



INTERNATIONAL TELECOMMUNICATION
UNION

TELECOMMUNICATION
STANDARDIZATION SECTOR

STUDY PERIOD 2013-2016

July 2015

English only

Original: English

Question(s):

STUDY GROUP 12

Source: Chemnitz University of Technology, Germany

Title: QoE Evaluation and Enforcement Framework for Internet Services

ABSTRACT:

In order to achieve acceptable service quality, the broad spectrum of Internet services requires differentiated handling and forwarding of the respective traffic flows within increasingly “Internet Protocol (IP)” based mobile networks. The “3rd Generation Partnership Project (3GPP)” standard based procedures allow for such service differentiation by means of dedicated “GPRS Tunnelling Protocol (GTP)” tunnels, which need to be specifically setup and potentially updated as the mixture of client initiated service consumption changes. The ISAAR (Internet Service quality Assessment and Automatic Reaction) framework augments existing quality of service functions in mobile networks by flow based network centric quality of experience monitoring and enforcement functions. The following chapters describe the ISAAR architecture in chapter 2 followed by the explanation of the monitoring entity QMON in chapter 3 with the QoE estimation and realisation of the rules entity in chapters 4 and the enforcement entity in chapter 5. In chapter 6 the summary and outlook are given.

1 Introduction

Internet based services have become an essential part of private and business life and the user experienced quality of such services is crucial for the users' decision to subscribe and stay with the service or not. However the experienced service quality results from the whole end-to-end line-up from participating entities. It is starting from the service generation, covering potentially several transport entities and finishing up in the application displaying or playing the result on the end device's screen or audio unit. However, the contributing performances of the individual service chain parties can often not be separately assessed from the end user perspective. Sluggish service behaviour can thus stem from slow server reaction, transport delay or losses due to congestion along the forwarding path as well as from the end device capabilities and load situation during the information processing and output. More insight can be gained from the mobile network perspective, which potentially allows for a differentiated assessment of the packet flow transport together with a transparent and remote Quality of Experience (QoE) estimation for the resulting quality on the end device. User satisfaction and user experienced service quality are strongly correlated and lead - from an Internet service provider point of view - either to an increase in subscription numbers or to customer churn towards competitors. Neither the capabilities and load situations on end devices nor the performance of content provider server farms nor the transport performance on transit links can be influenced by the operator of a mobile network. Therefore, this QoE framework will concentrate on the monitoring and enforcement capabilities of today's mobile and fixed operator networks in terms of differentiated packet flow processing and potentially SDN enabled forwarding. Since all competing providers will face similar conditions on either end of the service chain, the emphasis on the provider own match between service flow requirements and attributed network resources in a cost efficient manner will be key for the operator business success. The Internet Service quality Assessment and Automatic Reaction (ISAAR) quality of experience framework takes this situation into account and leverages the packet forwarding and traffic manipulation capabilities available in modern fixed and mobile networks. It focuses on LTE and LTE Advanced networks, but is applicable to the packet domains in 3G and even 2G mobile networks as well. Since different services out of the broad variety of Internet services will ideally require individual packet flow handling for all possible services, the ISAAR framework will focus only on the major service classes for cost and efficiency reasons. The set of tackled services is configurable and should sensibly be limited to only the major contributing sources in the overall traffic volume or the strong revenue generating services of the operator network. The current Sandvine Internet statistic report 0 for instance shows that only HTTP, Facebook and progressive download video services alone cover about 65% of the overall network traffic.

2 QoE Framework Architecture

The logical architecture of the ISAAR framework is shown in Figure 1. ISAAR consist of three functional entities which are the QoE Monitoring (QMON), the QoE Rules Unit (QRULE) and the QoE Enforcement (QEN). QMON provides modular service specific quality assessment functionality for selected classes of services combined with QRULE to create the per flow behaviour and QEN for the quality enforcement. The assessment as well as the enforcement is done for service flows on packet and frame level. It incorporates PCC mechanisms as well as packet and frame prioritisation in the IP, Ethernet, and the MPLS layer. MPLS as well as OpenFlow can also be used to perform flow based traffic engineering to direct flows in different paths. ISAAR's modular structure in the architecture elements allows for later augmentation towards new service classes as well as a broader range of enforcement means as they are defined and implemented. Service Flow Class Index and Enforcement Database register the available detection, monitoring and enforcement capabilities to be used and referenced in all remaining components of the architecture.

The framework architecture is 3GPP independent but can also interwork with the 3GPP PCC. If available, it also can make use of flow steering in SDN networks using OpenFlow. The interworking with 3GPP is mainly realized by means of the Sd interface 0 (for traffic detection support), the Rx interface (for PCRF triggering as application function and thus triggering the setup of dedicated bearers) and the Gx / Gxx interface 0 (for reusing the standardized Policy and Charging Enforcement Function (PCEF) functionality as well as the service flow to bearer mapping in the BBERF).

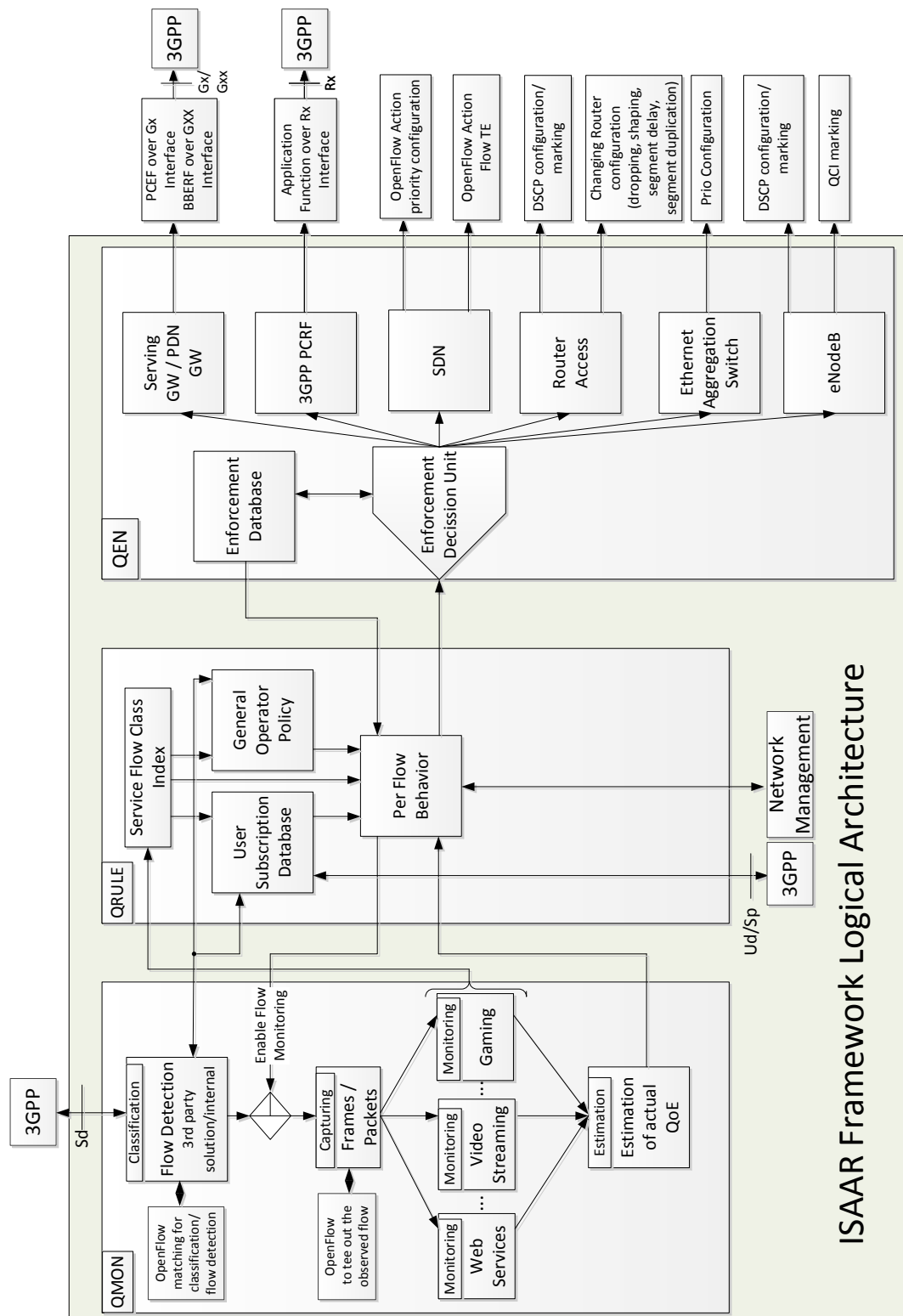


Figure 1 ISAAR framework

This independent structure generally allows for its application in non-3GPP mobile networks as well as in fixed line networks.

3 QoE Monitoring (QMON)

Today's mobile networks carry a mix of different services. Each traffic type has its own network transport requirements in order to live up to the user expectation. To observe the achieved transport quality and its resulting user service experience, network operators need to monitor the QoE of the respective services. Since the quality of service experienced by the user is not directly measurable within the network, a new method is required, which can calculate a QoE Key Performance Indicator (KPI) value out of measurable QoS parameters. The most challenging and at the same time most rewarding service QoE estimation method is the one for video streaming services. Therefore, the paper will focus on video quality monitoring and estimation, not limiting the more general capabilities of ISAAR for all sorts of service KPI tracking. YouTube is the predominant video streaming service in mobile networks and ISAAR is consequently delivering a YouTube based QoE solution first. Within this YouTube monitoring we are able to detect and evaluate the QoE of MP4, Flash Video (FLV) as well as WebM video in Standard Definition (SD) and High Definition (HD) format. There are some client based video quality estimation approaches around (e.g. the YoMo application [1]), but we consider such end device bound solutions as being cumbersome and prone to manipulation. Therefore, ISAAR will not incorporate client-side solutions but concentrates on simple, transparent and network-based functionality only. Some other monitoring solutions follow a similar way of estimation, like the Passive YouTube QoE Monitoring for ISPs approach [2]. However, they are not supporting such a wide range of video encodings as well as container formats. Another approach is the Network Monitoring in EPC [3] system, but this does not focus on flow level service quality. The flow monitoring which is used in the ISAAR framework is explained in chapter 3.2. However, before the QoE of a service can be estimated, the associated data flow needs to be identified. Chapter 3.1 explains the flow detection and classification in detail.

3.1 Flow Detection and Classification

Before the QoE of a video stream can be estimated, the stream has to be identified within the traffic mix. In the current QMON implementation a minimalistic built-in Deep Packet Inspection (DPI) algorithm is used to detect TCP based video streams. It is a very basic DPI implementation with some restrictions, e.g. the traffic needs to be unencrypted and the typical video data patterns in the payload have to be known by the measurement system. Therefore, in a simple demo implementation the build-in detection algorithm is limited to video streams which comprise supported video payloads in unencrypted TCP traffic flows. The use of QMON in a production network should not rely on this light weight DPI system, but rather make use of professional DPI solutions. The coupling of QMON with Sandvine DPI machines using the Gx interface is currently worked on.

3.2 Video Quality Measurement

Traditionally, video QoE monitoring solutions were focusing on fine grained pixel error and block structure errors. However such Key Performance Indicators (KPIs) are not suitable for progressive download video streams, since YouTube and other popular video portals are using the so called pseudo streaming scheme that downloads the video file without losses into a play out buffer first and plays it out from there. Due to the data correctness ensured by TCP and the equalized transport delays by the buffering, pixel errors due to bad QoS transport parameters can no longer occur. The main cause for bad quality of progressive download videos are therefore stalling events due to delayed data reception and resulting buffer depletion times. Thus, QMON focuses on the occurrence and duration of playback stalls only. To determine these events it is necessary to estimate the fill level of the play out buffer and to detect depletion events. Due to the fact that QMON does not have access to the user's end device, it

relies on the data which can be observed at a measurement point within the network. The required information needs to be extracted out of TCP segments since YouTube and other progressive download streaming services are based on HTTP/TCP transport. Therefore, the TCP segment information and the TCP payloads of the video flow have to be analysed. This analysis of the TCP based video download derives the estimated buffer fill level based on the video timestamps encoded within the video payload of the respective TCP flow. For this extraction it is necessary to decode the video data within the payload. After determining the play out timestamp it is compared to the observation timestamp of the corresponding TCP segment.

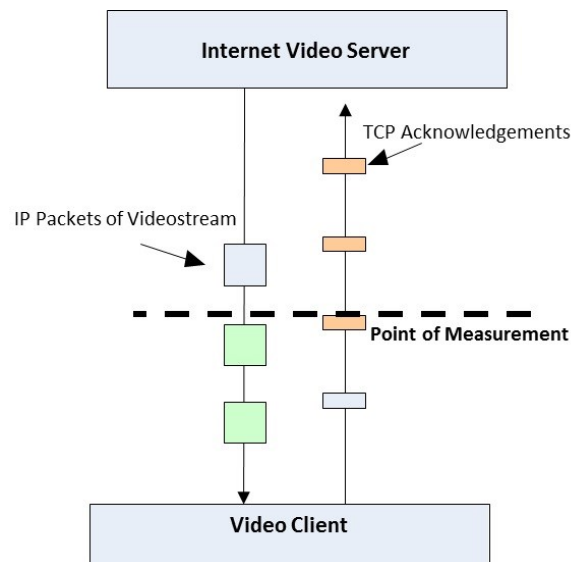


Figure 2: Video quality estimation scheme 0

The estimation process is shown in Figure 2. The result of this comparison is an estimate of the fill level of the play out buffer within the client's device. This estimation is done without access to the end device. The network based QoE measurement setup is shown in Figure 3.

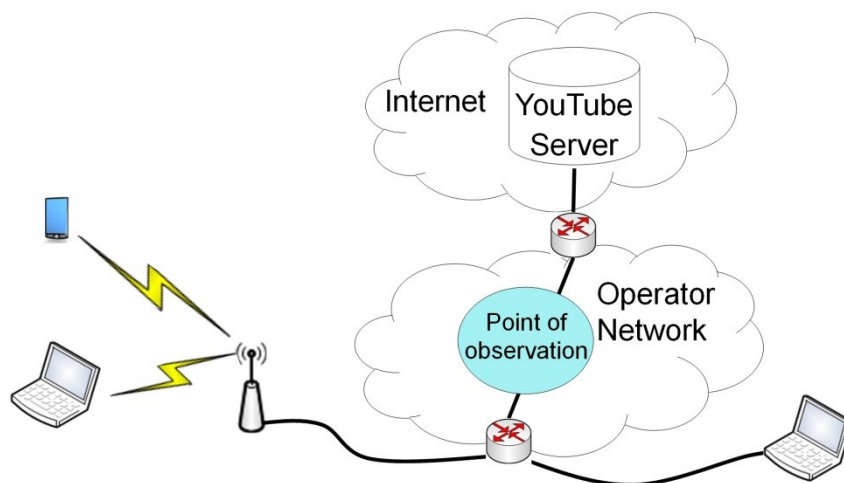


Figure 3: QoE measurement setup

3.3 Video Quality Rating

A 5 point Mean Opinion Score (MOS) is used as a common scale for user experience. The MOS is calculated due to the occurrence of stalling events. Each stall decreases the MOS. The

impairment decrease of a single stall event depends on the number of previously occurred stalls and follows an e-function which best reflects the quality perception of human beings. For each video an initial buffering stall is taken into account which does not influence the perceived quality if it is below 10 seconds. The exact quality estimation function is shown below where x represents the number of stalling events:

$$MOS = e^{-\frac{x}{5}+1,5} \quad (1)$$

The resulting scale can be found in Figure 4.

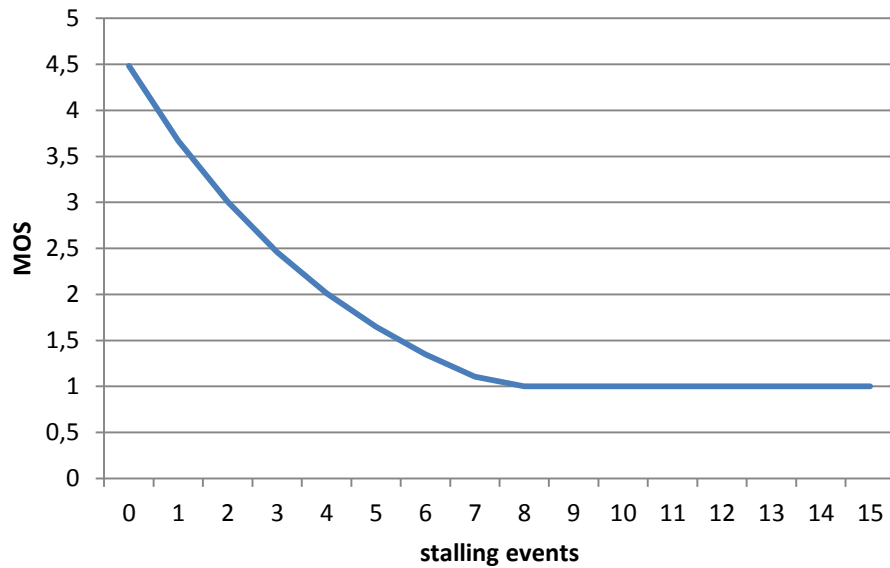


Figure 4: MOS vs. stalling events

Figure 5 displays the estimated buffer fill-level of a HD YouTube video. The amount of buffered video time hits the zero line 5 times therefore, five stalling events occurred during the video playback. The video stalling events take place at 18, 27, 45, 59 and 75 seconds and each stalling event decreases the video MOS according to equation (1).

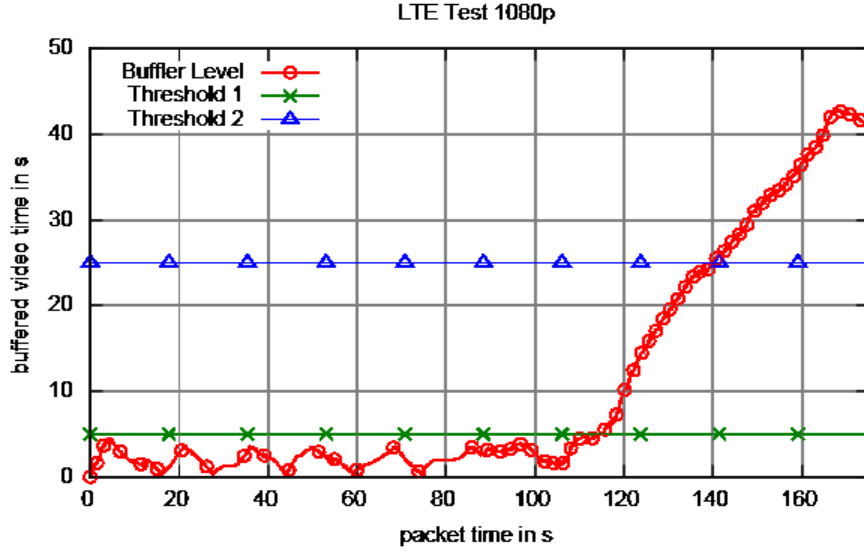


Figure 5: Example video buffer fill-level estimation

The resulting video quality is shown in Figure 6.

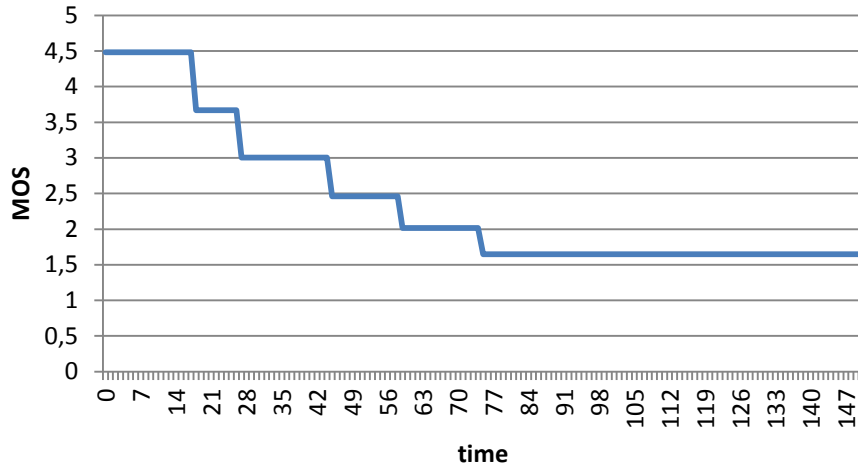


Figure 6: MOS video example according to equation (1)

However, the user experience in reality is not as simple as shown in the figures before. One of the problems is the memory effect of the QoE. That means the MOS is also improving over time if no further impairment has happened. Therefore, this effect has to be taken into account within the quality estimation formula. A time dependency of the influence of stalling events has been modelled again with a weighted e-function. That means the MOS estimate can recover if the video is running smoothly. To incorporate the memory effect portion, equation (1) has been changed as shown below in equation (2), where x represents again the number of stalling events, t depicts the time since the last stall happened in seconds and α is a dimensioning parameter which adjusts the influence of the memory effect. α has been set to 0.14 for the shown figures.

$$MOS = e^{-\frac{x}{5} + 1.5 - \alpha \cdot \sqrt{t}} \quad (2)$$

x = number of stalls

t = time since last stall

$\alpha = \text{memory parameter}$

Figure 7 shows the calculated video score respecting the memory effect as shown in equation (2).

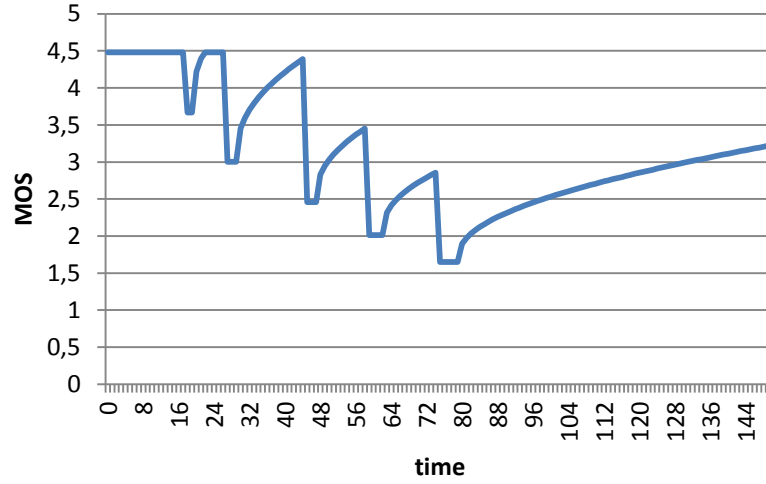


Figure 7: MOS video QoE according to equation (2)

3.4 Location aware monitoring

Due to the fact that it is not possible to measure all streams within an operator network, a subset of flows has to be chosen either randomly or in a policy based fashion. For example, the sample flows could be selected based on the criteria of which tracking area the flow goes into. If it is possible to map the eNodeB cell IDs to a tracking area, the samples also can be selected in a regionally distributed fashion. With that, it could be decided whether a detected flow is monitored or not due to the respective destination region. Over the time, this sample selection procedure can shift the policy focus to regions with poor QoE estimation results in order to narrow down the affected regions and network elements.

4 QoE POLICY AND RULES (QRULE)

In this chapter the QoE Policy and Rules entity of the ISAAR framework is presented. QRULE decides which priority is assigned to a certain service. Therefore, the flow information as well as the estimated QoE for each observed flow are obtained from QMON. It also contains a service flow class index where all measurable service types are stored. The enforcement actions for the required flow handling are determined based on the information obtained from QMON and listed in the subscriber Database as well as the general operator policy. Also the enforcement database within the QEN is taken into account. Combining all this information the QRULE maps observed QoE to the Per Flow Behaviour (PFB) for each data stream managed by ISAAR. PFBs are defined by appropriate marking of packets and frames. Each PFB has to be specified.

As an example the measured video information presented in Figure 5 are used to show the PFB for video streams. Here, three possible PFBs (corresponding to three different markings) are provided. These PFBs depend on the measured QoE for that the main KPI is the buffer fill level. Two buffer fill level thresholds are defined in the figure: $th1 = 5$ seconds and $th2 = 25$ seconds. If the QoE is poor, i.e. the video buffer fill level is below threshold 1 ($t < th1$), the EF class (101110) should be used. If the fill level is between threshold 1 and 2 ($th1 < t < th2$) a DSCP value like CS5 (101 000) should be chosen, because the video QoE is sufficient. Finally if the fill level exceeds threshold 2 ($th2 < t$) a DSCP value with a lower priority like BE (000 000) or LE (001 000) is taken, so that other flows might get preferred access to the resources.

QRULE also decides which kind of marking is deployed depending on the technology accessible in the network. It is possible to apply different kinds of priority markings like IP DiffServ, Ethernet priority, MPLS traffic class marking and for mobile networks the QCI tunnel mapping for GTP could be used. The rules unit has to ensure that there are no oscillating effects in the network. Oscillating could occur on flow level if one flow that is lifted up in priority causes quality impairments for the neighbouring flows. Thus, the second flow will also require enforcement actions, which in turn causes the first one to deteriorate again. To overcome this effect QRULE has to consider which flows were manipulated. Continuous action triggering is an early indication for such race conditions, which results in QRULE dampening of enforcement actions. Oscillating could also occur not only on flow level but on local area level within the network. Thus regional impairment mitigation should not cause increased levels of impairments in neighbouring regions. If this is being detected by location aware QMON, QRULE should also dampen enforcement actions. Close interworking of ISAAR with network management systems fosters this detection of oscillation situations and provides vital information for root cause analysis. If the majority of the traffic would need to be precededented in priority, ISAAR has simply hit its limitation. As these mechanisms are often used in combination, there must be a consistent mapping between them. This mapping is also performed by QRULE. Further details on the mapping can be found in 0. For future investigation ISAAR is prepared to incorporate the interworking of GTP and MPLS LSPs in a transparent fashion. Further details on the interworking can be found in 0.

5 QoE Enforcement (QEN)

The enforcement of the PFB is done in the third functional block of ISAAR, “QEN”. For data streams with a certain QoE QRULE determines the PFBs and QEN reacts accordingly by applying suitable mechanisms to influence the transmission of the involved data frames or packets. There are several ways to enforce the required behaviour. The first one is to use the PCRF/PCEF in mobile networks and trigger the setup of dedicated bearers via the Rx interface. A second option is to deploy layer 2 and layer 3 frame/packet markings. Based on these markings a differentiated frame/packet handling (scheduling, dropping) is enforced in the network elements which are traversed by the frames/packets (per hop behaviour). In case a consistent marking scheme across all layers and technologies is ensured by the QRULE entity, the QEN does not need to change the existing configuration of the network elements. With GTP tunnels in place, the priority marking has to be applied within the GTP tunnel as well as outside. The outside marking enables routers to apply differentiated packet handling also on GTP encapsulated flows without requiring a new configuration. For IPsec encrypted GTP the marking also has to be included into the IPsec header. The inner and outer IP markings are set in downstream and in upstream direction based on the flow information (five tuple) and the PFB obtained from QMON. As a third option - in case that the predefined packet handling configuration of the routers should not be used - the ISAAR framework is also able to perform a fully automated router configuration. With that, the QEN may explicitly change the router packet handling behaviour (e.g. packet scheduling and dropping rules) to influence the flows.

6 Summary

The ISAAR framework presented in this paper addresses the increasingly important quality of experience management for Internet based services in mobile and fixed networks. It takes the network operator’s position to optimize the transport of packet flows belonging to most popular video streaming, voice, Facebook and other web services in order to satisfy the customer’s service quality expectations. The framework is aware of the 3GPP standardized PCC functionality and tries to closely interwork with the PCRF and PCEF functional entities. However, ISAAR is also meant to work without the 3GPP functions. Therefore it sets up a three component logical architecture, consisting of a classification and monitoring unit (QMON), a decision unit (QRULE) and an enforcement unit (QEN) in order to selectively monitor and manipulate single service specific flows. The differentiated PFB is mainly achieved by deploying different priority markings. More sophisticated mechanisms for location aware service flow observation and steering as well as direct router respectively OpenFlow switch configuration access for traffic engineered flow routing are optionally available within the modular ISAAR framework. Due to the strong correlation between achieved video streaming QoE and customer satisfaction, the high traffic volume share of YouTube video streaming services is tackled first in the on-going ISAAR implementation activity. An optimized network-based precise video QoE estimation mechanism is coupled with automated packet flow shaping and dropping mechanisms guided by a QoE estimation. This way, a smooth play out with reduced network traffic demand can be achieved. Since ISAAR is able to work independently of 3GPP’s QoS functionality, it can be used with reduced functionality in any IP based operator network.

7 Acknowledgment

This work has been carried out within the CELTIC project MEVICO funded by the Federal Ministry of Education and Research (BMBF) of Germany (grant 01BU1013).

8 References

- [1] Sandvine: Global Internet Phenomena Report; 2011
- [2] 3GPP: TS 23.203 Policy and Charging Control Architecture. 3GPP standard (2012); http://www.3gpp.org/ftp/Specs/archive/23_series/23.203/23203-b60.zip
- [3] Ekström, H.: QoS Control in the 3GPP Evolved Packet System. In: IEEE Communications Magazine February 2009 pp. 76-83. IEEE Communications Society, New York (2009)
- [4] I Balbas, J. P., Rommer, S., Stenfelt, J.: Policy and Charging Control in the Evolved Packet System. In: IEEE Communications Magazine February 2009 pp. 68-74. IEEE Communications Society, New York (2009)
- [5] Ma, K. J., Bartos, R., Bhatia, S., Naif, R.: Mobile Video Delivery with HTTP. In: IEEE Communications Magazine April 2011 pp. 166-175. IEEE Communications Society, New York (2011)
- [6] Oyman, O., Singh, S.: Quality of Experience for HTTP Adaptive Streaming Services. In: IEEE Communications Magazine April 2012 pp. 20-27. IEEE Communications Society, New York (2012)
- [7] Ouellette, S., Marchand, L., Pierre, S.: A Potential Evolution of the Policy and Charging Control/QoS Architecture for the 3GPP IETF-Based Evolved Packet Core. In: IEEE Communications Magazine May 2011 pp. 231-239. IEEE Communications Society, New York (2011)
- [8] Alasti, M., Neekzad, B., Hui, L., Vannithamby, R.: Quality of Service in WiMAX and LTE Networks. In: IEEE Communications Magazine May 2010 pp. 104-111. IEEE Communications Society, New York (2010)
- [9] Sterle, J., Volk, M., Sedlar, U., Bester, J., Kos, A.: Application-Based NGN QoE Controller. In: IEEE Communications Magazine January 2011 pp. 92-101. IEEE Communications Society, New York (2011)
- [10] 3GPP: TS 29.212 Policy and Charging Control (PCC) over Gx/Sd reference point. 3GPP standard (2011);
- [11] Wamser F., Pries R., Staehle D., Staehle B., Hirth M.: YoMo: A YouTube Application Comfort Monitoring Tool; March 2010
- [12] Schatz R., Hossfeld T., Casas P.: Passive YouTube QoE Monitoring for ISPs. 2nd International Workshop on Future Internet and Next Generation Networks (Palermo, Italy): June 2012
- [13] Wehbi B., Sankala J.: Mevico D5.1 „Network Monitoring in EPC”, Mevico Project (2009-2012)
- [14] Rugel S., Knoll T. M., Eckert M., Bauschert T.: A Network-based Method for Measurement of Internet Video Streaming Quality; European Teletraffic Seminar Poznan University of Technology, Poland 2011; <http://ets2011.et.put.poznan.pl/index.php?id=home>
- [15] Tobias Hoffeld, Sebastian Biedermann, Raimund Schatz, Alexander Platzer, Sebastian Egger, Markus Fiedler: The Memory Effect and Its Implications on Web QoE Modeling, 2011
- [16] Knoll T. M.: Cross-Domain and Cross-Layer Coarse Grained Quality of Service Support in IP-based Networks; <http://archiv.tu-chemnitz.de/pub/2009/0165/>
- [17] Windisch, G.: Vergleich von QoS- und Mobilitätsmechanismen in Backhaul-Netzen für 4G Mobilfunk; 2008
- [18] Eckert M.: Analyse und automatisierte Konfiguration klassenbasierter Paketvermittlung; 2010

Contact:	Dr.-Ing. Thomas Martin Knoll Chemnitz University of Technology Germany	Tel: +49 (0) 371 531 33246 Fax: +49 (0) 371 531 833246 Email: knoll@etit.tu-chemnitz.de
Contact:	Dipl.-Ing. Marcus Eckert Chemnitz University of Technology Germany	Tel: +49 (0) 371 531 39581 Fax: +49 (0) 371 531 839581 Email: Marcus.eckert@etit.tu-chemnitz.de

Attention: This is not a publication made available to the public, but an internal ITU-T Document intended only for use by the Member States of ITU, by ITU-T Sector Members and Associates, and their respective staff and collaborators in their ITU related work. It shall not be made available to, and used by, any other persons or entities without the prior written consent of ITU-T.
--