Integrated QoS Management

by NGMN Alliance

Version:	**1.3**
Date:	**12-December 2012**
Document Type:	**Final Deliverable (approved)**
Confidentiality Class:	**P - Public**
Authorised Recipients: (for CR documents only)	

Working Group:	**P-BEV (Backhaul EVolution)**
Project:	**Integrated QoS Management**
Editor / Submitter:	**Jerry VERSCHOORE DE LA HOUSSAYE**
Contributors:	**Miguel Angel ALVAREZ BERNARDO, Jerry VERSCHOORE DE LA HOUSSAYE**
Approved by / Date:	**NGMN Board / 12-December 2012**

Commercial Address:

ngmn ltd.
Friedrich-Ebert-Anlage 58 • 60325 Frankfurt • Germany
Phone +49 69 / 9 07 49 98-0 • Fax +49 69 / 9 07 49 98-41
office@ngmn.org • www.ngmn.org

Registered Office:

ngmn ltd.
42–44 Portman Road • Reading • Berkshire • RG30 1EA • UK
Company registered in England and Wales n. 5932387
VAT Number GB 918713901

Abstract: Short introduction and purpose of document

The deployment of LTE networks is a key step in fulfilling the promise of high broadband anywhere and being connected all the time. With the increasing separation between peak and average bitrates, the use of advanced QoS management mechanisms becomes increasingly necessary in order to deploy cost-effective backhaul networks and avoid over-dimensioning.

A key element of better QoS management is a better integration of service and transport mechanisms and performance measurements. RAN nodes, being necessary points of passage, could be the ideal link between these two levels. However, no standard approach currently exists, and the mechanisms proposed by vendors are proprietary and opaque.

NGMN can provide a framework for defining and supporting the development of integrated QoS management solutions in three ways: clarifying their definition, facilitating interoperability and their further integration with transport-level mechanisms.

This will increase the operational agility of operators, enable vendors to expose their support of advanced features and improve the overall capacity of the network to handle high-value traffic.

The current timing (with developing alternatives but no generic frame of reference) is right for NGMN to help establish a consensus on the definition and characteristics of these solutions.

This framework is organised into two parts;
- Part one is about service flow classification and the associated inter-layer classes of service alignment and,
- Part two is about investigating potential integrated service/transport layer QoS mechanisms.

This document covers part one of this framework.

Document History

Date	Version	Author	Changes
05/3/2012	1.0	Jerry Verschoore de la Houssaye – Tellabs, Miguel Angel Alvarez Bernardo – Orange	Draft published for first review
26/4/2012	1.1	Jerry Verschoore de la Houssaye – Tellabs,	Draft published for TAC review
17/09/2012	1.2	Jerry Verschoore de la Houssaye – Tellabs	Integrated feedback
12/12/2012	1.3		Final Deliverable

Contents

1 INTRODUCTION TO INTEGRATED QUALITY OF SERVICE MANAGEMENT

This document aims to develop a framework for "Integrated Quality of Service Management" for mobile backhaul as part of the wider "Backhaul Evolution" project. This stream is focused mainly on service flow classification and inter-layer Class of Service alignment, continuing previous work started in NGMN project OSB (Optimized Solutions for mobile Backhaul). A further topic for analysis is the use of integrated QoS management mechanisms involving both radio and transport nodes, but this is out of scope of the current document, as will be detailed below.

Given that the previous OSB project already addressed QoS issues, it is useful to introduce the main reasons for continuing work on this subject. Even more so when taking into account the extreme popularity of QoS subjects in all industry organizations - what exactly does the NGMN bring to the table? The short answer is: a short-term, pragmatic QoS management approach based on cooperation between network layers.

The fact is that over the evolution of mobile networks, we've moved from an initial model based on landline telephony (one service and strict reservation of resources with no sharing once access has been granted) to an environment where we have multiple services (video streaming, internet browsing, email in addition to voice) running in parallel over a shared medium. Given that the peak bandwidth needs of some of them greatly exceed the average traffic that they generate, we're confronted with a situation that is radically different (and more akin to traditional Internet dimensioning) from the initial QoS model.

With services that are reaching a peak-to-average ratio of 100, and have very different traffic patterns, both the traditional "TDM transmission" model of strict resource reservation per user and the free-for-all model of IP statistical multiplexing seem hardly sufficient. Our approach goes rather in the direction of facilitating a cooperative approach, where all stakeholders (network operators, end-users and content providers) are aware of the resources available and the current needs of users, so that an optimal allocation of resources can be reached.

In particular, we consider that the target solution should fulfil three base requirements:

+ **Instant Awareness**: the mechanism shall be aware of the evolution of available resources.

+ **Coherent Allocation**: the target mechanism shall avoid over-allocating resources in one segment with regards to the rest of the network. In particular, radio-level acceptance and bearer dimensioning shall be aligned with transport-level congestion management.

+ **Responsiveness**: the mechanism shall react dynamically to the evolution of user needs and available resources. This means that it shall be possible to select a higher level of service or a higher share of resources, and that the system shall react to this demand. A way to prioritize services in a congestion scenario is to allow an informed choice by all the involved stakeholders

We call the solutions combining these three concepts "Integrated QoS" management, as it will involve all interested parties in the resource allocation process, instead of applying pre-determined rules that may not be adapted to all possible scenarios.

We further subdivide Integrated QoS mechanisms into two sub-groups, depending on the implicit or explicit nature of the coordination across segments. Implicit mechanisms are related to CoS marking and service flow classification, so that, even if QoS management is not identical across the network, at least a certain relative priority of different flows is respected. However, marking service flow packets is only one potential implementation, Explicit mechanisms go much further, and involve dynamic exchanges of information and coordinated decision-making for resource allocation, pre-emption, etc. This second category is out of scope for the current document and is for future study.

1.1 Introduction to this study: Implicit QoS Integration

The following sections introduce in more detail the scope, objectives and structure of the present study, as delimited in the introduction section. It also strives to recap previous reflections on this subject, adapting some results from previous studies (particularly in sections 1.1.1 and 1.1.2).

1.1.1 Rationale

Why should the NGMN address the topic of Quality of Service, which is also a focus of study at multiple other organisms, such as for example; the Third Generation Partnership Project (3GPP), the Internet Engineering Task Force (IETF) or the Metro Ethernet Forum (MEF)?

What room for improvement is there?

In order to answer this question, we'll first review the evolution of service flow classification in fixed and mobile networks, to justify our goal and the role of this study in achieving it.

The truth is that the current approach to QoS management, which has evolved from Internet standards and their statistical – "best effort" – approach, may be an improvement with regards to the previous situation in the fixed world, but it still is far worse than what we could guarantee in legacy networks. Because of the static configuration of ATM links, it was always possible to achieve some degree of statistical multiplexing while still guaranteeing the bandwidth of sensitive service flows. Many related features, like for instance Call Admission Control, were lost with the passage to IP and variable-bandwidth services. There is also a lower level of certainty linked to the use of statistical mechanisms, and all these factors contribute to slowing the adoption of all-IP architectures by operators.

In order to recover a tighter control of the end-user quality of experience, while keeping the efficiency of IP networks, the notion of differentiated classes of service, or "differentiated services" (DiffServ) was introduced by the IETF. The key idea is to mark priority traffic, so that it will remain unaffected by the possible congestion due to the fluctuation of the link usage level.

All the mechanisms derived from this approach have expanded the number of ways that this prioritization may be enacted, but without re-establishing a close link between service needs and (pseudo)-deterministic transport behaviours. However, this approach soon showed its limitation, being declined into a myriad of incoherent classes of service, and leading to scenarios where one of the two following extremes needs to be chosen: either over-dimensioning of the transport infrastructure, or reducing the quality down to best-effort due to saturation and its inefficient management.

We cannot go beyond the basic DiffServ approach without improving the understanding of the needs of each individual service, and abandoning the idea of a strict top-down ordering (from "best" to "worst"). We cannot avoid coarse measures without understanding the boundaries of a given traffic and which transport factors impact it.

Lastly, RAN service elements cannot take full advantage of the mobile backhaul network unless they are aware of its characteristics and structure.

1.1.2 Definitions

Throughout this document various QoS/CoS related terms are used which are defined below for clarity.

Service flow classification in this document is defined as the process of:

+ Firstly, defining a limited number of classes of service (CoS), each of which: has a predefined set of network performance expectations by implementing a different scheduling priority with respect to the other classes, resulting in a different service experience or obtained network performance for service flows in this class – in case of network resource contention.

+ Secondly, mapping a service flow according to its service-specific network behaviour requirements into one of the defined set of classes of service.

The CoS a service flow gets mapped into depends on:

+ The intrinsic service flow QoS requirements
+ The network performance expectations or service behaviours of the available classes of service defined in the network

The scope of the QoS requirements shown in this document covers to the end-to-end transport service delivered by network infrastructure between the service source (e.g., the originating application server) and destination (e.g., client in the UE for downstream traffic), unless otherwise stated.

However, the scope of inter-layer CoS alignment is limited to the mobile backhaul network segment. The transport layer considered in this document may consist of Ethernet, IP and/or MPLS layers. These layers support the service layer (for example, S1/X2 bearers for 4G/LTE networks) between the mobile gateway and the base station (see figure 1 below). The radio interfaces between the base station and the user equipment or between microwave backhaul links are out of scope.

For this document, it is assumed that the Service Manager of the QoS management function, which is responsible for the bearer service QoS mechanisms as stated in TS 23.107 chapter 4.3 "Technical Requirements for QoS" will be responsible for implementing the proposed service flow classification and that the associated Translation Function will convert the bearer service attributes to the proposed CoS markings of the underlying transport layers. It is therefore assumed that the transport layer CoS markings for applications coming from external networks are not transparently copied inside the PLMN.

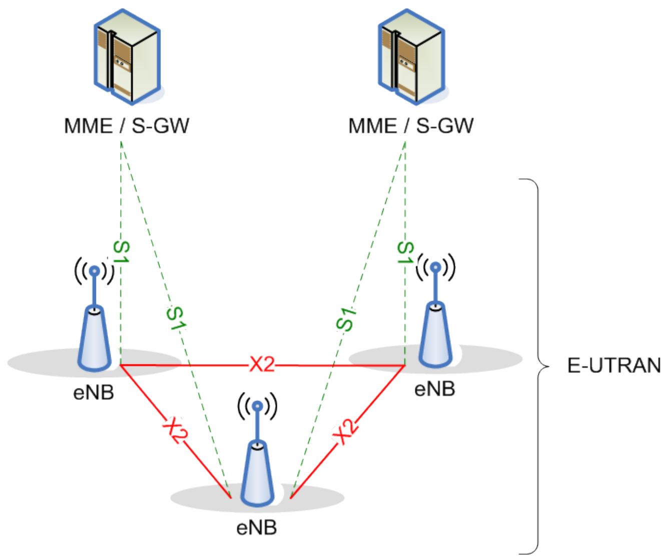

Figure 1: Mobile backhaul network for LTE from 3GPP TS 36.300 v8.11.0 (2009-12) – "Figure 4-1: Overall Architecture"

Out of scope are:

- The dynamic signalling between the service layer (RAN) and the underlying transport layer (L1/2/3); this is part of the explicit QoS mechanisms subgroup of "integrated QoS management" and is for future study.
- The impact of adaptive modulation of Microwave links; this is linked to the point above and is for future study
- The impact of encrypted service flows such as for example IPsec
- The impact of OAM service flows from transport network elements and radio network equipment
- The impact of LTE-A service delivery mechanisms on the service flow requirements, which is for future study

The service flow QoS requirements are not tied to any specific service- or transport layer technology.

In case of specific CoS marking, the specific service or transport layer is always mentioned or implicitly indicated (for example: PCP (Priority Code Point, formerly known as p-bits), DSCP (DiffServ Code Point), TC (Traffic Class) etc.).

Inter-layer CoS alignment deals with the mapping or aligning of the different service- and transport layers' CoS schemes to each other. For example aligning, IP layer's DiffServ per hop behaviour scheduling classes (PSC) and the CoS labels scheme (H, M or L) as recommended by the Metro Ethernet Forum (MEF) for the Ethernet layer and the TC field priority markings in MPLS labels.

See figure 2 below for an illustration of how the various QoS topics relate to each other.

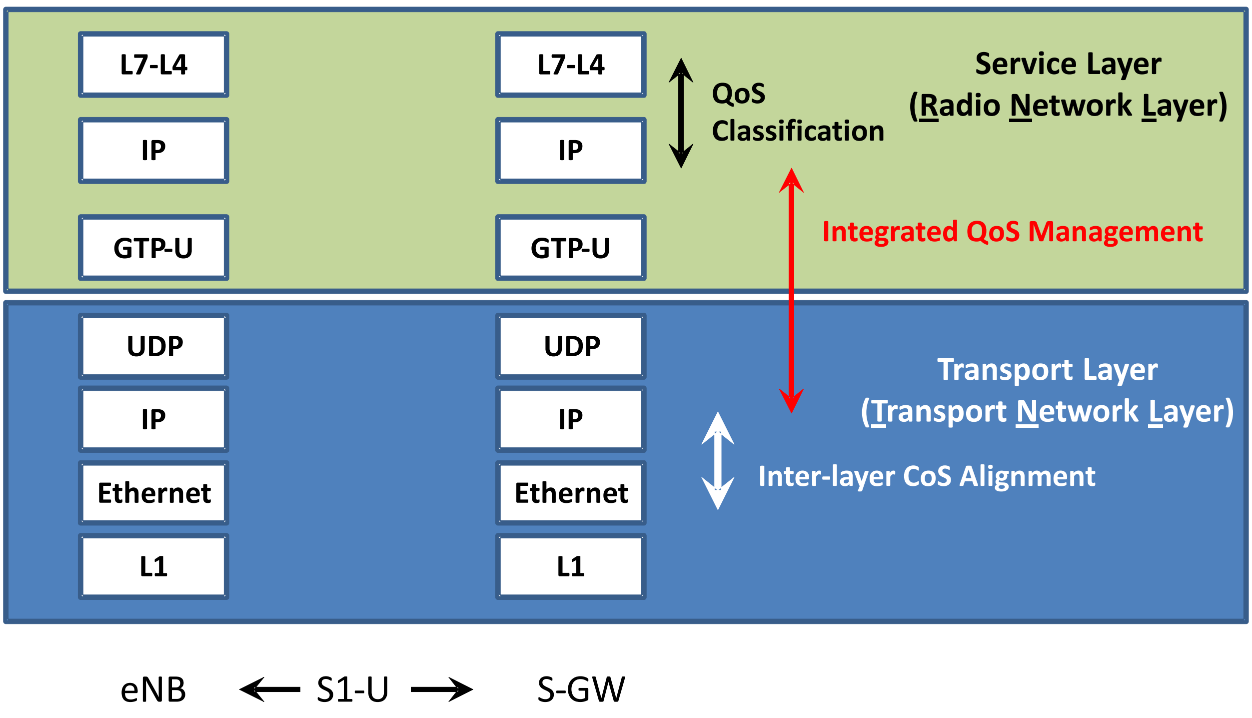

Figure 2: Integrated QoS Management components and scope

2 SERVICE FLOW CLASSIFICATION

2.1 Objective & Challenges

A template is proposed in this document to collect the relevant service flow characteristics and their network service requirements. The template is a tool to help operators to:

- + Group together the service flows whose network service tolerance values are within the same "range".
- + Determine the "ideal" number of classes of service to deploy in the network by adjusting the value of these "ranges".

The mechanism described seems simple and obvious, but the key lies in the choice of which relevant characteristics to prioritize, as not all flow characteristics may line up. For example, certain service flows might be grouped together based on their delay variation tolerances, but might not have been grouped together if their tolerance to loss was the key criterion.

However, if every packet ended up being marked as "highest priority", the QoS mechanism would collapse to a single level, namely; "Best Effort". The art of balancing these service flow priorities is what this *implicit Integrated QoS Management* section is all about. The key is to obtain the true service flow characteristics and to group the service flows intelligently into a limited number of service classes, allowing the backhaul network to deliver the best "compromise" in case of congestion.

In order to do this, the service flow QoS classification template is a start. This template will allow taking all the relevant characteristics into account. For example, it is not enough to simply look at the raw bandwidth requirements or at the delay/delay variation tolerance alone. The weight attributed to each of those relevant characteristics is up to operator adjustment – taking into account their commercial/marketing goals and the particular state of their network.

2.2 Service Flow Characteristics and Network Impairment Tolerances

Relevant or useful **characteristics** for our classification are those that:

- ➢ Give a high-level view of a service flow like for example, packet rate, direction, burstiness or dependencies on other service flows belonging to the same application session
- ➢ Can be used to classify a service flow into a category with similar types of service flows with regards to a certain network behaviour or tolerance to network impairments

Similarly, the most relevant or useful "**requirements**" (i.e. level tolerance to network impairments) will be those that are:

- ➢ Critical in order to respect the network behaviour that was used as reference for the service's design (e.g. delay, delay variation)
- ➢ Critical in order to stay within the targeted range of end-user quality of experience
- ➢ Used to prioritize the service flows relative to each other, for instance with regards to the network scheduling and queuing
- ➢ Used in conjunction with the high-level characteristics to map into service classes

In the template, the following service flow characteristics have been selected as the most relevant for our purposes:

- ➢ Packet Rate
- ➢ Elasticity
    - o The capability of a service flow to adapt its rate when congestion is encountered

- Direction
- Interdependencies
  - E.g., the association between signalling flows and application data service flows

And also the following service flow tolerances have been identified as the most relevant:

- Packet (error) Loss
- Packet Delay (end-to-end, one way)
- Packet Delay Variance (end-to-end, one way)

2.3 NGMN Service Flow QoS Classification Template

The following table shows the proposed Service Flow Template that could be used to collect the relevant service flow characteristics and the relevant tolerances to packet network impairments.

The values for the service flow characteristics and tolerances shown in the template below are illustrative and are currently being used as guidelines by the various standardisation organisations.

These values are from an end-to-end (one way) perspective and are specified only based on application / service level requirements: from "ear-to-mouth" in case of voice services or from server to client application for data applications. In case the scope is different from those mentioned here, the scope of the characteristics will be specified in the relevant "standardization organization" section.

Ideally, the values should be derived by general agreement/consensus, through practical experience/feedback of the telecommunications carriers' community and with input from application vendor/developer recommendations.

The sources for the values used in the example below come from the following organisations:

+ ITU-T: G.114, G.1010, H.264
+ 3GPP: TS 22.105, TS 23.107, TS 23.203
+ IETF: RFC 4594, RFC 5127
+ BBF: TR-126, TR-221
+ MEF : MEF23.1

Note 1) some of the recommendations for these values have been inherited from one organization to the other and sometimes are based on dated application requirements, especially for the video service flows.

Note 2) as stated previously in this chapter, the values shown in the template below are for the end-to-end application level requirements, the values for the delay and delay variation metrics for the mobile backhaul network should of course be less. According to chapter 3.2.1 "3GPP Classification", in table 6 which shows the "Standardized QCI characteristics" from 3GPP TS 23.203 version 8.6.0 – Table 6.1.7 (note 1), 3GPP suggests that 20ms can be considered a valid representative value for the delay in a mobile backhaul network between the PCEF (Policy & Charging Enforcement Function) and the base station.

Table 1: NGMN - Service Flow Template (example)

SERVICE FLOW (SF)				SF CHARACTERISTICS				SF TOLERANCE			SF CLASSIFICATION	
PLANE	*TYPE / CAT.*	*APPLICATION*	*NAME*	*DATA RATE* (kb/s)	*RATE ADAPTIVE* (Yes/No)	*DIRECTION* (Up/Down/Sym.)	*INTER-DEPENDENCY*	*(Error) LOSS* (10^{-x})	*(one-way) DELAY* (ms)	*DELAY VARIANCE* (ms or L/M/H)	*4 CoS Model*	*RATIONALE*
Data	Audio	VoIP		4 - 64	No	Sym.	Signalling	10^{-2}	150 - 400	< 1	C1	Voice
		2G CS			No	Sym.	Signalling	10^{-2}	100	Moderate	C1	
		3G CS			No	Sym.	Signalling	10^{-2}	100	Moderate	C1	
		Music Streaming		64 - 320	No	Sym.	Signalling		300	Low/Moderate		
	Video	Conversational Live Streaming	Fixed-Rate Conferencing		No	Sym.	Signalling	10^{-3}	100	Low		
		Non-Conversational Live Streaming	Broadcast		No	Down	Signalling	10^{-3}	100	Low		
			Rate-Adaptive		Yes	Down	Signalling	10^{-3}	150	Low		
		Buffered Streaming	Fixed-Rate		No	Down	Signalling	10^{-6}	300	Moderate		
			Rate-Adaptive		Yes	Down	Signalling	10^{-6}	300	Moderate		
	Data	High Throughput			Yes	Sym.		10^{-6}		Moderate		
		Low Latency			Yes	Sym.		10^{-6}		Low		
		Low Priority			Yes	Sym.		10^{-6}		High		
		Standard			Yes	Sym.		10^{-6}		High		
	Interactive	Web Browsing			Yes	Sym.		10^{-6}	300	Moderate		
		Real-Time Gaming			No	Sym.		10^{-3}	50	Very Low	C1	Real-Time / Marketing
Control	Signalling	RAN	GTP-C, S1-MME, etc.		No	Sym.				Moderate	C1	CP
		Router network	RSVP-TE, etc.		No	Sym.		10^{-6}	100	Moderate	C1	
	Routing	Router network	OSPF, BGP, etc.		No	Sym.				Moderate	C1	
	Synch.	Frequency	PTPv2, NTP		No	Down/Sym.				Very Low	C1	
		Phase/ Time of Day	PTPv2		No	Down/Sym.				Very Low	C1	
Management	OAM/P	RAN / Router network			No	Sym.				Moderate		
	Remote Access		SSH/Telnet, etc.		No	Sym.				Moderate		

3 CLASSIFICATION RECOMMENDATIONS FROM STANDARDISATION ORGANIZATIONS

The following chapters give an overview of the work done by the standardisation organisations and forums on the topic of QoS and the related CoS. The following organisations have been included in this overview for their relevance in the packet based network environment and to limit the scope to the most used or known QoS recommendations and practises:

+ GSMA (Global System for Mobile communications Association)
+ 3GPP (Third Generation Partnership Project)
+ IETF (Internet Engineering Task Force)
+ MEF (Metro Ethernet Forum)

3.1 GSM Association (GSMA)

The GSM Association deals with the service layer for the context of this document. The GSMA has published a recommendation used by mobile operators to interconnect their networks for roaming traffic. The recommendations we look at in this overview comes from document: IR.34 version 6.1: “Inter-Service Provider IP Backbone Guidelines”.

3.1.1 GSMA Classification

The GSMA uses the IETF’s DiffServ per hop behaviour (PHB) as a reference for classifying applications into classes of service. As can be seen in the table below an inter-layer CoS alignment or mapping is proposed between the GERAN service layer and the transport layer. In this case the transport layer is the IP layer (L3) of the client layer application.

Table 2: “Application mapping into DSCP values” from GSMA IR34 version 6.1 – table 7

APPLICATION	DiffServ PHB	SERVICE CLASS
Video Share	EF (Expedited Forwarding)	Conversational
VoIP	EF	Conversational
Push-to-Talk	AF4 (Assured Forwarding)	Streaming
Video Streaming	AF4	Streaming
Unrecognized GTP Traffic	AF3	Interactive
DNS	AF3	Interactive
Online Gaming	AF3	Interactive
Browsing	AF2	Interactive
Instant Messaging	AF1	Interactive
Remote Connection	AF1	Interactive
Email, MMS	BE (Best Effort)	Background

3.1.2 GSMA Marking

GSMA uses four traffic classes similar to 3GPP. The four traffic classes share six classes of service using IETF’s recommended DiffServ code points. Of note is that data services requiring a network service response better than best effort, due to more stringent delay requirements are classified into the Interactive traffic class and are differentiated using three PHB service classes, namely; AF31/21/11.

Table 3: "Traffic classes and their mapping to DSCP values" from GSMA IR34 version 6.1 – table 6

QoS INFORMATION		PHB	DSCP
TRAFFIC CLASS	THP (Traffic Handling Priority)		
Conversational	N/A	EF	101110 (46)
Streaming	N/A	AF41	100010 (34)
Interactive	1	AF31	011010 (26)
	2	AF21	010010 (18)
	3	AF11	001010 (10)
Background	N/A	BE	000000 (0)

3.2 Third Generation Partnership Project (3GPP)

In the context of this document, 3GPP deals with the service layer. In the chapters below we look at the service flow classification and CoS markings proposed by 3GPP.

3.2.1 3GPP Classification

At a high level the service flows belonging to applications are grouped into four classes of service, called traffic classes. The relevant characteristics or fundamental characteristics as called by 3GPP are; time relation, delay, delay variation and loss.

The table below shows the classification for UMTS.

Table 4: "UMTS QoS classes" from 3GPP TS 23.107 version 6.3.0 – Table 1

TRAFFIC CLASS	CONVERSATIONAL CLASS	STREAMING CLASS	INTERACTIVE CLASS	BACKGROUND
	Conversational RT	*Streaming RT*	*Interactive BE*	*Background BE*
Fundamental characteristics	- Preserve time relation (variation) between information entities of the stream - Conversational pattern (stringent and low delay)	- Preserve time relation (variation) between information entities of the stream	- Request/response pattern - Preserve payload content	- Destination is not expecting the data within a certain time - Preserve payload content
Application examples	- Voice	- Streaming video	- Web browsing	- Background download of emails

3GPP also details performance objectives for the four traffic classes. In addition to the "typical" relevant characteristics, some specific 3GPP CoS indicators are introduced; Traffic Handling Priority and Allocation/Retention Priority.

Table 5: “UMTS bearer attributes defined for each bearer traffic class” from 3GPP TS 23.107 version 6.3.0 – Table 2

TRAFFIC CLASS	CONVERSATIONAL CLASS	STREAMING CLASS	INTERACTIVE CLASS	BACKGROUND CLASS
Maximum Bit Rate	X	X	X	X
Delivery Order	X	X	X	X
Max SDU Size	X	X	X	X
SDU Format Information	X	X		
SDU Error Ratio	X	X	X	X
Residual Bit Error Ratio	X	X	X	X
Delivery of Erroneous SDUs	X	X	X	X
Transfer Delay	X	X		
Guaranteed Bit Rate	X	X		
Traffic Handling Priority			X	
Allocation/Retention Priority	X	X	X	X
Source Statistics Descriptor	X	X		
Signalling Indication			X	
Evolved Allocation/Retention Priority	X	X	X	X

3GPP TS 23.203 deals with policy and charging control architecture for both 3GPP access networks (GERAN, UTRAN/E-UTRAN) and non-3GPP access networks. The table below shows the service flow classification for LTE.

Table 6: “Standardized QCI characteristics” from 3GPP TS 23.203 version 8.6.0 – Table 6.1.7

QCI	RESOURCE TYPE	PRIORITY	PACKET DELAY BUDGET (ms) (Note 1)	PACKET ERROR LOSS RATE (Note 2)	EXAMPLE SERVICES
1 (Note 3)	GBR	2	100	10^{-2}	Conversational Voice
2 (Note 3)		4	150	10^{-3}	Conversational Video (live streaming)
3 (Note 3)		3	50	10^{-3}	Real-Time Gaming
4 (Note 3)		5	300	10^{-6}	Non-Conversational Video (buffered streaming)
5 (Note 3)	Non-GBR	1	100	10^{-6}	IMS Signalling
6 (Note 4)		6	300	10^{-6}	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)
7 (Note 3)		7	100	10^{-3}	Voice, Video (live streaming), Interactive Gaming
8 (Note 5)		8	300	10^{-6}	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)
9 (Note 6)		9	300	10^{-6}	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)

NOTES	
(1)	A delay of 20 ms for the delay between a PCEF and a radio base station should be subtracted from a given PDB to derive the packet delay budget that applies to the radio interface. This delay is the average between the case where the PCEF is located "close" to the radio base station (roughly 10 ms) and the case where the PCEF is located "far" from the radio base station, e.g. in case of roaming with home routed traffic (the one-way packet delay between Europe and the US west coast is roughly 50 ms). The average takes into account that roaming is a less typical scenario. It is expected that subtracting this average delay of 20 ms from a given PDB will lead to desired end-to-end performance in most typical cases. Also, note that the PDB defines an upper bound. Actual packet delays - in particular for GBR traffic - should typically be lower than the PDB specified for a QCI as long as the UE has sufficient radio channel quality.
(2)	The rate of non congestion related packet losses that may occur between a radio base station and a PCEF should be regarded to be negligible. A PELR value specified for a standardized QCI therefore applies completely to the radio interface between a UE and radio base station.
(3)	This QCI is typically associated with an operator controlled service, i.e., a service where the SDF aggregate's uplink / downlink packet filters are known at the point in time when the SDF aggregate is authorized. In case of E-UTRAN this is the point in time when a corresponding dedicated EPS bearer is established / modified.
(4)	If the network supports Multimedia Priority Services (MPS) then this QCI could be used for the prioritization of non-real-time data (i.e. most typically TCP-based services/applications) of MPS subscribers.

(5)	This QCI could be used for a dedicated "premium bearer" (e.g. associated with premium content) for any subscriber / subscriber group. Also in this case, the SDF aggregate's uplink / downlink packet filters are known at the point in time when the SDF aggregate is authorized. Alternatively, this QCI could be used for the default bearer of a UE/PDN for "premium subscribers".
(6)	This QCI is typically used for the default bearer of a UE/PDN for non-privileged subscribers. Note that AMBR can be used as a "tool" to provide subscriber differentiation between subscriber groups connected to the same PDN with the same QCI on the default bearer.

The QCI characteristics describe the packet forwarding treatment that a service flow aggregate receives end-to-end (one way) between the UE (User Equipment) and the PCEF (Policy and Charging Enforcement Function) and not between the application layer source and sink endpoints. Figure 3 below shows the scope of the QCI characteristics graphically:

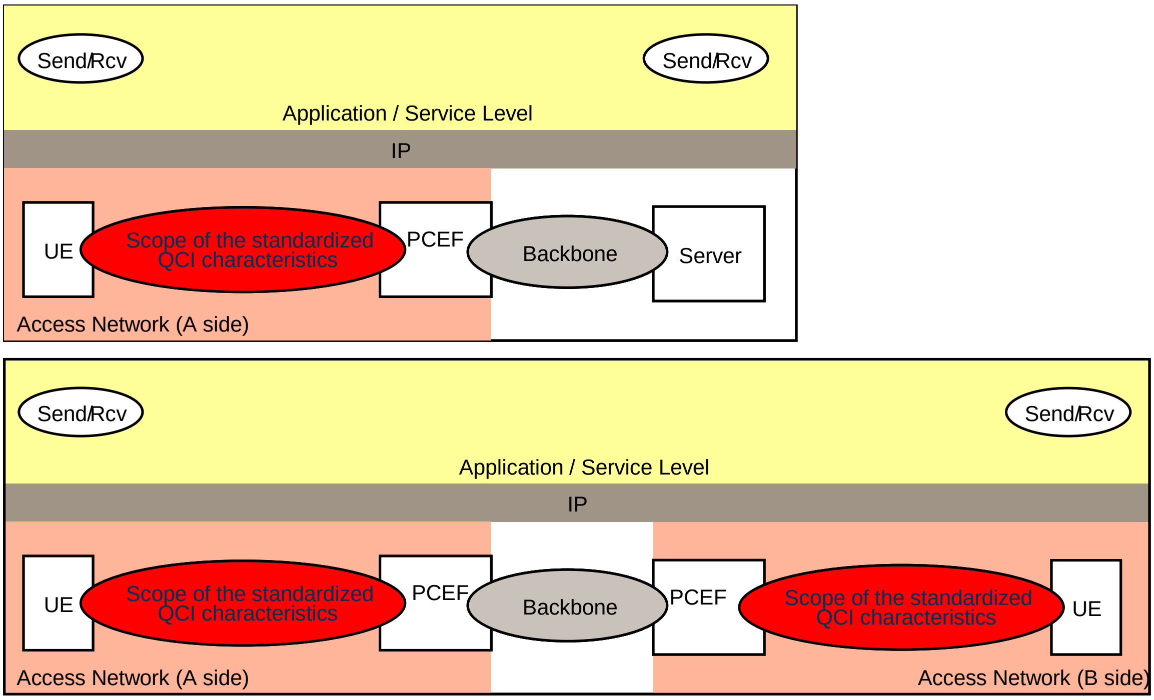

Figure 3: "Scope of the Standardized QCI characteristics for client/server (upper figure) and peer/peer (lower figure) communication" from 3GPP TS 23.203 version 8.6.0 – Figure 6.1.7-1

3.2.2 3GPP TS 23.207 End-to-End QoS Concept and Architecture

The overall scope of the QoS architecture work in 3GPP is of course the service layer (xRAN and mobile core) and the underlying mobile backhaul transport layer is not taken into account. As can be seen in the table below the underlying transport layer of the backhaul network is out of scope for 3GPP.

In 3GPP architectures, the backhaul network is assumed to be one-on-one aligned with the service layer QoS architecture and its associated number of classes of service. The backhaul network is modelled as a point-to-point connection without bottlenecks or resource contention.

As stated in the beginning of this document the reality of the backhaul network is such that bottlenecks in the aggregation network are an integral characteristic of packet networks. This is necessary in order to achieve statistical multiplexing gain, taking advantage of the variable service flow emission times and variable packet sizes.

Hence the reason to address the mobile backhaul network QoS in this NGMN P-BEV, Integrated QoS Management project.

The alignment of service flow QoS schemes requires special attention in order to assure that high priority service flows are protected in case of congestion.

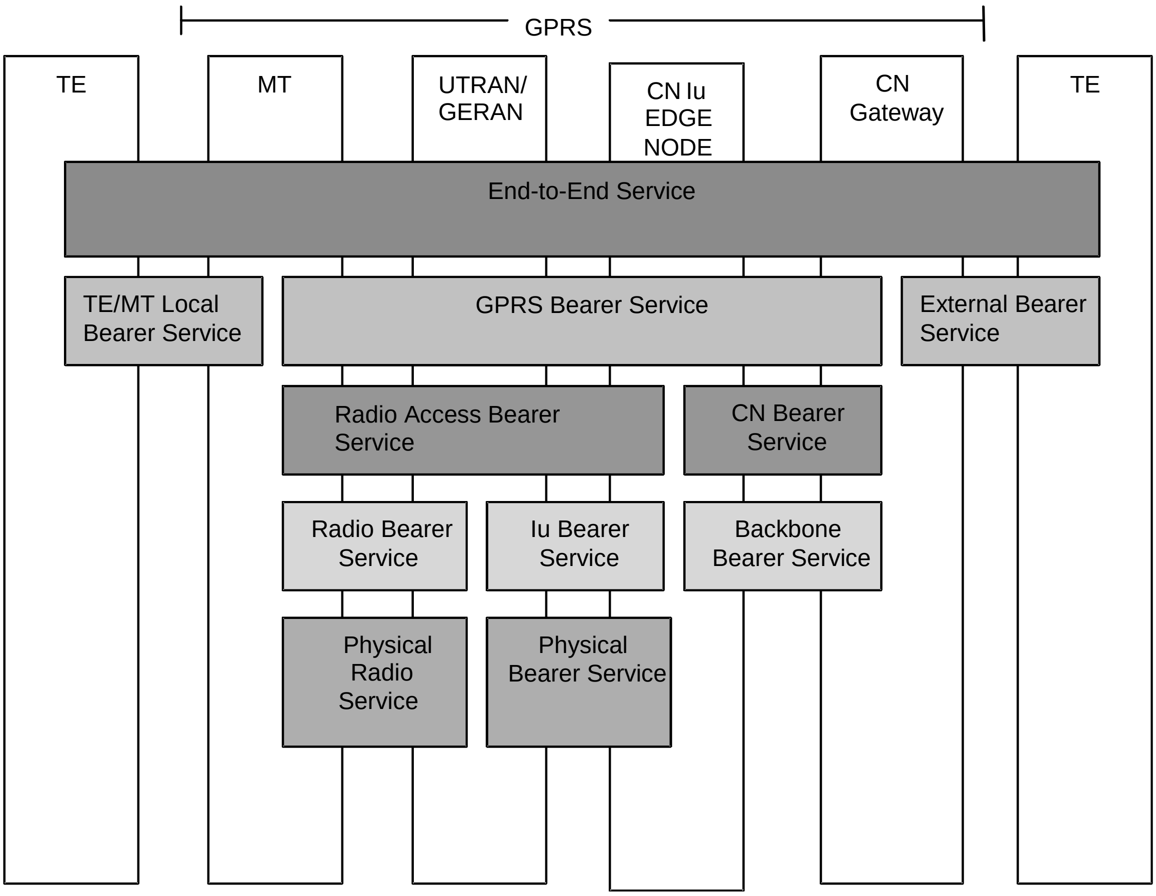

Figure 4: "UMTS QoS Architecture" from 3GPP TS 23.207 version 10.0.0 – Figure 1

3.3 Internet Engineering Task Force (IETF)

The IETF deals with L3 or network layer part of the "transport" layer network in the context of this document. IETF has published multiple Requests for Comments (RFC) on QoS topics. The differentiated service (DiffServ) concept is a well-known priority marking scheme used in IP packet networks and based on a set of forwarding behaviours called Per Hop Behaviour (PHB) who in turn are organized into a PHB Scheduling Classes (PSC) framework and are associated with relative priority markings, called DiffServ Code Points (DSCP), carried in the IP header.

Below the class of service marking, formerly known as TOS field, of the IPv4 header. DiffServ has reorganized the TOS field into a DSCP field and an ECN field.

0	1	2	3	4	5	6	7
DSCP						ECN	

DSCP	Differentiated Services Code Point
ECN	Explicit Congestion Notification

Figure 5: DiffServ Code Point (DSCP) Field in IPv4 Header

3.3.1 IETF Classification

3.3.1.1 RFC4594

RFC 4594 "Configuration Guidelines for DiffServ Service Classes" (august 2006) is an informational request for comments which provides information concerning service flow classification and mapping into classes of service or forwarding behaviours and can be seen as a cornerstone for IP QoS mechanisms.

Service flows are first categorized into four major application categories. These categories are then mapped against twelve different classes of service. The relevant service flow characteristics used by IETF are elasticity, rate adaptation, emission rates, flow duration and the well-known trio of: loss, delay and delay variation (jitter). Most of these relevant characteristics have been adopted in our NGMN service flow template as well (see chapter 2.3).

The table below shows the classification of service flows and their mapping into classes of service.

Table 7: "User/Subscriber Service Classes Grouping" from RFC 4594 – Figure 1

APPLICATION CATEGORIES	SERVICE CLASS	SIGNALED	FLOW BEHAVIOR	G.1010 RATING
Application Control	Signalling	N/A	Inelastic	Responsive
Media-Oriented	Telephony	Yes	Inelastic	Interactive
	Real-Time Interactive	Yes	Inelastic	Interactive
	Multimedia Conferencing	Yes	Rate Adaptive	Interactive
	Broadcast Video	Yes	Inelastic	Responsive
	Multimedia Streaming	Yes	Elastic	Timely
Data	Low-Latency Data	No	Elastic	Responsive
	High-Throughput Data	No	Elastic	Timely
	Low-Priority Data	No	Elastic	Non-Critical
Best Effort	Standard	Not Specified	Not Specified	Non-Critical

Table 8: “Service Class Characteristics” from RFC 4594 – Figure 2

SERVICE CLASS NAME	TRAFFIC CHARACTERISTICS	TOLERANCE to		
		LOSS	DELAY	JITTER
Network Control	Variable size packets, mostly inelastic short messages, but traffic can also burst (BGP)	Low	Low	Yes
Telephony	Fixed-size small packets, constant emission rate, inelastic and low-rate flows	Very Low	Very Low	Very Low
Signalling	Variable size packets, somewhat bursty short-lived flows	Low	Low	Yes
Multimedia Conferencing	Variable size packets, constant transmit interval, rate adaptive, reacts to loss	Low-Medium	Very Low	Low
Real-Time Interactive	RTP/UDP streams, inelastic, mostly variable rate	Low	Very Low	Low
Multimedia Streaming	Variable size packets, elastic with variable rate	Low-Medium	Medium	Yes
Broadcast Video	Constant and variable rate, inelastic, non-bursty flows	Very Low	Medium	Low
Low-Latency Data	Variable rate, bursty short-lived elastic flows	Low	Low-Medium	Yes
OAM	Variable size packets, elastic & inelastic flows	Low	Medium	Yes
High-Throughput Data	Variable rate, bursty long-lived elastic flows	Low	Medium-High	Yes
Standard	A bit of everything	Not specified		
Low-Priority Data	Non-real-time and elastic	High	High	Yes

Secondly, these CoS are mapped in turn to a set of Per Hop Behaviours (PHB) with the corresponding DCSP values. These PHBs are in essence representing the underlying QoS mechanism or packet forwarding treatment within a node. In IETF “jargon” the PHB are indicative of the classes of service rather than the twelve CoS that are mapped to them.

The table below shows the mapping of the twelve CoS to the PHB and DSCP with some application examples.

Table 9: "DSCP to Service Class Mapping" from RFC 4594 – Figure 3

SERVICE CLASS NAME	PHB	DSCP	APPLICATION EXAMPLES
Network Control	CS6	110000 (48)	Network Routing
Telephony	EF	101110 (46)	IP Telephony Bearer
Signalling	CS5	101000 (40)	IP Telephony Signalling
Multimedia Conferencing	AF41	100010 (34)	H.323/V2 Video Conferencing (adaptive)
	AF42	100100 (36)	
	AF43	100110 (38)	
Real-Time Interactive	CS4	100000 (32)	Video Conferencing, Interactive Gaming
Multimedia Streaming	AF31	011010 (26)	Streaming Video and Audio on demand
	AF32	011100 (28)	
	AF33	011110 (30)	
Broadcast Video	CS3	011000 (24)	Broadcast TV and Live Events
Low-Latency Data	AF21	010010 (18)	Client-Server Transactions, Web-based Ordering
	AF22	010100 (20)	
	AF23	010110 (22)	
OAM	CS2	010000 (16)	OAM&P
High-Throughput Data	AF11	001010 (10)	Store-and-Forward Applications
	AF12	001100 (12)	
	AF13	001110 (14)	
Standard	DF (CS0)	000000 (0)	Undifferentiated Applications
Low-Priority Data	CS1	001000 (8)	Any flow that has no BW assurance

The IETF also details the PHB further with packet conditioning recommendations. The table below is shown for indicative purposes. The packet conditioning and active queue management implementations are not in scope for this document.

Table 10: "Summary of QoS Mechanisms Used for Each Service Class" from RFC 4594 – extract of Figure 4

SERVICE CLASS NAME	PHB	CONDITIONING	QUEUING	ACTIVE QUEUE MANAGEMENT
Network Control	CS6	See RFC 2474 section 3.1	Rate	Yes
Telephony	EF	Police: single rate (sr) + burst size (bs)	Priority	No
Signalling	CS5	Police: sr + bs	Rate	No
Multimedia Conferencing	AF41	two-rate Three Colour Marker (trTCM)	Rate	Yes
	AF42	trTCM		Yes
	AF43	trTCM		Yes
Real-Time Interactive	CS4	Police: sr + bs	Rate	No
Multimedia Streaming	AF31	trTCM	Rate	Yes
	AF32	trTCM		Yes
	AF33	trTCM		Yes
Broadcast Video	CS3	Police: sr + bs	Rate	No
Low-Latency Data	AF21	trTCM	Rate	Yes
	AF22	trTCM		Yes
	AF23	trTCM		Yes
OAM	CS2	Police: sr + bs	Rate	Yes
High-Throughput Data	AF11	trTCM	Rate	Yes
	AF12	trTCM		Yes
	AF13	trTCM		Yes
Standard	DF (CS0)	N/A	Rate	Yes
Low-Priority Data	CS1	N/A	Rate	Yes

3.3.1.2 RFC 5127

RFC 5127 "Aggregation of DiffServ Classes" (2008) is an informational RFC which proposes to aggregate or map the twelve DiffServ classes of service into four aggregate classes of service.

Below the tables extracted from this RFC.

Table 11: "Treatment Aggregate and Service Class Performance Requirements" from RFC 5127 – Figure 1

SERVICE CLASS NAME	TOLERANCE to LOSS	TOLERANCE to DELAY	TOLERANCE to JITTER	PHB	4 Classes of Service AGGREGATE	TOLERANCE to LOSS	TOLERANCE to DELAY	TOLERANCE to JITTER
Network Control	Low	Low	Yes	CS6	Network Control	Low	Low	Yes
Telephony	Very Low	Very Low	Very Low	EF	Real-Time	Very Low	Very Low	Very Low
Signalling	Low	Low	Yes	CS5				
Multimedia Conferencing	Low-Medium	Very Low	Low	AF4x				
Real-Time Interactive	Low	Very Low	Low	CS4				
Broadcast Video	Very Low	Medium	Low	CS3				
Multimedia Streaming	Low-Medium	Medium	Yes	AF3x	Assured Elastic	Low	Low-Medium	Yes
Low-Latency Data	Low	Low-Medium	Yes	AF2x				
OAM	Low	Medium	Yes	CS2				
High-Throughput Data	Low	Medium-High	Yes	AF1x				
Standard	Not specified			DF	Elastic	Not specified		
Low-Priority Data	High	High	Yes	CS1				

The table below also introduces a recommendation for the inter-layer CoS alignment or mapping to the MPLS TC field (formerly known as EXP bits) of the MPLS label transporting the IP packet.

Table 12: "Treatment Aggregate and MPLS EXP Field Usage" from RFC 5127 – extract from Figure 3

4 Classes of Service AGGREGATE	SERVICE CLASS NAME	PHB	DSCP	MPLS TC
Network Control	Network Control	CS6	110000	110
Real-Time	Telephony	EF	101110 (46)	100
	Signalling	CS5	101000 (40)	
	Multimedia Conferencing	AF41	100010 (34)	
		AF42	100100 (36)	
		AF43	100110 (38)	
	Real-Time Interactive	CS4	100000 (32)	
	Broadcast Video	CS3	011000 (24)	
Assured Elastic	Multimedia Streaming	AF31	011010 (26)	010 (2)
		AF32	011100 (28)	011 (3)
		AF33	011110 (30)	
	Low-Latency Data	AF21	010010 (18)	010 (2)
		AF22	010100 (20)	011 (3)
		AF23	010110 (22)	
	OAM	CS2	010000 (16)	010 (2)
	High-Throughput Data	AF11	001010 (10)	010 (2)
		AF12	001100 (12)	011 (3)
		AF13	001110 (14)	
Elastic	Standard	DF	000000 (0)	000 (0)
	Low-Priority Data	CS1	001000 (8)	001 (1)

3.4 Metro Ethernet Forum (MEF)

The Metro Ethernet Forum deals with the Ethernet layer or the “Layer 2 part of the transport layer” in the context of this document. Various Ethernet services characteristics are defined including CoS requirements for various applications.

The following overview is from the Implementation Agreement (IA) MEF 23.1 “Carrier Ethernet Class of Service – Phase 2”. The main scope of MEF is to drive end-to-end consistency across multiple Carrier Ethernet service providers with regards to business service characteristics and their associated SLAs and QoS requirements.

MEF 23 (Phase 1) defines a set of 3 classes of service, called “CoS Labels” for UNI-to-UNI (EVC) services for both single MEN (Metro Ethernet Network) and multiple interconnected MENs. The CoS Identification (marking) and Colour Identification (drop eligibility) in MEF23.1 are applicable at External Interfaces (EIs), which can be either UNI or ENNI.

In MEF 23.1 (Phase 2), values for CoS Performance Objectives (CPOs), grouped in Performance Tier sets, has Performance Parameters and Performance metrics. The CPOs are applicable to CoS Frame Sets between the EIs as can be seen in Figure 6 below;

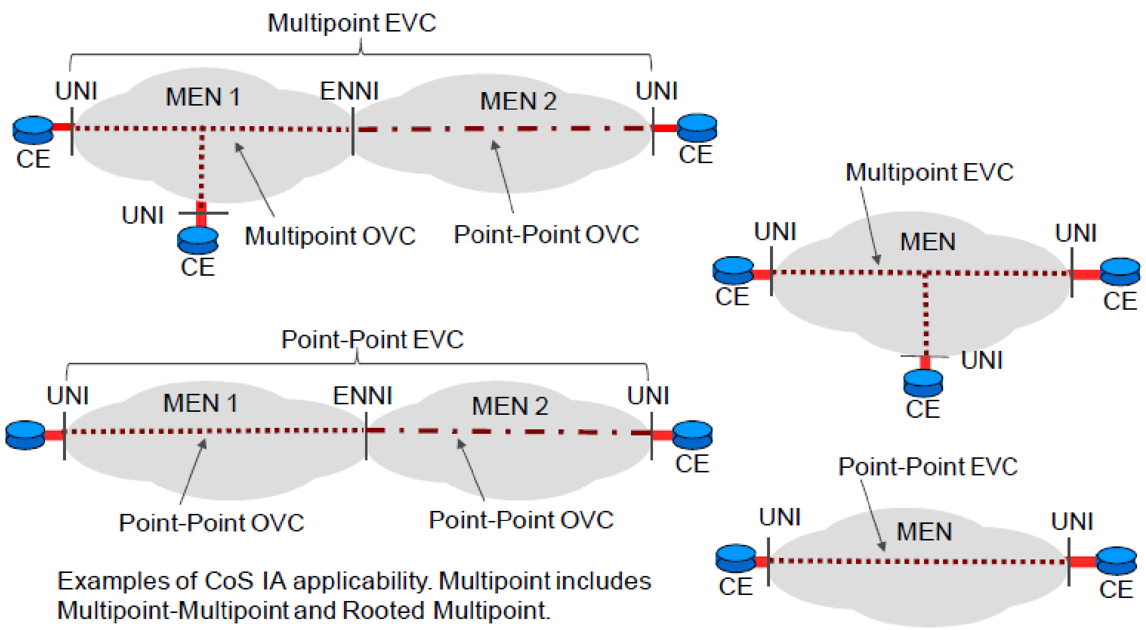

Figure 6: MEF23.1 – Applicability of CoS Frame Sets

The values for the CPOs for the classes of services in MEF23.1 are derived from publicly available documents from the IEEE, ITU and IETF standardization organizations for consistency.

The CoS Performance Objectives (CPOs) are: Frame Delay, Mean Frame Delay, Frame Delay Range, Inter-Frame Delay Variation and Frame Loss Ratio. Other CPOs have been introduced as placeholders for future phases of MEF23, namely; Availability, High Loss Interval and Consecutive High Loss Interval performance.

MEF 22.1 has a slightly different service flow classification or CoS mapping scheme. This latter Implementation Agreement adds the "H+" COS label that is for the exclusive use of synchronization traffic. Note that MEF 22.1 does not recommend that if packet synch is present there must be an H+ CoS Name added. H+ is recommended for packet based synch under certain conditions in 22.1. Per 22.1: "One issue that could influence the suitable number of Mobile Backhaul CoS Names is the presence of some traffic classes, such as packet-based synchronization traffic. For example, if the RAN BS oscillator is stable and of high-quality then performance requirements for the CoS Name can be less stringent compared to when using a lower quality oscillator. A set of CoS Names, such as one limited to the CoS Labels (H,M,L) and associated CPOs, is most clearly applicable if synchronization is achieved either using a non-packet based method (such as GPS, SyncE, or TDM); or using a packet based method augmented by a stable high quality oscillator at the RAN BS.

3.4.1 MEF 23.1 Classification

The following service flow categories or "applications" are used as input and classified into three classes of service or "CoS Labels" as they are called in MEF.

Table 13: "Explicit Application Mapping for Derivation of CPOs" from MEF 23.1 – Table 36

APPLICATIONS	COS LABEL
VoIP	H
VoIP & Video Signalling	M
Video Conferencing	M
Streaming Media	L
Interactive Gaming	H / M
Transactional Data	L
Mobile Backhaul H	H
Mobile Backhaul M	M
Mobile Backhaul L	L

While MEF 23.1 does not identify service flows for MBH, MEF 22.1 does imply one. This implication is from comparing table 7 with appendix B in MEF 22.1.

3.4.2 MEF CoS Marking

There are three classes of service defined in MEF 23.1 called CoS Labels; "H", "M" and "L". Unofficially these can be read as High, Medium and low. In addition a drop eligibility identifier (DEI) scheme is used using two colours; "green" and "yellow".

MEF proposes an inter-layer CoS alignment between Ethernet and IP using a subset of DSCP values. The operators are free to use the remainder of the DSCP and CoS values in any desired way, e.g., for additional classes of service.

MEF 23.1 suggests different marking schemes:

+ the PCP values are used to mark color only (just as the DEI bit can indicate color) and the CoS is implicitly derived from the EVC's C-VLAN ID, see table 14;
+ the PCP values are used to indicate CoS and color, table 15;
+ the PCP values are used to indicate CoS and the color is indicated by the DEI bit in the VLAN tag, see table 15

Table 14: "CoS ID Values when CoS ID is only EVC or OVC EP" from MEF 23.1 – Table 3

COS LABEL	CoS ID Types	COS IDENTIFIERS[1]			
		C-TAG PCP		PHB (DSCP)	
		COLOUR GREEN	*COLOUR YELLOW*	*COLOUR GREEN*	*COLOUR YELLOW*
H	EVC or OVC EP[2]	5, 3 or 1	N/S in phase 2	EF (46), AF31 (26) or AF11 (10)	N/S in phase 2
M	EVC or OVC EP[2]	5, 3 or 1	2 or 0	EF (46), AF31 (26) or AF11 (10)	AF32 (28), AF33 (30), AF12 (12), AF13 (14) or DF (0)
L	EVC or OVC EP[2]	5, 3 or 1	2 or 0	EF (46), AF31 (26) or AF11 (10)	AF32 (28), AF33 (30), AF12 (12), AF13 (14) or DF (0)

[1] Specifies only the PCP or DSCP values to be used for Color ID when CoS ID is limited to EVC or OVC EP (End Point). EVC and OVC EP indication for CoS ID is not constrained by CoS IA.
[2] EVC or OVC EP CoS ID would be different to differentiate CoS Labels H, M and L for different CoS Frame Sets on a given EI (External Interface).

Table 15: "CoS Identifiers and Colour Identifiers" from MEF 23.1 – Table 4

COS LABEL	COS & COLOUR IDENTIFIERS[1]						
	C-TAG PCP		PHB (DSCP)		S-TAG PCP without DEI supported		S-TAG PCP with DEI supported
	COLOUR GREEN	*COLOUR YELLOW*	*COLOUR GREEN*	*COLOUR YELLOW*	*COLOUR GREEN*	*COLOUR YELLOW*	
H	5	N/S in phase 2	EF (46)	N/S in phase 2	5	N/S in phase 2	5
M	3	2	AF31 (26)	AF32 (28) AF33 (30)	3	2	3
L	1	0	AF11 (10)	AF12 (12) AF13 (14) DF (0)	1	0	1

[1] Full CoS Identifier includes EVC or OVC End Point. Table specifies only the PCP or DSCP values to be used with EVC or OVC End Point to specify a CoS ID. EVC and OVC End Point indication is not constrained by CoS IA.

Not all CoS labels need to be used by operators and a scheme is proposed for aggregating classes of service. Care needs to be taken when different network sections from different metro Ethernet providers are used in the overall end-to-end mobile backhaul network. Especially when different service flows which are classified into "H" and "L" separately in one network segment are aggregated together in the next.

The tables below show two different classification schemes for the various combinations of aggregated CoS Labels, reported in the informative part of MEF 23.1. One scheme takes routers', the other takes Ethernet bridges' operational "best practises" into account for dealing with "well-known" PCP values. Again care needs to be taken when combining different metro Ethernet network segments or network service providers when different schemes are used.

The table below, shows PCP values used as "best practises" in the IP networking community. Note the PCP value of 5 for "H".

Table 16: "Example PCP Mapping for Multi-CoS Label EVC Supporting Only Standard CoS Labels at UNI – "Router-Application-Friendly" Mapping" from MEF 23.1 – Table 39

MEF COS LABEL COMBINATION	PCP MAPPING PER COS LABEL – COLOUR BLIND MODE		
	H	M	L
H+M+L	5	2 - 4, 6, 7	0, 1
H+M	5	0 - 4, 6, 7	N/A
H+L	5	N/A	0 - 4, 6, 7
M+L	N/A	2 - 7	0, 1

The table below, shows PCP values used as "best practises" in the LAN or Ethernet networking community. Note the PCP values between 4 and 7 for "H".

Table 17: "Example PCP Mapping for Multi-CoS Label EVC Supporting Only Standard CoS Labels at UNI – "Bridging-Application-Friendly" Mapping" from MEF 23.1 – Table 40

MEF COS LABEL COMBINATION	PCP MAPPING PER COS LABEL – COLOUR BLIND MODE		
	H	M	L
H+M+L	4 - 7	2, 3	0, 1
H+M	4 - 7	0 - 3	N/A
H+L	4 - 7	N/A	0 - 3
M+L	N/A	2 - 7	0, 1

The table below shows the inter-layer CoS alignment, between the PCP values in the VLAN header and the DSCP value in the IP header, for the aggregated CoS Label schemes.

Table 18: "Example DSCP Mapping for Multi-CoS Label EVC Supporting Only Standard CoS Labels at UNI" from MEF 23.1 – Table 41

MEF COS LABEL COMBINATION	DSCP MAPPING PER COS LABEL – COLOUR BLIND MODE		
	H	M	L
H+M+L	40 - 47	16 – 39, 48 - 63	0 - 15
H+M	40 - 47	0 – 39, 48 - 63	N/A
H+L	40 - 47	N/A	0 – 39, 48 - 63
M+L	N/A	16 - 63	0 - 15

In MEF 23.1 there is only one normative scheme with other examples/cases of possible alternatives. The examples in tables 16, 17 and 18 are not in the "Normative" part of MEF 23.1 but in the "Informative" appendices.

Caution needs to be taken when designing a mobile backhaul network spanning multiple network segments and/or network service providers.

3.5 Broadband Forum

The Broadband Forum has issued a Technical Report, TR-221 “Technical Specifications for MPLS in Mobile Backhaul networks” issue 1 in October 2011.

This TR-221 defines the use of MPLS and its associated services for the access and aggregation networks providing mobile backhaul services to 2G, 3G and LTE service flows. The MPLS-based services defined in TR-221 are:

+ VPWS (Virtual Private Wire Services)
+ VPLS or H-VPLS (Hierarchical - Virtual Private LAN Service),
+ BGP L3 VPN (IP Virtual Private Network) and
+ IP routing over MPLS

As MPLS is used as an underlying convergence technology for the different TNL (Transport Network Layer) such as TDM, ATM and IP, CoS alignment between the mobile TNL service flows and the MPLS service is required.

The Broadband Forum is aligned with the Metro Ethernet Forum (MEF) with regards to the Ethernet connectivity services. The QoS requirements in TR-221 are defined as follows for MPLS-based L2 VPN service:

+ The PE must support ingress bandwidth profile based on MEF 10.2
+ The PE must support at least 4 CoS and associated service metrics (e.g., delay, delay variation and packet loss) as defined in MEF 22.1
+ The PE should support Connection Admission Control to guarantee sufficient bandwidth is available to support new connections conforming to all SLA metrics defined in MEF 22.1
+ The ingress PE must map the PCP (in the PRI field of the 802.1Q VLAN tag) into the TC field of the MPLS label stack
+ For support of PTP synchronization over Ethernet, the network must support the synchronization performance metrics defined in “Performance for Synchronization Traffic Class” of MEF 22.1

In addition, TR-221 assumes that QoS markings are mapped from higher layers to the lower layers.

The QoS requirements for the underlying MPLS LSP tunnels are defined as follows:

+ The PE and P routers MUST support E-LSP as per Section 1.2/RFC 3270: LSPs which can transport multiple Ordered Aggregates, so that the TC field of the MPLS Shim Header conveys to the LSR the PHB to be applied to the packet (covering both information about the packet's scheduling treatment and its drop precedence)
+ The PE and P routers MAY support L-LSP as per Section 1.3/RFC 3270: LSPs which only transport a single Ordered Aggregate, so that the packet's scheduling treatment is inferred by the LSR exclusively from the packet's label value while the packet's drop precedence is conveyed in the TC field of the MPLS Shim Header

The requirements for the mapping or alignment between the mobile TNL service flows (encapsulated in pseudowires) and the MPLS LSP tunnels are defined as follows:

+ The PE MUST support COS marking in the TC bits of the LSP labels
+ The PE MUST support COS mapping between the QoS of TNL and TC bits of the LSP labels

+ The PE MUST support the Pipe model as per RFC 3270
+ The PE SHOULD support mapping of TNL COS to PW label TC bits
+ For multi-segment PW, the PE MUST support mapping of TNL COS to PW label TC bits
+ The PE SHOULD support marking of the PW label TC bits
+ For multi-segment PW, the PE MUST support marking of the PW label TC bits

The Broadband Forum does not impose any specific service flow classification, however recommends the alignment of CoS from the TNL service flows to the underlying transport mechanisms in a transparent way.

The proposed classification and alignment/mapping by NGMN are fully supported by the MPLS-based recommendations of TR-221.

4 EXAMPLES OF CLASSIFICATION SCHEMES IN LIVE NETWORKS

In order to put all the received inputs into context, and also to make sure that our proposals were based on realistic usage examples, we gathered some classification schemes employed in the industry. In order to do this, we took advantage of the representative nature of the NGMN's membership to ask both operators and equipment vendors what schemes they are applying in their network, plan to use for LTE, or recommend their clients and affiliates. For this comparison, we didn't want to concentrate on the type of CoS marking used (L2 PCP or L3 DSCP or MPLS TC), but rather on the number of classes of service and their usage. Thus, we proposed all contributors to use the following template for this exercise, with abstract CoS names:

Table 19: NGMN – Class of Service Scheme

CLASS OF SERVICE	
C1	Highest
C2	
C3	
C4	
C5	
C6	
C7	
C8	Lowest

For each input, we allowed the use of as many as 8 different CoS, and took into account (marked in white) whether a service flow was not classified at all (or, simply, not present in the relevant network). We also used a simplified service flow template, which does not identify individual services, but instead relies on the service flows that can be manipulated at the backhaul edge, mostly RAN bearers and operational flows (e.g., data, control and management plane).

The results can be seen below, with no differentiation made between vendors and operators in order to enhance its anonymity:

Table 20: NGMN – Service Flow Classification Poll

GROUP	NAME	DESCRIPTION	CLASS OF SERVICE					
			INPUT 1	*INPUT 2*	*INPUT 3*	*INPUT 4*	*INPUT 5*	*INPUT 6*
2G	Voice	Conversational Voice	C3	C3	C3		C1	C2
	Data	Data	C8	C4	C4	C8	C1	C8
	C P	Control Plane	C2	C1	C2	C2	C1	C2
	M P	Management Plane	C2	C4	C2		C3	C8
3G	Rt-DCH	Real-time Data	C4	C3	C3	C1	C1	C2
	Nrt-DCH	Standard Data	C8	C3	C6	C8	C8	C3
	HSxPA	High Speed Data	C8	C5	C7		C8	C8
	C P	Control Plane	C2	C2	C1	C1	C1	C2
	M P	Management Plane	C2	C4	C4	C6	C2	C8
LTE	QCI 1	Conversational Voice	C3	C3	C3	C1	C1	C2
	QCI 2	Conversational Video (live streaming)	C5	C3	C5	C7	C2	C3
	QCI 3	Real-Time Gaming	C6	C3	C5	C3	C2	C2
	QCI 4	Non-Conversational Video (buffered streaming)	C4	C3	C4	C4	C4	C3
	QCI 5	IMS Signalling	C2	C2	C3	C2	C1	C2
	QCI 6	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)	C8	C4	C6	C6	C4	C3
	QCI 7	Voice, Video (live streaming), Interactive Gaming	C7	C6	C6	C6	C4	C8
	QCI 8	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)	C8	C6	C7	C5	C4	C8
	QCI 9	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)	C8	C6	C8	C8	C8	C8
	C P	Control Plane	C2	C2	C3	C1	C1	C2
	M P	Management Plane	C2	C4	C3	C6	C3	C8
Synch	PTPv2	IEEE1588v2	C1	C1	C1	C1	C1	C2
	NTP	NTP	C3	C1	C1	C6		C2
Router	M P	Management Plane	C2	C4	C8		C3	C1
	tx M P	Management Plane	C2	C4	C8		C3	C1
	C P	Control Plane	C2	C2	C3	C3	C3	C1

Even with a relatively limited sample, we can already extract some interesting conclusions:

There is a high variance of service flow classification and their priorities between operators, both in the number of CoS used (from 4 up to 8) and the relative priority of some service flows.
However, some service flows receive a very homogeneous treatment: QCI 9 has the lowest priority in all cases, and packet synchronization the highest.

The most stark contrast between answers is found in the Management Plane service flows, with answers ranging from C1 (highest priority) to C8. This, however, is possibly due to a different definition of "MP" being used in different organizations (for instance, inclusion of detection mechanisms for protection algorithms or not).

In a later section, we have built on this analysis to develop a common classification proposal, making sure that the relative priorities expressed here are respected as much as possible.

5 INTER-LAYER CLASS OF SERVICE ALIGNMENT

5.1 Introduction

The previous chapters discussed QoS classification, as it deals with the mapping of service flows to specific classes of service belonging to a specific service layer (RAN bearers) or transport layers (Ethernet, MPLS and IP).

This chapter discusses the mapping of the classes of service of different service and transport layers to each other; what we call “inter-layer CoS alignment”, such as between the IP and Ethernet layer for example.

Inter-layer CoS alignment ideally would need to take into account all aspects of QoS such as; classification, marking and active queue management (AQM) techniques such as; metering, policing, shaping and scheduling. This all in order to assure the service flows performance objectives are met end-to-end, irrespective of all the different node types (switches (L2), routers (L3), RAN (L4-7)), interacting with the service flows at different layers, that can be encountered in the backhaul network.

However, as stated previously, for this part of the Integrated QoS study only the service flow classification and the inter-layer CoS alignment are in scope.

5.2 Inter-layer Class of Service Alignment Challenge

To illustrate in a graphical manner the challenge of aligning the CoS of the RAN service layer transported by an IP layer running on top of an Ethernet layer, the following diagrams use some video service flow examples extracted from the classification tables seen in chapter three.

A picture paints a thousand words...

Immediately the different granularities of the CoS schemes become apparent in figure 7 below.

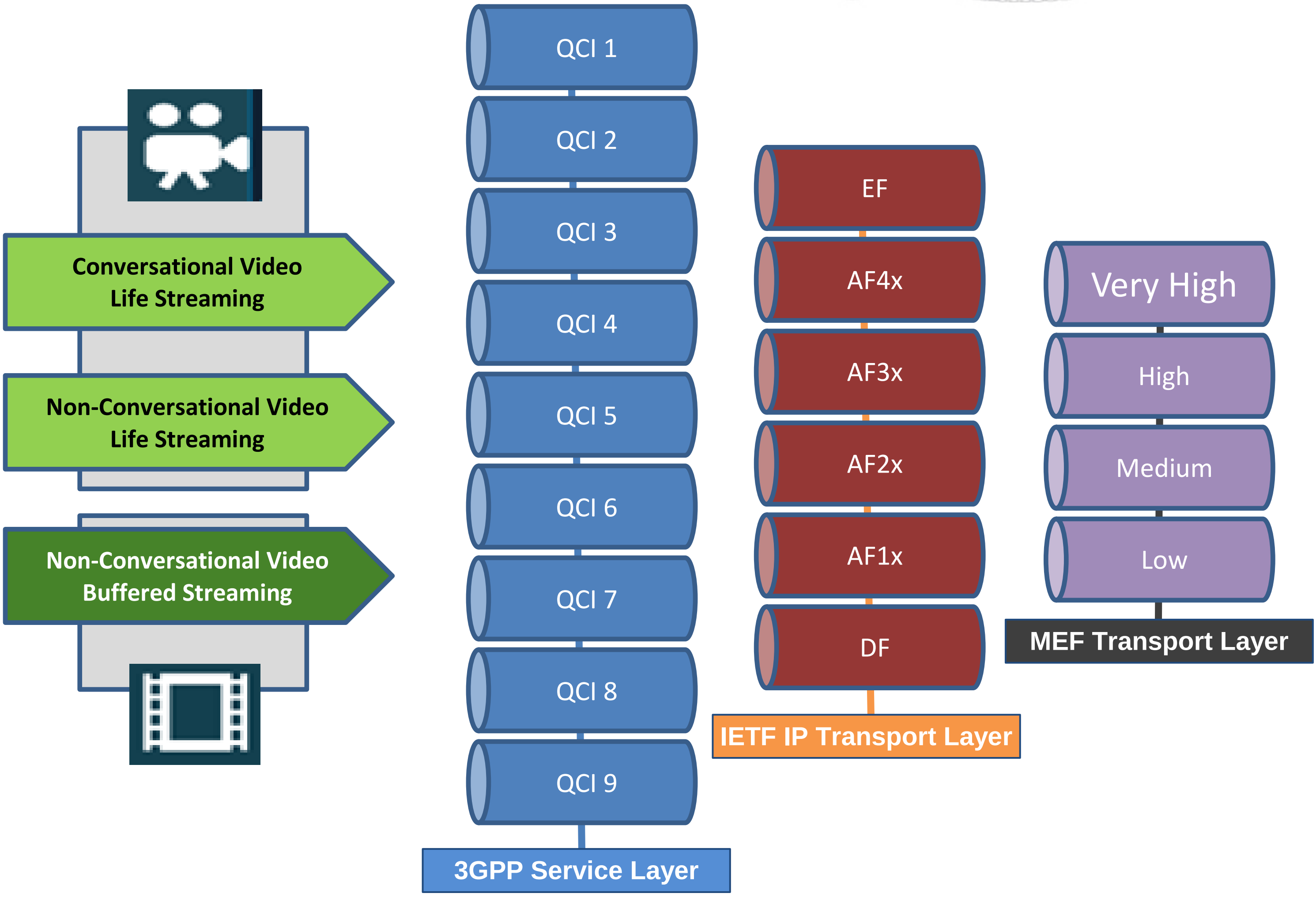

Figure 7: NGMN – Inter-Layer Class of Service Alignment Challenge for Video

And looking at the result by applying the recommendations of 3GPP, IETF and MEF we have the following option for conversational video live streaming in figure 8 below.

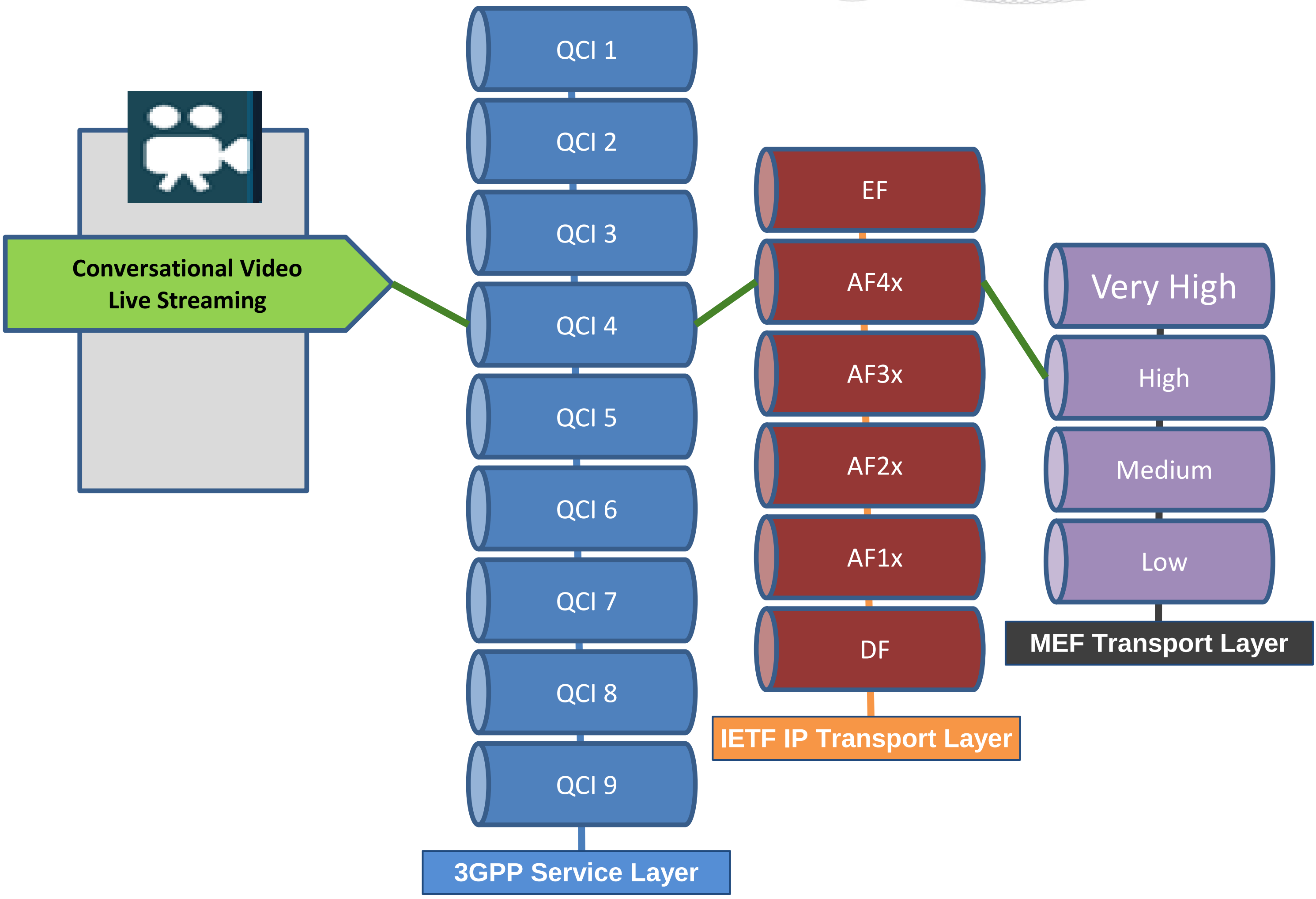

Figure 8: NGMN – Inter-Layer Class of Service Alignment Example for Conversational Video Live Streaming

For non-conversational live streaming the results are shown in figure 9 below.

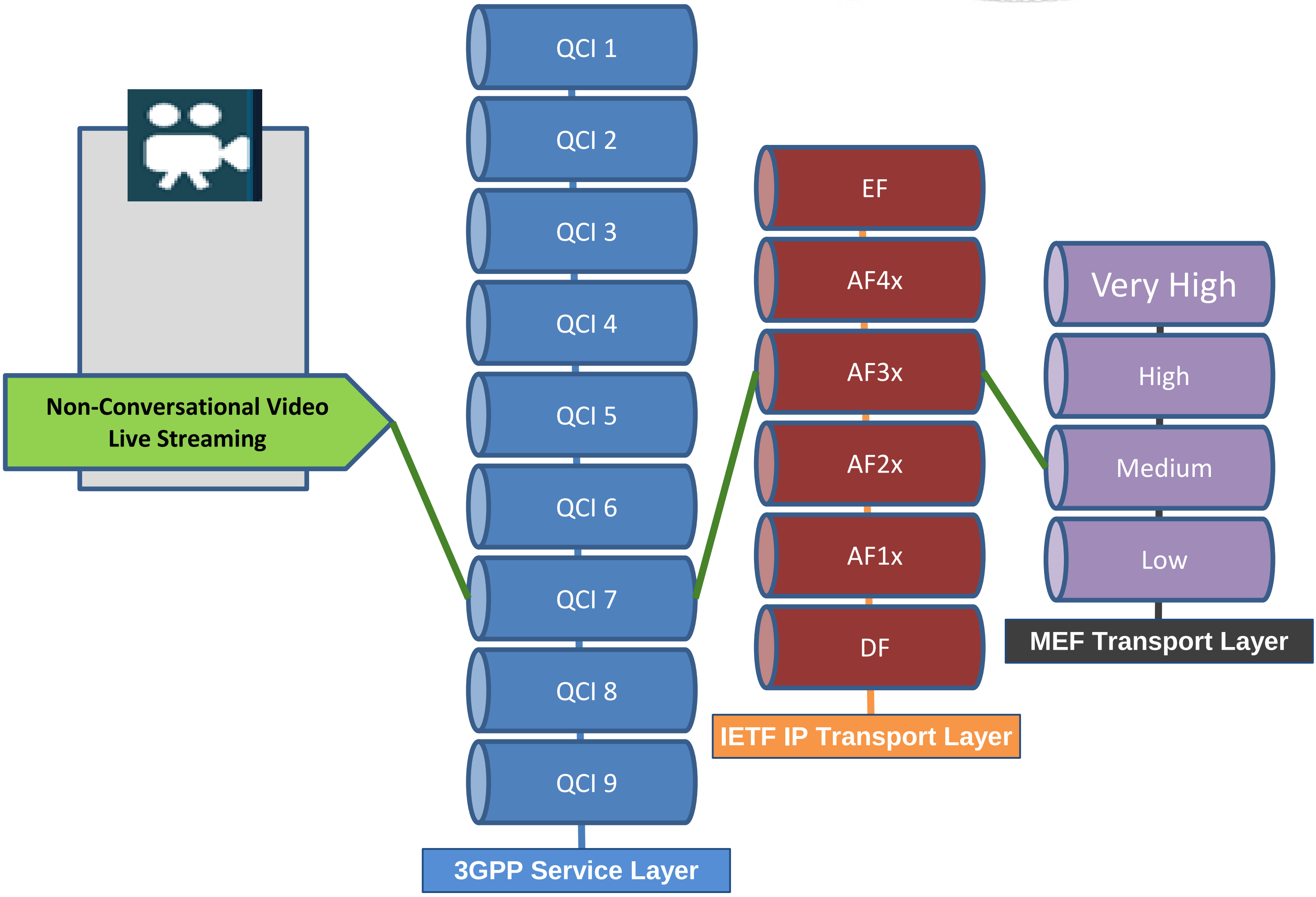

Figure 9: NGMN – Inter-Layer Class of Service Alignment Example for Non-Conversational Video Live Streaming

And for buffered video streaming the results are shown in figure 10 below.

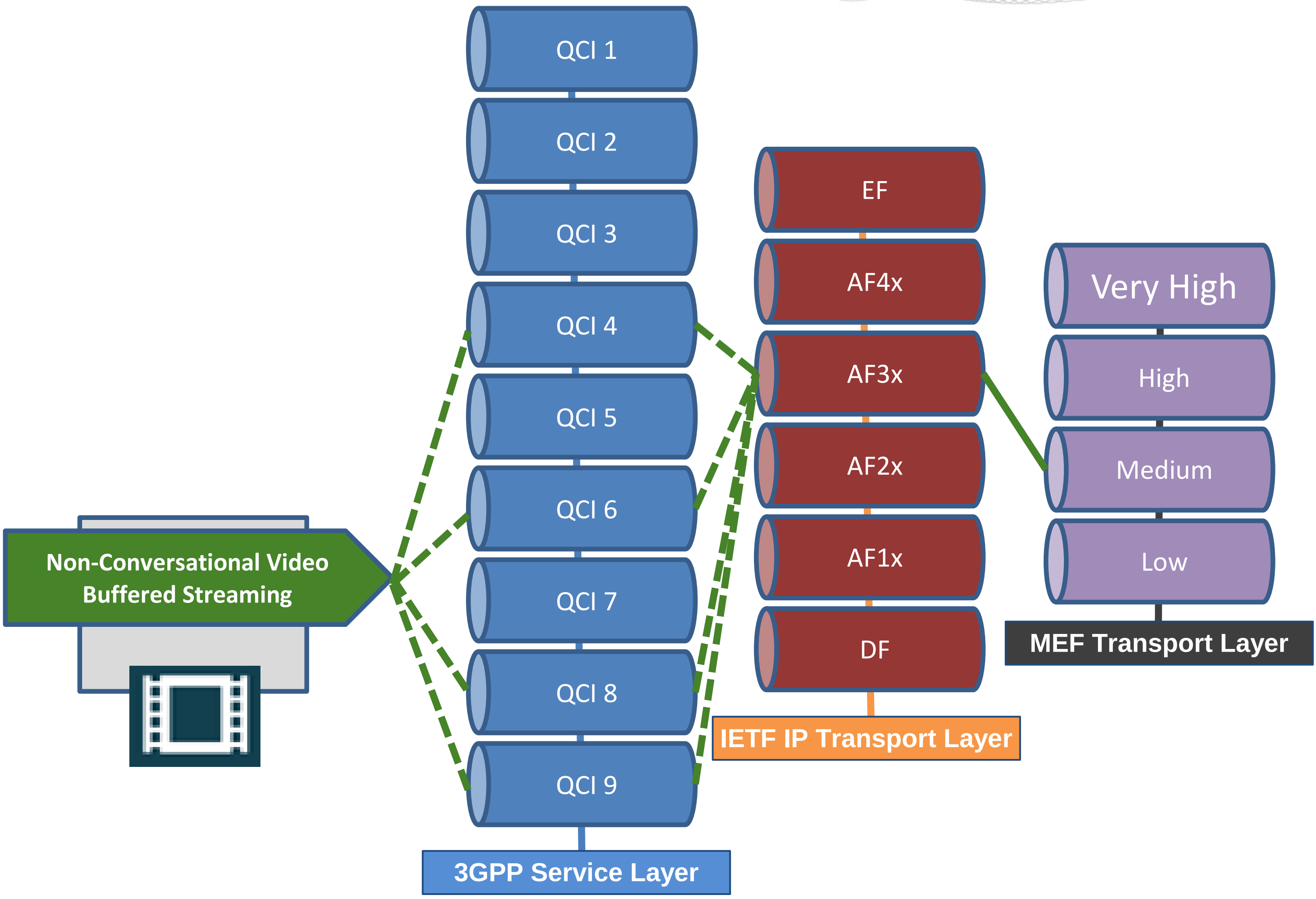

Figure 10: NGMN – Inter-Layer Class of Service Alignment Example for Non-Conversational Video Buffered Streaming

Figure 10 shows that the multiple choices for classification of non-conversational, buffered streaming video service flows in 3GPP are not reflected in the classification schemes of the underlying IP (IETF) and Ethernet (MEF) transport layers. This is a mismatch in the granularity between the layers.

6 NGMN SERVICE FLOW QOS CLASSIFICATION AND INTER-LAYER ALIGNMENT PROPOSITIONS

This chapter contains the main proposal and result of our work on QoS so far: a series of classification schemes that hopefully will facilitate coordination across network layers and between operators.

We have tried to take into account all the research and formalization work done by other Standard Development Organizations, while at the same time respecting the practical needs of operators and finding a pragmatic approach that can be applied with the equipment available today. Furthermore, as will be detailed below, our proposal leaves ample leeway to operators, so that they can adapt the generic proposal to their own situation and still benefit from improved interoperability, amongst other advantages. In summary, our main objectives with this proposal are:

- + Facilitate interconnection and roaming
- + Simplify RAN Sharing and Wholesale service negotiation

+ Provide a baseline reference for operator discussion with providers and equipment vendors

In order to build a progressive consensus, we present first the criteria that lead to the three selected classification schemes, which are then developed and explained in detail.

6.1 Basic Concepts

Given that the target is to propose a simpler, common service flow classification and mapping mechanism for the industry, we have traded some level of precision in the classification of the different **services** in order to facilitate the implementation of our scheme. We think that, until there's a standard definition of signalling and a standard mechanism to identify them (via a tag, for instance), all service-based approach will require DPI and are therefore too unwieldy. Thus, we've developed a classification scheme based on the **flows** that are visible at the backhaul network edge, as opposed to the end-user services. The idea behind this approach is that the first node that maps any given service to the Transport Network Layer (and often the source for this traffic) can itself mark (and effectively classify) the flow according to our guidelines. Some examples or flows are radio access bearers, router control- or management-plane flows and synchronization over packet.

For our classification, and in the same spirit as our summary of currently employed schemes, we have chosen 8 abstract classes of service. This allows us to separate the classification (and the hierarchy that is obtained as a result) from the "name" given to each class of service.

6.2 NGMN Mapping Criteria Proposition

Given the multitude of inputs, it was inevitable that, at some point during the process of coming to a generic proposal, there would be a need to arbitrate between opposed proposals. We developed the following criteria as an objective reference for that process:

Table 21: Mapping Criteria Proposition

1	When grouping flows, the associated KPIs (Key Performance Indicator) must respect the constraints of the most demanding flow ("group up")
2	If the performance of control flows can impact that of associated user flows, they must receive a priority at least equivalent to that of the associated flow.
3	Higher-priority flows must consume (in aggregate) less resources than lower priority flows
4	Resources that enable differentiation among different classes of services are: available bandwidth, lower delay probability (and more constant delay), drop probability. Higher priority for one resource doesn't mean higher priority for all of them.

Notes:

+ #1 obviously does not apply to the lowest class in any scheme, which as "best effort" implies no performance targets.

+ #3 is based on queuing theory (less total traffic in a class implies a lower delay probability), and the desire to avoid resource starvation for the lower classes, which will always partly rely on surplus capacity.

+ #4 is particularly relevant for classification schemes with a large number of classes of service, where more complex differentiations than just "higher priority" may be needed.

Finally, an unspoken criterion is that the number classes of service must be large enough to allow re-classification according to user profile and needs. For instance, some business users may have higher delay and peak bandwidth requirements, and request a higher classification of their traffic.

6.3 NGMN Service Flow Classification Proposal

Below can be found our proposed classification schemes, in increasing order of complexity, and applied to the high level service flow template introduced earlier in chapter 4. We have chosen to present three different schemes for several reasons: firstly, we wanted to be able to develop a coherent argumentation step-by-step. Secondly, it allows us to avoid recommending a single approach, while keeping a fundamental coherence between all approaches. Finally, we haven't gone further than three schemes, as we don't think that it will be easy (or even possible) to offer significantly different service levels in a scenario with too many classes of service. The three selected mechanisms (using up to four different classes of service) offer in our view sufficient variety, even taking into account the appearance of new value-added services.

Table 22: NGMN – Service Flow Classification Proposal

GROUP	NAME	DESCRIPTION	CLASS OF SERVICE		
			2 CoS	*3 CoS*	*4 CoS*
2G	Voice	Conversational Voice	C1	C1	C1
	Data	Data	C8	C2	C2
	C P	Control Plane	C1	C1	C1
	M P	Management Plane	C8	C8	C8
3G	Voice	Conversational Voice	C1	C1	C1
	Rt-DCH	Real-time Data	C1	C1	C1
	Nrt-DCH	Standard Data	C8	C8	C8
	HSxPA	High Speed Data	C8	C8	C8
	C P	Control Plane	C1	C1	C1
	M P	Management Plane	C8	C8	C8
LTE	QCI 1	Conversational Voice	C1	C1	C1
	QCI 2	Conversational Video (live streaming)	C8	C2	C2
	QCI 3	Real-Time Gaming	C1	C1	C1
	QCI 4	Non-Conversational Video (buffered streaming)	C8	C2	C2
	QCI 5	IMS Signalling	C1	C1	C1
	QCI 6	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)	C8	C8	C3
	QCI 7	Voice, Video (live streaming), Interactive Gaming	C8	C8	C3
	QCI 8	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)	C8	C8	C3

	QCI 9	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)	C8	C8	C8
	C P, OAM[1]	Control Plane, Operation Administration and Maintenance (Fault Management and Performance Monitoring)	C1	C1	C1
	M P	Management Plane	C8	C8	C8
Synch	PTPv2	IEEE1588v2	C1	C1	C1
	NTP	NTP	C1	C1	C1
Router	M P	Management Plane	C8	C8	C8
	tx M P	Management Plane	C8	C8	C8
	C P	Control Plane	C1	C1	C1

As shown, we start with just two classes of service (BE or “best effort” and C1), and then progressively move some flows out of the “best effort” category in order to improve their service level. Each stage reflects different constraints, and as such they are introduced below:

6.3.1 Two-CoS Classification Scheme

Table 23: NGMN – Two Classes of Service Classification Scheme

C1	Voice, Real-Time Gaming, Synchronization and Control Plane/OAM
C8	Everything Else

In this case, the only separation is for the most sensitive flows. We single out voice services, for which the expectation is maintaining the existing level of quality, particularly during the transition from circuit to packet transport networks. Other crucial flows such as synchronization and the control plane/OAM, which severely impact service availability and protection mechanisms, are in the highest class of service as well.

A little bit more surprising may be the inclusion of real-time gaming into C1, but it is in line with the characteristics of this flow: it is relatively low-bandwidth (we’re only exchanging game data, not images or video) and requires a very short latency in order to provide adequate responsiveness.

All other flows, which would consume significantly more resources, are considered as “best effort”. This scheme is adapted to scenarios where the basic service level expectation is analogous to internet service (which usually doesn’t have any QoS differentiation), while still allowing to maintain adequate quality for the most demanding services in spite of the mobile network’s constraints.

Target Performance Indicators for the 2-CoS Scheme can be derived from the service flow template in chapter 2.3 and are shown in the table below. As per the proposed classification criteria, the most stringent requirements of the service flows in the class of service are taken as target performance indicator.

[1] OAM messages may need to use the same CoS used from service flows they are associated to according to table 21 Mapping Criteria Proposition, criteria 2

Table 24: NGMN – Target Performance Indicators for 2-CoS Classification Scheme

CLASS OF SERVICE	APPLICATION-LEVEL TOLERANCES[(1)]		
	LOSS	DELAY	DELAY VARIANCE
	(10-x)	(ms)	(ms or L/M/H)
C1	10^{-6}	50	Very Low
C8	-	-	-

(1) Note: As stated in chapter 2.3, the values shown in the table above are for the end-to-end application level requirements; the values for the delay and delay variation metrics for the mobile backhaul network should of course be less. According to chapter 3.2.1 "3GPP Classification", in table 6 which shows the "Standardized QCI characteristics" from 3GPP TS 23.203 version 8.6.0 – Table 6.1.7 (note 1), 3GPP suggests that 20ms can be considered a valid representative value for the delay in a mobile backhaul network between the PCEF (Policy & Charging Enforcement Function) and the base station.

6.3.2 Three-CoS Classification Scheme

Table 25: NGMN – Three Classes of Service Classification Scheme

C1	Voice, Real-Time Gaming, Synchronization and Control Plane/OAM
C2	2G Data (EDGE) and Real-Time Video
C8	Everything Else

This scheme improves upon the previous one by splitting a new class of service (C2) out of the BE category. In it, we are grouping real-time video, which is increasingly becoming a reference value-added service with specific bandwidth availability and delay constraints, and also 2G data. There's a pragmatic reason for this, related to standard operator environments: users will not rely on 2G mobile data if they have access to 3G, HSPA or LTE. Therefore, and as a "data of last resort" service (usually with the best coverage of all mobile data services), it's interesting to prioritize it above the best effort class in order to avoid a 2G-only terminal, or 2G+3G or 2G+3G+LTE terminals falling back to 2G data services because of 2G only coverage, being starved by other terminals using 3G or LTE data services in the same RAN backhaul domain.

Given the services in the C2 class of service, it is straightforward that its total bandwidth consumption will be significantly higher than that of C1, and that, for both 2G data and real-time video, traffic needs will fluctuate more than voice (for example). Therefore, we consider that C1 services may be strictly prioritized (with a different queue type) and subject to acceptance control, while these two mechanisms will not be used with C2 flows.

Target Performance Indicators for the 3-CoS Scheme can be derived from the service flow template in chapter 2.3 and are shown in the table below. As per the proposed classification criteria, the most stringent requirements of the service flows in the class of service are taken as target performance indicator.

Table 26: NGMN – Target Performance Indicators for 3-CoS Classification Scheme

CLASS OF SERVICE	APPLICATION-LEVEL TOLERANCES[(1)]		
	LOSS	DELAY	DELAY VARIANCE
	(10-x)	(ms)	(ms or L/M/H)
C1	10^{-6}	50	Very Low
C2	10^{-6}	100	Low
C8	-	-	-

(1) Note: As stated in chapter 2.3, the values shown in the table above are for the end-to-end application level requirements; the values for the delay and delay variation metrics for the mobile backhaul network should of course be less. According to chapter 3.2.1 “3GPP Classification”, in table 6 which shows the “Standardized QCI characteristics” from 3GPP TS 23.203 version 8.6.0 – Table 6.1.7 (note 1), 3GPP suggests that 20ms can be considered a valid representative value for the delay in a mobile backhaul network between the PCEF (Policy & Charging Enforcement Function) and the base station.

6.3.3 Four-CoS Classification Scheme

Table 27: NGMN – Four Classes of Service Classification Scheme

C1	Voice, Real-Time Gaming, Synchronization and Control Plane/OAM
C2	2G Data (EDGE) and Real-Time Video
C3	Premium Data (buffered Video, non-GBR Real-Time)
C8	Everything Else

In this scheme, C3 is truly a “better-than-best-effort” CoS. We’ve chosen to include in C3 the three LTE QCIs that really involve a notion of quality of service expectation, while keeping the other three existing classes of service in place. This reflects the fact that LTE brings greater granularity to QoS management than 3G or 2G, as illustrated by the larger number of different bearer types that is available in this standard.

Of course, no guidelines are given with regards to the use of the three QCIs that map to C3 (QCI 6, QCI 7 and QCI 8), in order to give operators enough room to develop their own service and business models. For instance, different bearers could be used depending on user type, fair usage criteria, etc.

We hope that most operators will identify this four-colour CoS scheme as the one best adapted to their needs, and that it will form the basis for interconnection and wholesale services. At the very least, we think that wholesale networks should support this level of granularity. Otherwise, customers might need to (again) map into a smaller number of classes of service, which will end up being less efficient for the provider network, or offering an insufficient level of quality to the customer.

Target Performance Indicators for the 4-CoS Scheme can be derived from the service flow template in chapter 2.3 and are shown in the table below. As per the proposed classification criteria, the most stringent requirements of the service flows in the class of service are taken as target performance indicator.

Table 28: NGMN – Target Performance Indicators for 4-CoS Classification Scheme

CLASS OF SERVICE	APPLICATION-LEVEL TOLERANCES(1)		
	LOSS	DELAY	DELAY VARIANCE
	(10-x)	(ms)	(ms or L/M/H)
C1	10^{-6}	50	Very Low
C2	10^{-6}	100	Low
C3	10^{-6}	100	Moderate
C8	-	-	-

(1) Note: As stated in chapter 2.3, the values shown in the table above are for the end-to-end application level requirements; the values for the delay and delay variation metrics for the mobile backhaul network should of course be less. According to chapter 3.2.1 “3GPP Classification”, in table 6 which shows the “Standardized QCI characteristics” from 3GPP TS 23.203 version 8.6.0 – Table 6.1.7 (note 1), 3GPP suggests that 20ms can be considered a valid representative value for the delay in a mobile backhaul network between the PCEF (Policy & Charging Enforcement Function) and the base station.

6.3.4 The Limits of the NGMN Approach

In our three CoS schemes, we have chosen to stay voluntarily high-level, so it will be possible for users to adapt it to their situation. We also do not think that it is possible to achieve a sufficient consensus if the proposal is far too detailed, given the number of different positions on QoS and the intense nature of the debate. Therefore, this generic proposal only specifies the following points:

+ Four distinct classes of service and hierarchy levels, of which eventually only two or three may-be used (even though all four should be supported by the equipment)
+ Target performance indicators for each class of service in a mobile backhaul context. The values in this version of the document are derived from current consensus as indicated in chapter 2.3. The service flow template however could serve as a future input once updated or evolved values are obtained, especially for the delay variance values.
+ Mapping to tagging values to simplify interconnection

It is however possible to subdivide internally these classes of service into sub-sets, either for traceability or even traffic engineering purposes, as long as the performance targets are respected for all flows in the class of service.

Furthermore, we do not enter into the possible interactions of different flows within the same class of service, for instance when only some of the flows are subject to acceptance controls.

There is also a certain dependency on the transport infrastructure being used. For instance, the jitter induced by large packets will be much more relevant in a slower link, such as copper (rather than fibre). This point is only relevant for providers, and should be made transparent to the customer via the network's design.

These questions, and all other classification points that are related to a specific scenario, are left to be addressed by operators.

6.4 NGMN Service Flow Classification versus Some Current Practices Comparison

The NGMN mapping compared to the mapping that was received as input to the project is shown below:

Table 29: NGMN – Service Flow Classification Comparison

GROUP	NAME	DESCRIPTION	CLASS OF SERVICE								
			NGMN			LIVE EXAMPLES					
			2 CoS.	3 CoS.	4 CoS.	INPUT 1	INPUT 2	INPUT 3	INPUT 4	INPUT 5	INPUT 6
2G	Voice	Conversational Voice	C1	C1	C1	C3	C3	C3		C1	C2
	Data	Data	C8	C2	C2	C8	C4	C4	C8	C1	C8
	C P	Control Plane	C1	C1	C1	C2	C1	C2	C2	C1	C2
	M P	Management Plane	C8	C8	C8	C2	C4	C2		C3	C8
3G	Rt-DCH	Real-time Data	C1	C1	C1	C4	C3	C3	C1	C1	C2
	Nrt-DCH	Standard Data	C8	C8	C8	C8	C3	C6	C8	C8	C3
	HSxPA	High Speed Data	C8	C8	C8	C8	C5	C7		C8	C8
	C P	Control Plane	C1	C1	C1	C2	C2	C1	C1	C1	C2
	M P	Management Plane	C8	C8	C8	C2	C4	C4	C6	C2	C8
LTE	QCI 1	Conversational Voice	C1	C1	C1	C3	C3	C3	C1	C1	C2
	QCI 2	Conversational Video (live streaming)	C8	C2	C2	C5	C3	C5	C7	C2	C3
	QCI 3	Real-Time Gaming	C1	C1	C1	C6	C3	C5	C3	C2	C2
	QCI 4	Non-Conversational Video (buffered streaming)	C8	C2	C2	C4	C3	C4	C4	C4	C3
	QCI 5	IMS Signalling	C1	C1	C1	C2	C2	C3	C2	C1	C2
	QCI 6	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)	C8	C8	C3	C8	C4	C6	C6	C4	C3
	QCI 7	Voice, Video (live streaming), Interactive Gaming	C8	C8	C3	C7	C6	C6	C6	C4	C8
	QCI 8	Video (buffered streaming)	C8	C8	C3	C8	C6	C7	C5	C4	C8

		TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)									
	QCI 9	Video (buffered streaming) TCP-based (www, email, chat, ftp, ptp file sharing, progressive video, etc.)	C8	C8	C8	C8	C6	C8	C8	C8	C8
	C P	Control Plane	C1	C1	C1	C2	C2	C3	C1	C1	C2
	M P	Management Plane	C8	C8	C8	C2	C4	C3	C6	C3	C8
Synch	PTPv2	IEEE1588v2	C1	C1	C1	C1	C1	C1	C1	C1	C2
	NTP	NTP	C1	C1	C1	C3	C1	C1	C6		C2
Router	M P	Management Plane	C8	C8	C8	C2	C4	C8		C3	C1
	tx M P	Management Plane	C8	C8	C8	C2	C4	C8		C3	C1
	C P	Control Plane	C1	C1	C1	C2	C2	C3	C3	C3	C1

7 CONCLUSIONS AND LIAISON RECOMMENDATIONS

All in all, we have seen that a multitude of classification and inter-layer CoS alignment schemes exist for the RAN service layer and both the IP and Ethernet transport layers. We have also seen differences in the classification of application service flows by observing current practices in operator networks and vendor recommendations.

The service flow classification in itself is an operator decision, based on operational and commercial grounds, but alignment between operators and between operators and third-party content providers is especially important for a consistent end user experience. In all cases, the inter-layer CoS alignment is important in order to guarantee that the QoS requirements of a service flow are not compromised when traversing different backhaul network segments or third-party networks. Especially in inter-layer CoS alignment, some basic recommendations would be to: use the same number of CoS or alternatively the aggregation of certain CoS into one is consistently throughout the path of a service flow. In essence, the priority hierarchy is important, not the absolute marking values.

The results of this part of the Integrated QoS Management study led us to the following high level or conceptual recommendations:

- Service Flow Classification;
    - promote using a common classification scheme for service flows as per NGMN proposal
    - promote use of common service flow template as a tool to identify the QoS requirements such that a common benchmark can be established

- Inter-layer CoS Alignment;
    - promote using a common number of CoS for the service and transport layers
    - Get agreement with other standards bodies on using a common set of guidelines for the aggregation of classes of service, using this NGMN proposal as a starting point

- Start part two of Integrated QoS Management, namely Inter-layer QoS signalling as marking service flow packets is only one potential implementation, with its limits, as discussed in this document.

8 REFERENCES

This document has used the following documents as reference:

	SOURCE	DOCUMENT	TITLE
1	3GPP	TS 23.203 v11.2.0	Technical Specification Group Services and System Aspects; Policy and charging control architecture
2		TS 23.107 v10.1.0	Technical Specification Group Services and System Aspects; Quality of Service (QoS) concept and architecture
3		TS 23.207 v10.0.0	Technical Specification Group Services and System Aspects; End-to-end Quality of Service (QoS) concept and architecture
4	GSMA		
5	IETF	RFC 1633	Integrated Services in the Internet Architecture: an Overview
6		RFC 2205	Resource ReSerVation Protocol (RSVP) – version 1 Functional Specification
7		RFC 2474	Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers
8		RFC 2475	An Architecture for Differentiated Service
9		RFC 2597	Assured Forwarding PHB Group
10		RFC 2698	A Two Rate Three Colour Marker
11		RFC 3168	The Addition of Explicit Congestion Notification (ECN) to IP
12		RFC 3246	An Expedited Forwarding PHB (Per Hop Behaviour)
13		RFC 3260	
14		RFC 4594	Configuration Guidelines for DiffServ Service Classes
15		RFC 5127	Aggregation of DiffServ Service Classes
16		RFC 5865	A Differentiated Services Code Point (DSCP) for Capacity-Admitted Traffic
17	MEF	6.1	Ethernet Services Definitions – Phase 2
18		10.2	Ethernet Services Attributes Phase 2
19		10.2.1	Performance Attributes Amendment to MEF 10.2
20		22	Mobile Backhaul Implementation Agreement Phase 1
21		22.1	Mobile Backhaul Implementation Agreement Phase 2
22		23	Carrier Ethernet Class of Service - Phase 1 Implementation Agreement
23		23.1	Carrier Ethernet Class of Service - Phase 2 Implementation Agreement

END OF DOCUMENT