# Part IV

# Resource and Mobility Management

# 14

# QoE Management Framework for Internet Services in SDN-Enabled Mobile Networks

Marcus Eckert and Thomas Martin Knoll
*Chemnitz University of Technology, Chemnitz, Germany*

## 14.1 Overview

In order to achieve acceptable service quality, the broad spectrum of Internet services requires differentiated handling and forwarding of the respective traffic flows in particular within increasingly overloaded mobile networks. The 3GPP procedures allow for such service differentiation by means of dedicated GPRS Tunneling Protocol (GTP) tunnels, which need to be specifically set up and potentially updated based on the client-initiated service traffic demand. The software defined networking (SDN)-enabled quality monitoring (QMON) and enforcement framework for Internet services presented in this chapter is named Internet Service quality Assessment and Automatic Reaction framework and will be abbreviated as ISAAR herein. It augments existing quality of service functions in mobile as well as software defined networks by flow-based network-centric quality of experience monitoring and enforcement functions. The framework is separated in three functional parts, which are QMON, quality rules (QRULE), and quality enforcement (QEN). Today's mobile networks carry a mixture of different services. Each traffic type has its own network transport requirements in order to live up to the user expectations. 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 chapter will focus on

---

*Software Defined Mobile Networks (SDMN): Beyond LTE Network Architecture*, First Edition.
Edited by Madhusanka Liyanage, Andrei Gurtov, and Mika Ylianttila.

video QMON 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 nowadays, and ISAAR is consequently delivering a YouTube-based QoE solution first. The KPI extraction and mapping to a measurable QoE value like the Mean Opinion Score (MOS) is done by QMON. The QRULE is supplied with the flow information and the estimated QoE of the corresponding stream by the QMON entity. The QRULE module also contains a service flow class index in which all measurable service flow types are registered. The enforcement actions for the required flow handling are determined based on subscription and policy information from the subscriber database and the general operator policy rule set. The third functional block in the ISAAR framework is QEN where the flow manipulation is performed. QRULE, that is, requests to change the per-flow behavior (PFB) of data streams with low QoE and QEN, reacts accordingly by applying suitable mechanisms to influence the transmission of the respective data frames or packets of those flows. One possibility to influence the data transmission is to use the PCRF/PCEF 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. 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 the SDN approach, there is a fourth possibility to influence data flows by using OpenFlow capabilities.

The first two sections state the current situation followed by the explanation of the ISAAR architecture in Section 14.4 and its internal realization in Sections 14.5, 14.6, and 14.7. In Section 14.7, the SDN demonstrator is presented and the summary and outlook are given in Section 14.9.

## 14.2 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 lineup from participating entities. It starts from the service generation, covers potentially several transport entities and finishes 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 behavior can thus stem from slow server reaction and 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 service 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 toward 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 networks in terms of differentiated packet flow processing and potentially software defined networking

(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 mobile network resources in a cost efficient manner will be key for the mobile operator business success. That applies especially for SDN-enabled networks, where a split between control and data path elements is made. This way, functions traditionally realized in specialized hardware can now be abstracted and virtualized on general-purpose servers. Due to this virtualization, network topologies as well as transport and processing capacities can be easily and quickly adopted to the service demand needs under energy and cost constraints. One of the SDN implementation variants is the freely available OpenFlow (OF) standard [1]. With OF, the path of packets through the network can be defined by software rules. OF is Ethernet based and implements a split architecture between so-called OF switches and OF controllers. A switch with OF control plane is referred to as "OF switch." The switch consists of the specialized hardware (flow tables), the secure channel for communication between switch and OF controller, and the OF protocol, which provides the interface between them [2]. The Internet Service quality Assessment and Automatic Reaction (ISAAR) QoE framework takes this situation into account and leverages the packet forwarding and traffic manipulation capabilities available in modern 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 [3], for instance, shows that only HTTP, Facebook, and YouTube services alone cover about 65% of the overall network traffic.

## 14.3 State of the Art

The standardization of mobile networks inherently addresses the topic of quality of service (QoS) and the respective service flow handling. The 3GPP-defined architecture is called Policy and Charging Control (PCC) architecture, which started in Release 7 and applies now to the Evolved Packet System (EPS) [4]. The Policy and Charging Rules Function (PCRF) is being informed about service-specific QoS demands by the application function (AF). Together with the Traffic Detection Function (TDF) or the optionally available PCRF intrinsic Application Detection and Control (ADC), traffic flow start and end events are detected and indicated to the PCRF. This in turn checks the Subscription Profile Repository (SPR) or the User Data Repository (UDR) for the permission of actions as well as the Bearer Binding and Event Reporting Function (BBERF) for the current state of already established dedicated bearers. As can be seen here, the 3GPP QoS control relies on the setup of QoS by reserving dedicated bearers. These bearers need to be set up, torn down for service flows, or modified in their resource reservation, if several flows are being bundled into the same bearer [5]. Nine QoS Class IDs (QCI) have been defined by 3GPP for LTE networks, which are associated with such dedicated bearers. Today, IP Multimedia Subsystem (IMS)-based external services and/ or provider own services make use of this well-defined PCC architecture and setup dedicated service flow-specific reservations by means of those bearers. Ordinary Internet services, however, are often carried in just one (default) bearer without any reservations and thus experience considerable quality degradations for streaming and real-time services. Therefore, network operators need to address and differentiate service flows besides the standardized QoS mechanisms of the 3GPP. HTTP-based adaptive streaming video applications currently amount the highest traffic share (see Ref. [3]). They need to be investigated for their application behavior, and appropriate actions should be incorporated in any QoS enhancing framework architecture. An overview of HTTP-based streaming services can be found in [6, 7]. There are many approaches found in the literature, which address specific services and potential enhancements. HTTP Adaptive Streaming Services (HAS) [8], for instance, is a new way to adapt the video streaming quality based on the observed transport quality. Other approaches target the increasing trend of fixed–mobile convergence (FMC) and network sharing concepts, which inherently require the interlinking of PCRF and QoS architecture structures and mechanisms (see, e.g., Ref. [9]). This architectural opening is particularly interesting for the interlinking of 3GPP and non-3GPP QoS concepts, but has not yet been standardized for close QoS interworking. The proposed interworking of WiMAX and LTE networks [10] and the Session Initiation Protocol (SIP)-based next-generation network (NGN) QoE controller concept [11] are just examples of the recent activities in the field. The ISAAR framework presented in this chapter follows a different approach. It aims for service flow differentiation either within single bearers without PCRF support or PCRF-based flow treatment triggering dedicated bearer setups using the Rx interface. This way, it is possible to use ISAAR as a stand-alone solution as well as aligned with the 3GPP PCRF support. The following sections document the ISAAR framework structure and work principle in detail.

## 14.4 QoE Framework Architecture

The logical architecture of the ISAAR framework is shown in Figure 14.1. The framework architecture is 3GPP independent but closely interworks with the 3GPP PCC. If available, it also can make use of flow steering in SDN networks using OF. This independent structure generally allows for its application in non-3GPP mobile networks as well as in fixed line networks. ISAAR provides modular service-specific quality assessment functionality for selected classes of services combined with a QoE rule and enforcement function. 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 prioritization in the IP, the Ethernet, and the MPLS layer. MPLS as well as OF can also be used to perform flow-based traffic engineering to direct flows in different paths. Its modular structure in the architecture elements allows for later augmentation toward 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. ISAAR is divided into three functional parts, which are the QMON unit, the QoE rules (QRULE) unit, and the QEN unit. These three major parts are explained in detail in the following sections. The interworking with 3GPP is mainly realized by means of the Sd interface [10] (for traffic detection support), the Rx interface (for PCRF triggering as AF and thus triggering the setup of dedicated bearers), and the Gx/Gxx interface [11] (for reusing the standardized Policy and Charging Enforcement Function (PCEF) functionality as well as the service flow to bearer mapping in the BBERF).

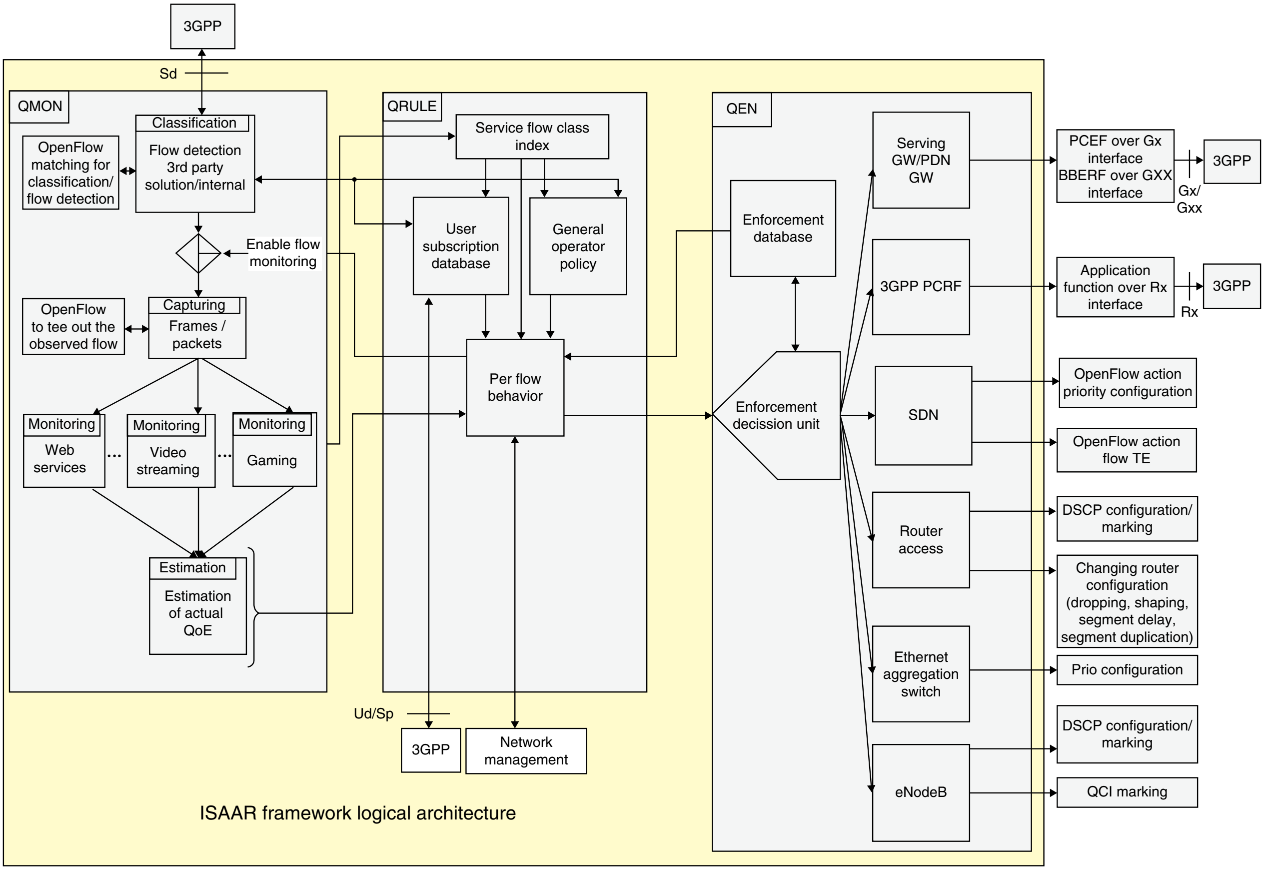

**Figure 14.1** SDN-enabled ISAAR framework.

Since ISAAR is targeting default bearer service flow differentiation also, it makes use of DiffServ code point (DSCP) markings, Ethernet prio markings, MPLS traffic class (TC) markings, as well as OF priority changes as available. This is being enforced within the QEN by Gateway and Base Station (eNodeB)-initiated packet header priority marking on either forwarding direction inside or outside of the potentially deployed GTP tunnel mechanism. This in turn allows all forwarding entities along the packet flow path through the access, aggregation, and backbone network sections to treat the differentiated packets separately in terms of queuing, scheduling, and dropping. The modular structure of the three ISAAR units (QMON, QRULE, and QEN) allows for a centralized as well as a decentralized deployment and placement of the functional elements.

## 14.5 Quality Monitoring

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 expectations. 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, this chapter will focus on video QMON 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 [12]), 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 QMON for ISPs approach [13]. 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 [14] system, but this does not focus on flow level service quality. The flow monitoring that is used in the ISAAR framework is explained in this section. However, before the QoE of a service can be estimated, the associated data flow needs to be identified. Section 14.5.1 explains the flow detection and classification in detail.

### *14.5.1 Flow Detection and Classification*

The ISAAR framework is meant to work with and without support of an external deep packet inspection (DPI) device. Therefore, it is possible to use a centralized DPI solution like the devices provided by Sandvine [15]. For unencrypted and more easily detectable traffic flows, the cheaper and more minimalist DPI algorithm that is built in the ISAAR framework can be used. In the first demo implementation, the build in classification is limited to TCP traffic, focusing on YouTube video stream detection within the operator's network. Extended with SDN support, there is a third possibility: given the proper configuration, the matching function from OF could be used to identify the supported service flows within the traffic mix. In the centralized architecture, the flow detection and classification is most suitably done by a commercial DPI solution. In this case, the QMON units have to be informed that a data stream was found and the classification unit has also to tell them the data stream-specific "five-tuple." Contained in the five-tuple are the source and destination IP address as well as the source and destination port and the used transport protocol. The QoE measurement starts as soon as the flow identification information (five-tuple) is available. Due to the new SDN features provided by OF, it is not only possible to identify specific data flows within the Internet. OF is also capable of teeing out a stream, which matches a specific pattern. Thereby, the QoE estimation could be distributed to different monitoring units, for example, depending on the specific Internet application. OF disposes the right flows to the right monitoring unit.

### *14.5.2 Video Quality Measurement*

Traditionally, video QMON solutions were focusing on fine-grained pixel error and block structure errors. However, such 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 playout 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 playout 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 that 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 analyzed. 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 playout timestamp, it is compared to the observation timestamp of the corresponding TCP segment [16].

The estimation process is shown in Figure 14.2. The result of this comparison is an estimate of the fill level of the playout 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 14.3.

### *14.5.3 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

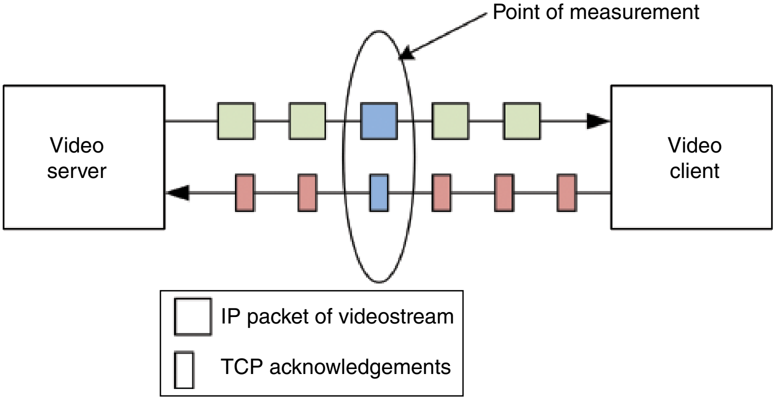

**Figure 14.2** Video quality estimation scheme.

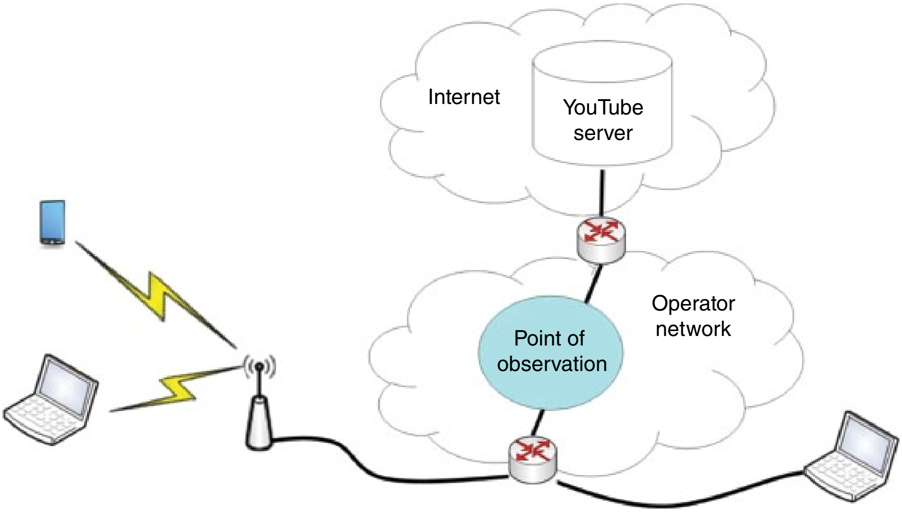

**Figure 14.3** QoE measurement setup.

stalls and follows an *e*-function that 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 s. The exact quality estimation function is shown below where *x* represents the number of stalling events:

$$\text{MOS} = e^{-x/5 + 1{,}5} \tag{14.1}$$

The amount of buffered video time hits the zero line five times; therefore, five stalling events occurred during the video playback. The video stalling events take place at 18, 27, 45, 59, and 75 s, and each stalling event decreases the video MOS according to Equation 14.1. The resulting video quality is shown in Figure 14.4.

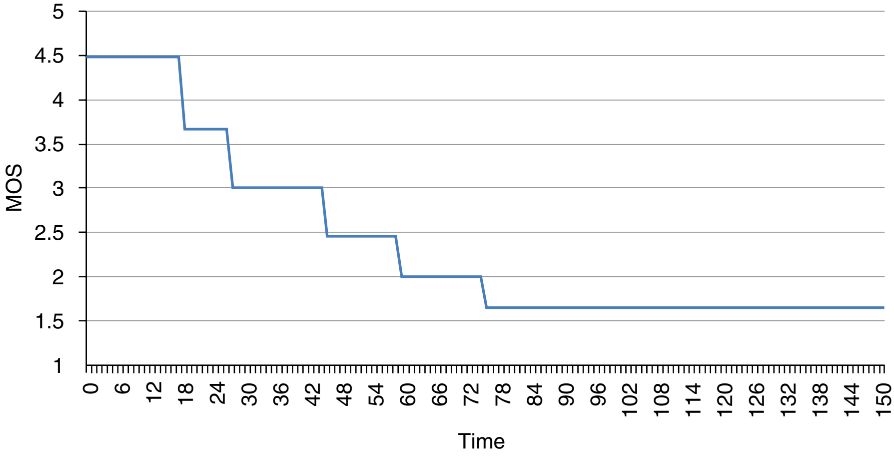

**Figure 14.4** MOS video example according to Equation 14.1.

However, the user experience in reality is not as simple as shown in the figures before. One of the problems is the memory effect [17] 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 modeled 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 14.1 has been changed as shown in Equation 14.2, 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:

$$\mathrm{MOS} = e^{x/5+1,5-\alpha\sqrt[e]{t}} \tag{14.2}$$

where $x$ = number of stalls; $t$ = time since last stall; and $\alpha$ = memory parameter.

Figure 14.5 shows the calculated video score respecting the memory effect as shown in Equation 14.2.

### 14.5.4 Method of Validation

Two methods of validation have been used to compare the QMON estimates with real user experiences. First, a group of test persons had been involved in the evaluation of the estimation method and the demonstrator. A test consisting of 17 YouTube videos (in all available resolutions) was set up. The videos were watched on laptops with mobile network access by the test users, and the data traffic was recorded at the Gi interface as the measurement point within the mobile operator's network. During the assessment, the users had to note down the occurrence of stalling events as well as their duration. Later, the recorded packet capture (PCAP) traces

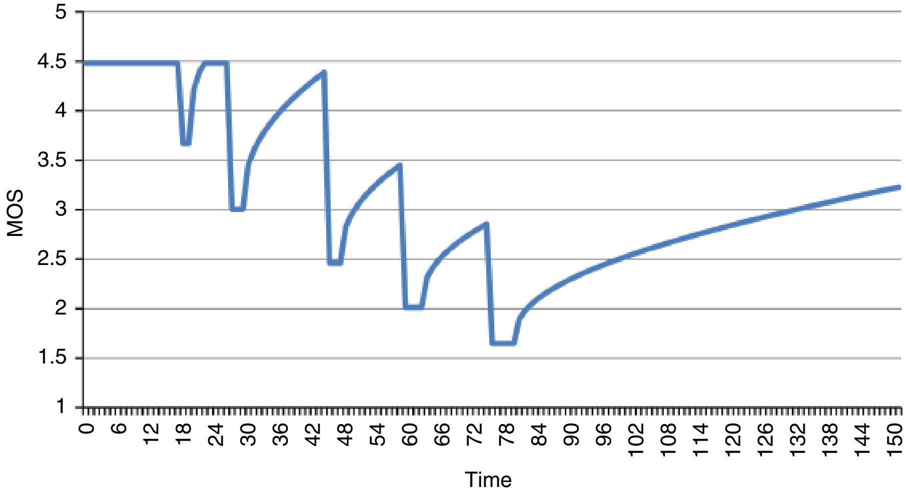

**Figure 14.5** MOS video QoE according to Equation 14.2.

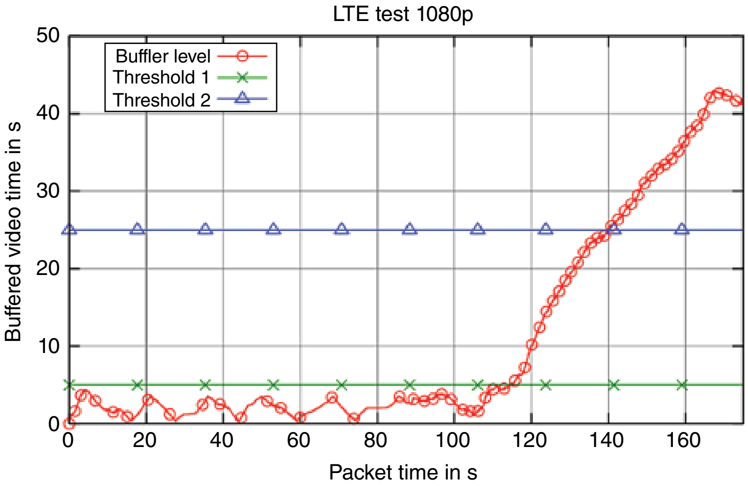

**Figure 14.6** Example video buffer fill-level estimation.

were processed with QMON. The results of both, the user assessment and the outcomes of the QMON calculation, were compared to each other to validate the functionality of QMON. In a second step, online monitoring was deployed (mainly in Long-Term Evolution (LTE) networks), where the live watching of videos was augmented with the QMON graph of the estimated buffer fill level (see, e.g., Figure 14.6) as well as the respective quality score (Figure 14.5). The comparison of observed stalling events in the video player and the zero-level buffer estimates in the QMON graph was used for evaluation.

### *14.5.5 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.

## 14.6 Quality Rules

In this section, the QRULE entity of the ISAAR framework is presented. The QRULE gets the flow information and the estimated QoE of the corresponding stream form the QMON entity. It also contains a service flow class index in which all measurable service flow types are stored. The enforcement actions for the required flow handling are determined based on information from the subscriber database and the general operator policy. Also, the enforcement database within the QEN is taken into account. Combining all this information, the QRULE maps the KPIs to the per-flow behavior (PFB) for each data stream managed by ISAAR. PFBs are defined by appropriate marking of packets and frames. Each PFB has to be specified. For video streams, three possible PFBs (corresponding to three different markings) are provided. These PFBs depend on the buffer fill level. In the example (Figure 14.7), two

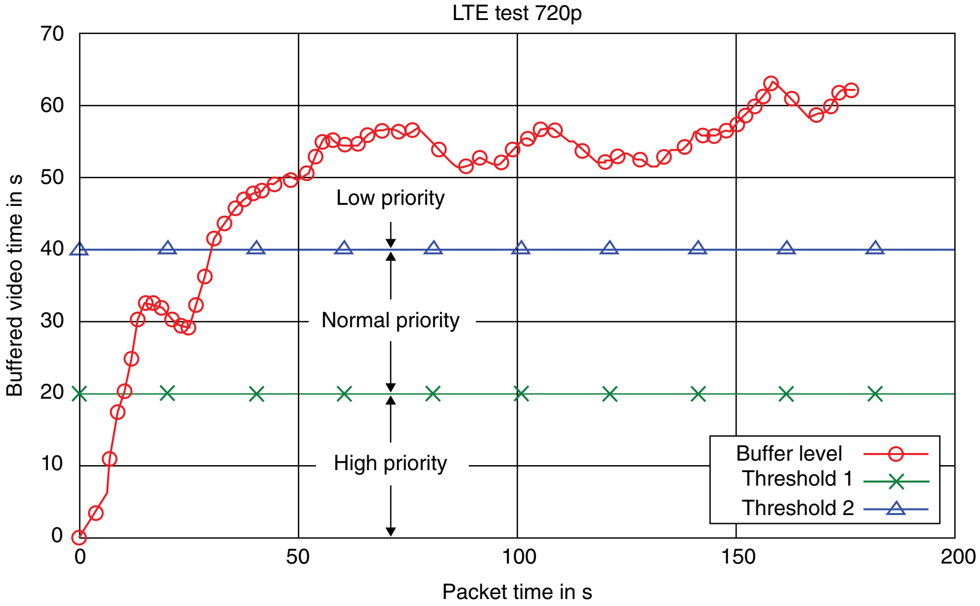

**Figure 14.7** Per-flow behavior dependent on the buffer fill level (YouTube example).

buffer fill-level thresholds are defined: th1 = 20 s and th2 = 40 s. If the QoE is poor, that is, the video buffer fill level is below threshold 1 ($t$ < th1), the EF class (101 110) should be used. If the fill level is between thresholds 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 networking technology. It is possible to apply IP DiffServ, Ethernet priority, MPLS TC marking, and QCI tunnel mapping for GTP. 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 neighboring 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 and in which location they are. Continuous action triggering is an early indication for such race conditions, which results in QRULE dampening of enforcement actions. That is, the transport impairment is such that ever-increasing priority is simply not solving the issue. 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 neighboring 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 precedented in priority, ISAAR has simply hit its limitation. If there are OF-enabled switches within the network, it is also possible to influence the priority of the frames belonging to a critical flow by changing the OF actions for that stream. As these mechanisms are often used in combination, there must be a consistent mapping between them. This mapping is also performed by the QRULE. Further details on the mapping can be found in Ref. [18]. For future investigation, ISAAR is prepared to incorporate the interworking of GTP and MPLS LSPs in a transparent fashion [19].

## 14.7 QoE Enforcement (QEN)

The enforcement of the PFB is done in the third functional block of ISAAR, "QEN." For data streams with a certain quality 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 behavior. 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 and dropping) is enforced in the network elements, which are traversed by the frames/packets (per-hop behavior). 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 behavior (e.g., packet scheduling and dropping rules) to influence the flows. With the SDN approach, a fourth possibility to influence data flows is realized by using OF features. For example, the priority of a flow can be changed in the forwarding configuration directly in an OF switch action list configuration. Furthermore, flow-specific traffic engineering could be realized. In order to use the OF features for flow enforcement, ISAAR is connected to the control interfaces of the SDN switches.

## 14.8 Demonstrator

To illustrate the QoE measurement, a demonstrator was used to process an example HD YouTube video for MOS calculation. The demo setup consisting of three laptops that are forming the SDN switch and SDN controller, another laptop where the QoE monitor was running, and two PCs that are generating background traffic is shown in Figure 14.8. The video is streamed from the video server to the video client through the SDN setup. The video traffic is copied out to the QMON device, which is evaluating the QoE of the video flow as described in Section 14.5. The video detection is done by matching rules within the SDN

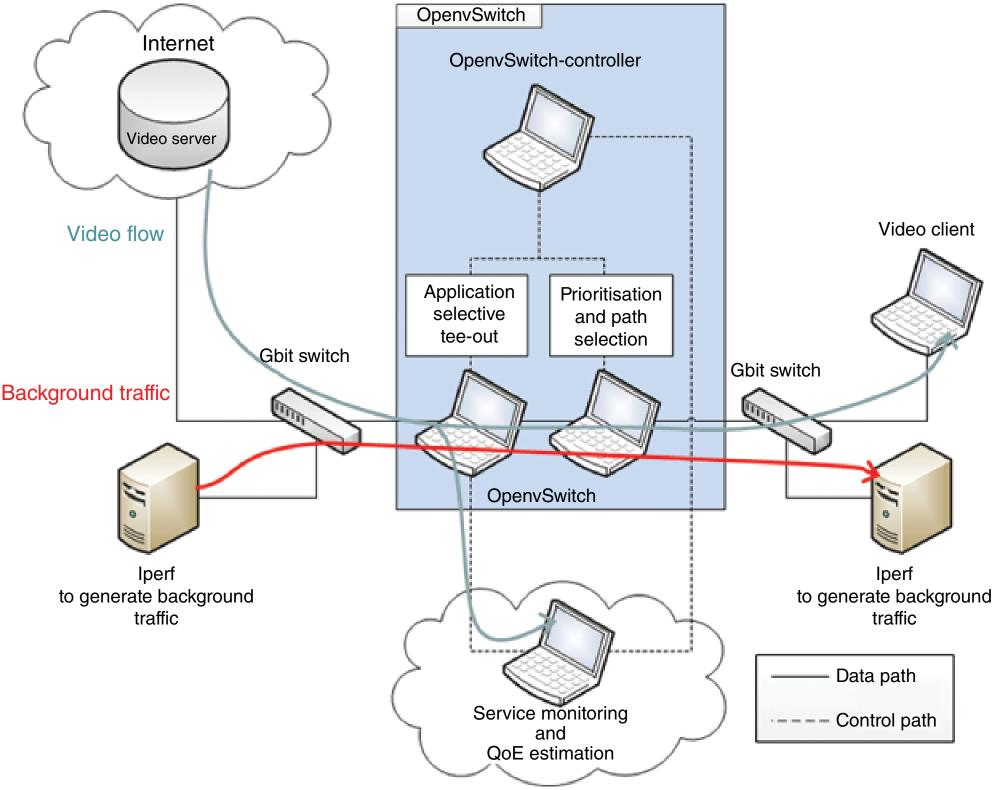

**Figure 14.8** SDN demonstrator setup.

switch. The switch is also used to change the priority of the video flow in case of high traffic loads. Therefore, two queues have been created inside the switches: one for the video stream and one for the other traffic. The SDN switch sorts the data packets into the right queue due to matching and action rules, which are configured by the controller.

Within the test, the video buffer was set to 10 s. The outgoing line to the video client has a data rate of 2 Mbps, the used video has an average bit rate of 800 kbps, and the background traffic is set to 1.4 Mbps. Therefore, without any traffic engineering, the line has to get congested due to an overuse of 200 kbps. In this experiment, no background traffic is applied to the network; only the video was transmitted. The SDN matching as well as the SDN enforcement had been switched off in that test. In the figure, it can be seen that the video buffer is filled with a plenty amount of data during the whole video playback, due to the 2 Mbps line that is only used by 800 kbps. Hence, there have no stalling events occurred and the QoE was not decreased. The second test is driven out without the SDN functionality but with applied background traffic. The results are shown in Figure 14.9. It can be seen that after the initial buffer event, the video playout is consuming the buffered data until the buffer level hits the zero line. In this moment, the video gets stuck and the MOS value and with it the QoE is decreasing. The stalling event itself reduced the QoE and the negative impact gets even higher each second the video is not playing. Therefore, the MOS value is falling until the video playback is restarted. After the playback is resumed, the memory effect kicks in and the MOS value is increasing as long as the video is playing. In the figure, you can see three major and one shorter stall of the video playback. As shown there the negative impact of each occurring stall is heavier as the impact of the previous. For a high video quality, such stalls have to be prevented.

However, now we applied the SDN QEN; the results can be found in Figure 14.10. The line is still limited to 2 Mbps and the background traffic is set to 1.4 Mbps; the video bit rate is not changed, too. But the video traffic can be put to another "high-quality" queue by the SDN

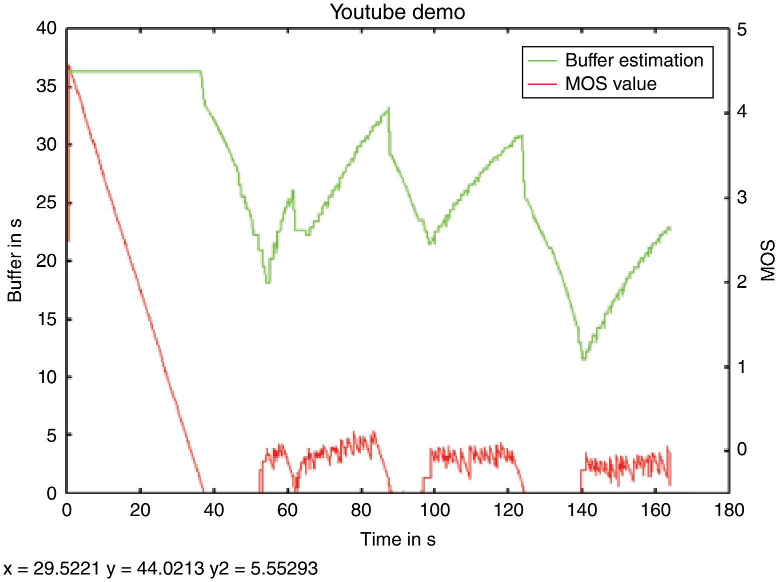

**Figure 14.9** Buffer fill level with background traffic without SDN.

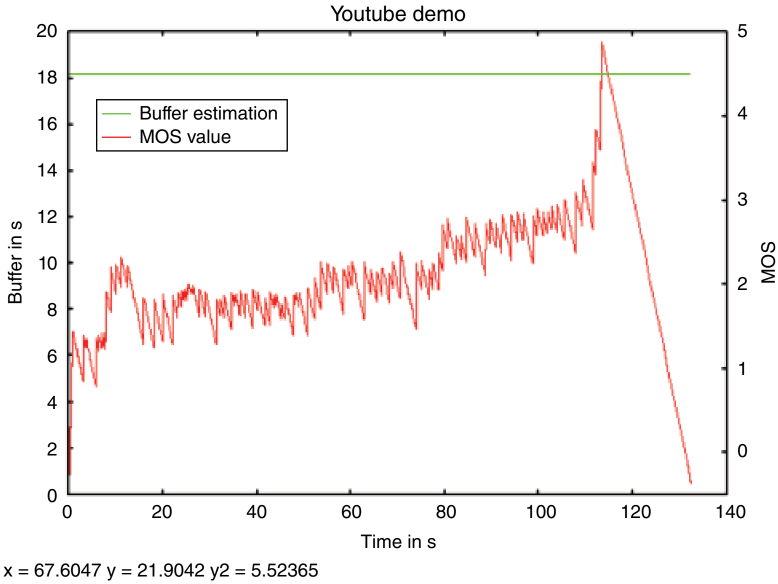

**Figure 14.10** Buffer fill level with background traffic with SDN.

controller. Therefore, the video buffer is filled with sufficient data over the whole video playback and no stalling events occurred. The demonstrator shows that it is possible to use SDN functions to detect specific traffic, copy it out, and enforce the needed QoE to it.

## 14.9 Summary

The ISAAR framework presented in this chapter addresses the increasingly important QoE management for Internet-based services in mobile 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, 3GPP QoS control is mainly based on dedicated bearers and observations in today's networks reveal that most Internet services are carried undifferentiated within the default bearer only. ISAAR therefore 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 with or without the standardized 3GPP QoS support. This is mainly achieved by priority markings on (potentially encapsulated) service flow packets making use of the commonly available priority and DiffServ capabilities in layer two and three forwarding devices. In the case of LTE networks, this involves the eNodeBs and SGWs/PGWs for selectively bidirectional marking according to the QRULE-determined service flow behavior. More sophisticated mechanisms for location-aware service flow observation and steering as well as direct router respectively OF 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 for mobile data services, the high traffic volume share of YouTube video streaming services is tackled first in the ongoing ISAAR implementation activity. An optimized network-based precise video QoE estimation mechanism is coupled with automated packet flow shaping and dropping means guided by a three-level playout buffer fill-level estimation. This way, a smooth playout with reduced network traffic demand can be achieved. To prove the functionality of the network-based video QoE estimation, a demonstrator has been implemented, which is capable of offline packet trace analyses from captured traffic as well as real-time online measurements. 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. In such setups, the service flow QoS enforcement would rely on IP DiffServ, Ethernet priority, and MPLS LSP TC marking as well as SDN-based flow forwarding only.

## References

[1] The OpenFlow Switch Specification. Available at http://OpenFlowSwitch.org. Accessed February 16, 2015.

[2] IBM; OpenFlow: The next generation in networking interoperability; 2011.

[3] Sandvine: Global Internet Phenomena Report; 2011.

[4] 3GPP: TS 23.203 Policy and Charging Control Architecture. 3GPP standard (2012).

[5] 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).

[6] 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).

[7] 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).

[8] 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).

[9] 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).

[10] 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).

[11] 3GPP: TS 29.212 Policy and Charging Control (PCC) over Gx/Sd reference point. 3GPP standard (2011).

[12] Wamser F., Pries R., Staehle D., Staehle B., Hirth M.: YoMo: A YouTube Application Comfort Monitoring Tool; March 2010.

[13] 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.

[14] Wehbi B., Sankala J.: Mevico D5.1 "Network Monitoring in EPC," Mevico Project (2009–2012). http://www.mevico.org/Deliverables.html. Accessed February 16, 2015.

[15] Sandvine Incorporated ULC: Solutions Overview. (2012); http://www.sandvine.com/solutions/default.asp. Accessed February 16, 2015.

[16] 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. Accessed February 16, 2015.

[17] Hoßfeld T., Biedermann S., Schatz R., Platzer A., Egger S., Fiedler M.: The Memory Effect and Its Implications on Web QoE Modeling; ITC '11 Proceedings of the 23rd International Teletraffic Congress 2011, pp. 103–110.

[18] 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/. Accessed February 16, 2015.

[19] Windisch, G.: *Vergleich von QoS- und Mobilitätsmechanismen in Backhaul-Netzen für 4G Mobilfunk*. Technische Universität Chemnitz, Chemnitz; 2008.