International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Series G

Supplement 61

(09/2017)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

ITU-T G.1020 – Internet protocol aware quality of service management

ITU-T G-series Recommendations – Supplement 61

ITU-T

International Telecommunication Union

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
MULTIMEDIA QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS	G.1000–G.1999
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000–G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000–G.8999
ACCESS NETWORKS	G.9000–G.9999

For further details, please refer to the list of ITU-T Recommendations.

Supplement 61 to ITU-T G-series Recommendations

ITU-T G.1020 – Internet protocol aware quality of service management

Summary

Supplement 61 to Recommendation ITU-T G.1020 documents a packet-oriented Internet protocol (IP)-centric quality of service (QoS) management model. This model is applicable to a wireless access point (e.g., Wi-Fi AP, evolved Node B (eNode B), Node B), referred to as an IP aware access point. The Supplement describes a possible cross-layer design for this model in long-term evolution/evolved packet core (LTE/EPC) networks. Some implementation challenges are highlighted, together with possible solutions implying only minor modifications in the eNode B. Performance of this proposal compared to various implementations of the third generation partnership project (3GPP) QoS model is evaluated using the network simulator-3 (ns-3) in realistic scenarios.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T G Suppl. 61	2017-09-28	12	11.1002/1000/13393

Keywords

IP aware, LTE, QoS management.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this publication, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this publication is voluntary. However, the publication may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the publication is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the publication is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this publication may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the publication development process.

As of the date of approval of this publication, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this publication. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Introduction

The mobile data landscape is rapidly changing. Current mobile devices like smartphones and tablets have connectivity capabilities that are close to or even better than fixed devices. Usage and traffic patterns have also changed. For example, video and music streaming are available on a widespread basis today, but scarcely existed some years ago. Furthermore, the long-term evolution/evolved packet core (LTE/EPC) network offers increased possibilities thanks to higher data rates and lower latency, which favours the emergence of new services.

The joint evolution of mobile usages and services [e.g., voice over LTE (VoLTE)], results in an important challenge for operators. Indeed, current mobile standards have not yet fully integrated this expectation of open, cheap and flexible web-oriented internet access. Notwithstanding the claimed all Internet protocol (IP) basis of the LTE system, current third generation partnership project (3GPP) quality of service (QoS) management for LTE/EPC networks is still circuit oriented. Namely, LTE QoS management is based essentially on virtual tunnels called bearers that provide for end-to-end transport service with specific QoS attributes. Contrary to usual QoS models of the fixed internet (IP networks), such as Diffserv, additional signalling procedures are then required in order to establish a dedicated bearer for each desired QoS level. This circuit-oriented model was inherited from transport based on time-division multiplexing (TDM), and it was well adapted to early mobile data service deployments when the amount of data exchanged was low and multitasking almost inexistent.

However, managing the QoS of mobile networks using a circuit-based model raises some tricky issues considering the high volumes and variability of mobile data traffic generated by smartphone applications of today, such as: scalability of the architecture due to the number of tunnels to be maintained, additional signalling traffic required to establish or modify the parameters of the tunnels and time to establish or modify a tunnel.

As an alternative, a lightweight QoS model has been introduced in [b-ITU-T SG12-255], [b-Hamchaoui, 2013], [b-Hamchaoui, 2014] and [b-Jobert] mainly inspired by IP policies commonly found in fixed networks. This model offers loose QoS differentiation in a heterogeneous traffic environment that solves the aforementioned limitations. The so-called IP-centric approach has already drawn interest among some major actors in the mobile industry.

Note, however, that both models could be simultaneously deployed: a circuit-based model could be used for managed services when signalling traffic is exchanged prior to starting the service (a typical example might be VoLTE), while the proposed packet-oriented model could be suitable to differentiate among the other services (e.g. Internet services).

Supplement 61 to ITU-T G-series Recommendations

ITU-T G.1020 – Internet protocol aware quality of service management

1 Scope

Managing the quality of service (QoS) of mobile networks using a circuit-based model (as is the case for services like voice over long-term evolution (VoLTE), see [ITU-T G.1028]) raises some tricky issues considering the high volumes and variability of mobile data traffic generated by current smartphone applications, such as: scalability of the architecture due to the number of tunnels to be maintained, additional signalling traffic required to establish or modify the parameters of the tunnels and time to establish or modify a tunnel.

As an alternative, this Supplement introduces a possible lightweight cross-layer design for the IP-centric model, which implies only minor modifications of an IP aware access point (AP). This model offers loose QoS differentiation in a heterogeneous traffic environment that solves the aforementioned limitations.

2 References

[ITU-T G.107]	Recommendation ITU-T G.107 (2015), *The E-model: A computational model for use in transmission planning*.
[ITU-T G.1020]	Recommendation ITU-T G.1020 (2006), *Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks*.
[ITU-T G.1028]	Recommendation ITU-T G.1028 (2016), *End-to-end quality of service for voice over 4G mobile networks*.
[ITU-T SG12-75]	ITU-T SG12 (Study period 2013) Contribution 75 (2013). *Performance of mobile networks*.
[ITU-T SG12-76]	ITU-T SG12 (Study period 2013) Contribution 76 (2013). *Performance of voice over LTE (VoLTE)*.
[ITU-T SG12-255]	ITU-T SG12 (Study period 2013) Contribution 255 (2015). *Packet oriented QoS management model for a wireless access point*.
[ITU-T SG12- TD-GEN 410]	Chairperson of the Quality of Service Development, Report of the Workshop on *Monitoring Quality of Service and Quality of Experience of Multimedia services in Broadband/Internet Networks*: 2014-04-14/16 Maputo Mozambique, TD-GEN 410, ITU-T SG12, Geneva, 2014-09-02/11.
[Alexa]	Alexa (2017). *Top sites in France*. https://www.alexa.com/topsites/countries/FR
[Bojovic]	Bojovic, B., Baldo, N. (2014), *A new channel and QoS aware scheduler to enhance the capacity of voice over LTE systems. In: Systems, Signals and Devices (SSD)*, 2014 11th International Multi-Conference on, p. 6. https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6802828
[COST-231]	COST 231 (1998). Final report. *Brussels: Co-operative for Scientific and Technical research*. http://www.lx.it.pt/cost231/final_report.htm

[Hamchaoui, 2013] Hamchaoui, I., Jobert, S., Boufelja, S. (2013), *IP aware radio scheduling – introducing IP QoS management in LTE networks*. Communications Workshops (ICC), 2013 IEEE International Conference on, pp. 1238-1242. https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6637583

[Hamchaoui, 2014] Hamchaoui, I., Diego, W., Jobert, S. (2014), IP centric QoS model for mobile networks – Packet based QoS management for Intra-bearer arrangements. In: *Wireless Communications and Networking Conference* (WCNC), 2014 IEEE, pp. 2653-2658.

[Jobert] Jobert, S., Hamchaoui, I., Diego, W. (2013), *Packet-oriented QoS management model for a wireless access point*, IETF, 11 July. https://tools.ietf.org/html/draft-jobert-iccrg-ip-aware-ap-00

[Pries] Pries, R., Magyari, Z., Tran-Gia, P. (2012), An HTTP web traffic model based on the top one million visited web pages. In: 2012 8th *Euro-NF Conference on Next Generation Internet* (NGI), IEEE. pp. 133-139. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6252145

[Ramos-Munoz] Ramos-Munoz, J., Prados-Garzon, J., Ameigeiras, P., Navarro-Ortiz, J., Lopez-Soler, J. (2014), Characteristics of mobile YouTube traffic. *IEEE Wireless Communications*, **21**, pp. 18-25.

[Yang] Yang, R., Son, H.J. (2013), YouTube's Live TV streaming in mobile – HLS & adaptive. https://www.netmanias.com/en/post/blog/6035/google-hls-http-adaptive-streaming-iptv-video-streaming-youtube/youtube-s-live-tv-streaming-in-mobile-devices-hls-adaptive

[Zhou] Zhou, D., Baldo, N., Miozzo, M. (2013), Implementation and validation of LTE downlink schedulers for ns-3. In: SimuTools '13 *Proceedings of the 6th International ICST Conference on Simulation Tools and Techniques*, pp. 211-218.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Supplement

None.

4 Abbreviations and acronyms

This Supplement uses the following abbreviations and acronyms:

AM Acknowledged Mode

AMC Adaptive Modulation and Coding

AMR-NB Adaptive Multirate Narrow-Band

AP Access Point

ARQ Automatic Repeat request

BE-PF Best Effort (per-flow QoS support)

CQA-PF	Channel and QoS Aware (per-flow QoS support)
CQI	Channel Quality Indicator
DL	Downlink
DRB	Data Radio Bearer
DSCP	Differentiated Services Code Point
eNB	evolved Node B
eNode B	evolved Node B
EPA	Extended Pedestrian A (model)
EPC	Evolved Packet Core
EPS	Evolved Packet System
FTP	File Transfer Protocol
GBR	Guaranteed Bit Rate
GPRS	General Packet Radio Service
GTP	GPRS Tunnelling Protocol
HARQ	Hybrid Automatic Repeat request
HAS	HTTP Adaptive Streaming
HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
LA	Link Adaptation
LENA	LTE EPC Network simulator
LTE	Long-Term Evolution
MAC	Media Access Control
MOS	Mean Opinion Score
ns-3	network simulator-3
PDCP	Packet Data Convergence Protocol
PF	Proportional Fair
P-GW	Packet data network Gateway
PLR	Packet Loss Rate
PLT	Page Load Time
PSS-PF	Priority Set Scheduler (per-flow QoS support)
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
RB	Radio Bearer
RLC	Radio Link Control
RoHC	Robust Header Compression
SINR	Signal to Interference and Noise Ratio

TCP	Transport Control Protocol
TDM	Time-Division Multiplexing
TTI	Transmission Time Interval
TX	Transmission
UDP	User Datagram Protocol
UL	UpLink
UM	Unacknowledged Mode
VoIP	Voice over IP
VoLTE	Voice over LTE
Wi-Fi	Wireless Fidelity

5 Conventions

None.

6 Quality of service model in third generation partnership project mobile networks

A main virtual circuit, called the evolved packet system (EPS) bearer, should be set up between the mobile terminals and the packet data network gateway (P-GW) before any traffic can be exchanged between them. The EPS bearer is constituted of several local bearers established between neighbouring network elements. For example, a data radio bearer (DRB) is set up between the eNB and the mobile terminal. The EPS bearer provides a transport service with specific QoS attributes. When a mobile terminal attaches to the network, a default bearer with a "best effort" QoS is established.

Other bearers could be further set up, one per required QoS level. Bearers are operated in connected mode, i.e., they are established, modified or disconnected via mobile signalling protocols. Each EPS bearer is associated with a QoS profile mainly defined by the QoS class identifier (QCI) and possibly a guaranteed bit rate (GBR). The QCI is associated with a specific forwarding behaviour applying to any packets supported by this bearer. The QCI includes parameters like the layer 2 (L2) packet delay budget and packet loss rate (PLR). Note that the radio scheduler plays a crucial role in the QoS over the radio segment, and consequently, it dramatically impacts the end-to-end customer experience. However, Third generation partnership project (3GPP) specifications do not define any scheduling algorithm, leaving open its design and implementation. Various examples of radio scheduling algorithms can be found in the literature. However, vendors generally implement proportional fair (PF) algorithms. Such algorithms propose a trade-off between cell throughput optimization and fairness between the mobile terminals.

6.1 Quality of service procedures for downstream traffic

As an entry point of traffic to an LTE network, the P-GW provides connectivity from mobile terminal to external packet data networks. Downstream traffic flows arrive at the P-GW that directs them on to their respective EPS bearer, thanks to a filter describing the matching between the traffic flows and their corresponding bearers (identified by their IP headers, for example).

Downstream IP packets are then carried through the LTE network to the eNB via a general packet radio service (GPRS) tunnelling protocol (GTP) tunnel (named S1 bearer); the eNB removes the GTP header of each IP packet before delivering it to the radio interface (Uu). Figure 1 shows the eNB protocol stack for the downstream radio interface (Uu). The top of the protocol stack is the location of the packet data convergence protocol (PDCP) sublayer. 3GPP specifications stipulate the creation of independent PDCP and radio link control (RLC) entities for each EPS bearer.

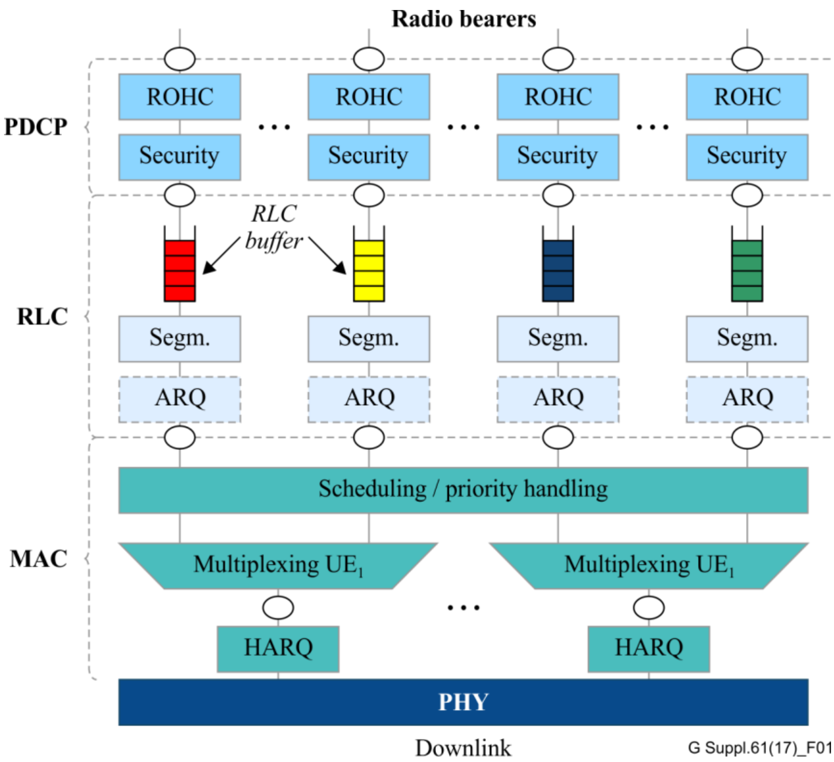

Figure 1 – Evolved Node B radio downstream interface protocol stack (simplified)

The RLC sublayer performs segmentation/concatenation and reassembly and can be operated in two modes: unacknowledged mode (UM) and acknowledged mode (AM). The choice of the RLC transmission mode depends on whether delay or integrity is favoured at the radio level. Each PDCP entity is associated with one or two RLC entities (i.e., uni-directional or bi-directional), which depends on the radio bearer (RB) characteristic and the RLC transmission mode (UM or AM). Each RLC entity implements a buffer in order to store packets coming from the PDCP sublayer and reports periodically the amount of buffered data to the media access control (MAC) layer. Every transmission time interval (TTI), the radio scheduler distributes the available radio resources among all mobile terminals based on the scheduler strategy. The scheduling strategy is mainly based on the QoS information, the RLC buffer status and the link adaptation (LA) algorithm.

6.2 Discussion on the 3GPP quality of service management

The 3GPP QoS model can prove to be tricky and complex to use. In particular, it raises issues in terms of the following.

- **Scalability**: In a centralized architecture such as the long-term evolution/evolved packet core (LTE/EPC) network, the P-GW manages a huge number of bearers. Note that each mobile terminal uses an independent bearer per QoS level, in addition to its default bearer.
- **Efficiency**: The establishment, modification, release and mobility management of each bearer generate a non-negligible signalling load in the network. Furthermore, the creation and destruction of PDCP and RLC entities imply an important processing load on the eNB.
- **Performance**: Bearers' establishment implies a processing time that results in network latency. This delay is not negligible and may severely impact web-oriented applications.
- **Costs**: A QoS-aware radio scheduler is the cornerstone of the 3GPP QoS model. However, they break the virtuous trade-off of PF algorithms as they allocate more resources to "premium" terminals regardless of their radio conditions. This may severely impact the overall cell capacity, particularly when prioritized mobile terminals are in poor radio conditions. This phenomenon is emphasized when premium mobile terminals are strongly

favoured. As a consequence, tuning QoS-aware schedulers to offer an appropriate QoS level without starving other mobile terminals proves to be a difficult task.

Mobile operators can conclude that the 3GPP QoS scheme is too complex and costly in view of the likely revenue on 3G and 4G networks, except for VoLTE. Finally, 3G/4G traffic is often supported through a unique best effort default bearer per customer or through two bearers when VoLTE is offered. Note that some performance issues have been reported in [ITU-T SG12- TD-GEN 410].

In fact, the 3GPP QoS model is probably not the appropriate answer to the somewhat contradictory expectations for cheap, efficient and flexible QoS on the mobile internet.

7 IP-centric architecture

Based on the experience acquired on the fixed Internet, an alternative QoS architecture for the mobile network has been proposed. In this IP-centric model, a unique EPS bearer (default bearer) per mobile terminal is established so as to manage the connectivity and other specific features of mobile networks. All the flows belonging to a given user are transported on that user's unique multi-QoS bearer. QoS is further managed at the packet level, based on the differentiated services code point (DSCP) field, for example. In this case, the DSCP should have been marked by a trusted entity, such as the P-GW or a trusted router.

A set of priority queues per mobile terminal should further be implemented in the eNB. For a given mobile terminal, dowstream packets arriving at the eNB through the aforementioned bearer are added to the relevant queue, based on their DSCP marking. As it is located in the eNB, this queuing system can be served at a rate taking into account very accurately the lower layer status (e.g., radio conditions, cell load or available radio resources), which generally change every TTI.

Figure 2 illustrates the proposed packet-oriented QoS management model, with an example where three classes have been configured in the IP aware AP (e.g., one for traffic requiring low delay/jitter, one for traffic requiring low loss and one for Best Effort traffic). The dotted IP layer at the IP aware AP indicates that this node may not implement a full IP layer (e.g., only the DSCP marking in the IP header has to be analysed to classify the packets in the relevant queues, but the forwarding of the packets may be based on traditional relaying functions between the tunnel on a fixed network and the radio layers).

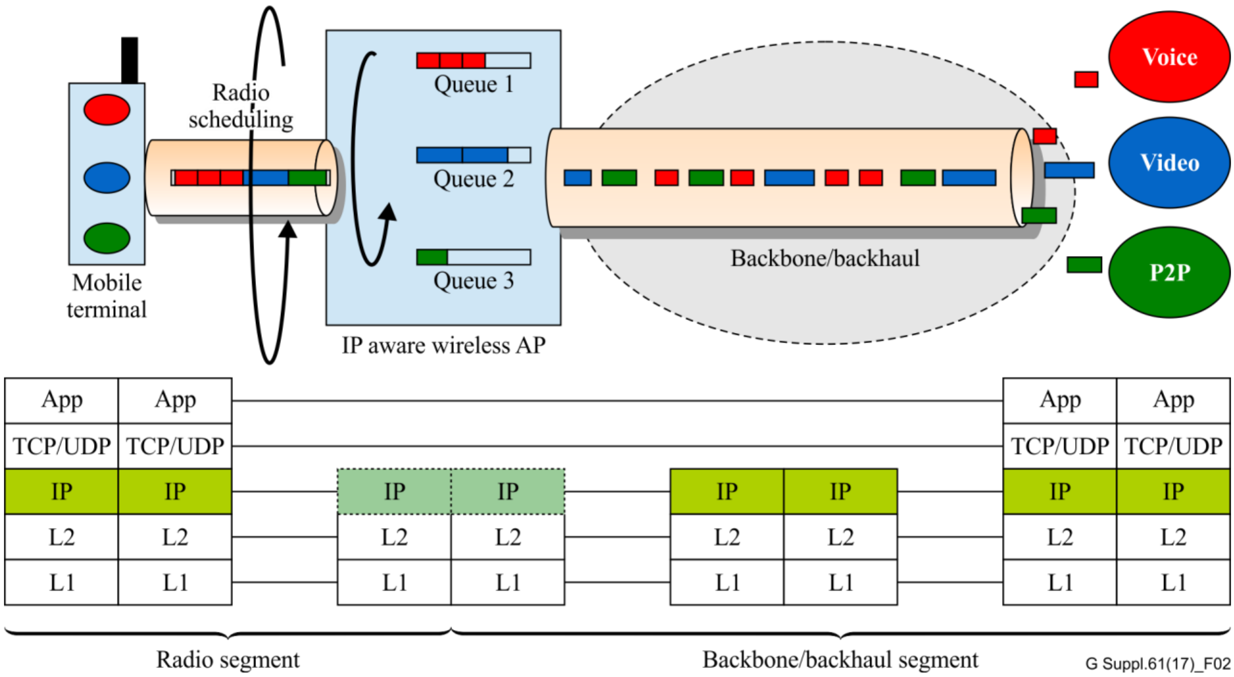

Figure 2 – Packet-oriented quality of service management model in LTE/EPC networks

a) **Intra-bearer arrangements**: the radio scheduler does not take into account the traffic mix waiting for transmission when allocating radio resources to this mobile terminal. This means that a basic scheduler algorithm can be used (e.g., PF). In [Hamchaoui], an intra-bearer arrangement model is introduced.

b) **Inter-bearer arrangements**: the radio scheduler takes into account the traffic mix of each mobile terminal waiting for transmission when allocating radio resources to this mobile terminal. In this case, the scheduling algorithm should know the queue state of the mobile terminal. For this purpose, several approaches are possible (e.g., weighting the allocation according to the prioritized traffic volume/priority queue backlog, ensuring a maximum latency for specific classes).

Clause 8 discusses the cross-layer design of the intra-bearer arrangement model, which provides an interesting trade-off between fairness and efficiency in heterogeneous traffic scenarios, such as those that exist today.

8 Cross-layer design for an IP aware access point

8.1 Design challenges

As explained in clause 7, downstream IP packets of a given mobile terminal arrive at the eNB via a unique GTP tunnel (S1 bearer). Then the IP aware AP removes the GTP header of each packet and places the packet in the IP queuing system associated with this mobile terminal. The eNB should then implement a pseudo IP layer atop the radio protocol stack, the main features of which come down to packet classification and queueing.

Many different designs may then be considered in order to address the interactions between this new IP layer and lower radio layers. A description follows of a solution that implies only minor modifications in the eNB and none in the mobile terminal. In this case, a common PF scheduler can be used in the eNB, and only minor changes in the RLC layer are required besides our pseudo IP layer. The main challenges concerning the RLC layer are described as follows.

- **RLC buffer management**: The RLC buffer should contain no more data than the exact amount required by the MAC sublayer on each TTI. Otherwise, on the next scheduling cycle (next TTI); low priority data that remain in the RLC buffer would be served before high priority packets arrived in the interim.
- **Data volume waiting for transmission**: In 3GPP standard implementations, the radio scheduler should know the amount of data queued in the RLC buffer in order to allocate no more radio resources than needed. In our IP aware AP, the RLC buffer is managed so as to be generally empty (see RLC buffer management in the previous entry), as data are mainly queued in the IP layer queuing system. The MAC layer should then take into account all the data waiting for transmission, either buffered in the RLC or in the IP layers.

8.2 Design description

Figure 3 shows our IP aware AP. Each IP aware AP implements an IP layer with an IP queueing system per mobile terminal, where IP packets are classified and queued, and then a unique DRB per mobile terminal is used.

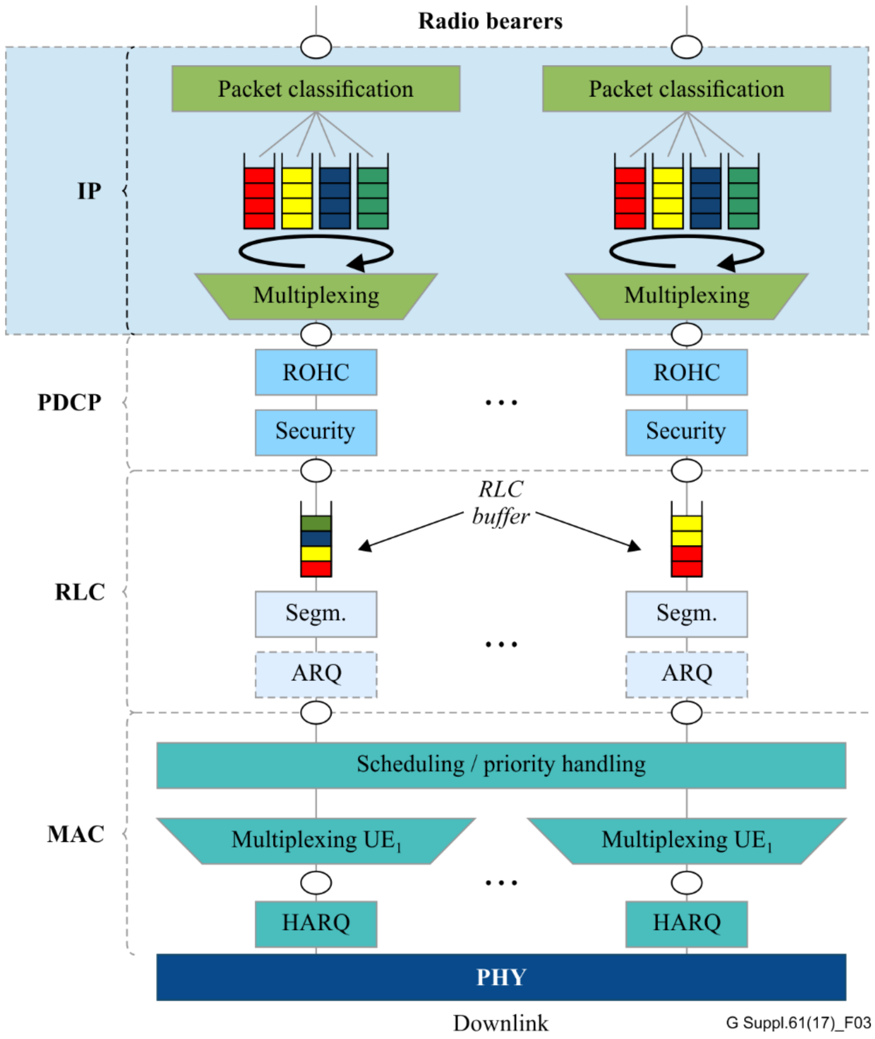

Figure 3 – IP aware access point – radio interface downlink protocol stack (simplified)

At the MAC layer, users are chosen in each TTI, together with their respective radio resources and coding schemes, depending upon the radio scheduler strategy. This strategy also takes into account a queueing status provided by the RLC layer, so that the credits allocated to a mobile terminal do not exceed its real needs. In our IP-centric model, this queueing status should take into account not only the RLC queue – which is supposed to be empty most of the time thanks to the IP/RLC synchronization previously described – but also principally the IP queues devoted to this mobile terminal.

In return, packets are picked up from the IP queueing system by the RLC layer in accordance with the credits allocated to this mobile terminal by the MAC layer (at a so called transmission opportunity) at each scheduling cycle. Based on this credit, each IP entity provides packets to its corresponding PDCP entity. Then the PDCP layer adds a sequence number, after which it compresses headers using the robust header compression (RoHC) protocol and ciphers the packet; finally, it adds a PDCP header before to sending it to its corresponding RLC entity.

The RLC entity segments or concatenates packets according to the data credit and an RLC header is added containing, amongst others, the corresponding sequencing number. As only one DRB per mobile terminal is implemented, the RLC transmission mode is the same for all packets belonging to this mobile terminal. A possible option is to set a fixed RLC transmission mode (i.e., AM or UM), but no advantage of the variety of RLC transmission modes is taken. Another option (more complex), is to develop a dynamic RLC transmission mode based on packet loss, radio conditions or even both.

Both solutions are entirely compatible with our IP-centric model. Finally, the RLC entity sends required data to the MAC sublayer, which performs its usual treatments.

9 Simulation and performance evaluation

9.1 Network parameters

The proposed cross-layer design has been evaluated thanks to the network simulator-3 (ns-3), version 3.21. The LTE EPC network simulator (LENA) module has been modified in order to implement an IP aware AP, as described in clause 8.2. A typical outdoor scenario was considered with 20 mobile terminals attached to a single eNB; thus intercell interference is not taken into account. The IP aware AP is equipped with an omnidirectional antenna and mobile terminals experience varying channel conditions. The RLC layer in the eNB is configured in UM. A realistic channel model with path loss and fading was used. The path loss is simulated as described in [COST-231] and the fast fading as described by the extended pedestrian A (EPA) model using a Rayleigh multi-path fading model. Given the path loss model and the other network parameters, we have a wide range of signal to interference and noise ratio (SINR) values, which provide channel quality indicator (CQI) values between 1 and 15.

At the beginning of each run, mobile terminals are placed randomly in a disc representing the cell within a distance ranging between 30 m and 500 m. Then, mobile terminals move within the disc according to a random walk model, at a fixed speed of 3 km/h. The simulation parameters are shown in Table 1 and the system configuration is as follows. The cell is connected via the PDN Gateway (P-GW) to the Internet. Two servers are implemented, one for voice over Internet protocol (VoIP) and the other for file transfer protocol (FTP), YouTube and web services. Both servers are connected to the P-GW via an over-provisioned point-to-point link in order to avoid congestion in this segment of the network. In the IP-centric scenario, each server performs the DSCP marking depending on the application.

Table 1 – Simulation parameters

Number of terminals	20 (random positions)
Users mobility model	Random walk (3 km/h)
Bandwidth	50 RB (10 MHz)
Cell coverage radius	500 m
Path loss Model	Cost231
eNB TX power/noise figure	46 dBm/5 dB
Mobile terminal TX power / noise figure	24 dBm/5 dB
Fading loss model	EPA 3 km/h (urban scenario)
Adaptive modulation and coding (AMC) model	PiroEW2010
Downlink/uplink (DL/UL) carrier frequency	2 120 MHz/1 930 MHz
RLC transmission mode	UM (unacknowledged mode)
RLC buffer size	100 kbytes (70 FTP packets)

9.2 Traffic description

Each mobile terminal uses one, two, three or four applications as described in Table 2. One mobile terminal uses all applications, which corresponds to a tethering configuration.

Table 2 – Traffic distribution

Traffic mix	No of mobile terminals	Distribution			
		VoIP	**YouTube**	**Web**	**FTP**
VoIP only	1	1			
YouTube only	1		1		
Web only	1			1	
FTP only	1				1
FTP + web	4			4	4
FTP + YouTube	4		4		4
FTP + VoIP	4	4			4
FTP + VoIP + web	3	3		3	3
Tethering (all applications)	1	1	1	1	1
Total	**20**	**9**	**6**	**9**	**17**

i) **VoIP traffic**: The arrivals of new voice calls are assumed to follow a Poisson process with an arrival rate, $\lambda = 7.101$. The call duration is estimated using a LogNormal distribution with parameters $\mu = 3.894\,530\,9$ and $\sigma = 1.004\,1$. VoIP traffic is based on adaptive multirate narrow-band (AMR-NB) codec. In order to measure the quality of experience (QoE) of VoIP, the E-model specified in [ITU-T G.107] is used.

ii) **Web traffic**: A realistic hypertext transfer protocol (HTTP) traffic model is used, as proposed by [Pries], which was also implemented on an ns-3, as well as five top French websites based on 2014 ranks of [Alexa] in order to simplify and harmonize the performance measurements in terms of page load time (PLT), see Table 3.

Table 3 – Website details

Website [viewed 2017-11-11]	**Average size (bytes)**	**Total objects**
1. boutique.orange.fr	2 366 380	129
2. Leboncoin.fr	728 592	78
3. Facebook.com	10 47 690	185
4. Lefigaro.fr	1 054 458	50
5. pole-emploi.fr	1 205 991	320

iii) **YouTube traffic**: A specific module has been implemented in ns-3 in order to emulate the delivery of YouTube traffic in HTTP adaptive streaming (HAS) mode. For this purpose, the study has taken advantage of the traffic model described in [Ramos-Munoz] and [Yang]. In HAS, the video object is broken out into chunks, and the server maintains several profiles for each chunk. In our simulation, the chunk duration is 5 s and four profiles are defined and identified by their *itag* values (see Table 4) depending on their encoding rate. Each time the client requests a chunk from the server, it also selects a profile in accordance with the underlying network conditions. A dual-threshold buffering strategy has been implemented using parameters estimated in [Ramos-Munoz] in order to dynamically adjust the requested chunk profiles to the client playback queue size. A maximum timeout for chunk request has also been set to avoid blocking situations during simulation; this parameter is fixed at 100 s. When the timeout is exceeded, the video session is stopped and a new video session starts. The duration of the whole video is fixed at 150 s with an exponential inter-video interval with 10 s of mean.

Table 4 – YouTube video quality information

itag	Resolution	Encoding rate (kbit/s)
151	128 × 72	64
132	426 × 240	266
92	426 × 240	395
93	640 × 360	758

iv) **FTP traffic (background traffic)**: Some mobile terminals support background best effort traffic (see Table 2) which is represented by FTP sessions. The size of these files follows a uniform law between 1 Mbyte and 5 Mbyte. The arrivals of new FTP sessions follow a Poisson process with an arrival rate, λ = 10 s.

9.3 Scenario description

In order to evaluate the performance of the IP-centric model, four scenarios are considered.

- **Scenario 1 – Mono bearer, best effort scheme, PF scheduler**: This scenario addresses the case of current deployments in LTE networks supported by a single bearer in best effort with a basic PF scheduler.
- **Scenarios 2 – Multi-bearer, CQA-PF scheduler** or **Scenarios 3 – Multi-bearer, PSS-PF scheduler**: Scenarios 2 and 3 simulate standard 3GPP multi-bearer configurations; in these cases, the mobile terminals establish dedicated EPS bearers for non-best effort applications (i.e., VoIP, YouTube and web). To support these dedicated bearers, two QoS-aware schedulers available in the ns-3 were selected, along with the channel and QoS aware (CQA) scheduler defined in [Bojovic] and the priority set scheduler (PSS) described in [Zhou].

  Whether in scenario 2 or scenario 3, VoIP bearers are affected by a QCI equal to 1 and a GBR of 44 kbit/s. YouTube bearers are assigned a QCI equal to 4 and a GBR equal to 1.5 Mbit/s, and finally, web is supported by non-GBR dedicated bearers with a QCI equal to 9.

  However, QoS-aware scheduler implementations in ns-3 are rather limited, as they do not support multi-service mobile terminals: If several dedicated bearers are established from a given mobile terminal, the scheduler will apply the same traffic parameters (QCI, GBR) to any of them. To get around this problem, a unique mobile terminal is emulated that uses *N* QoS levels through *N* mobile terminals with the same mobility pattern.
- **Scenario 4 – Mono-bearer, IP-centric model, PF scheduler**: This scenario simulates an IP aware AP as described in the previous clauses.

9.4 Performance analysis

Performance results related to the scenarios described previously are presented here. Each simulation run lasts for 1 000 s, with a warm-up time of 5 s during which statistics are not collected, and is replicated three times with different seeds. Applications are started at a random time uniformly distributed between 1 s and 5 s.

Figure 4 depicts the cumulative distribution functions (CDF) of cell throughput in the analysed scenarios [Best Effort, CQA-PF, priority set scheduler (per-flow QoS support) (PSS-PF) and IP aware]. Note that the cell throughput in the CQA-PF scenario is particularly degraded, because the CQA algorithm strongly favours the GBR bearer, even in bad radio conditions, at the expense of others bearers. In contrast, the best effort PF scheme achieves the highest cell throughput, closely followed by the IP-centric scenario, as they do not interfere in PF resource allocation. The cell throughput of the PSS scenario is only moderately degraded as the PSS-PF algorithm is less aggressive than the CQA-PF.

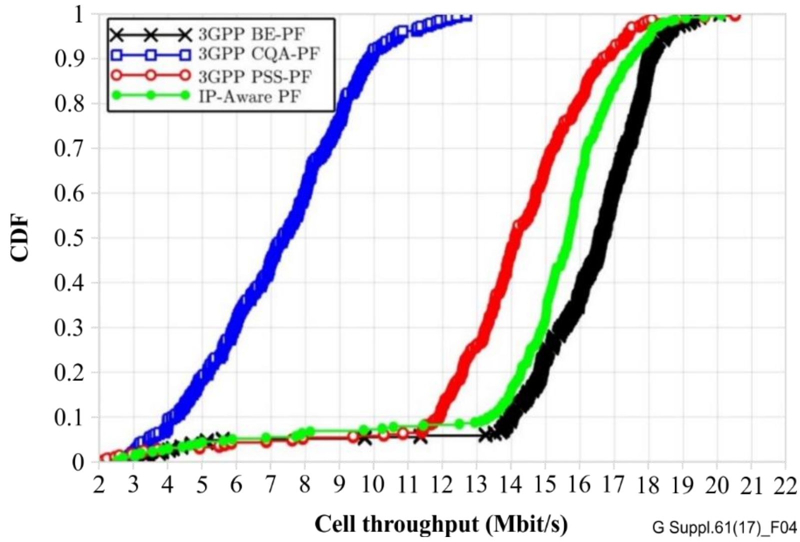

Figure 4 – Cell throughput performance

Figure 5 provides the boxplots of the VoIP mean opinion score (MOS). The median is given by the central mark and the borders of the box are the 0.25 and 0.75 percentiles. Figure 5 shows that the MOS for the best effort scheme is poor, as its median is around 1.8 (poor quality), whereas the CQA, PSS and IP aware schemas have similar MOS, around 4.1 (good quality).

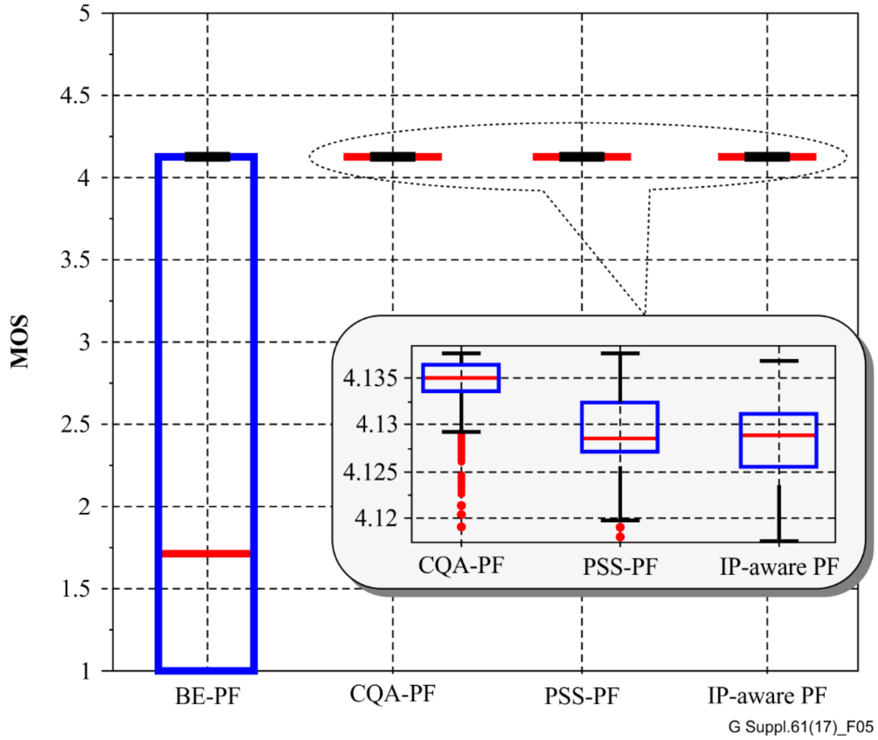

Figure 5 – Voice over Internet protocol quality of experience

Regarding YouTube traffic (Figure 6), almost 70% of chunks are delivered in the highest quality in the CQA scenario; the other schemes obtain a rather mixed distribution in chunk quality. Figure 7 provides a boxplot display of the YouTube flow throughputs. While the PSS and the IP aware schemes obtain a median throughput around 1.2 Mbit/s, the best effort median throughput is limited to around 400 kbit/s. The CQA scheme is far ahead, with an impressive median throughput of 3.8 Mbit/s.

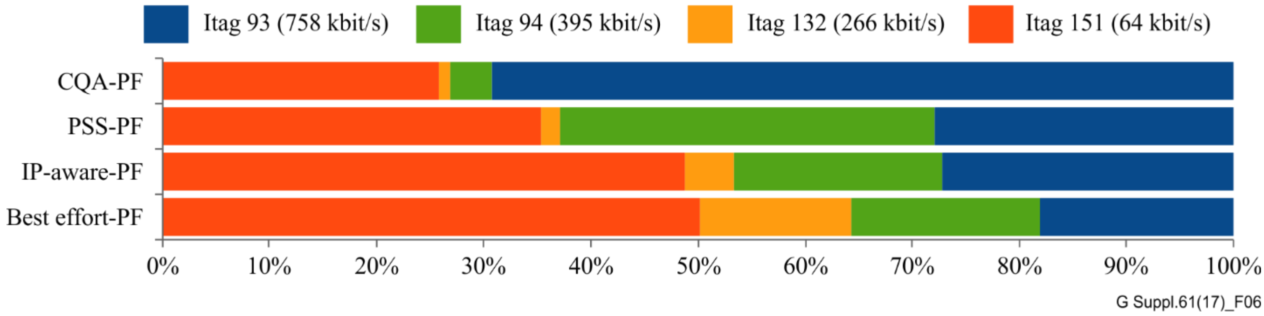

Figure 6 – YouTube chunk distribution

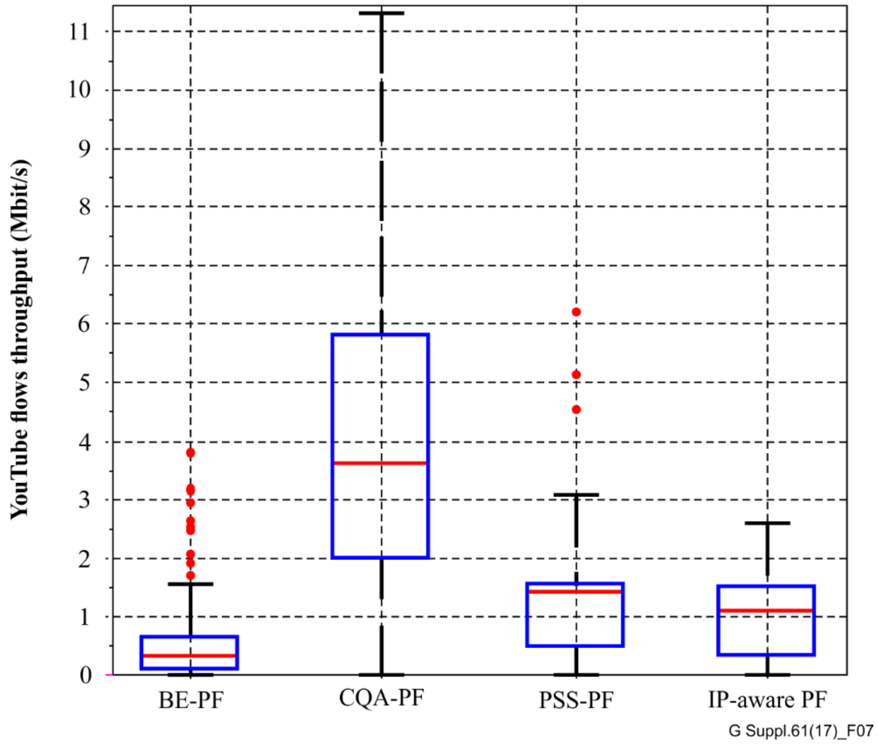

Figure 7 – YouTube flow throughput

Figure 8 provides the PLT of the websites listed in Table 3. As mentioned previously, the CQA algorithm strongly favours GBR bearers at the expense of web traffic, supported by non-GBR bearers. Figure 8 shows that the PLT systematically exceeds the timeout fixed at 30 s in the CQA scenario. Note also that the best effort scheme obtains the longer PLT, but no timeout, whereas in most cases the IP aware scenario achieves better performance than the PSS scheme, except for the largest website (boutique.orange.fr), where the PSS scheme has an advantage of 1 s in PLT.

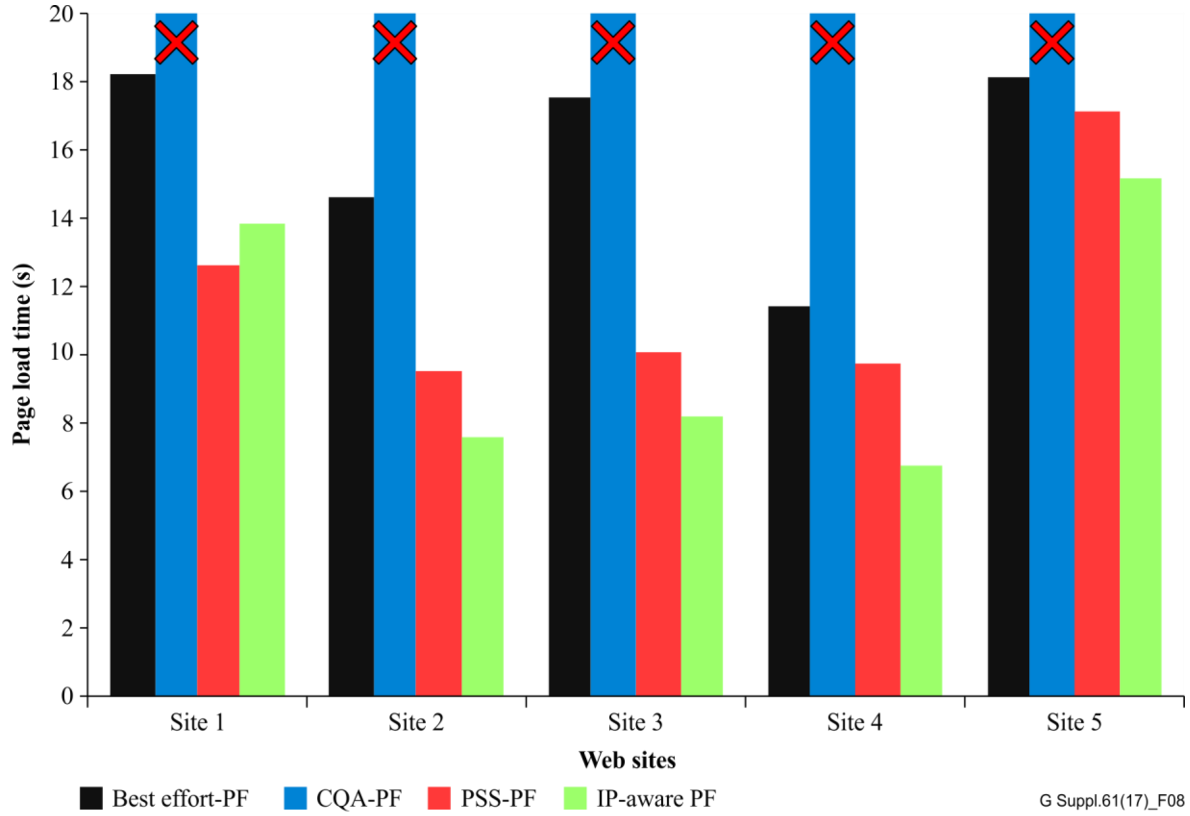

Figure 8 – Web page load time

10 Conclusions and perspectives

In this Supplement, a cross-layer design for an IP-centric QoS model was presented together with its performance evaluation in a realistic environment.

Simulations have highlighted the efficiency of the IP aware scheme notably when multiple flows with different QoS requirements are simultaneously supported by a terminal. In particular, it has been shown that the VoIP MOS obtained in the IP aware scenario is not significantly different from those obtained with standardized multi-bearer scenarios. Moreover, the IP aware scheme preserves the global throughput of the cell, contrary to multi-bearer scenarios with GBR bearers, which may strongly degrade the global cell capacity.

For these reasons, the IP-centric model is thought to offer a promising scheme for mobile networks as it provides for cost-effective QoS management with a performance level similar to those of standardized solutions.

Finally, note that the underlying IP-centric model is compatible with QoS management in fixed internet networks. To this extent, the IP-centric model allows for a unified solution in a fixed/mobile convergent world, which opens the door to access-agnostic applications.

To complete the results of this study, it is necessary to:

– define the necessary details of the packet-oriented QoS management model presented in this contribution; this may include: recommended number of classes of service, details of IP aware AP behaviour when the allocation of the radio resource depends on the downstream traffic mix, details of the interactions between IP queues and radio layer.

– Derive relevant performance parameters based on this model. New metrics may be defined (see for instance the discussion in [ITU-T SG12-75] and [ITU-T SG12-76]).

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	**Transmission systems and media, digital systems and networks**
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2017