) field which replaces the Type of service byte in the IP header (no bit added)

Sophisticated classification, marking, policing, and shaping operations need only be implemented at network boundaries of a Diffserv domain or hosts.

Packets are classified and marked to receive a particular per-hop-forwarding behaviour (PHB)

Diffserv considers uni-directional flows

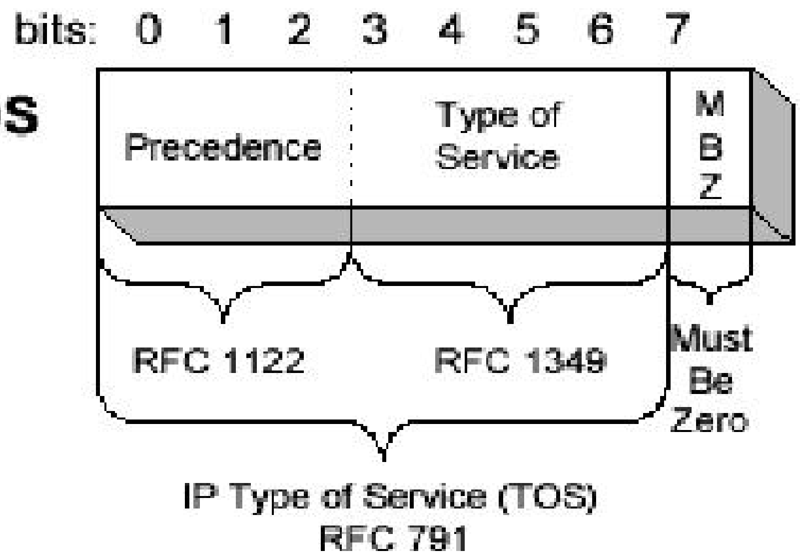
Differentiated Services (continued)

Two basic services (PHB) are defined:

- Expedited Forwarding (EF): to provide a building block for low loss, low delay, and low delay variation services

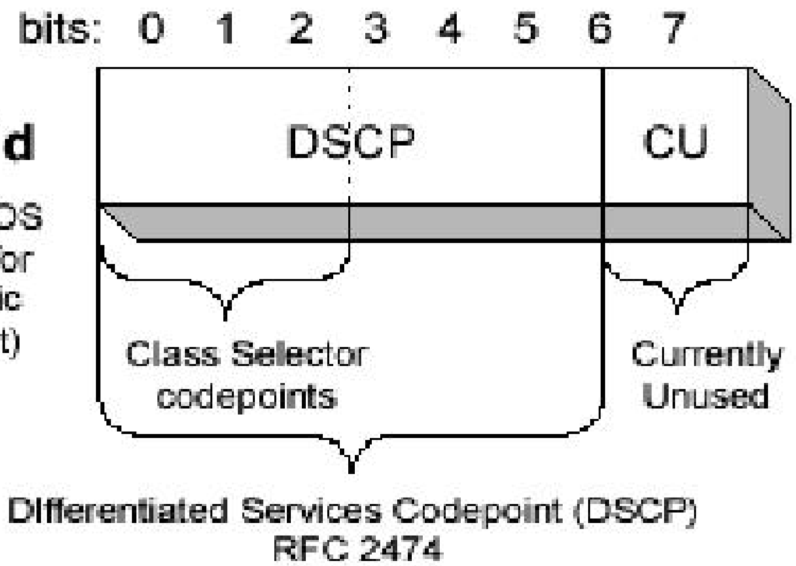
- Assured Forwarding (AF) : to provide a service with different discard probabilities among the same aggregation of flows

IPv4 TOS byte



DS-Field

(for **IPv4**, TOS octet, and for **IPv6**, Traffic Class octet)



IP v4

Version == 4	4 bits IHL	8 bits Type of Service	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

Premium IP Specification

- ⇒ Differentiated Services Architecture and use the expedited forwarding per hop behavior (EF PHB)
- ⇒ interface definition between domains that behaves as an EF PHB
- ⇒ do not starve best effort traffic (limited percentage of link capacity devoted to Premium IP)
- ⇒ initial provisioning structure: static, no dynamic signaling
- ⇒ IETF IPPM QoS parameters measurement framework
- ⇒ QoS parameters monitoring system is a key element

Agenda

- ✓ Qualitative definition of Quality of Service
- ✓ Users' requirements
- ✓ Quantitative definition of QoS
- ✓ Which QoS service
- ✓ Premium IP model
- Proposed implementation scheme
- Service Level Agreement and Specification

Premium IP
Service Implementation

Basic principles

minimize number of action per node
do not use a signaling protocol
modular approach that allows different implementation schemes at every hop or domain and allows domain to join the service when ready
Do not try to solve the most general problem, but rather develop a model that can be implemented in short time using available tools

Simplifying the actions for each node

In principle, each node might perform an awful lot of tasks:

- admission control and classification

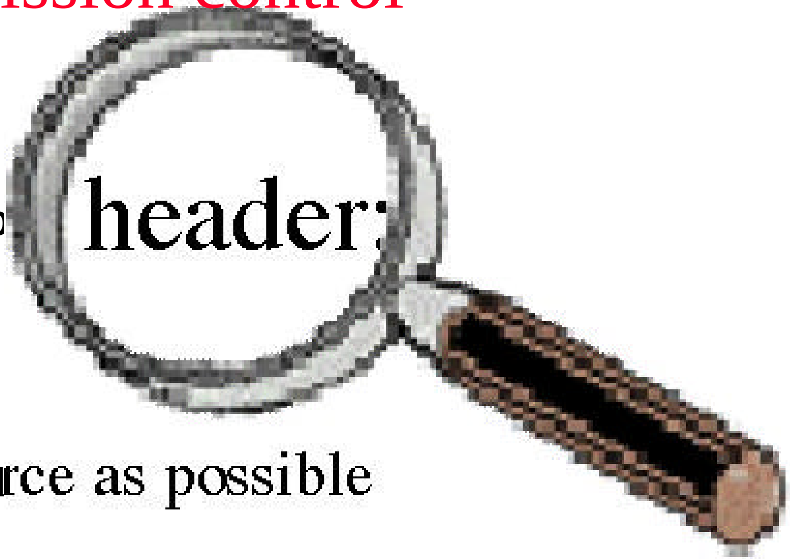
- marking
- policing
- scheduling

- shaping
- congestion control

- QoS rules propagation
- monitoring and accounting

Admission control

Use the information in the IP header:



- IP source *and* destination
  (prefixes) as near to the source as possible

- the DSCP (or IP precedence equivalent value) along the path

- perform an optional, suggested, admission control based on AS
  source and destination at inter-domain links (safety measure)

- rules might be based on additional parameters, as time-of-day

Admission control (continued)

The consequences are:

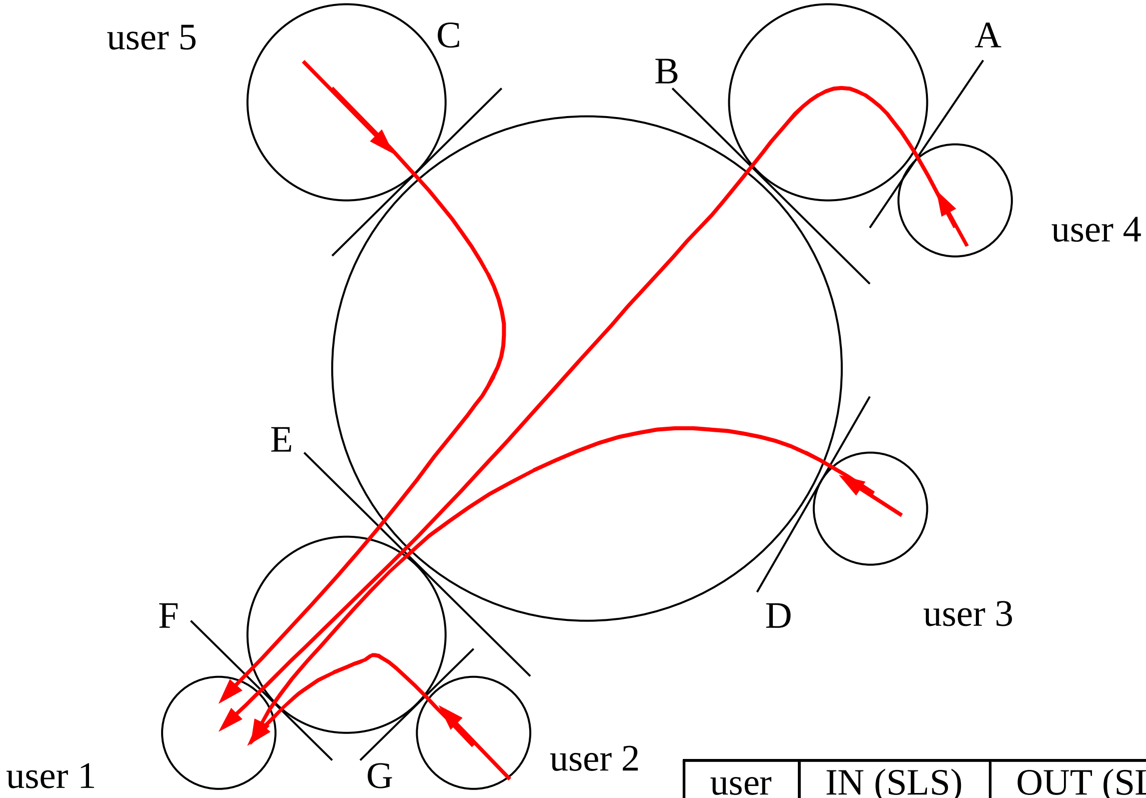
- allowing the computation of total requested Premium IP capacity at each network node in the default case (and for main backup cases too)

- short access list near users’ premise (few users)

- simple control at backbones (IP addresses are not propagated)

- choosing destination aware service (next slide)

Admission control (continued)



Destination Aware
Vs destination Unaware

user	IN (SLS)	OUT (SLS)
1	10	10
2	10	10
3	10	10
4	10	10
5	10	10

Examining the tasks for each node

In principle, each node might perform an awful lot of tasks:

- admission control and classification always

- marking

- policing

- scheduling

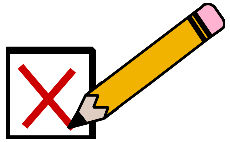
- shaping

- congestion control

- QoS rules propagation

- monitoring and accounting

Marking



- Mark each “EF” legal packet at first classification point
- Use the same DSCP value on all domains (Class selector 5 - decimal 40 [RFC 2474] to have interoperability with ToS-only capable hardware) - strongly suggested -
- valid DSCP coupled to invalid IP addresses implies discard to allow easy debugging
- packets with other DSCP values are left untouched

Marking is mandatory at the first classification point, remarking is optional.

Examining the tasks for each node

- admission control and classification — always

- marking — Selected locations

- policing

- scheduling

- shaping

- congestion control

- QoS rules propagation

- monitoring and accounting

Policing

Microflow policing should be done as close as possible to the source according to agreed (through SLA) Premium IP capacity. This step is mandatory

Policing will be done using a token bucket. The depth of the token bucket will be two MTU close to the source and increase to 5 or more along the path if additional policing is required

It is suggested to perform only one additional policing stage at the ingress to GÉANT from an NREN, with a larger aggregated capacity value than the sum of the agreements.

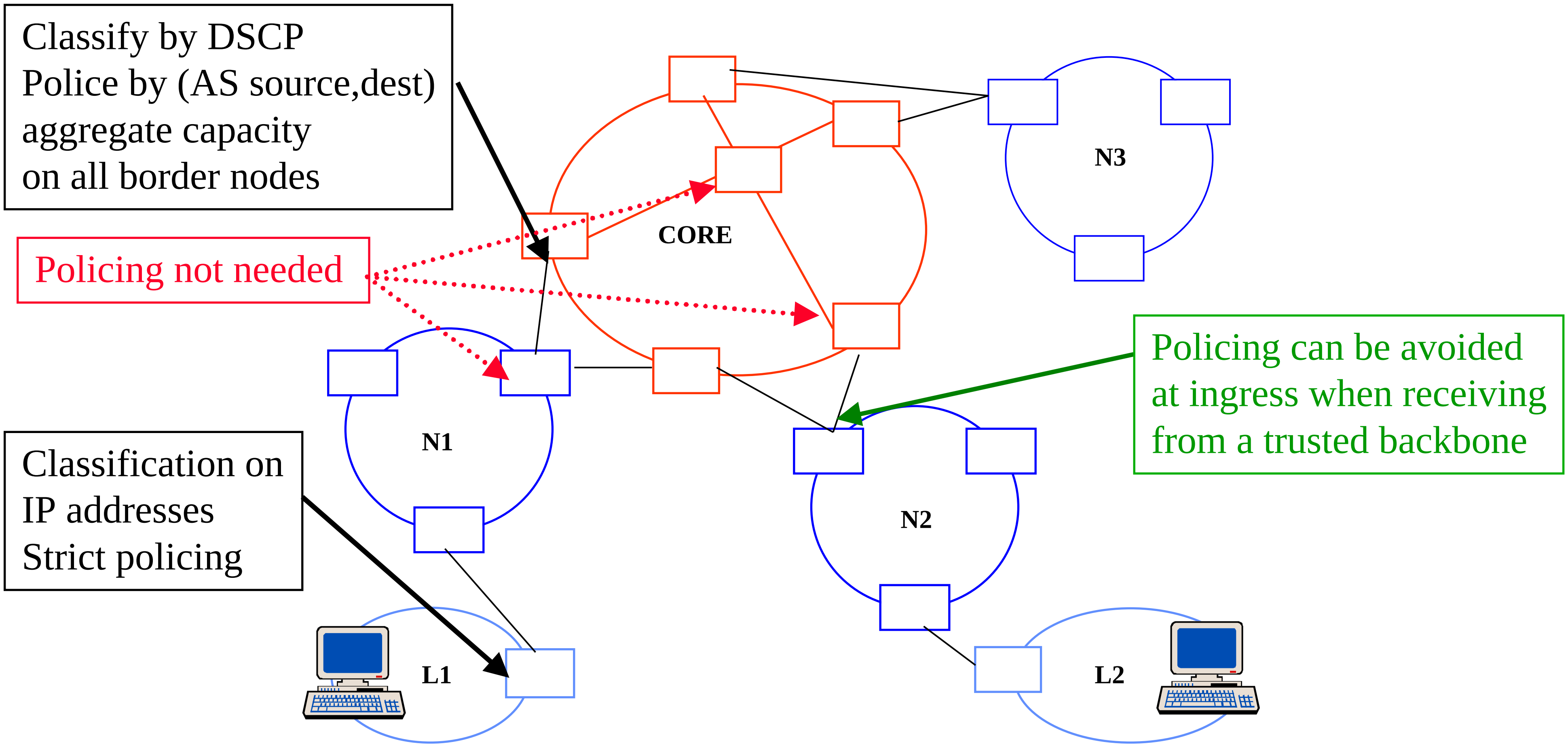
“Avoid unwanted packet loss” is the motto.

Policing (continued)

The additional policing stage at the ingress to GÉANT from an NREN serves the purpose of protecting Premium IP traffic from misconfiguration/DoS coming from a single source.

It creates virtual “pipes” for the aggregated Premium flows from each NREN to each other (when needed). The failure of one “pipe” does not influence the others.

Sample multidomain network



L1, L2	: end user domain (for example LANs)
N1, N2, N3	: intermediate transport domains (for example NRENs backbones)
CORE	: interconnection domain (for example GÉANT)
▭	: router/switch

Examining the tasks for each node

- admission control and classification — always

- marking — Selected locations
- policing — Selected locations
- scheduling — always

- shaping
- congestion control

- QoS rules propagation
- monitoring and accounting

Shaping

The compliance of the Premium user flow to the contracted capacity is the key for the result of the service.

Shaping is intended here as limiting the rate of transmission to a specific value.

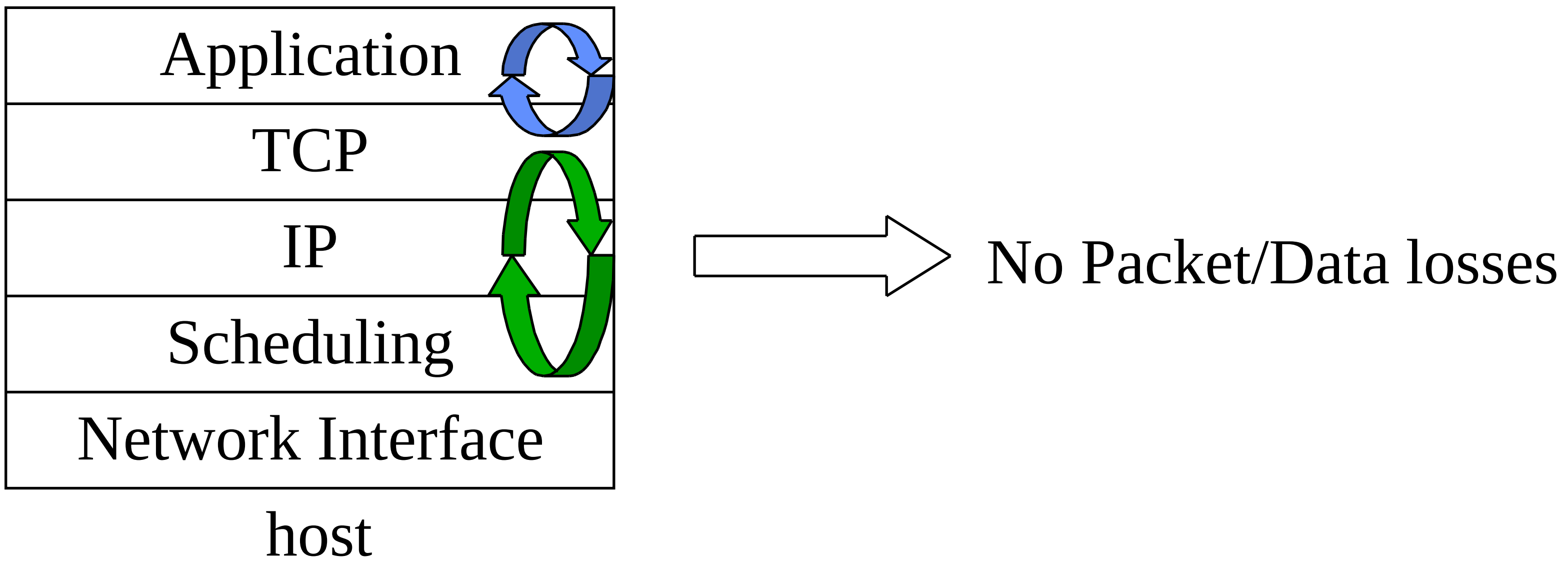
The speed of the core link and the highest priority in scheduling for the packets of the IP Premium service make delay variation small even at aggregation points.

At 2.5 Gb/s the transmission time of a 1500 bytes packet is about 5 microseconds. The consideration suggests to start the service without enabling shaping in the core and it shaping may be optional also at the border, provided the sources produce a well shaped flow.

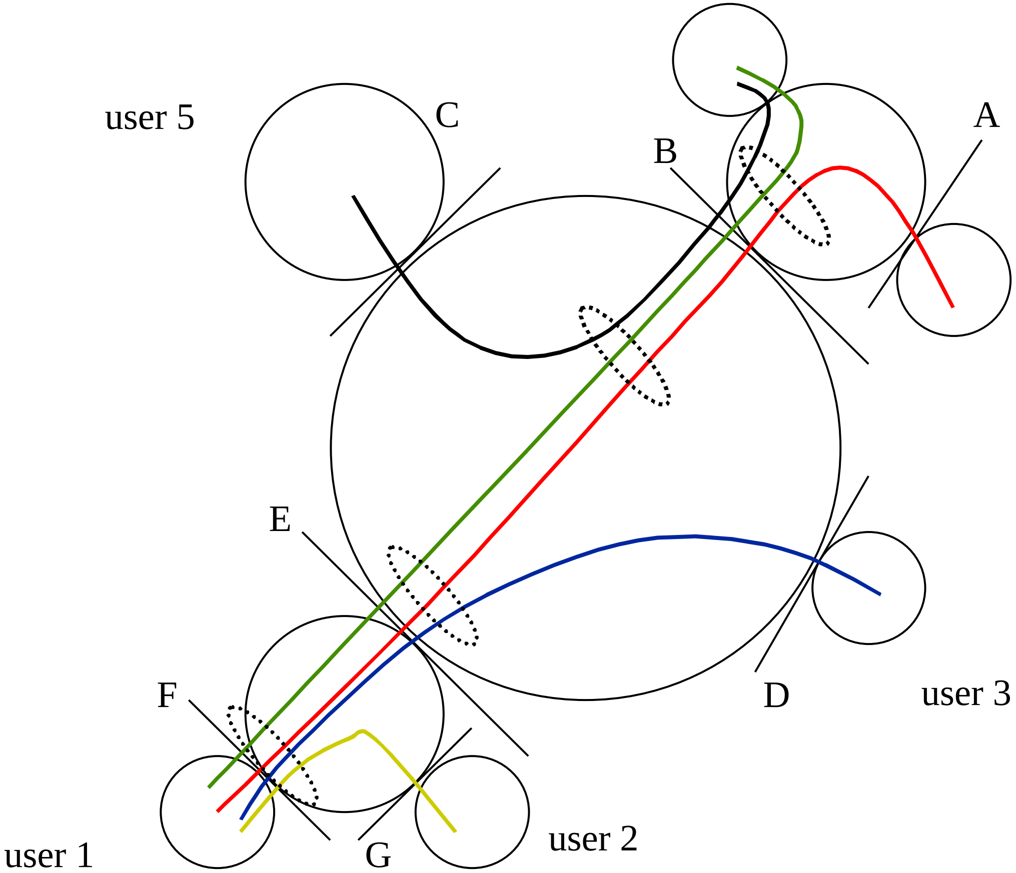
Shaping

The sending source is hence required to shape the traffic it produces.

Shaping inside the sending host itself is the preferred way, shaping by the network will in most case lead to packet losses



Shaping



Multiple aggregation-separation points and link speed changes.

Examining the tasks for each node

- admission control and classification — always

- marking — Selected locations
- policing — Selected locations
- scheduling — always

- shaping — NO Done by source
- congestion control — not needed

- QoS rules propagation — Selected locations
- monitoring and accounting — Selected locations

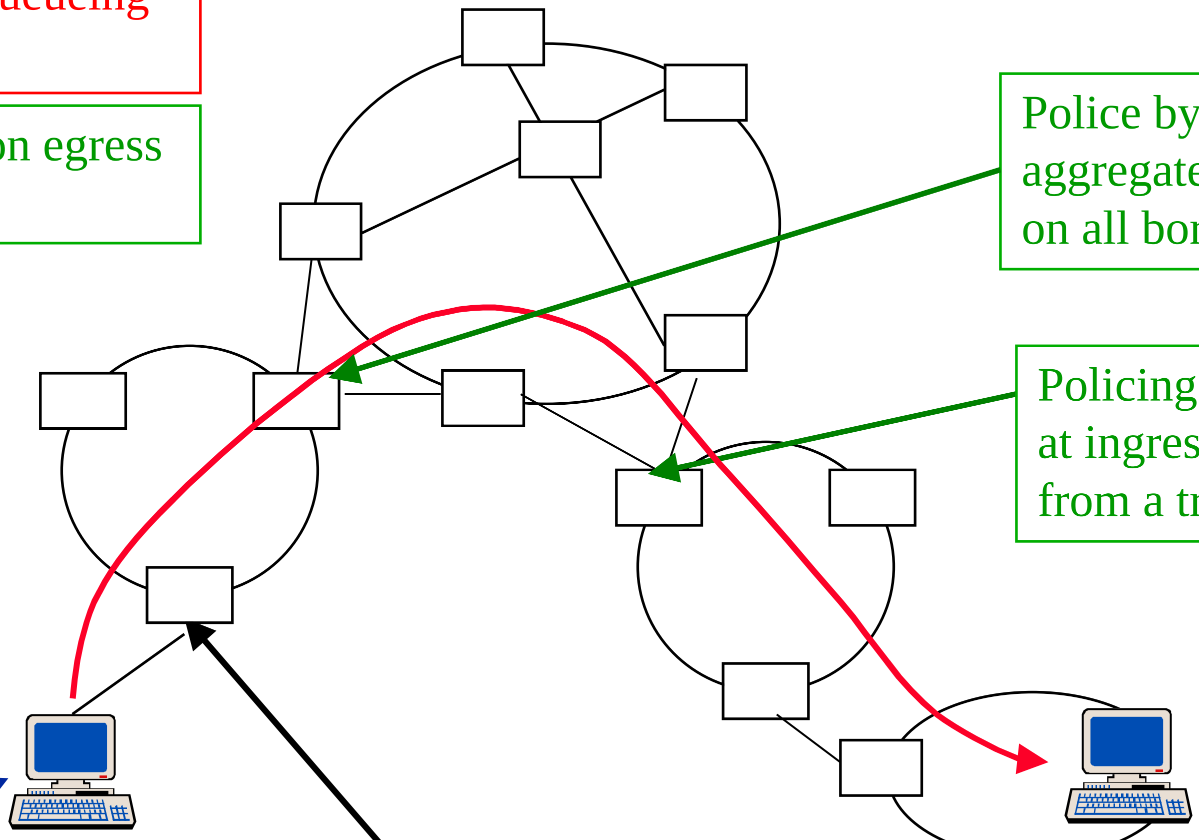
Summary

Classify (DSCP)
High priority queueing
on all nodes

Do not police on egress
Do not shape

Police by (AS source,dest)
aggregate capacity
on all border nodes

Policing can be avoided
at ingress when receiving
from a trusted backbone



Shape ONLY here

Classify (IP pair prefixes)
Police - Strict, Capacity
Mark

Grey Areas

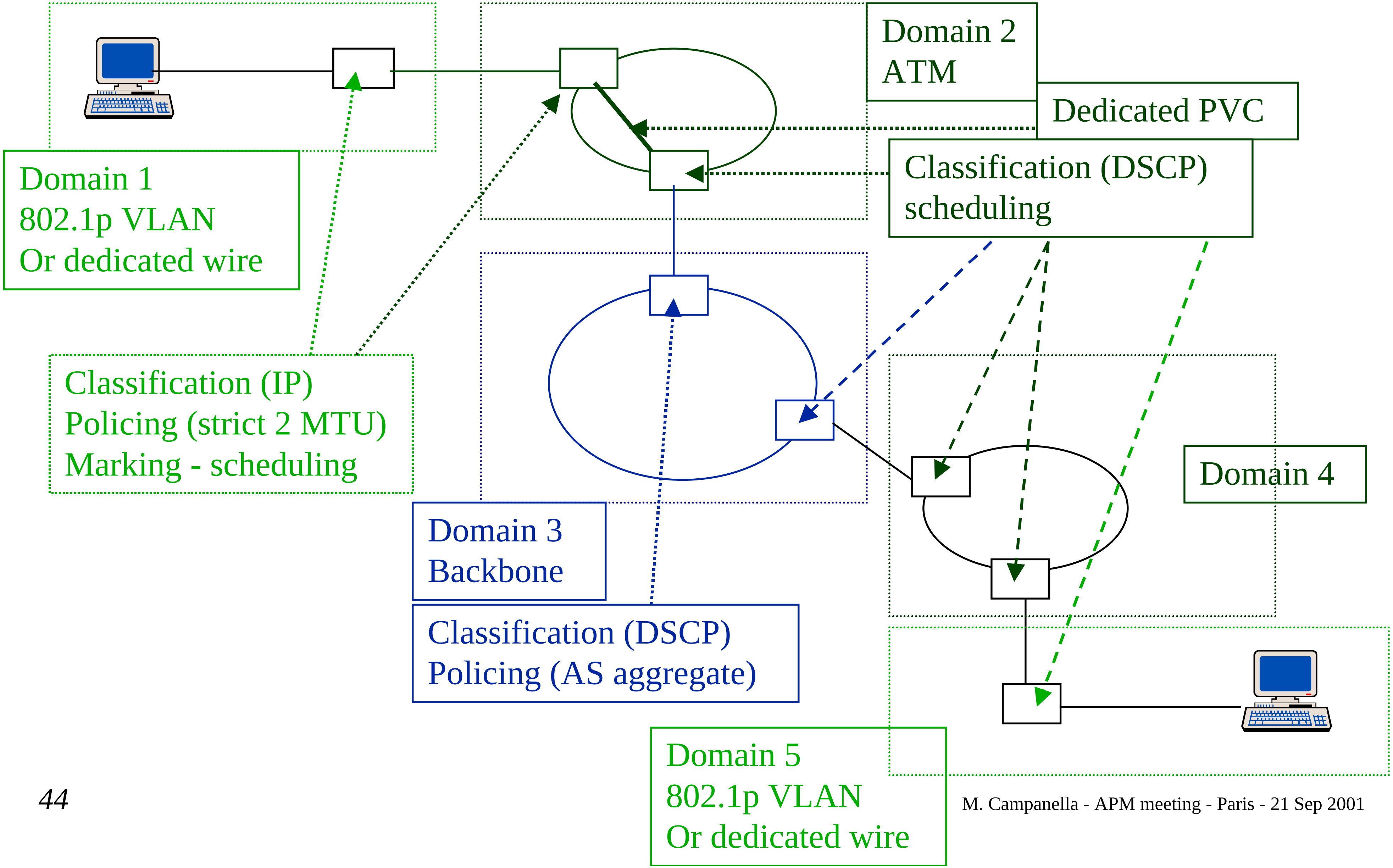
Exact configuration of buffering and token bucket depth in routers. As a rule of thumb the token bucket depth can be assumed to be 1.2 * (Diffserv active interfaces on router)

Scalability
- the maximum amount of aggregated Premium IP capacity the network can offer
- hardware capabilities

Fast provisioning of the service

Widespread availability and tuning of “last mile” (LANs)

Example (one direction)



Final remarks

The network can be set-up from the beginning to provide QoS.

Provisioning is then just a matter of adding appropriate ACLs near users' sources after defining and signing SLAs.

References

Sequin Deliverable D2.1 "Quality of Service definition"
http://www.dante.net/tf-ngn/SEQ-D2.1.pdf

GÉANT Deliverable D9.1 "Specification and implementation plan for a Premium IP service":
http://www.dante.net/tf-ngn/GEA-01-032.pdf

GÉANT Deliverable D9.1 - Addendum 1 "Implementation architecture specification for the Premium IP service":
http://www.dante.net/tf-ngn/D9.1-addendumv2.pdf

GÉANT Deliverable D9.1 - Addendum 2 "Service Level Agreement specification for Premium IP service":
to be available soon in http://www.dante.net/tf-ngn/

TF-NGN public documents: http://www.dante.net/tf-ngn/

Agenda

- ✓ Qualitative definition of Quality of Service
- ✓ Users' requirements
- ✓ Quantitative definition of QoS
- ✓ Which QoS service
- ✓ Premium IP model
- ✓ Proposed implementation scheme
- \- Service Level Agreement and Specification

SLA/SLS
Fundamental concepts

In the first phase the SLS negotiation will be performed manually (no bandwidth broker).

The analytical computation of the QoS metric in a IP based network is extremely complex.
The SLA specification requires extensive testing of the available infrastructure. Usually only upper bounds of QoS parameter ranges can be specified.

There are always two SLA, one for each direction. The contracted values might be different (asymmetric capacity for example)

Fundamental concepts (continued)

Users must understand and know the application requirements in term of QoS parameters.

There is the need of a central database to keep up to date track of allocate resources and check their availability.

The SLA/SLS is in reality a chain of SLA/SLS between neighbour domains and a final end-to-end one.

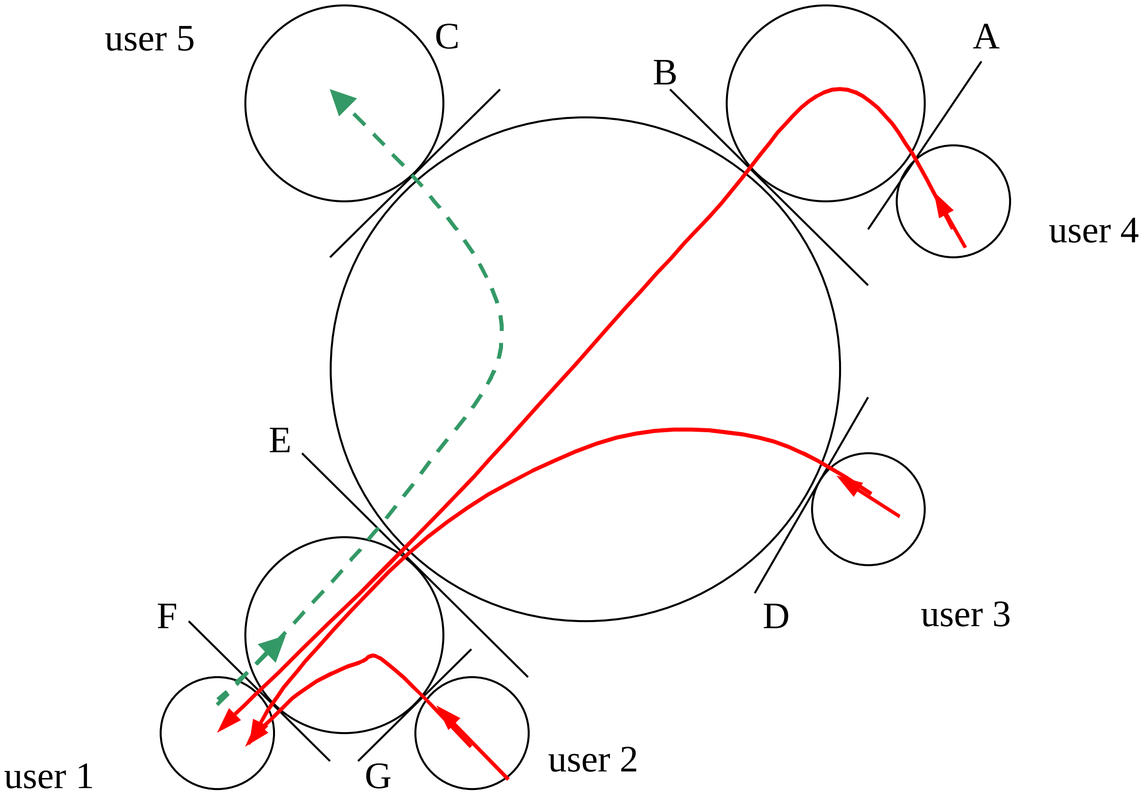
Fundamental concepts (continued)

The IP Premium service is aimed at providing end to end QoS. To fulfil this goal the establishment of a particular service instance must be made known to all domains involved.

The service must be defined both as an end to end service level agreement and be accepted as a modification in the chain of service level agreements between all involved domains. For example the capacity requested between user 1 and user 4 will be seen by involved domains as an increase of the premium capacity agreed between them.

Debugging can be assigned to a single specific entity.

Local Vs Global agreements



Suppose user 1 wants to speak IP Premium with user 5 only. Users 2, 3, 4 wants to speak with 1

If the destination address is known, then it is possible to dimension boundary F, but user 1 will send and receive more IP Premium traffic then he expects.

The SLA should be propagated end to end

Thank you !