

CESNET Technical Report 17/2008

QoS Design and Implementation in the CESNET2 MPLS-based Backbone

PAVEL ŠMRHA, JOSEF VERICH

Received 4.12.2008

Abstract

The paper deals with the design and implementation of QoS in the CESNET2 MPLS backbone composed of Cisco 7600 and Cisco CRS-1 routers. The technical approach is to implement DiffServ within the MPLS-based CESNET2 backbone using the E-LSP point-to-cloud (destination unaware) QoS model for 6 classes of traffic aggregates with different QoS characteristics. To ensure the DSCP transparency for transported IPv4/IPv6 user traffic the so called “Short-Pipe” MPLS DiffServ tunneling mode is deployed. The QoS implementation for IPv4/IPv6 traffic currently uses 6 classes of service: *Real-Time* (EF-PHB), *Network Control* (AF-PHB), *Video* (AF-PHB), *Critical Traffic* (AF-PHB), *Best Effort* (default PHB), and *Less than Best Effort* (AF-PHB).

Keywords: Quality of Service, DiffServ, MPLS

Keywords: QoS Design and Implementation in the CESNET2 MPLS-based Backbone

1 Introduction

The basic goal of Quality of Service (QoS) is to provide service differentiation among IP packets in the network. Such a service differentiation is noticeable during periods of network congestion (i.e. in case of contention for resources) and results in different levels of network performance.

The initial phase of QoS implementation in the CESNET2 network was completed in 2005 [15]. Current document describes the second network QoS deployment phase in CESNET2 MPLS backbone, the Czech National Research and Educational Network (NREN), It covers mainly service class restructuring, IPv6 support and unification of QoS configuration for new hardware platforms or modules (e.g. CRS-1 and Cisco 7600 ES20 modules). Such an implementation of preferential service for a class of traffic not only requires technical support to achieve appropriate network behavior, but also raises other issues related to establishing an acceptable overall QoS policy framework [8].

The main goals of this network QoS stage of implementation of in the CESNET2 MPLS backbone are the following ones:

- To define a simple yet consistent QoS policy framework for CESNET2 MPLS backbone, which is also easy to implement, provision, and manage.
- To provide the necessary QoS support for CESNET2 IPv4/IPv6 transit traffic to enable end-to-end QoS-based connectivity for users within different network domains (e.g. university MANs, regional networks etc.).

- To ensure DSCP transparency for all SLS-conformant IPv4/IPv6 user traffic while transiting the CESNET2 MPLS backbone (i.e. the original value of the DSCP field within the customer in-profile IPv4/IPv6 packets is not modified as they are transported through the CESNET2 MPLS backbone).
- To ensure *compliance* with the GÉANT2 Premium IP class of service.
- To be compatible with other GÉANT2 interdomain classes of service (e.g. Best Effort and Less than Best Effort), while providing new additional useful classes of service not (yet) implemented in GÉANT2.
- To design as unified QoS configuration as possible across all network devices in the CESNET2 MPLS backbone (Cisco 7600 OSR/Sup720/PFC3BXL/CXL and Cisco CRS-1).

The technical approach is to implement DiffServ within the CESNET2 MPLS-based backbone using the E-LSP point-to-cloud (destination unaware) QoS model for a few classes of traffic aggregates with different QoS characteristics. To ensure the DSCP transparency for transported IPv4/IPv6 user traffic the so called “Short-Pipe” MPLS DiffServ tunneling mode has been deployed¹.

In every case (with/without congestion) there should be guarantees for all traffic classes for their (provisioned) bandwidth as well as other qualitative characteristics of their class (e.g. delay, jitter, packet loss). In case of no congestion it is required that all classes² can use the otherwise unused available bandwidth in proportion to their weight (even if they utilize more bandwidth than guaranteed for their class by provisioning).

2 Quality of Service on GÉANT2

This chapter summarizes current implementation of QoS on GÉANT2 using information available from the DANTE/GÉANT2 web site [13].

Currently, three service classes are offered that can be used for traffic transiting the GÉANT2 network:

- The Premium IP service provides guarantees on one-way delay, jitter and packet loss. Applications with real time constraints may benefit from this service; for instance, video conferencing.
- The Best Effort service does not provide any performance guarantees. However, the goal is to give a fair share to each traffic flow. Best effort is the default class of service and is useful for traffic from non-real-time applications such as e.g. FTP.
- The Less than Best Effort (*scavenger*) service utilizes the remaining capacity that is not used by best effort and premium IP traffic. This service may be useful to transfer vast amounts of (non real-time) data for applications such as GRID computing without adversely affecting other traffic.

¹ Due to the possibility of penultimate hop popping (PHP) in case of not using MPLS/BGP VPN technology (when only one MPLS label is present).

² With the possible exception of specific hardware or configuration limitations (e.g. various Cisco IOS “priority” classes implementations) or other requirements (e.g. for security purposes).

Both Best Effort (BE) and Less than Best Effort (LBE) services are available to peer networks. The Premium IP (PIP) service, however, is provided only upon a reservation.

Table 1. DSCP to service class mapping on GÉANT2.

Service Class Name	DSCP
Premium IP (PIP)	EF
Best Effort (BE)	0
Less than Best Effort (LBE)	CS8

3 Quality of Service on CESNET2

3.1 DiffServ Marking Schemes

Currently there are at least two consistent DiffServ marking schemes available for academic user traffic: the so called *Cisco QoS Baseline* [9], which was adopted by some universities in the past, and RFC 4594 *Configuration Guidelines for DiffServ Service Classes* [2].

3.1.1 Cisco QoS Baseline

While the IETF RFC standards provided a consistent set of per-hop behaviors for applications marked to specific DSCP values, they had never specified which application should be marked to which DSCP value until 2006. This confusion led Cisco to put forward a standards-based marking recommendation in their strategic architectural *QoS Baseline* document [9] in 2002. Eleven different application classes that could exist within the organization were examined and extensively profiled, and then matched to their optimal RFC-defined Per-Hop Behaviors (PHBs). The QoS Baseline is a strategic document designed to unify QoS within Cisco. It provides uniform, standards-based recommendations to help ensure that QoS products, designs, and deployments are unified and consistent. The QoS Baseline defines up to 11 classes of traffic that may be viewed as critical to a given organization. A summary of these classes and their respective standards-based markings and recommended QoS configurations are shown in Table 2:

- The **IPRouting** class is intended for IP routing protocols, such as Border Gateway Protocol (BGP), Open Shortest Path First (OSPF).
- **Voice** refers to VoIP bearer traffic only (and does not include call-signaling traffic).
- **Interactive-Video** refers to IP video-conferencing.
- **Streaming Video** is aimed for either unicast or multicast uni-directional video.
- The (Locally-Defined) **Mission-Critical Data** class is intended for a subset of *Transactional Data* applications that contribute most significantly to the business objectives (this is a nontechnical assessment).
- The **Call-Signaling** class is intended for voice and/or video signaling traffic, such as Skinny, SIP, H.323.

- The **Transactional Data** class is intended for foreground, user-interactive applications such as database access, transaction services, interactive messaging, and preferred data services.
- The **Network Management** class is intended for network management protocols, such as SNMP, Syslog, DNS.
- The **Bulk Data** class is intended for background, non-interactive traffic flows, such as large file transfers, content distribution, database synchronization, email and backup operations.
- The **Best Effort** class is the default class. Unless an application has been assigned a preferential/deferential service, it remains in this class. Most organizations have hundreds of applications on their networks; the majority of which will fall into the Best Effort service class.
- The **Scavenger** class is based on RFC 3662 and an Internet 2 draft that defines a “Less-than-Best Effort” service. In the event of link congestion, this class will be dropped the most aggressively.

Table 2. Summary of QoS Mechanisms Used for Each Cisco Baseline Service Class.

Service class	DSCP	Conditioning	PHB used	Queuing	AQM
routing	CS6	sr+bs	RFC 2474	rate	RED
voice	EF	sr+bs	RFC 3246	priority	No
interactive video	AF4x	trTCM	RFC 2597	rate	WRED/DSCP
streaming video	CS4	sr+bs	RFC 2474	rate	RED
mission-critical data	AF3x	srTCM	RFC 2597	rate	WRED/DSCP
call signaling	CS3	sr+bs	RFC 2474	rate	RED
transactional data	AF2x	srTCM	RFC 2597	rate	WRED/DSCP
network management	CS2	sr+bs	RFC 2474	rate	RED
bulk data	AF1x	trTCM	RFC 2597	rate	WRED/DSCP
best effort	CS0	-	RFC 2474	rate	RED
scavenger	CS1	-	RFC 3662	rate	RED

Standard-based marking recommendations allow for better integration with service-provider offerings as well as other internetworking scenarios. In Cisco IOS Software, *rate-based* queuing translates to CBWFQ; *priority* queuing is LLQ. DSCP-Based WRED (based on RFC 2597) drops AFx3 before AFx2, and in turn drops AFx2 before AFx1. RSVP is recommended (whenever supported) for *Voice* and/or *Interactive-Video* admission control in Cisco products that support QoS features and will use these *QoS Baseline* recommendations for marking, scheduling, and admission control.

The *QoS Baseline* recommendations are intended as a standards-based guideline for users - not as a mandate. Users do not have to deploy all 11 traffic classes, but may start with simple QoS models and expand over time as business needs arise.

3.1.2 RFC 4594: Configuration Guidelines for DiffServ Service Classes

Informational RFC 4594 describes service classes configured with DiffServ (DS) and recommends how they to use them and how to construct them using Differentiated Services Code Points (DSCPs), traffic conditioners, Per-Hop Behaviors (PHBs), and Active Queue Management (AQM) mechanisms. There is no intrinsic

requirement that particular DSCPs, traffic conditioners, PHBs, and AQM be used for a certain service class, but as a policy and for interoperability it is useful to apply them consistently.

Table 3 defines the recommended relationship between service classes and DSCP assignment with application examples. It is recommended that this relationship be preserved end to end.

Table 3. DSCP to service class mapping [2].

Service class	DSCP	Application example
network control	CS6	network routing: BGP, OSPF, ISIS, RIP, CR-LDP, RSVP-TE
telephony	EF	IP telephony bearer: VoIP, CEdIP, virtual wire
signaling	CS5	IP telephony signaling: SIP, H.323, H.248, MGCP, SIP-T
multimedia conferencing	AF4x	H.323/V2 video conferencing (adaptive): H.323/v2
real-time interactive	CS4	video conferencing and interactive gaming
multimedia streaming	AF3x	streaming video and audio on demand
broadcast video	CS3	broadcast TV & live events
low-latency data	AF2x	client/server transactions, Web-based and ERP: SNA/DLSw, SAP
OAM	CS2	OAM&P: SNMP, TFTP, FTP, Telnet, SSH
high-throughput data	AF1x	store and forward applications: FTP, email
Standard	CS0	Undifferentiated applications: DNS, DHCP/Bootp, NTP
low-priority data	CS1	any flow that has no bandwidth assurance (Scavenger)

Table 4 provides a summary of DiffServ QoS mechanisms that should be used for the defined service classes. According to what applications/services need to be differentiated, network administrators can choose the service class(es) that need to be supported in their network.

3.1.3 Recommended DSCP Marking Scheme

When comparing both recommendations, *Cisco Baseline* and *RFC 4594*, we can find certain incompatibilities, both in the number and types of the classes proposed and in their marking, as Table 5 shows.

The major changes include one application class missing, two marking changes, and two new application class additions [7]:

- The *QoS Baseline* locally-defined *Mission-Critical Data* class has been deleted from RFC 4594.

Table 4. Summary of QoS Mechanisms Used for Each Service Class [2].

Service class	DSCP	Conditioning	PHB used	Queuing	AQM
network control	CS6	sr+bs	RFC 2474	rate	yes
telephony	EF	sr+bs	RFC 3246	priority	no
signaling	CS5	sr+bs	RFC 2474	rate	no
multimedia conferencing	AF4x	trTCM	RFC 2597	rate	yes, per DSCP
real-time interactive	CS4	sr+bs	RFC 2474	rate	no
multimedia streaming	AF3x	trTCM	RFC 2597	rate	yes, per DSCP
broadcast video	CS3	sr+bs	RFC 2474	rate	no
low-latency data	AF2x	srTCM	RFC 2597	rate	yes, per DSCP
OAM	CS2	sr+bs	RFC 2474	rate	yes
high-throughput data	AF1x	trTCM	RFC 2597	Rate	Yes per DSCP
standard	CS0	-	RFC 2474	rate	yes
low-priority data	CS1	-	RFC 3662	rate	yes

- The *QoS Baseline* marking recommendation of CS4 for *Streaming Video* has been changed in RFC 4594 to mark *Multimedia Streaming* to AF31.
- The *QoS Baseline* marking recommendation of CS3 for *Call Signaling* has been changed in RFC 4594 to mark *Call Signaling* to CS5.
- A new video class has been added to RFC 4594: *Real-Time Interactive*, which is to be marked CS4. This was done to differentiate between lower-grade desktop video telephony (referred to as *Multimedia Conferencing*) and higher-grade videoconferencing and telepresence.
- Another new video class has been added to RFC 4594: *Broadcast Video*, which is to be marked CS3. This was done to differentiate between lower-grade desktop video streaming (referred to as *Multimedia Streaming*) and higher-grade *Broadcast Video* applications. *Multimedia Streaming* uses the AF3 class and is subject to markdown policies, while *Broadcast Video* uses the CS3 class and is not subject to markdown.

Table 5. RFC 4594 vs. Cisco Baseline Service Class Comparison.

RFC 4954 service class	DSCP	Cisco baseline service class	DSCP
Network Control	CS6	Routing	CS6
Telephony	EF	Voice	EF
Signaling	CS5	Call Signaling	CS3
Multimedia Conferencing	AF4x	Interactive Video	AF4x
Real-Time Interactive	CS4	-	-
-	-	Mission-Critical Data	AF3x
Multimedia Streaming	AF3x	Streaming Video	CS4
Broadcast Video	CS3	-	-
Low-Latency Data	AF2x	Transactional Data	AF2x
OAM	CS2	Network Management	CS2
High-Throughput Data	AF1x	Bulk Data	AF1x
Standard	CS0	Best Effort	CS0
Low-Priority Data	CS1	Scavenger	CS1

Some incompatibilities of marking can be partially eliminated under certain

circumstances through the support of dual marking, e.g. for the aggregated signaling class, it is possible (under an absence of support of conflict *Broadcast Video* class marked also as CS3) to support simultaneously CS3 and CS5 within the transit DS domain according to the *Cisco Baseline* model because CS5 is not used for any basic class in that model. Unfortunately, it is then necessary to perform reclassification/remarking to the selected semantics at the border of DS domain for conflict marking in general (CS3, CS4 and AF3x). As the informational RFC 4594 recommendation offers finer granularity of support for multimedia applications and is rather oriented to general providers of wide scale of services, we have chosen the variant of *Cisco Baseline* model for the proposed QoS classes of DS domain of the CESNET2 network environment. On one hand it seems to correspond better to the requirements for QoS in current academic user traffic thanks to the concept of the usually required *Mission-Critical Data* service class, on the other hand it supports more consistent implementation of QoS in the environment of IP and MPLS networks implemented on Cisco Systems devices (which can be very important particularly because of some limitations in implementation of MPLS QoS in different hardware platforms of this vendor).

3.2 CESNET2 Service Classes

We propose six service classes for traffic transiting the CESNET2 MPLS-based backbone:

- The **Real-Time** (RT) service provides guarantees on bandwidth, one-way delay, jitter and packet loss. This is intended to be the same service as the one referred to in the context of the GÉANT2 *Premium IP* (PIP) service. Applications with real time constraints may benefit from this service; for instance, Voice over IP (VoIP), real-time data from remote sensors.
- The **Network Control** (NC) service provides guaranteed bandwidth for necessary network control traffic (e.g. routing updates) with relatively low delay and low packet loss. This service is intended to be considered rather as a CESNET2 intradomain service, but in special cases it can also be used as an interdomain working to ensure stability when required (e.g. to be more resilient to congestion when using dynamic routing protocols with neighboring DS domains of interconnected universities).
- The **Video** service is intended as a preferential service for inelastic applications requiring guaranteed bandwidth, (much) better delay performance and very low packet loss than the ones offered by BE but for which the tight constraints on delay and jitter offered by RT are not necessary. Some kind of videoconferencing, streaming delivery of non-interactive audio-video material are candidate applications. Traffic of this type is expected to form a relatively large proportion of total traffic.
- The **Critical Traffic** (CT) service is intended as a preferential service for elastic (data) applications requiring guaranteed bandwidth with better delay performance and lower packet loss than the ones offered by BE but for which the tight constraints on delay and jitter offered by RT are not necessary. Interactive and transactional data applications and various signaling protocols are

candidate applications. Traffic of this type is expected to form a relatively large proportion of total traffic.

- The **Best Effort** (BE) service does not provide any performance guarantees. However, the goal is to give a fair share to each traffic flow. This is intended to be the same service as the one referred to in the context of the GÉANT2 BE service. Best effort is the default class of service and is useful for traffic from non-real-time applications such as FTP or email.
- The **Less than Best Effort** (LBE) service will utilize the capacity that is not used by other traffic classes. This is intended to be the same service as the one referred to in the context of the GÉANT2 LBE service or as the *Scavenger Service* (QBSS) in QBone/Internet2/Abilene. This service may be useful to transfer vast amounts of (non real-time) data for applications such as GRID computing or backups without adversely affecting other traffic. This service can also be used to carry some excess traffic based on results of various policing instead of dropping it, just to give it a “last resort” chance to be transferred through the network if (and only if) the network resources are available without affecting other higher service classes.

3.3 CESNET2 MPLS-backbone QoS Policy Framework

3.3.1 Network Model

The policy framework defines Exp-LSP DiffServ technology with point-to-cloud destination-unaware model as a basis for QoS services to be deployed as part of the second phase of the CESNET2 QoS project. To ensure the DSCP transparency the so called “Short-Pipe” MPLS DiffServ tunneling mode will be deployed: the ingress PE router at the CESNET2 MPLS DiffServ domain border derives the EXP bits from the original DSCP of the incoming trusted IP packet, the egress PE router uses this original DSCP of the underlying IP packet (which remains unchanged while transiting the CESNET2 MPLS DiffServ domain) instead of (possibly changed) EXP marking used by intermediate P routers. Due to the possibility of penultimate hop popping (PHP) in case of not using MPLS/BGP VPN technology (when only one MPLS label is present) even the second-to-last router inside the CESNET2 MPLS DiffServ domain may use this original DSCP of the underlying IP packet for its forwarding decision.

The main feature of this model is that each router in the network processes packets in a differentiated manner: the differentiation being made on the basis of the value of DSCP field of a packet at ingress/egress network points and EXP field at intermediate network points, disregarding a destination address (path) of a packet. This feature results in a simplicity of deploying and managing QoS services but at the same time it has a negative side effect when a great amount of higher QoS classes’ flows go to the same output interface of the router, so that an interface becomes congested by these particular classes of traffic. The point-to-cloud model cannot prevent this situation completely, as routers and network administrators don’t check paths of flows and network users have the right to direct their flows wherever they like. To decrease the probability of this situation and instead

to increase the probability of providing a stable QoS service, the border routers of the CESNET2 MPLS DiffServ domain should police the higher level QoS classes to maintain them at relatively low levels conforming agreed traffic contracts (SLA).

CESNET2 compliant DSCP marking must be provided by the neighboring DiffServ domain at CESNET2 ingress to ensure an easy QoS implementation, provisioning and management within the CESNET2 DiffServ domain. A preventive admission control method to keep DiffServ model working properly is to limit higher level QoS traffic within each aggregated class of service according to the SLA by policing the exceeding/violating portion of such traffic (and typically remarking it to lower level QoS classes such as BE or LBE) at the ingress (or egress when required) of the CESNET2 MPLS DiffServ domain. As a result no tedious management of otherwise necessary access control lists of authorized sources for each particular QoS traffic class is required.

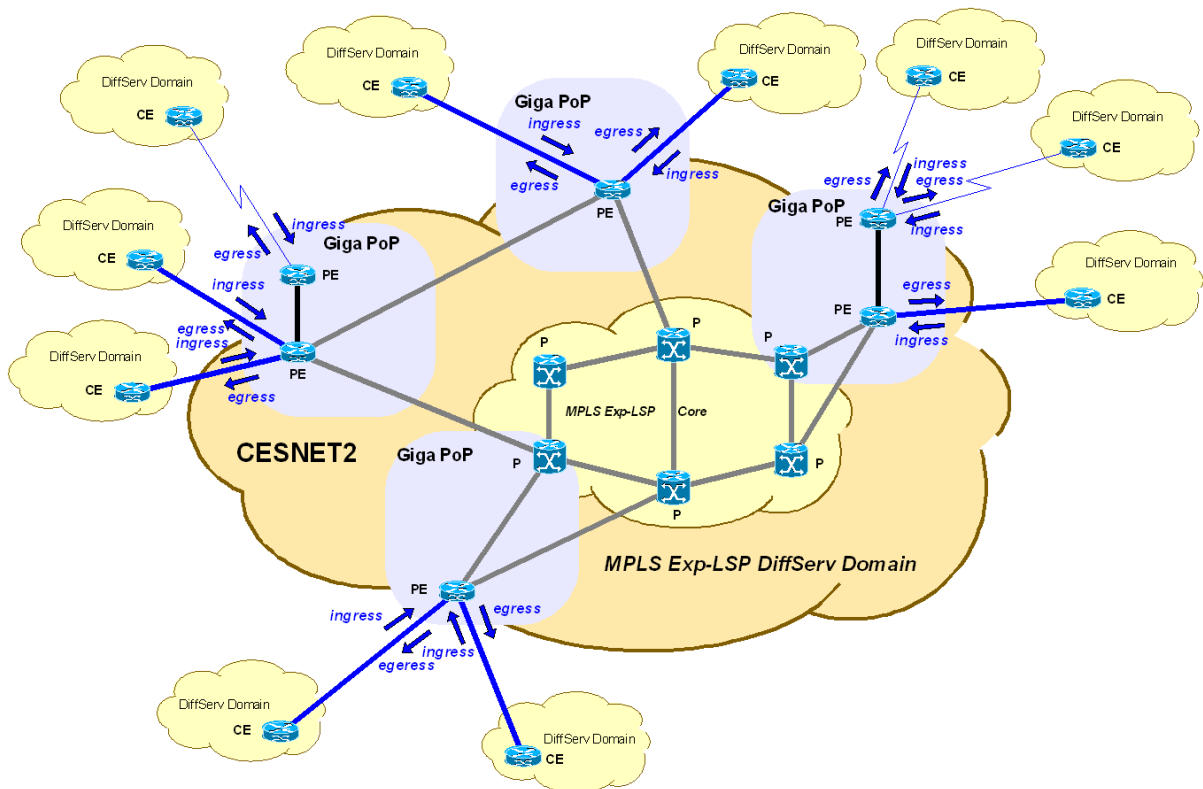


Figure 1. Point-to-cloud destination unaware MPLS DiffServ CESNET2 QoS model with aggregated ingress (or egress) class-based admission control at the border.

3.3.2 Provisioning

In considering the provisioning of the links constituting intra- and inter-domain boundaries, several capacity limits can be distinguished [8]:

- *Absolute Link Capacity (ALC)*: The transmission capacity of the physical link.
- *CESNET2 DiffServ Limit (C2DSL)*: The maximum proportion of a given traffic class allowed on any link. If too much traffic of a given class is allowed onto

a link, then the qualitative characteristics (e.g. jitter and delay) on the individual constituent flows will increase and the behavior will tend to become more like BE. In the extreme case with all the traffic allowed to be e.g. EF, the behavior would revert to BE. The CESNET2 DiffServ traffic limit is the maximum proportion of a traffic class which should be admitted on a link, and it is determined by the need for this traffic class to operate correctly for the traffic mix for which this traffic class is intended. This limit is important because if more than the absolute amount represented by this proportion is needed on a given link, then a larger capacity link is needed in order to maintain the particular characteristics class of service.

- *Operational Allocation (OA)*: The operating proportion determined as being available for the given traffic class on a given link. At any particular time, the provisioned traffic of this class, i.e. the amount admitted for the link, may be less than the CESNET2 DiffServ limit for this traffic class, since the demand within the CESNET2 DiffServ domain may not require the maximum allocation, or it may be decided to limit it for other reasons.
- *Operational Policing Limit (OPL)*: The operating policing parameters in force for the traffic class on a link. The purpose of policing traffic as it arrives is to ensure that the agreed provision of capacity for traffic using a particular service is not exceeded. Even in an environment where networks trust each other, such policing is generally desirable at least as the first level of defense against malicious traffic starving lower level service traffic.

Dimensioning intradomain provisioning

The following intradomain provisioning of the *CESNET2 DiffServ Limit* is required within the CESNET2 MPLS DiffServ domain:

- At most 33% of the *absolute link capacity* can be allocated for the Real-Time class of service.
- At least 25% of *absolute link capacity* must be left for the Best Effort class of service (i.e. remain unallocated to other service classes).

Table 6 shows the CESNET2 MPLS DiffServ intradomain operational allocation of reserved bandwidth and proposed PHB for supported QoS classes of services.

Table 6. CESNET2 intra-domain initial operational allocation of reserved bandwidth and PHB for QoS classes of services.

Service class name	PHB	OA bandwidth
Real-Time	EF	25%
Network Control	AF	2%
Video	AF	20%
Critical Traffic	AF	25%
Best Effort	Default	25%
Less than Best Effort	AF	3%

Dimensioning inter-domain provisioning

Dimensioning inter-domain provisioning of the CESNET2 MPLS DiffServ domain with other QoS domains must be compliant with CESNET2 intra-domain provisioning policy. It mainly means that the aggregated incoming and outgoing traffic of the particular QoS class must not negatively influence the required PHB of any network node within the CESNET2 MPLS DiffServ domain.

The QoS SLS should contain the following mandatory traffic parameters for *each higher level class of service*:

- Ingress Committed Rate within the *class*: ICR_{class} ,
 - Ingress Burst Conform within the *class*: IBC_{class} ,
 - Ingress Burst Exceed within the *class*: IBE_{class} ,
- and the following optional traffic parameters:
- Egress Committed Rate within the *class*: ECR_{class} ,
 - Egress Burst Conform within the *class*: EBC_{class} ,
 - Egress Burst Exceed within the *class*: EBE_{class} ,

and *aggregated traffic* (the whole traffic for all classes of service including higher level service classes and lower level service classes):

- Ingress Committed Rate: ICR,
 - Ingress Burst Conform: IBC,
 - Ingress Burst Exceed: IBE,
- optionally:
- Egress Committed Rate: ECR,
 - Egress Burst Conform: EBC,
 - Egress Burst Exceed: EBE.

Such an SLS is capable of providing the necessary parameters for properly adjusting the srTCM (*single rate three color marker*) based policers deployed at the border of the CESNET2 MPLS DiffServ domain to ensure required operational policing limits.

For TCP traffic with unknown characteristics the following initial parameter settings are recommended:

- $IBC \text{ [Byte]} = ICR \text{ [bit/s]} * 1 \text{ [Byte]} / 8 \text{ [bit]} * RTT \text{ [s]}$ (where RTT means Round Trip Time).
- $IBE \text{ [Byte]} = 2 * IBC \text{ [Byte]}$.

3.3.3 Classification and Marking

The implementation of QoS on the CESNET2 MPLS-based backbone uses wherever possible a DSCP to Exp value mapping that corresponds to the Cisco IOS implicit MPLS ingress behavior: the first three bits of the DSCP are copied to the Exp field during MPLS label imposition.

Table 7 shows the CESNET2 MPLS DiffServ domain compliant DSCP mapping for the particular supported service classes.

It is a responsibility of the neighboring DiffServ domain to ensure proper marking of the IPv4/IPv6 packets entering the CESNET2 MPLS DiffServ domain to be

Table 7. CESNET2 compliant DSCP/Exp/CoS marking for service classes.

Service class	PHB	DSCP	Exp MPLS	802.1p CoS
real-time	EF	EF, CS5	5	5
network control	AF	CS6, CS7	6 (7)	6 (7)
video	AF	AF4x, CS4	4	4
critical traffic	AF	AF3x, CS3, AF2x, CS2	3 (2)	3 (2)
best effort	default	0 (others)	0	0
LBE	AF	AF1x, CS1	1	1

compliant with the CESNET2 QoS services. Owing to the “Short-Pipe” MPLS DiffServ tunneling mode the DSCP transparency of the in-profile user IPv4/IPv6 packets (SLA conformant) is provided while being transported through the CESNET2 MPLS DiffServ domain.

3.3.4 Authentication, Authorization, and Trust

QoS admission control must be applied by means of policing at all ingress points of the CESNET2 MPLS DiffServ domain in the following way:

- Ingress srTCM policer checks all the incoming traffic within each particular QoS service class against an SLA.
- All the IPv4/IPv6 packets of the conforming traffic of the particular QoS class remain *unchanged* and the corresponding Exp mapping (determining the PHB within the CESNET2 MPLS DiffServ domain) is derived from their original DSCP values (see Table 7).
- The DSCP of IPv4/IPv6 packets representing the exceeding/violating (out-of-profile) traffic of the particular QoS service class *may be changed* and the corresponding Exp mapping (determining the PHB within the CESNET2 MPLS DiffServ domain) is derived from the implied policing rules valid for this type of exceeding/violating traffic (see Table 8).

Policing

QoS admission control *must be applied* by means of policing at all *ingress* points of the CESNET2 MPLS DiffServ domain to enforce the SLA with other DiffServ domains for each supported service class. Traffic policing (or shaping) is *optional* at *egress* points of the CESNET2 MPLS DiffServ domain.

The ingress QoS operational policing limits are enforced by the srTCM based policers. The goal is to reclassify/markdown input user traffic according to the Table 8. If ICR/IPR (the total sum of the traffic across all service classes) equals the absolute link capacity an ordinary srTCM policer is sufficient to check the SLS compliance for each particular service class. Otherwise, if ICR/IPR is less than the absolute link capacity, a hierarchical srTCM policer is required: the first level policer checks the conformity of the total traffic (the whole traffic regardless the service classes) and the second level policer checks the conformity of each particular service class being provided. Optionally, the same can be applied to EBC, if the egress QoS operational policing limits are required.

Table 8 shows the proposed reclassification policy of the ingress QoS admission

control at the edge of the CESNET2 MPLS DiffServ domain. It is strongly recommended that the Network Control class of incoming traffic is checked against the ACL of the authorized external sources of the neighboring QoS domain to allow only authorized control traffic (e.g. BGP and/or MSDP) to transit (or end inside) the CESNET2 MPLS DiffServ domain to avoid unwanted malicious competition of unauthorized customer traffic within the privileged Network Control class.

Table 8. Ingress admission control and conditioning at the edge of the CESNET2 DiffServ domain.

Service class	Conform ($t < IBC_t$)	Exceed ($IBC_t > t < IBE_t$)	Violate ($t > IBE_t$)
real-time	real-time	best effort	LBE
network control	network control	best effort	LBE
video	video	best effort	LBE
critical traffic	critical traffic	best effort	LBE
best effort	–	–	–
best effort	–	–	–

3.4 QoS Implementation on the CESNET2 MPLS-based Backbone

Technical implementation of the proposed CESNET2 QoS Policy Framework heavily depends on hardware platform: not only on the particular Cisco router model, but also on the particular module in case of Cisco 7600:

- Cisco 7600 OSR Sup720-3BXL/3CXL
- equipped with OSM³, SIP⁴ and ES20⁵ modules can use a proper variant of the LLQ/CBWFQ queuing disciplines for EF/AF PHB on output (configured via the Cisco IOS MQC) with WRED if supported,
- equipped with LAN modules⁶ can use WRR scheduling with a priority queue on output (typically 1p7q8t for 10 Gigabit Ethernet, 1p3q8t or 1p2q2t for Gigabit Ethernet and 1p3q1t for Fast Ethernet output interfaces in case of CESNET2 MPLS backbone) with WRED if supported.
- Cisco CRS-1 can use the LLQ/MDRR queuing disciplines for EF/AF PHB on output (configured via the Cisco IOS XR MQC).

Figure 2 shows the classification of the types of interfaces of the CESNET2 MPLS Exp-LSP DiffServ domain from the QoS point of view:

- IP-to-IP: IP based QoS interfaces (the external/neighboring IPv4/IPv6 DiffServ domain):
- CE-PE-out: outgoing CE⁷ interface facing to PE,

³ OSM-1OC48-POS-SL+, OSM-1OC48-POS-SS+, OSM-2+4GE-WAN+

⁴ 7600-SIP-400/SPA-1XOC48POS/RPR

⁵ 7600-ES20-10G3CXL

⁶ LAN modules typically denote the internal queuing structure of their interfaces by notation of the number of priority (p) and WRR (q) queues, each with the number of configurable WRED thresholds (t). Thus the abbreviation 1p2q2t means: 1 priority queue, 2 WRR queues each with 2 thresholds.

⁷ External CE of a neighboring IPv4/IPv6 DiffServ domain.

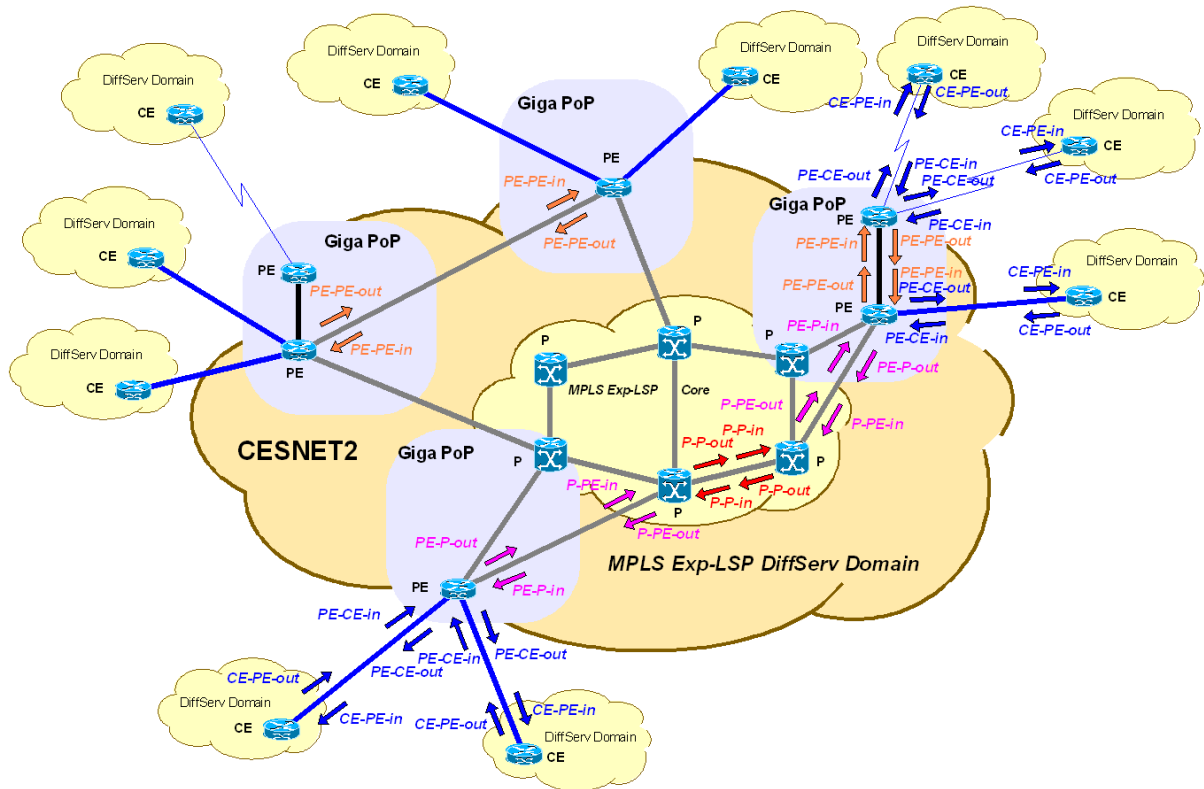


Figure 2. Types of QoS interface in the CESNET2 MPLS DiffServ domain.

- CE-PE-in: incoming CE⁸ interface from PE.
- IP-to-MPLS: IP based QoS interfaces:
- PE-CE-in: incoming PE interface from CE.
- MPLS-to-IP: IP based QoS interfaces:
- PE-CE-out: outgoing PE interface facing to CE.
- MPLS-to-MPLS: MPLS⁹ based QoS interfaces:
- PE-PE-out: outgoing PE interface¹⁰ facing to another PE,
- PE-PE-in: incoming PE interface¹¹ from another PE,
- PE-P-out: outgoing PE interface facing to P,
- PE-P-in: incoming PE interface¹² from P,
- P-P-out: outgoing P interface facing to another P,
- P-P-in: incoming P interface from another P,
- P-PE-out: outgoing P interface¹³ facing to PE,
- P-PE-in: incoming P interface from PE.

⁸ External CE of a neighboring IPv4/IPv6 DiffServ domain.

⁹ When penultimate hop popping occurs, pure IPv4/IPv6 packets can traverse the last link within the MPLS cloud.

¹⁰ Penultimate hop popping can occur.

¹¹ Penultimate hop popping can occur.

¹² Penultimate hop popping can occur.

¹³ Penultimate hop popping can occur.

3.4.1 QoS classification

For traffic classification of supported service classes the following IOS/XR class maps can be used:

IP-to-MPLS IP-based QoS interfaces:

```
class-map match-any CESNET2-RealTime-DSCP
  description Real-Time
  match dscp cs5 ef
class-map match-any CESNET2-NetworkControl-DSCP
  description Network Control
  match dscp cs6 cs7
class-map match-any CESNET2-Video-DSCP
  description Video
  match dscp cs4 af41 af42 af43
class-map match-any CESNET2-CriticalTraffic-DSCP
  description Critical Traffic
  match dscp cs2 af21 af22 af23 cs3 af31 af32 af33
class-map match-any CESNET2-LessEffort-DSCP
  description Less than Best Effort
  match dscp cs1 af11 af12 af13
class-map match-any CESNET2-BestEffort-DSCP
  description Best Effort
  match dscp 0
```

MPLS-to-MPLS or MPLS-to-IP MPLS-based or IP-based QoS interfaces:

```
class-map match-any CESNET2-RealTime
  description Real-Time
  match dscp cs5 ef
  match mpls experimental topmost 5
class-map match-any CESNET2-NetworkControl
  description Network Control
  match dscp cs6 cs7
  match mpls experimental topmost 6 7
class-map match-any CESNET2-Video
  description Video
  match dscp cs4 af41 af42 af43
  match mpls experimental topmost 4
class-map match-any CESNET2-CriticalTraffic
  description Critical Traffic
  match dscp cs2 af21 af22 af23 cs3 af31 af32 af33
  match mpls experimental topmost 2 3
class-map match-any CESNET2-LessEffort
  description Less than Best Effort
  match dscp cs1 af11 af12 af13
  match mpls experimental topmost 1
class-map match-any CESNET2-BestEffort
  description Best Effort
  match dscp 0
  match mpls experimental topmost 0
```

3.4.2 Cisco 7600 Sup720-3BXL/3CXL QoS Implementation

Hardware QoS acceleration on Cisco 7600 Sup720-3BXL/3CXL running IOS version 12.2(33)SRB2 is configured by

```
mls qos
```

Configure DSCP or Exp field packet rewrite on outgoing interfaces according to the internal DSCP (necessary for PFC MPLS QoS compatibility):

```
mls qos rewrite ip dscp
```

The trust state of the incoming physical IP interfaces with the QoS support enabled must be set to DSCP (the trust state of MPLS encapsulated interfaces does not matter¹⁴, but in case of possible PHP the DSCP trust state is also necessary):

```
interface gigabit ...
mls qos trust dscp
```

QoS admission control and reclassification/markdown

QoS admission control for IP-to-MPLS traffic is configured only on the PE-CE-in input IP interfaces of ingress PE routers in the CESNET2 MPLS DiffServ domain using the hardware based predefined/preconfigured policed DSCP maps. It is important to note that classification (based on the original received IP header) and marking (done to the internal DSCP) do not distinguish between IP-to-IP traffic and IP-to-MPLS traffic. The commands that you use to mark IPv4/IPv6 DSCP and mark Exp have the same result as when you mark the internal DSCP. The (PFC) internal DSCP values of the *out-of-profile* policed packets (corresponding to the exceeding/violating portion of incoming traffic) can be changed by global internal PFC maps. This results in a possible real DSCP packet rewrite on non-MPLS output IP interfaces of the same PE router; however, the Exp value of the newly imposed MPLS header on MPLS output interfaces for all in-profile and out-of-profile packets (corresponding to the dscp-exp map) is derived from the internal (PFC) DSCP which reflects the result of input policing:

```
!
! Redefine policed-dscp maps for input admission control
!
! Exceed traffic: normal-burst policed-dscp map
no mls qos map policed-dscp normal-burst
! Reclassify/remark RealTime (EF, CS5) DSCP to BestEffort (0)
mls qos map policed-dscp normal-burst 40 46 to 00
! Reclassify/remark NetworkControl (CS6, CS7) DSCP to BestEffort (0)
mls qos map policed-dscp normal-burst 48 56 to 00
! Reclassify/remark Video (CS4, Af4x) DSCP to BestEffort (0)
mls qos map policed-dscp normal-burst 32 34 36 38 to 00
! Reclassify/remark CriticalTraffic (CS3, AF3x, CS2, AF2x) DSCP
```

¹⁴ Unless explicitly using no mls qos mpls trust experimental.

```

! to BestEffort (0)
mls qos map policed-dscp normal-burst 24 26 28 30 to 00
mls qos map policed-dscp normal-burst 16 18 20 22 to 00
!
! Violate traffic: max-burst policed-dscp map
no mls qos map policed-dscp max-burst
! Reclassify/remark RealTime (EF, CS5) DSCP to LessEffort (CS1)
mls qos map policed-dscp max-burst 40 46 to 8
! Reclassify/remark NetworkControl (CS6, CS7) DSCP to LessEffort (CS1)
mls qos map policed-dscp max-burst 48 56 to 8
! Reclassify/remark Video (CS4, Af4x) DSCP to LessEffort (CS1)
mls qos map policed-dscp max-burst 32 34 36 38 to 8
! Reclassify/remark CriticalTraffic (CS3, AF3x, CS2, AF2x) DSCP
! to LessEffort (CS1)
mls qos map policed-dscp max-burst 24 26 28 30 to 8
mls qos map policed-dscp max-burst 16 18 20 22 to 8

```

Ingress IP4/IPv6 admission control and reclassification/markdown example for a 10 Gigabit Ethernet PE-CE-in input IP interface on a PE router follows:

```

!
! QoS ingress admission control (LAN ports)
! -----
!
! QoS ingress admission control policy definition from MAN/university/Inet
! - 10 Gigabit Ethernet LAN ports
!
policy-map CESNET2-IP2MPLS-CE2PE-10GE-in
! bc(d)=CIR*d/8, be(d)=CIR*d/8
class CESNET2-NetworkControl-DSCP
! CIR: 10Mbps, bc: 500ms, be: 1000ms
police cir 10000000 bc 625000 be 1250000
conform-action set-mpls-exp-transmit 6
exceed-action policed-dscp-transmit
violate-action policed-dscp-transmit
class CESNET2-RealTime-DSCP
! CIR: 40Mbps, bc(d=125ms)=CIR*d/8=40000000*0.125/8=625000,
! be(d=250ms)=CIR*d/8=40000000*0.250/8=1250000
police cir 40000000 bc 625000 be 1250000
conform-action set-mpls-exp-transmit 5
exceed-action policed-dscp-transmit
violate-action policed-dscp-transmit
class CESNET2-Video-DSCP
! CIR: 60Mbps, bc: 250ms, be: 500ms
police cir 60000000 bc 1875000 be 3750000
conform-action set-mpls-exp-transmit 4
exceed-action policed-dscp-transmit
violate-action policed-dscp-transmit
class CESNET2-CriticalTraffic-DSCP
! CIR: 100Mbps, bc: 500ms, be: 1000ms
police cir 100000000 bc 6250000 be 12500000
conform-action set-mpls-exp-transmit 3
exceed-action policed-dscp-transmit

```

```

        violate-action policed-dscp-transmit
class CESNET2-LessEffort
    set mpls experimental imposition 1
class CESNET2-BestEffort
    set mpls experimental imposition 0
class class-default
    set mpls experimental imposition 0

!
! PE-CE-in physical ingress admission control for port-based interface
! from MAN/university/Inet
!
interface TenGigabitEthernet ...
mls qos trust dscp
...
service-policy input CESNET2-IP2MPLS-CE2PE-10GE-in

```

Implementation of DiffServ EF/AF PHB on ES20, OSM, and SIP/SPA modules

The required DiffServ EF/AF PHB is configured on the output interfaces of the ES20 modules using the LLQ/CBWFQ queuing disciplines via Cisco IOS MQC in the following manner:

```

! ES20 modules output QoS policy definition (IP/MPLS), WRED not supported
! -----
! 7600-ES20-10G3CXL
!
policy-map CESNET2-DiffServ-ES20-out
class CESNET2-RealTime
    priority
    ! CIR: 2.5Gbps (bc, be: not supported on ES20)
    police 2500000000 conform-action transmit exceed-action drop
class CESNET2-NetworkControl
    bandwidth percent 2
class CESNET2-Video
    bandwidth percent 20
class CESNET2-CriticalTraffic
    bandwidth percent 25
class CESNET2-LessEffort
    bandwidth percent 3
class class-default
    bandwidth percent 25

```

The required DiffServ EF/AF PHB is configured on output interfaces of the OSM/SIP/SPA modules using the LLQ/CBWFQ queuing disciplines via Cisco IOS MQC in the following way:

```

!
! OSM/SIP/SPA PoS OC48 modules output QoS policy definition (IP/MPLS)
! -----
! OSM-10C48-POS-SL+, OSM-10C48-POS-SS+, 7600-SIP-400/SPA-1XOC48POS/RPR
!

```

```

policy-map CESNET2-DiffServ-OSM-POS-out
  class CESNET2-RealTime
    priority percent 25
  class CESNET2-NetworkControl
    bandwidth percent 2
    random-detect dscp-based
  class CESNET2-Video
    bandwidth percent 20
  class CESNET2-CriticalTraffic
    bandwidth percent 25
    random-detect dscp-based
  class CESNET2-LessEffort
    bandwidth percent 3
    random-detect dscp-based
  class class-default
    bandwidth percent 25
    random-detect dscp-based

```

The required DiffServ EF/AF PHB is configured on output interfaces of the OSM GE-WAN modules using the LLQ/CBWFQ queuing disciplines via Cisco IOS MQC as follows:

```

!
! GE-WAN modules output QoS policy definition (IP/MPLS)
! -----
! OSM-2+4GE-WAN+
!
policy-map CESNET2-DiffServ-OSM-GEWAN-out
  class CESNET2-RealTime
    priority
    ! CIR: 250Mbps, bc(d=125ms)=CIR*d/8=250000000*0.125/8=3906250,
    !      be(d=250ms)=CIR*d/8=250000000*0.250/8=7812500
    police cir 250000000 bc 3906250
    conform-action transmit
    exceed-action drop
  class CESNET2-NetworkControl
    bandwidth percent 2
    random-detect dscp-based
  class CESNET2-Video
    bandwidth percent 20
  class CESNET2-CriticalTraffic
    bandwidth percent 25
    random-detect dscp-based
  class CESNET2-LessEffort
    bandwidth percent 3
    random-detect dscp-based
  class class-default
    bandwidth percent 25
    random-detect dscp-based

```

```

!
! OSM modules output QoS policy configuration (IP/MPLS)
! -----
!
! Physical routed GE/WAN MPLS interface
interface GE-WAN ...
 mls qos trust dscp
 service-policy output CESNET2-DiffServ-OSM-GEWAN-out

```

Implementation of the DiffServ EF/AF PHB on LAN modules

Tables 9, 10, 11 and 12 show the initial DiffServ EF/AF PHB queue scheduling and Exp/CoS mapping for all 1p7q8t, 1p3q8t, 1p2q2t, and 1p3q1t output ports of the LAN modules (Exp/CoS=Tx denotes the Exp/CoS values to the particular threshold number assignment, Tx denotes the threshold number).

Table 9. The IP/MPLS DiffServ compliant output queuing for 1p7q8t LAN 10GE ports. Abbreviations: Q=Queue, W=Weight, QL=Queue-limit.

Q	Type	Exp/CoS=Tx	W	QL	T1min	T1max	T2-8min	T2-8max
1	WRR	1=T1	3	5%	80%	100%	100%	100%
2	WRR	0=T1	25	25%	80%	100%	100%	100%
3	WRR	2=T1, 3=T1	25	25%	80%	100%	100%	100%
4	WRR	4=T1	20	15%	80%	100%	100%	100%
5	WRR	6=T1, 7=T1	2	5%	80%	100%	100%	100%
6	WRR	-	-	-	80%	100%	100%	100%
7	WRR	-	-	-	80%	100%	100%	100%
8	Priority	5	-	25%	-	-	-	-

Table 10. The IP/MPLS DiffServ compliant output queuing for 1p3q8t LAN 10GE ports. Abbreviations: Q=Queue, W=Weight, QL=Queue-limit.

Q	Type	Exp/CoS=Tx	W	QL	T1min	T1max	T2min	T2max	T3min	T3max
1	WRR	1=T1	3	5%	0%	100%	100%	100%	100%	100%
2	WRR	0=T1	25	25%	0%	100%	100%	100%	100%	100%
3	WRR	4=T1, 2=T2, 3=T2, 6=T3, 7=T3	72	45%	70%	80%	80%	90%	90%	100%
4	Priority	5	-	25%	-	-	-	-	-	-

Table 11. The IP/MPLS DiffServ compliant output queuing for 1p2q2t LAN GE ports. Abbreviations: Q=Queue, W=Weight, QL=Queue-limit.

Q	Type	Exp/CoS=Tx	W	QL	T1min	T1max	T2min	T2max
1	WRR low	1=T1, 0=T2	25	40%	40%	80%	80%	100%
2	WRR high	2=T1, 3=T1, 4=T1, 6=T2, 7=T2	75	30%	70%	90%	90%	100%
3	Priority	5	-	30%	-	-	-	-

The required DiffServ EF/AF PHB is configured on all IP and MPLS output interfaces of the LAN modules using the following WRR/priority queues specific commands:

Table 12. The IP/MPLS DiffServ compliant output queuing for 1p3q1t LAN FE ports. Abbreviations: Q=Queue, W=Weight, QL=Queue-limit.

Q	Type	Exp/CoS=Tx	W	QL	T1min	T1max
1	WRR	1=T1	3	80%	100%	
2	WRR	0=T1	25	80%	100%	
3	WRR	2=T1, 3=T1, 4=T1, 6=T1, 7=T1	72	90%	100%	
4	Priority	5	-	0%	0%	

1p7q8t LAN 10 Gigabit Ethernet output ports:

```

!
! 1p7q8t LAN 10GE output ports
! -----
! WS-X6704-10GE
!
! PQ: RealTime25%EF, CS5Exp5Cos5
! Q5T1: NetWorkControl 2%CS7, CS6Exp6(7)Cos6(7)
! Q4T1: Video20%CS4, AF4xExp4Cos4
! Q3T1: CriticalTraffic25%CS3, AF3x, CS2, AF2xExp3(2)Cos3(2)
! Q2T1: BestEffort25%0Exp0Cos0
! Q1T1: LessEffort3%CS1, AF1xExp1Cos1
!
!-----
!
!>>> interface Ten1/4
!
default wrr-queue bandwidth
default wrr-queue cos-map
default wrr-queue queue-limit
default wrr-queue random-detect 1
default wrr-queue random-detect 2
default wrr-queue random-detect 3
default wrr-queue random-detect 4
default wrr-queue random-detect 5
default wrr-queue random-detect 6
default wrr-queue random-detect 7
default wrr-queue threshold 1
default wrr-queue threshold 2
default wrr-queue threshold 3
default wrr-queue threshold 4
default wrr-queue threshold 5
default wrr-queue threshold 6
default wrr-queue threshold 7
! size: q1% q2% q3% q4% q5% q6% q7%
wrr-queue queue-limit 5 25 25 15 5 0 0
! Allocates 5% to Q1, 25% to Q2, 25% to Q3, 15% to Q4,
! Allocates 5% to Q5, 0% to Q6 and 0% to Q7
! weight: q1 q2 q3 q4 q5 q6 q7
wrr-queue bandwidth 3 25 25 20 2 0 0
! Sets the WRR weights for 3:25:25:20:2:0:0 (Q1 through Q7)
! enable WRED: q1, q2, q3, q5

```

```

wrr-queue random-detect 1
wrr-queue random-detect 2
wrr-queue random-detect 3
wrr-queue random-detect 5
! threshold: q# t1% t2% t3% t4% t5% t6% t7% t8%
wrr-queue random-detect min-threshold 1 80 100 100 100 100 100 100 100
wrr-queue random-detect min-threshold 2 80 100 100 100 100 100 100 100
wrr-queue random-detect min-threshold 3 80 100 100 100 100 100 100 100
wrr-queue random-detect min-threshold 4 80 100 100 100 100 100 100 100
wrr-queue random-detect min-threshold 5 80 100 100 100 100 100 100 100
wrr-queue random-detect min-threshold 6 80 100 100 100 100 100 100 100
wrr-queue random-detect min-threshold 7 80 100 100 100 100 100 100 100
wrr-queue random-detect max-threshold 1 100 100 100 100 100 100 100 100
wrr-queue random-detect max-threshold 2 100 100 100 100 100 100 100 100
wrr-queue random-detect max-threshold 3 100 100 100 100 100 100 100 100
wrr-queue random-detect max-threshold 4 100 100 100 100 100 100 100 100
wrr-queue random-detect max-threshold 5 100 100 100 100 100 100 100 100
wrr-queue random-detect max-threshold 6 100 100 100 100 100 100 100 100
wrr-queue random-detect max-threshold 7 100 100 100 100 100 100 100 100
! q: queue# threshlod# cos#...
wrr-queue cos-map 1 1 1
wrr-queue cos-map 2 1 0
wrr-queue cos-map 3 1 2 3
wrr-queue cos-map 4 1 4
wrr-queue cos-map 5 1 6 7
! p: queue# cos#...
priority-queue cos-map 1 5

```

1p3q8t LAN Gigabit Ethernet output ports:

```

!
! 1p3q8t LAN GE output ports
! -----
! WS-X6724-SFP, WS-X6748-GE-TX
!
! PQ: RealTime25%EF, CS5Exp5Cos5
! Q3T3: NetWorkControl 2%CS7, CS6Exp6(7)Cos6(7)
! Q3T1: Video20%CS4, AF4xExp4Cos4
! Q3T2: CriticalTraffic25%CS3, AF3x, CS2, AF2xExp3(2)Cos3(2)
! Q2T1: BestEffort25%0Exp0Cos0
! Q1T1: LessEffort3%CS1, AF1xExp1Cos1
!
!-----
!
!>>> interface Giga4/4
default wrr-queue bandwidth
default wrr-queue cos-map
default wrr-queue queue-limit
default wrr-queue random-detect 1
default wrr-queue random-detect 2
default wrr-queue random-detect 3
default wrr-queue threshold 1

```

```

default wrr-queue threshold 2
default wrr-queue threshold 3
! size: q1% q2% q3%
wrr-queue queue-limit 5 25 45
! weight: q1 q2 q3
  wrr-queue bandwidth 3 25 72
! enable WRED: q1, q2, q3
wrr-queue random-detect 1
wrr-queue random-detect 2
wrr-queue random-detect 3
! threshold: q# t1% t2% t3% t4% t5% t6% t7% t8%
wrr-queue random-detect min-threshold 1 80 100 100 100 100 100 100 100
wrr-queue random-detect min-threshold 2 80 100 100 100 100 100 100 100
wrr-queue random-detect min-threshold 3 70 80 90 100 100 100 100 100
wrr-queue random-detect max-threshold 1 100 100 100 100 100 100 100 100
wrr-queue random-detect max-threshold 2 100 100 100 100 100 100 100 100
wrr-queue random-detect max-threshold 3 80 90 100 100 100 100 100 100
! q: queue# threshlod# cos#...
wrr-queue cos-map 1 1 1
wrr-queue cos-map 2 1 0
wrr-queue cos-map 3 1 4
wrr-queue cos-map 3 2 2 3
wrr-queue cos-map 3 3 6 7
! p: queue# cos#...
priority-queue cos-map 1 5

```

1p2q2t LAN Gigabit Ethernet output ports:

```

!
! 1p2q2t LAN GE output ports
! -----
! WS-SUP720-3BXL, WS-X6516-GBIC, WS-X6516A-GBIC, WS-X6816-GBIC,
! WS-X6548-GE-TX,
! OSM-10C48-POS-SL+, OSM-10C48-POS-SS+, OSM-2+4GE-WAN+
!
! PQ: RealTime25%EF, CS5Exp5Cos5
! Q2T2: NetWorkControl 2%CS7, CS6Exp6(7)Cos6(7)
! Q2T1: Video20%CS4, AF4xExp4Cos4
! Q2T1: CriticalTraffic25%CS3, AF3x, CS2, AF2xExp3(2)Cos3(2)
! Q1T2: BestEffort25%0Exp0Cos0
! Q1T1: LessEffort3%CS1, AF1xExp1Cos1
!
!-----
!
!>>> interface Giga4/4
default wrr-queue bandwidth
default wrr-queue cos-map
default wrr-queue queue-limit
! size: q1% q2% [=p1%]
wrr-queue queue-limit 40 30
! weight: q1 q2
wrr-queue bandwidth 25 75

```

```

! threshold: q# t1% t2%
wrr-queue random-detect min-threshold 1 40 80
wrr-queue random-detect min-threshold 2 70 90
wrr-queue random-detect max-threshold 1 80 100
wrr-queue random-detect max-threshold 2 90 100
! q: queue# threshlod# cos#...
wrr-queue cos-map 1 1 1
wrr-queue cos-map 1 2 0
wrr-queue cos-map 2 1 2 3 4
wrr-queue cos-map 2 2 6 7
! p: queue# cos#...
priority-queue cos-map 1 5

```

1p3q1t LAN Fast Ethernet output ports:

```

!
! 1p3q1t LAN FE output ports
! -----
! WS-X6548-RJ-45
!
! PQ: RealTime25%EF, CS5Exp5Cos5
! Q3T1: NetWorkControl 2%CS7, CS6Exp6(7)Cos6(7)
! Q3T1: Video20%CS4, AF4xExp4Cos4
! Q3T1: CriticalTraffic25%CS3, AF3x, CS2, AF2xExp3(2)Cos3(2)
! Q2T1: BestEffort25%0Exp0Cos0
! Q1T1: LessEffort3%CS1, AF1xExp1Cos1
!
!-----
!
!>>> interface Fast8/3
default wrr-queue bandwidth
default wrr-queue cos-map
! weight: q1 q2 q3
wrr-queue bandwidth 3 25 72
! threshold: q# t1%
wrr-queue random-detect min-threshold 1 80
wrr-queue random-detect min-threshold 2 80
wrr-queue random-detect min-threshold 3 90
wrr-queue random-detect max-threshold 1 100
wrr-queue random-detect max-threshold 2 100
wrr-queue random-detect max-threshold 3 100
! q: queue# threshold# cos#...
wrr-queue cos-map 1 1 1
wrr-queue cos-map 2 1 0
wrr-queue cos-map 3 1 2 3 4 6 7
! p: queue# cos#...
priority-queue cos-map 1 5

```

3.4.3 Cisco CRS-1 QoS Implementation

QoS admission control and reclassification/markdown

QoS admission control for IP-to-MPLS traffic is configured only on the PE-CE-in input IP interfaces of ingress PE routers into the CESNET2 MPLS DiffServ domain using the Cisco IOS XR MQC in the following way:

```

!
! QoS ingress admission control policy definition from MAN/university/Inet
! - Gigabit Ethernet LAN ports
!
policy-map CESNET2-IP2MPLS-CE2PE-GE-in
  ! bc(d)=CIR*d/8, be(d)=CIR*d/8
  class CESNET2-NetworkControl-DSCP
    ! CIR: 10Mbps, bc: 500ms, be: 1000ms
    police cir 10000000 bc 625000 be 1250000
      conform-action set mpls experimental imposition 6
      exceed-action set mpls experimental imposition 0
      violate-action set mpls experimental imposition 1
  class CESNET2-RealTime-DSCP
    ! CIR: 40Mbps, bc(d=125ms)=CIR*d/8=40000000*0.125/8=625000,
    !     be(d=250ms)=CIR*d/8=40000000*0.250/8=1250000
    police cir 40000000 bc 625000 be 1250000
      conform-action set mpls experimental imposition 5
      exceed-action set mpls experimental imposition 0
      violate-action set mpls experimental imposition 1
  class CESNET2-Video-DSCP
    ! CIR: 60Mbps, bc: 250ms, be: 500ms
    police cir 60000000 bc 1875000 be 3750000
      conform-action set mpls experimental imposition 4
      exceed-action set mpls experimental imposition 0
      violate-action set mpls experimental imposition 1
  class CESNET2-CriticalTraffic-DSCP
    ! CIR: 50Mbps, bc: 500ms, be: 1000ms
    police cir 50000000 bc 3125000 be 6250000
      conform-action set mpls experimental imposition 3
      exceed-action set mpls experimental imposition 0
      violate-action set mpls experimental imposition 1
  class CESNET2-LessEffort
    set mpls experimental imposition 1
  class CESNET2-BestEffort
    set mpls experimental imposition 0
  class class-default
    set mpls experimental imposition 0

```

Ingress IP4/IPv6 admission control and reclassification/markdown example for a Gigabit Ethernet PE-CE-in input IP interface on a PE router follows:

```

!
! PE-CE-in physical ingress admission control for port-based interface from
! MAN/university/Inet
!

```

```
interface GigabitEthernet ...
  service-policy input CESNET2-IP2MPLS-CE2PE-GE-in
```

Implementation of the DiffServ EF/AF PHB

The required DiffServ EF/AF PHB is configured on output interfaces using the LLQ/MDRR queuing disciplines via Cisco IOS XR MQC in the following way:

```
!
! Output backbone policy (MPLS/IP)
! =====
!
policy-map CESNET2-DiffServ-out
  class CESNET2-RealTime
    ! CIR: 2.5Gbps, bc(d=125ms)=CIR*d/8=2500000000*0.125/8=39062500,
    !     be(d=250ms)=CIR*d/8=2500000000*0.250/8=78125000
    police rate percent 25 burst 125 ms peak-burst 250 ms
    conform-action transmit
    exceed-action set mpls experimental topmost 0
    violate-action drop
  !
  priority
  !
  class CESNET2-NetworkControl
    bandwidth percent 2
    random-detect default
  !
  class CESNET2-Video
    bandwidth percent 20
  !
  class CESNET2-CriticalTraffic
    bandwidth percent 25
    random-detect default
  !
  class CESNET2-LessEffort
    bandwidth percent 3
    random-detect default
  !
  class class-default
    bandwidth percent 25
    random-detect default
  !
end-policy-map

!
! Output QoS policy configuration (IP/MPLS)
!-----
!
! Physical routed GE MPLS interface
interface GigabitEthernet ...
  service-policy output CESNET2-DiffServ-out
```

4 Conclusion

Currently there is only a little need for QoS coming from general academic users. This can be justified by the fact that CESNET2 MPLS-based backbone operational behavior is based on an overprovisioned network approach for common user traffic. On the other hand we believe that some emerging extensive applications like real-time data acquisition from remote sensors, high definition video and GRID computation will need delay, jitter and packet loss guarantees in the near future no matter what the increase of common user traffic will be.

It is good to remember that QoS introduces a system of managed unfairness. The successful QoS policy rollout should be followed by ongoing monitoring of service levels and periodic adjustments and tuning of QoS policies. As traffic conditions change, we will have to adapt to these changes and maybe start the QoS deployment cycle anew, by redefining the objectives, tuning and testing corresponding designs, rolling out new designs and monitoring them to check if they match the redefined objectives.

The major drawback of the proposed CESNET2 QoS policy implementation is the fact that real QoS PHB on egress IPv4/IPv6 edge interfaces of the CESNET2 MPLS DiffServ domain towards other neighboring QoS domains may be based on the original DSCP of user packets entering the CESNET2 MPLS DiffServ domain (due to “Short-Pipe” MPLS DiffServ tunneling mode because of the possibility of penultimate hop popping). This means that the real egress traffic mix may not obey the ingress QoS admission control rules. A question remains whether it is worth another limited output QoS policing to enforce the egress QoS admission rules once the user packets have already been successfully transported throughout the CESNET2 MPLS DiffServ domain or whether they should rather obey the (usually mandatory) ingress QoS admission control rules when entering the neighboring QoS domain. Penultimate hop popping (in case of not using MPLS/BGP VPN technology, when only one MPLS label is present) can be avoided by using the `mpls ldp explicit-null global` configuration command, but unfortunately in case of Cisco 7600 Sup720-3BXL/3CXL PFC QoS only at the cost of possible drastic forwarding performance degradation (by 50%).

5 Glossary of Terms

- AF – Assured Forwarding
- AF-PHB – Assured Forwarding Per Hop Behavior
- BGP – Border Gateway Protocol
- CE – Customer Equipment (router)
- CBR – Constant Bit Rate
- CBWFQ – Class-Based Weighted Fair Queuing
- CoS – Class of Service (IEEE 802.1p bits)
- DiffServ – Differentiated Services
- DS – Differentiated Services
- DSCP – Differentiated Services Code Point
- EBC – Egress Burst Conform for the whole traffic (regardless QoS classes)

- EBC_{class} – Egress Burst Conform within the *class*
- EBE – Egress Burst Exceed for the whole traffic (regardless QoS classes)
- EBE_{class} – Egress Burst Exceed within the *class*
- ECR – Egress Committed Rate for the whole traffic (regardless QoS classes)
- ECR_{class} – Egress Committed Rate within the *class*
- EF – Expedited Forwarding
- EF-PHB – Expedited Forwarding Per Hop Behavior
- EPR – Egress Peek Rate for the whole traffic (regardless QoS classes)
- EPR_{class} – Egress Peek Rate within the *class*
- Exp – Experimental field of the MPLS header
- Exp-LSP – Label Switch Path based on the Experimental field
- E-LSP – Label Switch Path based on the Experimental field
- IBC – Ingress Burst Conform for the whole traffic (regardless QoS classes)
- IBC_{class} – Ingress Burst Conform within the *class*
- IBE – Ingress Burst Exceed for the whole traffic (regardless QoS classes)
- IBE_{class} – Ingress Burst Exceed within the *class*
- ICR – Ingress Committed Rate for the whole traffic (regardless QoS classes)
- ICR_{class} – Ingress Committed Rate within the *class*
- IETF – Internet Engineering Task Force
- IP – Internet Protocol
- IPv4 – Internet Protocol version 4
- IPv6 – Internet Protocol version 6
- IPR – Ingress Peek Rate for the whole traffic (regardless QoS classes)
- IPR_{class} – Ingress Peek Rate within the *class*
- LAN – Local Area Network
- LBE – Less than Best Efforts
- LLQ – Low Latency Queuing
- MAN – Metropolitan Area Network
- MBS – Managed Bandwidth Service
- MIB – Management Information Base
- MPLS – Multiple Protocol Label Switching
- MQC – Modular Quality of Service Command-Line Interface
- MDRR – Modified Deficit Round Robin
- NREN – National Research and Education Network
- OSM – Optical Service Module
- P – Provider (router)
- PE – Provider Edge (router)
- PHB – Per Hop Behavior
- PHP – Penultimate Hop Popping
- PFC – Policy Feature Card
- PIP – Premium IP
- PoP – Point of Presence
- PoS – Packet over Sonet/SDH
- QoS – Quality of Service
- RED – Random Early Detection
- SLA – Service Level Agreement

- SLS – Service Level Specification
- sr+bs – General policing mechanism that provides single rate with burst size control
- srTCM – single rate Three Color Marker [10]
- trTCM – two-rate Three Color Marker [10]
- TE – Traffic Engineering
- VLAN – Virtual Local Area Network
- VPN – Virtual Private Network
- WRED – Weighted Random Early Detection
- WRR – Weighted Round Robin

References

- [1] ALVAREZ, S. *QoS for IP/MPLS Networks*. Cisco Press, 2006, ISBN 1-58705-233-4
- [2] BABIARZ, J.; CHAN, K.; BAKER, F. *Configuration Guidelines for DiffServ Service Classes*. RFC 4594¹⁵, IETF, August 2006.
- [3] BLACK, D.; BLAKE, S.; CARLSON, M.; DAVIES, E.; WANG, Z.; WEISS, W. *An Architecture for Differentiated Services*. RFC 2475¹⁶, IETF, December 1998.
- [4] CAMPANELLA, M. *Implementation architecture specification for the Premium IP service*. SEQUIN Project, Deliverable 2.1, Addendum 1. Available online¹⁷.
- [5] CAMPANELLA, M.; Sevasti, A. *Service Level Agreements specification for IP Premium Service*. SEQUIN Project, Deliverable 2.1, Addendum 2. Available online¹⁸.
- [6] CAMPANELLA, M.; CHIVALIER, P.; SEVASTI, A.; SIMAR, N. *Quality of Service Definition*. SEQUIN Project, Deliverable 2.1, March 2001. Available online¹⁹.
- [7] CISCO SYSTEMS, *Cisco TelePresence Network Systems 2.0 Design Guide*. Cisco Systems, 2008, OL-14133-01. Available online²⁰.
- [8] COOPER, C. *Policy Framework for Introduction of Network QoS into JANET*. UKERNA, 2004
- [9] CISCO SYSTEMS. *Cisco QoS Baseline*. Cisco Systems, 2002. Available online²¹.

¹⁵ <http://tools.ietf.org/html/rfc4594>

¹⁶ <http://tools.ietf.org/html/rfc2475>

¹⁷ <http://www.dante.net/sequin/deliverables/SEQ-01-044.pdf>

¹⁸ <http://www.dante.net/sequin/deliverables/SEQ-01-043.pdf>

¹⁹ <http://www.dante.net/sequin/deliverables/SEQ-01-030.pdf>

²⁰ <http://www.cisco.com/en/US/docs/solutions/Enterprise/Video/TP-Book.html>

²¹ http://www.cisco.com/en/US/technologies/tk543/tk759/...technologies_white_paper0900aecd80295a9b.pdf

- [10] HEINANEN, J.; GUERIN, R. *A Single Rate Three Color Marker*. RFC 2697²², IETF, September 1999.
- [11] HEINANEN, J.; GUERIN, R. *A Two Rate Three Color Marker*. RFC 2698²³, IETF, September 1999.
- [12] OLIFER, V. *JANET QoS Development Project Report*. UKERNA, 2004
- [13] *Quality of Service on GÉANT2*. Available online²⁴.
- [14] SABATINO, R.; CAMPANELLA, M. *QoS Implementation Plan SEQUIN Project, Deliverable 4.2*, May 2002. Available online²⁵.
- [15] ŠMRHA, P. *Introduction of Network QoS into CESNET2*. Internal technical report, Activity 606, CESNET, 2005.

²² <http://tools.ietf.org/html/rfc2697>

²³ <http://tools.ietf.org/html/rfc2698>

²⁴ <http://www.geant2.net/server/show/conWebDoc.1582>

²⁵ <http://archive.dante.net/sequin/deliverables/SEQ-02-010.pdf>