Quality of Experience and Access Network Traffic Management of HTTP Adaptive Video Streaming

Michael Thomas Seufert

Würzburger Beiträge zur Leistungsbewertung Verteilter Systeme

Herausgeber

Satz

Reproduktionsfähige Vorlage des Autors.
Gesetzt in LaTeX Linux Libertine 10pt.

Quality of Experience and Access Network Traffic Management of HTTP Adaptive Video Streaming

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Julius–Maximilians–Universität Würzburg

vorgelegt von

Michael Thomas Seufert

geboren in

Schweinfurt

Würzburg 2017

Danksagung

Dankbarkeit ist das Gedächtnis des Herzens.

– Jean Massieu (1772–1846)

Nicht nur meine Doktorarbeit, sondern auch ein großes Kapitel meines Lebens ist zu Ende gegangen und es gibt viele Personen, die mich in dieser Zeit besonders begleitet haben und deshalb meinen Dank verdienen.

Zuerst möchte ich dabei meinen Doktorvater Prof. Dr.-Ing. Phuoc Tran-Gia nennen, der mir die Möglichkeit gab in einem spannenden Thema zu promovieren und mir dabei immer mit Rat und Tat zur Seite stand. Er vertraute mir interessante Forschungsprojekte an und lies mir dabei trotzdem immer genügend Freiheiten, um meinen Forschungsdrang auch außerhalb der Projekte in verschiedenste Richtungen auszuleben. Er ermöglichte mir viele Dienstreisen zu Projekttreffen und Konferenzen, die meinen beruflichen und privaten Horizont erweiterten und mir neue Forschungsanreize und Bekanntschaften mit Kollegen und Freunden verschafften. Schließlich stellte er ein hervorragendes Team am Lehrstuhl zusammen, mit dem die freundschaftliche und produktive Zusammenarbeit auch oft über die Arbeit hinausging. Für alle diese kleinen und großen Dinge möchte ich mich bei Prof. Dr.-Ing. Phuoc Tran-Gia ganz herzlich bedanken.

Ich danke Prof. Dr. Burkhard Stiller, den ich als Koordinator meines ersten europäischen Forschungsprojektes schon bei meiner ersten Dienstreise kennenlernen durfte, dass er sich als Zweitgutachter meiner Dissertation zur Verfügung gestellt hat. Er hat mich mich mit wertvollen Kommentaren zu meiner Arbeit unterstützt und seine ausgezeichnete, organisierte Leitung des EU FP7

SmartenIT-Projekt wird mir immer ein Vorbild bleiben. Außerdem möchte ich Prof. Dr. Reiner Kolla danken, der sich bereit erklärte als Vorsitzender der Prüfungskommission und Prüfer bei der öffentlichen Verteidigung zu fungieren. Ebenso möchte ich Prof. Dr.-Ing. Samuel Kounev danken, dass er sich als Prüfer bei meiner Disputation zur Verfügung gestellt hat.

Einer der größten Förderer meiner wissenschaftlichen Ausbildung ist Prof. Dr. Tobias Hoßfeld. Er weckte während meiner Diplomarbeit das Interesse an der Forschung und setzte sich dafür ein, dass ich in seiner Gruppe am Lehrstuhl anfangen konnte. Seine guten Kontakte ermöglichten es mir davor schon internationale berufliche Erfahrungen am Forschungszentrum Telekommunikation Wien zu sammeln. Wie kein anderer verstand er es mich für neue Themen zu motivieren und mich aktiv zu unterstützen. Seine unzähligen Zeilen MATLAB-Code und unsere endlosen Diskussionen bleiben dabei unvergessen. Er hatte nicht nur als Gruppenleiter für alle meine Probleme ein offenes Ohr, sondern steht mir auch jetzt noch gern als Freund mit guten Ratschlägen zur Seite. Dafür möchte ich ihm ganz herzlich danken. Mein herzlicher Dank gebührt auch Dr. Florian Wamser, der als Gruppenleiter meinen Forschungshorizont für neue Themen öffnete und mir ein guter Freund wurde. Er setzte sich stets für mich ein und lies mich dabei auch immer hinter die Kulissen der akademischen Welt blicken. Ich schätze sein ausgeprägtes technisches Verständnis und möchte unsere ausschweifenden Diskussionen nicht missen.

Ich bedanke mich auch bei meinen anderen aktuellen und ehemaligen Kollegen am Lehrstuhl für Kommunikationsnetze Kathrin Borchert, Dr. Valentin Burger, Lam Dinh-Xuan, Dr. Steffen Gebert, Stefan Geißler, Nicholas Gray, Alexej Grigorjew, Dr. Matthias Hartmann, Dr. Matthias Hirth, Dr. David Hock, Dr. Michael Jarschel, Dr. Dominik Klein, Stanislav Lange, Dr. Frank Lehrieder, Frank Loh, Christopher Metter, Anh Nguyen-Ngoc, Dr. Christian Schwartz, Susanna Schwarzmann, Anika Schwind und Dr. Thomas Zinner. Das immer freundschaftliche und produktive Arbeitsklima machte es mir leicht den täglichen Gang zur Arbeit anzutreten. Darüber hinaus bin ich ihnen für die entstandenen Freundschaften und vielen Aktivitäten auch abseits der Arbeit dankbar. Ich

danke allen Beteiligten der Projekte für die gemeinsamen Diskussionen und interessanten Forschungsarbeiten.

Zu guter Letzt möchte ich mich bei meiner Familie und meinen Freunden für die tolle Unterstützung während meiner Doktorarbeit bedanken. Mein grenzenloser Dank gebührt meinen Eltern Christiane und Manfred Seufert, meiner Schwester Kathrin Seufert, sowie meinen Großeltern Erna und Hermann Stürmer. Sie haben mir zu jeder Zeit verlässlich den Rücken frei gehalten, alle meine Entscheidungen mitgetragen und mich mit kleinen und großen Taten größtmöglich unterstützt. Es gibt keine Worte, die beschreiben könnten, wie sehr ich dafür dankbar bin, dass sie meine Familie sind. Während der Doktorarbeit lernte ich auch meine Freundin Anika Schwind kennen, der ich für die schönen Stunden auch abseits des Arbeitsplatzes danke und dafür, dass wir über wirklich alles reden können. Auch wenn nun einige Veränderungen anstehen, freue ich mich schon weiterhin mit ihr gemeinsam in dieselbe Richtung zu blicken.

Contents

1 Introduction

Video streaming over the Internet is pervasive and constitutes a huge market. It embraces providers of contents, streaming solutions, and telecommunications. Content providers typically offer quality productions of live events or on-demand videos, such as movies, documentaries, sports events, television programs, or advertisements. While, in the past, such content was only watched in movie theaters or on television, videos can now also be consumed on other Internet devices with screens, such as personal computers and smartphones. Moreover, the ubiquity of Internet access and the increased functionality of Internet devices allow end users to record, upload, and distribute video content themselves, e.g., via specialized web portals for user-generated content. This enables also the implementation of live video conferencing, which features real time creation, transmission, and consumption of video content. Despite the wide availability, frequent usage, and plentiful possible applications of video streaming, the major challenge is to understand and improve the quality of video streaming, which is tackled in this thesis.

The goal of video streaming services and Internet Service Providers (ISPs) is to achieve a high Quality of Experience (QoE), which is a concept used to describe the subjectively perceived quality of end users with an Internet service [82]. A high QoE results in customer satisfaction and a reduced churn rate of the service. Therefore, the most popular video streaming services, including Netflix[1], YouTube[2], Amazon Video[3], and many Internet Protocol Television

---

[1] `https://www.netflix.com` [Online] – Accessed: 24.08.2017
[2] `https://www.youtube.com` [Online] – Accessed: 24.08.2017
[3] `https://www.amazon.com/Amazon-Video/b?&node=2858778011` [Online] – Accessed: 24.08.2017

1

(IPTV) providers, store the video content in distributed Content Delivery Networks (CDNs) and adopt HTTP Adaptive Video Streaming (HAS) technology for transmitting the video content to the end users. The HAS adaptation logic aligns the streamed video to the network conditions by switching among different video qualities, i.e., different bit rates, to ensure a smooth playback. However, it is not completely understood how bit rate adaptation affects the subjectively perceived quality, and a holistic QoE model for HAS is missing.
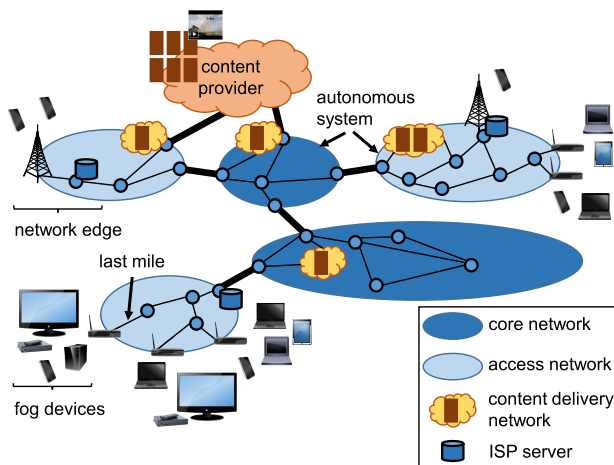
Figure 1.1: Video content delivery on the Internet.

A schematic illustration of the video content delivery on the Internet is depicted in Figure 1.1. End user devices in fixed or mobile access networks request videos, which are available on the data center or cloud infrastructure of the content provider, or can be distributed over a CDN to balance the load and bring them closer to the end users. The videos are transmitted from the closest point of presence via different Autonomous Systems (ASs), i.e., networks of ISPs, to the end user. As video traffic is very demanding in terms of high data volumes, bandwidth requirements, and delay constraints, ISPs additionally apply traffic

management to efficiently deliver video content through their networks. They can, for example, influence the routing or bandwidth allocation of streaming flows or cache popular videos on own servers in their networks, to avoid load and transit traffic, i.e., charged downstream traffic from other ASs, and ensure a high QoE of end users.

While CDNs and core networks have high capacities, the bottleneck of video traffic is often the last mile in the access network of the end users, i.e., the last link that reaches the end user's household or the base station. Especially operators of cellular access networks face huge loads and costly data transmissions, and thus, promote offloading of traffic flows to fixed networks. Offloading is mostly implemented via Wi-Fi, which is a term coined by the Wi-Fi Alliance[4] that comprises technology for Wireless Local Area Networks (WLANs) based on the Institute of Electrical and Electronics Engineers (IEEE) 802.11 standards[5][6]. Complementary Wi-Fi access is provided by public Wi-Fi hotspots with a high capacity fixed network backhaul link. However, Wi-Fi networks also prevail in home environments to connect the mobile devices of end users to the Internet. Thus, there is a growing Wi-Fi infrastructure, which can be utilized to relieve cellular networks, and deliver video traffic more efficiently. But although cellular operators take advantages, it is not clear if also end users can benefit from Wi-Fi offloading of video sessions.

To overcome the last mile problem, also the new edge and fog computing paradigms are considered, i.e., the instantiation of services and placement of data at the network edge close to end users, or even directly on end user devices. This attracts notice to the utilization of always-on and always-connected devices in home environments, such as residential gateways, Network-attached Storages (NASs), or Set-top Boxes (STBs). Virtualization technologies can be used to migrate service instances and run them on edge or fog devices. It also allows to move service instances and data, e.g., video content, according to the

---

[4]`http://www.wi-fi.org/` [Online] – Accessed: 24.08.2017

[5]`http://standards.ieee.org/about/get/802/802.11.html` [Online] – Accessed: 24.08.2017

[6]`http://grouper.ieee.org/groups/802/11/` [Online] – Accessed: 24.08.2017

mobility of the users. Thus, personalized service instances can always be run close to end users and can be accessed with short delays and high bandwidths. Still, mechanisms are needed, which incentivize and leverage the provision and sharing of end user resources to improve the QoE.

New traffic management solutions are designed to improve the QoE for Internet services, such as HAS. Cross-layer traffic management utilizes information from different layers, e.g., it considers network and application layer to both align the network management to the application needs, and adjust the application demands to the network conditions. Collaborative traffic management is based on information exchange and coordinated actions among different stakeholders. Additionally, social information, like user location, usage preferences, or content interests, can be gathered and exploited to enhance traffic management. This results in Socially-aware Traffic Management (SATM), which can be used to customize the service and its delivery to the individual needs of an end user to reach a high satisfaction with the service. However, it is not yet widely understood how SATM can be implemented and what performance gains can be expected.

Thus, there are several open challenges with respect to the QoE of video streaming. This monograph contributes to solving these challenges by presenting research on QoE, QoE-aware traffic management, and socially-aware traffic management for HAS. In the following, the scientific contribution is highlighted and the outline of this thesis is presented.

1.1 Scientific Contribution

This thesis focuses on understanding and improving the QoE of HAS. After identifying the most important QoE factors, possibilities for traffic management are considered to improve the subjectively perceived quality of video streaming in access networks. Therefore, an accurate QoE monitoring is required to notice quality degradations of the customers and appropriate traffic management actions have to be applied. As this vast research field cannot be covered

completely in a single monograph, only selected topics are studied. This thesis investigates different QoE monitoring approaches on network and application layer, and also evaluates the performance of selected QoE-aware and socially-aware traffic management approaches for HAS in access networks. Thereby, the focus is on general concepts and methodologies, and the presented research abstracts from specific technologies, if possible, to not limit the findings to systems, which could soon be outdated in today's fast-paced networking world. This means, streaming or networking technologies will not be investigated from a technology-centric perspective, for example, focusing on the performance of current video codecs or transmission protocols, but rather from a conceptual perspective. Thereby, the obtained results are more general and apply to all HAS systems.

Figure 1.2 visualizes the research activities, identifies the selected topics of this thesis, and embeds them into the research fields. Thereby, the cartography is split vertically into different layers, which loosely resembles the Open Systems Interconnection (OSI) model of telecommunication systems. However, the layers up to transport layer have been combined into the *network* layer, and, typical for the Internet model (TCP/IP model), the higher layers have been combined into the *application* layer. Today's highly interactive Internet services require the addition of a *user* layer, and the ubiquitous social networking and interactions among different users are reflected in the *social* layer. Horizontally, the cartography is divided into the three main research goals, which are *monitoring*, *traffic management*, and *performance evaluation*. Note that some of the publications might fit to more than one category. In this case, the publications are listed in the category, which yields the largest overlap. The works have been conducted within the European Union (EU) Framework Programme 7 (FP7) project *Socially-aware Management of New Overlay Application Traffic combined with Energy Efficiency in the Internet (EU FP7 SmartenIT)*, the EU Horizon 2020 (H2020) project *In-network Programmability for Next-generation Personal Cloud Service Support (EU H2020 INPUT)*, the sub-project *Monitoring and Analysis of Quality of Experience in Mobile Broadband Networks (EU H2020 MONROE/Mobi-*

QoE) of the project *Measuring Mobile Broadband Networks in Europe (EU H2020 MONROE)*, the Deutsche Forschungsgemeinschaft (DFG) projects *Entwurf und Bewertung neuer Mechanismen für das Internet der Zukunft - Neue Paradigmen und ökonomische Aspekte (DFG ÖkoNet)*, *Analyse und Optimierung des Trade-offs zwischen QoE und Energieeffizienz in Datenzentren (DFG QoE-DZ)*, and *Design und Bewertung neuer Mechanismen für Crowdsourcing als neue Form der Arbeitsorganisation im Internet (DFG Crowdsourcing)*, the European Cooperation in Science and Technology (COST) action *European Network on Quality of Experience in Multimedia Systems and Services (COST Qualinet)*, the project *Service Quality Definition and Measurement (NGMN P-SERQU)* of the Next Generation Mobile Networks (NGMN) Alliance, and several industry funded projects.

Figure 1.2: Cartography of the conducted research. The notation $[x]^y$ indicates that the scientific publication $[x]$ is discussed in Chapter y of this thesis.

The research activities can be roughly integrated into six fields. First, the QoE of Internet applications was investigated to find the most important factors that affect the subjectively perceived quality of such services. A special attention

was given to video streaming services, which are very popular and have high network demands. In Figure 1.2, these studies are highlighted in blue, and can be attributed to performance evaluation on user level. In this thesis, the results of the QoE research of [13–15, 39, 52] are presented. Based on the findings of previous QoE studies, the impact of adaptation-related parameters on the QoE of HAS is investigated. The impact of quality and time on each layer is identified and quantified, which allows to formulate and solve an optimization problem for HAS adaptation. It can be used to benchmark HAS adaptation logics with respect to QoE-optimal adaptation for given network conditions and video characteristics. A frequently used methodology for QoE research is crowdtesting, i.e., conducting subjective QoE studies via crowdsourcing. To obtain reliable results from a large, diverse, remote, unsupervised crowd of micro job workers is not trivial. Thus, appropriate methodology for crowdtesting had to be developed. The field of crowdtesting is categorized as monitoring on the user level and the corresponding publications are highlighted in violet, but crowdtesting methodology will not be detailed in this thesis.

After the QoE factors of Internet services, such as HAS, are known, traffic management can be applied to improve the perceived service quality. Thereby, the concepts of application-aware networks and network-aware applications were investigated and exploited. This field of QoE-aware traffic management is highlighted in green, and covers all research goals, i.e., monitoring, traffic management, and performance evaluation. This thesis studies QoE-aware traffic management for HAS in access networks, which does not require to operate on traffic aggregates, but allows to identify and manage single video flows. In this context, QoE monitoring approaches on application and network layer are presented following the publications [34, 45, 48, 61, 80]. Moreover, an analytical and simulative performance evaluation is conducted, which compares the QoE gain of different traffic management algorithms for a scenario, in which video streaming and web browsing flows share a bottleneck link [12], e.g., on a the backhaul link of a shared Wi-Fi home network or a mobile base station. Some other studies applied analytical methods to obtain accurate and scalable perfor-

mance evaluation results on abstract system models, and are highlighted in red but are not included in this thesis. Dedicated performance evaluation studies were also conducted on edge computing, which is a new paradigm that extends cloud computing by additionally utilizing computing resources at the network edge, e.g., servers at mobile base stations or even devices within the homes of end users. Personalized services can be instantiated or migrated at the network edge close to end users to support user mobility and achieve a high QoE. As edge computing goes beyond access network traffic management, these results are not included in this thesis.

QoE-aware traffic management was extended by additionally considering the end user, his shared resources, as well as social information, e.g., about his preferences, his interests, or his interactions with other users. This constitutes the new research field of Socially-aware Traffic Management (SATM), which is highlighted in orange. Similar to QoE-aware traffic management, SATM not only considers traffic management decisions and actions but also the monitoring of social information. The basic concept of SATM is introduced based on [1]. Afterwards, the focus is on Wi-Fi offloading as a SATM mechanism, which is not only widely considered by ISPs to reduce the load on stressed mobile access networks, but is also well suited in the home environment due to the prevalence of residential Wi-Fi networks. This thesis presents a hotspot location model for public Wi-Fi offloading [10], a SATM system for private Wi-Fi offloading in home environments [42, 56], and a performance evaluation on the QoE end energy efficiency of public and private Wi-Fi offloading of HAS flows [29, 44, 47].

In summary, this thesis is focused on the following research questions:

- How to model the impact of adaptation on the QoE of HAS?
- How to monitor the QoE of HAS on network and application layer?
- How to improve the QoE of HAS by resource allocation on bottleneck links?
- How to consider the end user, his shared resources, and social information for traffic management?
- How to utilize existing Wi-Fi infrastructure for improving the QoE of HAS?

1.2 Outline of the Thesis

The structure of this thesis is illustrated in Figure 1.3. After this introductory chapter, the three selected topics, which were presented above, are tackled in separate chapters. The studies within each chapter form sections and are visualized as boxes. The arrows between the boxes represent the relations between the sections, i.e., an arrow indicates that background or findings are utilized in the later section. Sections, which are visualized as blue boxes, start each chapter and introduce the background and related work of each research field. Green boxes represent result sections. In the end of each chapter, the lessons learned are summarized, which is indicated by orange boxes. In the following, the organization of this monograph is described.

Chapter 2 investigates the QoE of HAS. In order to understand HAS, Section 2.1.1 describes the underlying concept and technology. Previous studies and results on the QoE of video streaming are presented in Section 2.1.2. Section 2.2 investigates and quantifies the impact of adaptation-related parameters and time on layer on the QoE of HAS. These QoE findings from subjective crowdsourced QoE studies are used to formulate and solve an optimization problem for quality adaptation in Section 2.3. It can be used to benchmark HAS adaptation logics with respect to the QoE-optimal adaptation for a given network scenario and video characteristics. An exemplary performance evaluation of four different HAS adaptation logics is conducted. Section 2.4 summarizes this chapter.

Based on the QoE findings, Chapter 3 discusses QoE-aware traffic management for HAS for individual video sessions, e.g., in access networks. The relevant background is provided in Section 3.1. QoE monitoring is followed up in Section 3.2, which presents monitoring approaches on network and application layer, as well as an estimation of QoE factors from network parameters, and the corresponding results of subjective QoE studies. Section 3.3 investigates QoE-aware traffic management on a bottleneck link, which is shared by video streaming and web browsing flows, and presents analytic and simulative performance evaluation results. The lessons learned are described in Section 3.4.

In Chapter 4, SATM for HAS is tackled. After defining the concept of SATM in Section 4.1.1, the remainder of this chapter focuses on Wi-Fi offloading as a SATM mechanism, which can be well applied due to ubiquitous public Wi-Fi infrastructures and Wi-Fi network deployments in home environments. Section 4.2.1 presents a simple model for the locations of public Wi-Fi hotspots. In Section 4.2.2, a SATM platform for home routers is described, which can improve the QoE of HAS by offering Wi-Fi offloading, caching, prefetching, and content delivery. Section 4.3 evaluates the performance of Wi-Fi offloading in terms of QoE and energy efficiency, and additionally confirms the applicability of the hotspot location model of Section 4.2.1. Section 4.4 presents the lessons learned from this chapter.

Finally, Chapter 5 concludes this thesis and provides an outlook to future research activities.
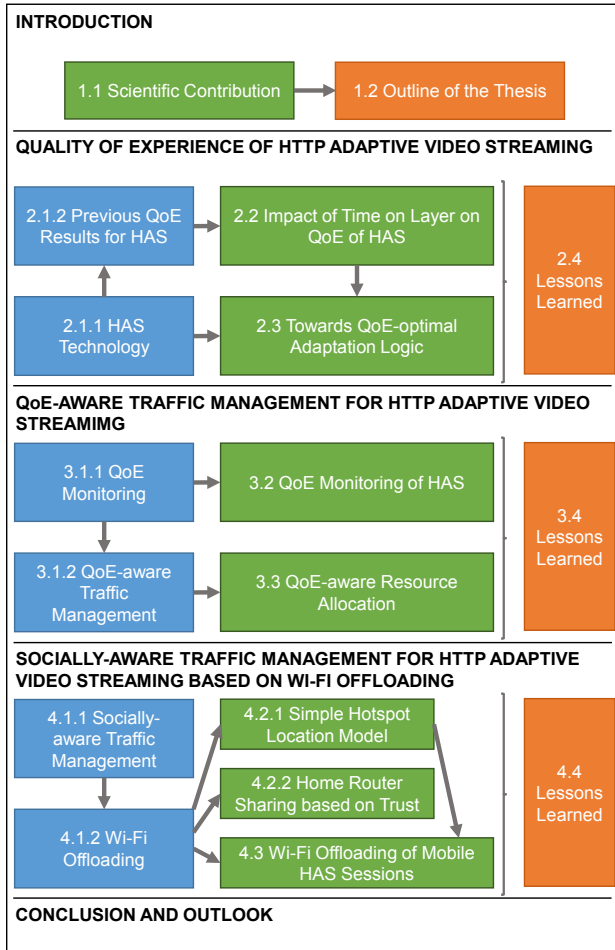
Figure 1.3: Organization and contribution of this monograph.

2 Quality of Experience of HTTP Adaptive Video Streaming

Video services on the Internet have evolved from offering mere downloads of video files to progressive downloads and streaming, which both describe the concurrent download and play out of media files. The currently prevailing streaming technology is HTTP Adaptive Video Streaming (HAS), which relies on standard web protocols (Hypertext Transfer Protocol (HTTP) and Transmission Control Protocol (TCP)) to promote a simple service implementation and high availability. The most important feature of HAS is the possibility to adapt the video bit rate to the network conditions. It is intended to ensure a smooth streaming when end users face throughput fluctuations, e.g., in mobile networks.

To understand and eventually improve Internet services like video streaming, application providers and Internet Service Providers (ISPs) need concepts to quantify the experience and satisfaction of their customers. Therefore, Quality of Service (QoS) was introduced, which is defined as "totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service" [83]. In practice, it describes the service delivery by objective network parameters, such as throughput, packet loss, delay, or jitter. However, for many services, QoS is not perfectly correlated with the customers' perception. With video streaming, for example, the user does not explicitly notice packet delays because they can be absorbed by the application buffer of the video player, but only perceives propagated network issues on application layer, e.g., playback interruptions when the buffer is empty (see

Section 2.1). Thus, the application behavior, which obviously depends on the underlying network, and its impact on customers' experience cannot be captured by pure QoS.

Thus, Quality of Experience (QoE), which is a concept of subjectively perceived quality, was introduced to complement QoS. QoE is defined as "the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state. [...] In the context of communication services, QoE is influenced by service, content, network, device, application, and context of use" [82]. QoE is highly application-dependent and, in order to understand all influence factors, extensive subjective studies have to be conducted. The results of these studies typically express QoE in terms of Mean Opinion Score (MOS) on an ordinal scale ranging from 1 (bad) to 5 (excellent) (cf. Absolute Category Rating (ACR) scale [84]), or in terms of acceptability scores.

This chapter focuses on the QoE of HAS and investigates the impact of quality adaptation on the subjectively perceived quality. Therefore, crowdsourced QoE studies are conducted to identify and quantify the influence of different adaptation-related parameters. The results allow to formulate and solve an optimization problem to compute the QoE-optimal adaptation for a given network scenario and video characteristics, which can be used to benchmark HAS adaptation logics in different evaluation scenarios.

Works on the QoE of HAS are summarized in [14], from which an overview on HAS technology and the most important QoE results will be presented in Section 2.1. Afterwards, the impact of the quality adaptation, namely, the amplitude and frequency of quality switches and the time on each quality layer, will be investigated in Section 2.2, which is based on [39, 52]. The results are used in Section 2.3, to compute QoE-optimal adaptation for benchmarking HAS adaptation logics on the basis of [13, 15]. Section 2.4 recaps the results of this chapter, which are the basis for the remainder of this work.

2.1 Background and Related Work

The first HAS solution was introduced by Move Networks in 2007[1] [85]. Afterwards, many proprietary streaming solutions were developed, the most popular solutions being Microsoft Silverlight Smooth Streaming (MSS) [86] by Microsoft Corporation in 2008, Apple HTTP Live Streaming (HLS)[2] by Apple Inc. in 2009, and Adobe HTTP Dynamic Streaming (HDS)[3] [87] by Adobe Systems Inc. in 2010. Although these solutions shared a similar technology (see Section 2.1.1), they were mutually incompatible. In 2009, first standardization approaches were started by the Third Generation Partnership Project (3GPP) for use in Long Term Evolution (LTE) networks [88]. This standard was continued in collaboration with the Moving Picture Experts Group (MPEG) [89], which eventually evolved into the Dynamic Adaptive Streaming over HTTP (DASH) standard in 2012 [90]. Based on the standard, an industry forum[4] of multimedia and network companies has been formed to catalyze the adoption of DASH in various services.

Like classical video streaming, HAS is based on the concurrent download of video data from a streaming server and playback of the downloaded video data in a client player. The data is requested via HTTP and transmitted via TCP to the client and stored in an application buffer. The client starts to play out the video after the buffer contains a sufficient amount of data. This waiting time between the video request and the start of playback is referred to as **initial delay** and influences the QoE of HAS. Although TCP ensures the transmission of undisturbed video data, network issues (e.g., insufficient bandwidth, packet loss, delay, and jitter) can decrease the throughput of the TCP transmission. If the throughput is lower than the bit rate of the video, the application buffer

---

[1] http://www.theguardian.com/media-network/media-network-blog/ 2013/mar/01/history-streaming-future-connected-tv [Online] – Accessed: 24.08.2017

[2] https://developer.apple.com/library/ios/documentation/ networkinginternet/conceptual/streamingmediaguide/ Introduction/Introduction.html [Online] – Accessed: 24.08.2017

[3] http://www.adobe.com/products/hds-dynamic-streaming.html [Online] – Accessed: 24.08.2017

[4] http://dashif.org [Online] – Accessed: 24.08.2017

decreases. When the buffer is depleted, the video playback stalls and can only be resumed after sufficient data have been received. As QoE is greatly influenced by these **stalling** (or rebuffering) events, streaming systems try to avoid stalling by **adaptation** of the video bit rate to the current network conditions. To change the video bit rate, the transmitted video has to be altered, e.g., in terms of resolution, frame rate, or compression, which introduces an additional impact on QoE. In the following, the technology behind HAS will be described in Section 2.1.1, and Section 2.1.2 will present related works on the impact of initial delay, stalling, and adaptation on the QoE of HAS.

2.1.1 HTTP Adaptive Video Streaming Technology

Figure 2.1 visualizes the HAS concept. To enable adaptation of the video streaming to the current network conditions, the HAS server, which can be part of a Content Delivery Network (CDN), stores the video content encoded in different representations, i.e., in different bit rates. The representations are split into segments (also referred to as chunks) and corresponding segments of different representations contain the same frames, such that the bit rate can be seamlessly switched at each segment boundary. Typically, all segments contain a fixed amount (i.e., 1 to 15 s) of video playback time and can either be extracted at runtime from the single representation file or are stored as separate files on the server. Apart from single-layer codecs (e.g., MPEG/Video Coding Experts Group (VCEG)'s H.264 Advanced Video Coding (H.264/AVC) [91] or H.265 High Efficiency Video Coding (H.265/HEVC) [92]), which encode each bit rate separately, multi-layer codecs (e.g., H.264 Scalable Video Coding (H.264/SVC) [93]) can be used, which enable bit stream switching. For example, H.264/SVC uses difference coding of the video content to predict data of higher layers (also referred to as enhancement layers) from lower representations. The quality can be increased incrementally to a higher layer by transmitting and adding the missing difference data. This is an advantage over single-layer codecs, for which a whole new segment has to be downloaded, and already transmitted data of lower

representations have to be discarded. In contrast, multi-layer codecs introduce overhead in terms of file size [94–97] and might require a higher amount of signaling as several HTTP requests might be needed for each segment [98].
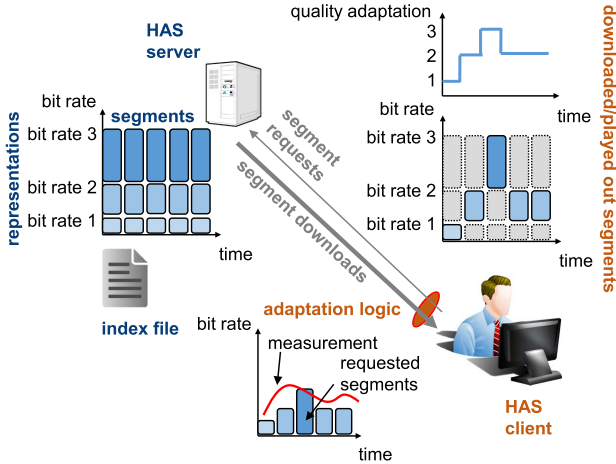
Figure 2.1: Concept of HAS.

When the client wants to start the video streaming, it sends a HTTP request to the HAS server to download an index file (also referred to as media presentation description, manifest, or playlist). The index file contains a list of available audio and video segments, their respective properties (e.g., codec, resolution, duration), and a means to send HTTP requests for a certain segment (e.g., Uniform Resource Locator (URL)). After the first request of a segment, before the decoding of the video and the playback can start, a certain amount of video data has to be transmitted from the server to the client. This initial delay is always present in a HAS service, and its minimal value depends on the encoder settings, the round-trip time, the throughput, and the speed of the decoder. Usually, the start of the video playback is delayed more than technically necessary to fill the playout buffer up to a preconfigured threshold of video data or video playtime. Thereby,

17

the playback buffer can be used to compensate for short term throughput variations. However, there is a trade-off between the amount of buffered playtime and the length of the initial delay (more buffered playtime results in a longer initial delay) and the risk of stalling (more buffered playtime means a higher robustness to short term throughput variations). A similar threshold is used in case of stalling, when the playback is interrupted until the buffer is filled up to the threshold. Here again, there is a trade-off between the amount of rebuffered playtime and the length of the interruption (more buffered playtime means a longer stalling duration) and the risk of a shortly recurring stalling event (more buffered playtime means a longer playback until potential next stalling event).

The adaptation logic at the client selects from the list of available segments in the index file which segments to download next. These decisions mostly take into account segment characteristics (e.g., bit rate), current playout statistics (e.g., buffer fill level), and current network conditions (e.g., bandwidth measurements or estimations). Several methods for bandwidth estimation are discussed in literature, e.g., [99–102]. By adapting the video bit rate, the adaptation logic aims at low initial delays, avoidance/minimization of stalling, and playing out the video in a high visual quality. Consequently, it is the adaptation logic, which predominantly influences the network demands of the video streaming and the resulting QoE of the end user. Four adaptation logics were selected and will be compared in terms of resulting QoE in Section 2.3. In the following, these algorithms will be introduced and short descriptions of the basic concepts are given. The comprehensive technical details can be obtained from the respective publications.

Reference [103] proposed a first prototype implementation of a DASH adaptation algorithm, which decides based on the estimated bandwidth for downloading the last segment, the current buffer level, and the average bit rate of each representation. The algorithm, which will be referred to as Klagenfurt University Adaptation Logic (KLU) throughout this work, selects the representation with the highest bit rate less or equal to the estimated bandwidth. The estimation is based on the throughput measured during the download of the last segment,

but is increased or decreased if the buffer level is less than 35% or higher than 50%, respectively. The algorithm presented in [104], which will be referred to as Technical University Berlin Adaptation Logic (TUB), aimed to maximize the played out quality, while avoiding unnecessary quality switches, and minimize the initial delay. It uses the same input parameters, but additionally three configurable buffer thresholds. In contrast to KLU, it tries to adapt only one quality level at a time when the buffer level is below or above configured thresholds. For example, if the buffer level is below a low threshold, the algorithm will estimate the bandwidth based on the throughput of recent segments and compare it to the bit rate of the current representation. If the bandwidth is lower, a segment of the next lower representation is requested. Switching up the quality is implemented in a similar fashion. Additionally, segment downloads can be delayed to avoid too large buffers, and a fast-start phase is used to minimize initial delay. Reference [105] adapted the video download strategy of the peer-to-peer streaming mechanism Tribler [106] for H.264/SVC-based HAS and aimed to stream a video quality that can be supported by the network. The algorithm, which will be called Tribler-based Adaptation Logic (TRI), uses a buffer threshold, and will only request the next higher quality layer if all segments of all lower layers are available up to the buffer threshold. This means, first, only the base layer of each segment is downloaded to reach the buffer threshold. After the threshold is reached, the algorithm will additionally download enhancement layers for the already buffered segments if possible. The Bandwidth Independent Efficient Buffering Adaptation Logic (BIEB) algorithm [107] aims to optimize the perceived QoE factors (cf. Section 2.1.2) and was also designed for H.264/SVC, but uses a more dynamic segment selection algorithm by relying on size ratios between segments of different representations. The desired buffer level for each representation consists of a base number of segments and a variable number of segments computed from the size ratios. The algorithm loops from the lowest to the current quality level until it encounters a representation for which the buffered number of segments is less than desired. If all desired segments have been downloaded, the algorithm increases the desired buffer level and switches

to the next higher quality level. Table 2.1 summarizes the different characteristics of the presented adaptation logics.

Table 2.1: Characteristics of the four investigated HAS adaptation logics.

	Bandwidth estimation	Buffer level	Conservative up-switch	Segment size ratio
KLU	x	x		
TUB	x	x	x	
TRI		x	x	
BIEB		x	x	x

Apart from these four algorithms, which will be benchmarked in Section 2.3 with respect to QoE-optimal adaptation, many other HAS adaptation logics have been proposed in related works. Reference [108] described an algorithm to minimize stalling. In [109], the authors proposed adaptation algorithms using feedback controllers based on the buffered video time as feedback. In [110], the authors proposed a video rate adaptation algorithm considering only the playback buffer. Reference [111] described an adaptation logic based on feedback control theory, which did not generate on-off traffic patterns, and thus, avoided unfairness and underutilization when many video flows a bottleneck link. References [112, 113] presented an adaptation logic, which tackled video bit rate oscillation. Reference [114] used only buffer occupancy for adaptation decisions, while [115] additionally considered throughput in a control-theoretic model. The algorithm presented in [116] considered the measured throughput, buffer occupancy, and the variation in segment sizes to better manage the users' QoE. Finally, [117] modeled adaptation control systems to gain design insights.

2.1.2 Previous QoE Results for HAS

The subjectively perceived quality of a HAS session depends on many technical and perceptual influence factors [14]. Based on the results of previous studies,

the research community agrees that initial delay, stalling, and quality adaptation have the biggest impact on the QoE. In the following, the most important QoE results for these impact factors will be presented.

Initial Delay

As the impact of initial waiting times on QoE strongly depends on the application [118], the results for other services cannot be directly transferred to HAS. However, works on web page load times [119], Internet Protocol Television (IPTV) channel zapping times [120], or Universal Mobile Telecommunications System (UMTS) connection setup times [121] suggested a logarithmic relationship between waiting times and QoE. For video streaming, initial delay is an expected waiting time well known from the everyday usage of video applications. References [118, 122] found that initial delay was fundamentally different from stalling, which is a sudden unexpected interruption within the service. 90% of the users in the study of [118] preferred initial delays over stalling. It was confirmed by [123] that initial delays were also considered less important for mobile video users and less critical for having a high QoE. Reference [124] confirmed the low impact of initial delays for high resolutions. Reference [118] quantified the influence of initial delay on QoE and found that initial delays up to 16 s only marginally reduced the MOS. Moreover, the results suggested that the impact of initial delay on QoE depended only on the length of the initial delay, but not on the length of the video.

When considering not only single videos, but rather sessions of videos, initial delays could become more relevant. References [125, 126] noted that users often did not intend to watch a specific video, but instead browsed through videos in order to find some content they were interested in. Thereby, they started many videos, but often watched only the first seconds. Reference [127] found a positive correlation of abandoned views and initial delay. Especially, if the initial delay exceeded 2 s, viewers started to abandon the video completely. They also found that the abandonment rate depending on initial delay was higher for short

videos than for long videos, and higher for fixed access users (fiber, cable, DSL) than for mobile users.

To sum up, for HAS, initial delays are not a major performance issue, especially if the users intend to watch the video. However, like for all services, initial delays should be kept short to decrease the probability that a user aborts the video and to increase the QoE of the users.

Stalling

In contrast to initial delay, stalling is perceived much worse because it is processed differently by the human sensory system [128]. Reference [129] showed that an increased duration of stalling deteriorated the QoE, but the position of stalling was not important. In contrast, [130] reported an impact of the position. Reference [129] also found that a single long stalling event was preferred to multiple short stalling events. Reference [131] showed that stalling at irregular intervals was perceived worse than periodic stalling. Moreover, the authors compared stalling to frame rate reduction and found that stalling was worse. Reference [132] presented a random neural network model to predict QoE based on stalling and the quantization parameter of the video encoder. In the corresponding subjective studies, users were more sensitive to stalling than to increased quantization, especially for less compressed videos. In [63], a model for mapping regular stalling to MOS was obtained from subjective crowdsourcing studies with regular stalling patterns. The results indicated an exponential relationship between stalling parameters, i.e., length and number of stalling events, and QoE. Users rated the quality to be good, if at most one stalling event occurred with a duration of a few seconds only. More stalling resulted in highly dissatisfied users. Reference [133] trained a more complex Hammerstein-Weiner model to predict the QoE and achieved higher median correlations than previous models. Reference [134] found that the degradation of stalling was worse when the presentation quality was higher and fitted a simple cubic relationship to describe this trend.

The results of previous studies suggest that stalling severely deteriorates the QoE of video streaming, and thus, should be avoided to the greatest extent. This means, HAS services need efficient adaptation logics, which align the video bit rate to the current network conditions. However, reducing the bit rate to avoid stalling will result in playing out a lower quality layer, which also impacts the end users' QoE.

Adaptation

Video streaming was investigated under vehicular mobility in [135] and it was found that the quality adaptation of HAS could effectively reduce stalling by 80% when bandwidth decreased. Moreover, HAS better utilized the available bandwidth when bandwidth increased. Thus, HAS trades off stalling for adaptation, which is generally expected to improve the QoE of end users. Reference [136] compared (controlled) resolution reduction to (uncontrolled) packet loss artifacts and confirmed that it was better to control the quality than to suffer from uncontrolled effects (provisioning-delivery hysteresis). However, also the controlled adaptation of HAS has an impact on the QoE of end users. The quality switch is a degradation [137] and a high frequency of quality adaptation will be annoying for end users [138]. Apart from the switch itself, the QoE changes according to the adaptation direction, however, switching down the video quality will have a stronger impact on QoE [137]. The adaptation amplitude is the most dominant factor and a high amplitude leads to a low acceptance, while low amplitudes might not be detectable [138]. However, abrupt up-switching might increase the QoE [139]. Reference [140] observed the recency effect, i.e., higher quality in the end lead to a higher QoE. Confirming the recency effect, [141] additionally found that segments in the beginning had a higher impact on QoE (i.e., primacy effect). Reference [142] concluded that the temporal aspects of adaptation became important. They also found that a higher base layer allowed for longer impairments to be accepted.

Although, adaptation is a less severe degradation than stalling [108], its impact on QoE must not be neglected. Additional to the general results presented

above, each adaptation dimension (e.g., resolution, frame rate, quantization) has a specific impact on the perceived quality. These QoE results for the different adaptation dimensions and their trade-offs can be found in [14]. Furthermore, several studies exist, which present models for time varying video quality. Reference [143] proposed linear models based on per-frame and per-segment objective metrics. The QoE is estimated based on average and standard deviation of the objective metrics and the adaptation frequency. Reference [144] followed a similar approach but utilized data mining. Reference [145] considered frequency, type, and temporal location of the quality adaptation. Reference [146] used a histogram of per-segment quality scores and the histogram of quality gradients to model the overall QoE. Reference [147] summarized subjective studies and discussed open research questions towards QoE models for HAS.

General QoE Models for HTTP Adaptive Video Streaming

Several models for incorporating different key performance indicators into a single QoE score have been proposed. In [108], QoE metrics from the 3GPP DASH specification TS 26.247 [89] were considered, namely, HTTP request/response transactions, representation switch events, average throughput, initial delay, buffer level, play list, and information from the media presentation description. However, the corresponding evaluation took only stalling as the most dominating impairment into account. Reference [132] considered the impact of the quantization parameter and stalling on the perceptual quality. Reference [148] found that adaptation strategy-related parameters (stalling, representation switches) had to be considered on a large time scale of up to some minutes, while video encoding-related parameters (resolution, frame rate, quantization parameter, bit rate) only influenced in the order of a few seconds. References [149, 150] used all impact factors, namely, initial delay, stalling, and adaptation, for their models.

2.2 Impact of Adaptation on the QoE of HAS

The related works foreshadow that the quality level of the video and the adaptation frequency may have a significant impact on the perceived QoE of HAS. Still, the results of these studies are not unanimously naming and quantifying the most important influence factors, which indicates that it is not well understood how quality adaptation affects the QoE. Therefore, this section investigates the impact of adaptation-related parameters with two crowdsourced subjective QoE studies.

Crowdsourcing is the outsourcing of small tasks to a large group of workers, and has become increasingly popular in the context of subjective QoE assessments, e.g., [63, 151]. In typical crowdsourced QoE studies, which are sometimes referred to as crowdtesting, quality rating tasks are distributed to a group of remote participants, e.g., via dedicated crowdsourcing platforms. The participants access the rating task via the Internet and consume the test stimuli on their own devices. After they upload their quality ratings, they receive a monetary compensation depending on the complexity and length of the rating task. The advantages of crowdsourcing studies over classical laboratory studies are their speed, price, number of participants, and the more realistic setting of service consumption in terms of context and system factors. Nevertheless, the heterogeneous environment and unsupervised nature of such QoE studies can introduce pitfalls, which have to be avoided by proper study design and the filtering of unreliable results [77]. Therefore, in the field of crowdsourced QoE studies, research deals with motivation and incentives of participants (e.g., [59, 152]), methods for screening the reliability of participants (e.g., [54]), mechanisms for asserting a high quality of results (e.g., [32, 36, 153]), and the development of crowdsourcing frameworks and platforms (e.g., [154]). A comprehensive report of best practices and lessons learned for crowdsourced QoE studies can be found in [77].

To assess the impact of adaptation on the QoE of HAS, the study in Section 2.2.1 [52] considers a HAS system with two quality layers. The results show

that the quality switch amplitude and the playback time on high quality layer are the most important influence factors. The results are generalized for video content with more than two quality layers in the study in Section 2.2.2 [39], indicating that also times on intermediate layers have to be taken into account.

2.2.1 Influence of Adaptation-related Parameters

First, the influence of switch amplitude (i.e., quality level difference), switching frequency, and recency effects on QoE of HAS is investigated. Therefore, different crowdsourcing campaigns are conducted to assess the subjective quality of different adaptation-related test conditions. The results of the conducted campaigns allow to quantify the influence of the adaptation-related parameters.

Study Description

As the duration of crowdsourced QoE evaluations should be in the order of minutes [155], the QoE study was split in different crowdsourcing campaigns. The campaigns were published on the Microworkers.com portal[5], a provider of crowdsourcing services with international user base. Employers can create campaigns with a certain number of identical tasks on the platform, and workers can browse and accept the available tasks. The employers pay monetary compensations via the platform to the workers for each satisfactorily fulfilled task. For all campaigns of this study, the task of the workers was to access the web-based subjective QoE study and rate the quality of different HAS sequences. In total, 710 unique workers (based on the Microworkers.com account ID) reliably participated in the campaigns and rated between 7–9 test sequences. The QoE study was implemented with the QualityCrowd2[6] framework [156], which conforms to the best practices for crowdsourced QoE studies introduced in [77].

The QoE study started with a short demographic survey and task instructions. These instructions explained, in simple English and illustrated with pictures,

---

[5]`https://microworkers.com/` [Online] – Accessed: 24.08.2017
[6]`https://www.ldv.ei.tum.de/forschung/videolabor/qualitycrowd/` [Online] – Accessed: 24.08.2017

how to use the framework to watch and rate the video sequences. Afterwards, the participant had to watch the tested video sequences in random order. The sequences were first downloaded to the browser cache to avoid problems related to insufficient bandwidth (e.g., stalling). Moreover, the user had to watch the entire sequence and rate it before continuing to the next video to prevent deceiving behavior of the workers (e.g., fast skipping through the sequences) and ensure a high rating quality. After the playback of the video sequence, the participant was asked *Did you notice any changes in quality during playback? If yes, did you feel annoyed by them?* and had to rate on a continuous 5-point Degradation Category Rating (DCR) scale with the options *Imperceptible (did not notice any)*, *Perceptible but not annoying (did notice, but did not care)*, *Slightly annoying*, *Annoying*, and *Very annoying*.

Table 2.2: Summary of tested effects in first crowdsourcing study.

Effect	Test design
Amplitude	Switch amplitude high or low
Last quality level	End on high or low quality level
Recency time	Different time after last quality switch
Frequency	Different number of switches for constant time on high layer
Time on high layer	Different time on high layer for the same number of switches

As content plays a key role in QoE (e.g., [157–159]) and the focus of this study is to investigate adaptation-related parameters, only one video was used. The video sequence was taken from the open-source short movie "Tears of Steel"[78] and had a length of 15 s (360 frames, starting from 0:25), which was cropped to 14 s (336 frames) in some campaigns. In this scene, which contains a low level of

---

[7]`https://www.youtube.com/watch?v=41hv2tW5Lc4` [Online] – Accessed: 24.08.2017
[8]`https://mango.blender.org/` [Online] – Accessed: 24.08.2017

detail and motion (SI: 8.5, TI: 5.37), two persons stand on a small bridge and discuss. Each test sequence was encoded from the source video using H.264/AVC (libx264) with Quantization Parameter (QP) 24, 24 Frames per Second (fps), and the original audio channel. Three quality levels were created by downscaling the video resolution to 640x360 (high), 320x180 (intermediate), and 160x90 (low), respectively. In the QoE study, the three quality levels were all displayed in a video player with size 640x360, thus, the sequences with lower resolution were upscaled in the browser of the participant. The encoded videos had an average bit rate between 0.64 Mbps and 0.75 Mbps and a size about 1.1 MB to 1.3 MB per video. For each investigated effect shown in Table 2.2, multiple test sequences with different adaptation patterns were created. For example, to investigate the switching frequency, a different number of quality switches (1, 2, 3, 4, 5, 6, 8, and 14 switches) were included in the sequence. Note that the sequences only included adaptation between two layers, i.e., either between high and intermediate (low switch amplitude), or between high and low (high switch amplitude). Detailed information on the adaptation patterns and the conducted campaigns can be found in [78].

In crowdsourced QoE studies with little control over remote and anonymous participants, it is necessary to design and deploy methods to counteract cheating and unreliable test participants, which want to use minimal efforts to receive the task payment and thereby pollute the quality of the ratings [77]. The QualityCrowd2 framework already implements some anti-cheating mechanisms, e.g., it prevents participants to shorten the study duration by ensuring that all video sequences are played out. In addition to the built-in mechanisms, one simple content question was included in the QoE study to check if participants had actually watched the video sequences. The participants were asked *Where did the protagonists stand on?* with answer possibilities *A building*, *A large field*, *A small bridge*, and *Riding on an elephant*. 11% of the test participants failed to correctly answer the content questions and their ratings were discarded from the results.

Results of the Study on Adaptation-related Parameters

As reference for the evaluation of the results, also three sequences, i.e., one for each quality level, with no quality switches were included in the conducted campaigns. The high quality layer sequence, i.e., resolution 640x360, was included in all tasks and received a MOS of 4.14 (95% confidence interval of 4.09 to 4.18). For the sequence with resolution 320x180, which was not included in all campaigns, a MOS rating of 3.52 (95% confidence interval of 3.42 to 3.62) was observed based on 310 user ratings. The lowest quality sequence (resolution 160x90) showed a MOS value of 2.51 (95% confidence interval of 2.37 to 2.66) based on 267 user ratings. The other sequences included adaptation patterns to assess the influence of the adaptation-related parameters. All patterns were rated by at least 82 reliable participants and they received on average 106 reliable ratings over all campaigns. In the following, these ratings will be evaluated.

Table 2.3: Quantification of main effects of adaptation-related parameters on MOS based on one-way ANOVA.

(a) All sequences.

	SROCC	F	p	η_p^2	f^2
Amplitude	-0.266	63.101	< 0.001	0.013	0.011
Last quality level	0.103	5.500	0.004	0.002	0.002
Recency time	0.109	0.900	0.480	0.001	0.001
Switches	-0.221	1.736	0.139	0.001	0.001
Time on high layer	0.295	17.742	< 0.001	0.043	0.037

(b) Sequences with quality switches only.

	SROCC	F	p	η_p^2	f^2
Amplitude	-0.302	66.350	< 0.001	0.017	0.015
Last quality level	0.018	2.508	0.082	0.001	0.001
Recency time	0.100	0.635	0.638	0.001	0.001
Switches	-0.088	1.750	0.136	0.002	0.002
Time on high layer	0.143	7.799	< 0.001	0.022	0.020

First, statistical analyses of effects of the investigated adaptation-related parameters were conducted. Table 2.3 shows the results of these analyses both for the set of all sequences and for the subset of sequences, which contain quality switches. Each row contains one of the investigated adaptation parameters and its effect on the subjective quality ratings. The first column shows the Spearman Rank Order Correlation Coefficient (SROCC), which indicates how much the influence factor and the subjective quality ratings are associated. Showing only absolute values smaller than 0.3, no high correlation can be observed for any adaptation parameter. Second, a one-way Analysis of Variance (ANOVA) was conducted to compare the MOS of different test conditions for statistical significance. The second and third column present the corresponding F-test statistic, i.e., the ratio of the mean squared errors, and p-value, i.e., the probability that the F-test statistic can take a value greater than or equal to the computed test statistic. The p-value is very small for both amplitude and time on high layer, in particular smaller than typical significance levels α (e.g., $\alpha = 0.05$), and thus, indicates that these parameters are significant. To measure the size of the effect, partial Eta-squared (η_p^2) and Cohen's f^2 were computed, which are shown in the last columns. Both amplitude and time on high layer have small effects according to η_p^2, and the time on high layer also has a small effect for f^2. However, no effect is visible for the influence of last quality level, recency time, and number of switches.

The results of the statistical analyses contradict the findings of several works (e.g., [137, 138]), which suggested an impact of switching frequency on the QoE. This effect was not found in the conducted QoE study, but switching frequency and time on layer can be correlated, as a quality switch typically changes the time on each layer. Figure 2.2 shows exemplary quality switching patterns, which are used during this study, and will be described with the corresponding results below. Figure 2.3a and 2.3b display the MOS and 95% confidence intervals for the number of quality switches (a) and the corresponding time on the high layer (b), for a campaign, in which both parameters change simultaneously because the quality switches are equidistantly spaced over the whole video length
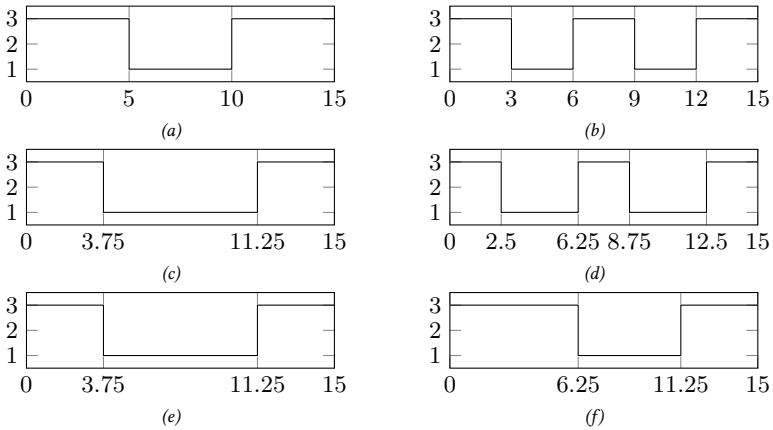
Figure 2.2: *Exemplary quality switching patterns for investigating the effects of switching frequency and time on high layer. The quality layer is depicted on the y-axis and the x-axis shows the video time in seconds. The patterns in the first row change number of quality switches and time on high layer simultaneously due to equidistantly spaced switches. The patterns in the second row change only number of quality switches, the patterns in the last row change only time on high layer.*

as can be seen in the first row of Figure 2.2. This means, for example, for two quality switches (cf. Figure 2.2a), the time on high layer was two third of the video length, while for four quality switches, the time on high layer was three fifth of the video length (cf. Figure 2.2b), etc. Figure 2.3 shows that both parameters seem to have a significant effect on the resulting QoE. Therefore, two campaigns of the QoE study are revisited to investigate in detail the effect of the switching frequency and the corresponding time on the high layer.

Figure 2.4a shows the MOS and 95% confidence intervals for different numbers of switches, when the total time on the high layer was kept constant as can be seen in the second row of Figure 2.2. This means, the quality switches were not spaced equidistantly, but such that the time on high layer was 50%
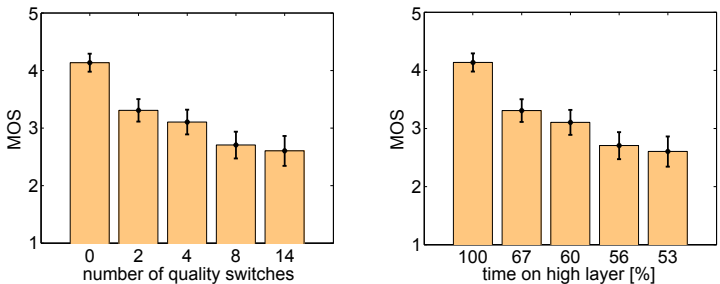
(a) MOS depending on number of quality switches.

(b) MOS depending on time on high layer.

Figure 2.3: MOS results when frequency and time on high layer change simultaneously due to equal time between switches.

of the video length for all sequences with switches (cf. Figures 2.2c and 2.2d). In this campaign, the sequence without switches shows a MOS of 3.93. For the sequences with quality switches, the MOS ranges slightly below 3 (from 2.71 to 2.91) and the confidence intervals overlap. Thus, no effect of the number of switches is visible. The one-way ANOVA for this campaign returns a p-value of $p = 0.0129$ for all sequences and $p = 0.7190$ for the subset of sequences with quality switches, i.e., $N > 0$. Figure 2.4b depicts the result of a different campaign, in which the time on high layer is varied for two different switching frequencies, i.e., $N = 2$ and $N = 4$ switches. Such patterns can be seen in the last row of Figure 2.2. In Figure 2.4b, the time on the high layer in percent is plotted on the x-axis, and the y-axis displays the MOS and 95% confidence intervals. The sequence without switches, i.e., 100% time on high layer, reaches a MOS of 4.05. For the other sequences, the confidence intervals for each pair of bars ($N = 2$ dark, $N = 4$ light) overlap, such that no impact of the number of switches can be seen. Also a two-sample t-test does not reveal significant differences ($p = 0.1055$) between the two groups with Cohens' $d = 0.1398$. However, when the time on the high layer decreases, the MOS decreases. Thus,

(a) No effect of switching frequency.

(b) Time on high layer is a main influence factor.
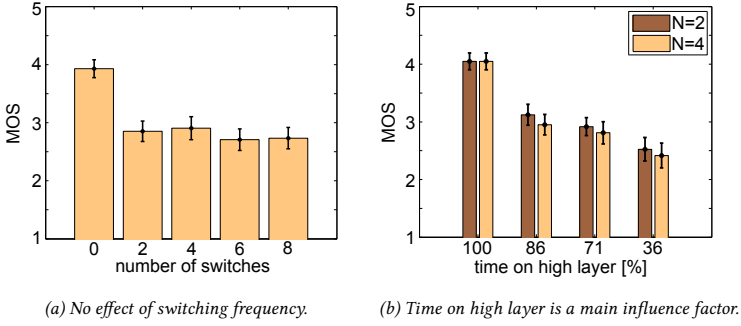
Figure 2.4: Impact of switching frequency and time on high layer on MOS.

an effect of the time on the high layer is visible here, which is significant according to a one-way ANOVA and a p-value of $p = 5.44 \cdot 10^{-37}$.

In the following, these insights are quantified in a new simplified QoE model for HAS sessions with two quality layers without initial delay or stalling. Figure 2.5 shows the proposed model, which considers the two main effects, amplitude and time on high layer. The y-axis depicts the MOS, while the time on high layer in percent is plotted on the x-axis. The MOS is bounded by the quality $y_H = f(100)$ of the high layer and $y_L = f(0)$ of the low layer, respectively, whose actual values express the amplitude effect between both layers. The gray boxes represent the MOS and 95% confidence intervals of the ratings of the video clip with constant high ([4.09;4.18]) or low ([2.37;2.66]) quality and were obtained in a separate campaign. The black data points indicate the MOS and 95% confidence intervals of the time on high layer campaign (cf. Figure 2.4b). Following the Exponential Interdependency of QoE and QoS (IQX) hypothesis [160], the bounds y_H and y_L and the data points are fitted with an exponential function $f(t) = \alpha e^{\beta t} + \gamma$ of the percentage of time on high layer t. The resulting fitting $f(t) = 0.003 \cdot e^{0.064 \cdot t} + 2.498$ reaches a high coefficient of determination $R^2 = 0.98$, which indicates that the exponential model is well suited to quantify the impact of the time on high layer.
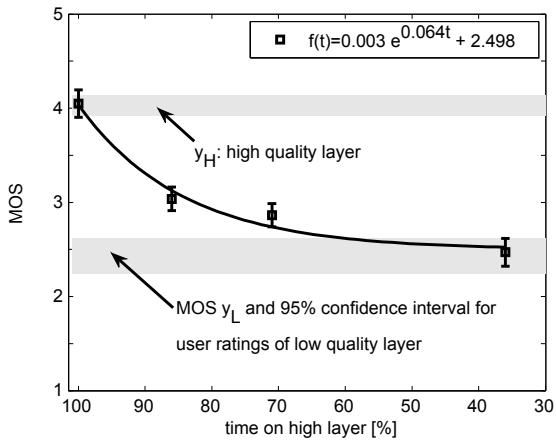
Figure 2.5: Simplified QoE model for two layers.

In summary, this study investigated the influence of adaptation-related parameters on the QoE of HAS. The results show that only the time on high layer and the switching amplitude, which relates to the quality of the other layer, are main influence factors. Contradictory to the results of related works, last quality level, recency time, and switching frequency have no significant impact.

2.2.2 Impact of Time on Intermediate Layer

The results of Section 2.2.1 show that time on high layer is a main influence factor on QoE of HAS. As this study only investigated the influence of adaptation between HAS content with two quality layers, it is still unclear if there is only an impact of the time on highest layer or the time on each layer. Thus, the second study will clarify this open issue by investigating a HAS system with three quality layers.

Study Description

The second study was also conducted as a crowdsourced QoE evaluation and was designed similar to the first study. This means, again a campaign was created on the Microworkers.com platform with tasks to rate the quality of video sequences with the QualityCrowd2 framework. 146 workers participated in the crowdsourcing campaign and watched nine sequences in random order. After each sequence, the user was asked *Did you notice any changes in quality during playback? If yes, did you feel annoyed by them?* and rated on a continuous 5-point DCR scale with options *Imperceptible (did not notice any)*, *Perceptible but not annoying (did notice, but did not care)*, *Slightly annoying*, *Annoying*, and *Very annoying*.

The content was a popular movie scene[9] from the video portal YouTube, which has a duration of 14 s (336 frames) and shows a monologue of one character in front of another in a living room. The original 720p sequence was first decoded to individual uncompressed images and encoded using H.264/AVC (libx264), 24 fps, and the original audio channel. The videos were downscaled to resolution 640x480, i.e., the size of the player in the QoE study, and three quality layers were created by adjusting the QP to 28 for high quality, 36 for intermediate quality, and 44 for low quality, respectively. The file sizes of the final sequences ranged from 444 KB for constant high quality to 266 KB for constant low quality.

The campaign was designed to evaluate nine different adaptation patterns, which are illustrated in Figure 2.6. Three constant quality profiles a (high), b (intermediate), and c (low) were tested as reference profiles. Moreover, three pairs of patterns (d/e, f/g, and h/i) with three quality switches were created. They differ between pairs in the time on intermediate layer, and within pairs in the position of the intermediate layer. Thus, they can be used to assess both effects. The sequences start with 2 s of the high quality layer, then switch down to either the intermediate (d, f, h) or the low (e, g, i) quality layer. After some

---

[9] `https://www.youtube.com/watch?v=IFSAbxfLBYU` [Online] – Accessed: 24.08.2017, starting from 1:05

time, the quality is switched to the low (d, f, h) or the intermediate (e, g, i) layer, respectively. Patterns d/e play out the intermediate layer for 8 s, patterns f/g for 5 s, and patterns h/i for 2 s. The third quality adaptation at 12 s switches back to the high quality layer.

Unreliable participants were filtered out with a simple content question (4 participants) and with the help of inter- and intra-rater reliability scores. These scores were computed as the SROCC of a participant's ratings and the mean of all ratings (inter-rater) and the SROCC of a participant's ratings and the time-averaged quality level of the patterns (intra-rater). If the correlation coefficients were below 0.5, the participants were filtered out (69 participants). 73 reliable participants (50%) remained after the filtering and their ratings were used for the evaluation.
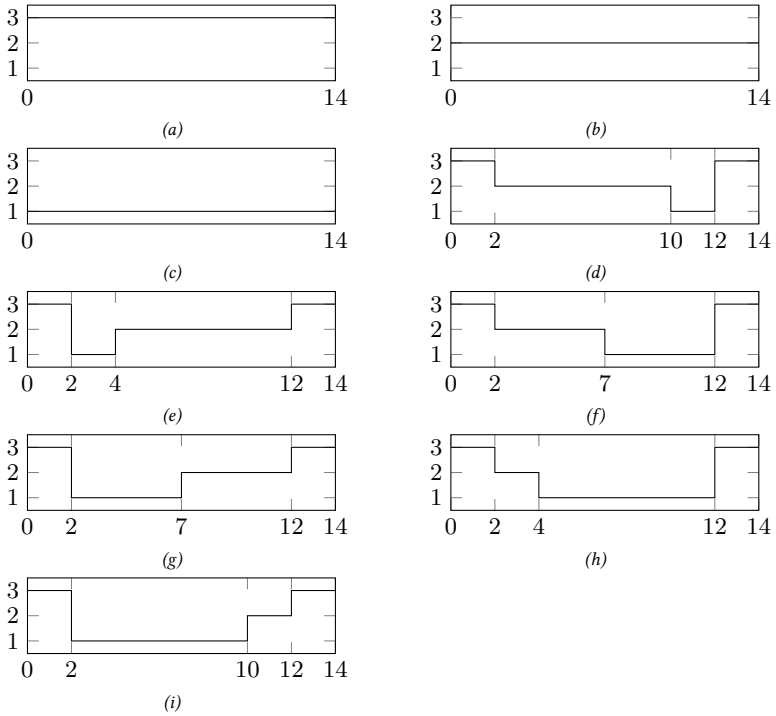
Figure 2.6: Quality switching patterns with the quality layer on the y-axis and the video time in seconds on the x-axis.

Results of the Study on Time on Intermediate Layer

Figure 2.7 shows the MOS and 95% confidence intervals of all sequences based on the ratings of the 73 reliable users. The constant quality sequences (a-c) achieve MOS values of 3.93 (high), 2.99 (intermediate), and 2.21 (low). The clear separation of these scores indicates that the participants are able to visibly distinguish the selected three quality layers. Moreover, the constant high and low sequences form the limits of the perceived quality for this study. Investigating the results for the sequences with quality adaptation, it is obvious that the sequences within each pair of symmetrical sequences (d/e, f/g, and h/i) achieve similar MOS values. Thus, the effect of the position of the intermediate layer is negligible. However, the MOS between the pairs is different, which indicates that the time on the intermediate layer influences the QoE of HAS systems.
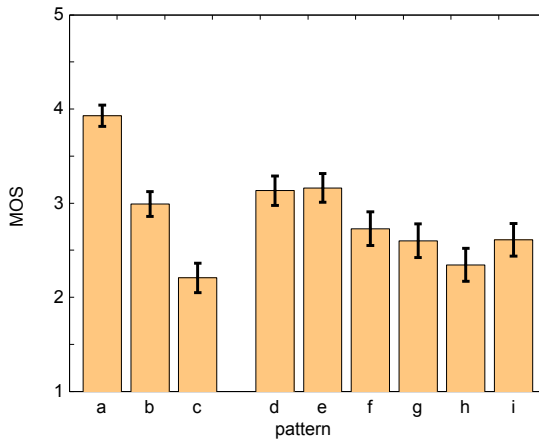
Figure 2.7: MOS and 95% confidence intervals of overall quality. Each test condition was rated by 73 different users.
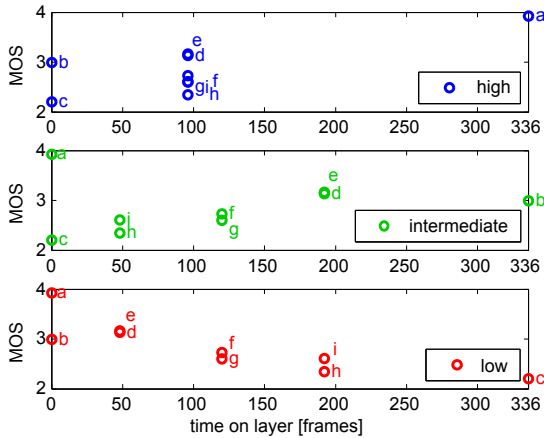
Figure 2.8: MOS values of the tested adaptation profiles with respect to the time on each layer.

Figure 2.8 investigates this effect by plotting the MOS of each sequence depending on the time on each layer in frames. The x-axes show the time on the high layer (top plot), the time on intermediate layer (middle plot), and the time on low layer (bottom plot). The y-axes depict the MOS values of the different sequences. The constant sequences a-c have all frames on the same quality layer and can be used as a reference. The top plot shows that the time on the high layer is not the only influence factor. All sequences with adaptation d-i have the same time on high layer but achieve different MOS values. The other two plots again nicely show that the qualities of the pairs of symmetric sequences are rated similarly. Additionally, a positive correlation of the time on intermediate layer (middle plot) and a negative correlation of the time on low layer (bottom plot) can be observed.

To sum up, it became evident that the quality of adaptation sequences are bounded by the MOS values of the constant high and low quality sequences.

Furthermore, the quality of each layer and the time on each layer are the only adaptation-related parameters, which have a significant impact on the QoE of HAS.

2.3 Towards a QoE-optimal Adaptation Logic

The previous findings show that the quality adaptation has a big impact on the resulting QoE of a streaming session. As the HAS adaptation logic directly influences the adaptation of the played out video, it is the key component of a HAS streaming system. In order to compare and eventually improve adaptation logics in terms of QoE, a QoE-optimal segment download strategy for given network conditions and video characteristics is obtained by solving a two-step Mixed Integer Linear Program (MILP). It can be used to benchmark HAS adaptation logics with respect to the QoE-optimal adaptation in different evaluation scenarios. After the optimization problem is formulated, an exemplary scenario is considered to demonstrate the applicability of the benchmark. Therefore, four different adaptation strategies from literature are compared to the QoE-optimal adaptation computed from the MILP [13, 15].

2.3.1 Optimization Problem for QoE-optimal Adaptations

The QoE-optimal adaptation can be computed for given network conditions and video characteristics. As stalling is the worst quality degradation, the optimal adaptation has to avoid stalling and ensure a smooth playback. Moreover, the initial delay T_0, i.e., the time after which the playback starts, is fixed. $V(t)$ represents the download volume at time t, i.e., the amount of data, which a user can download during the time $[0, t]$. The video is split in n segments and available in $R = \{1, \ldots, r_{max}\}$ representations. Each segment i of representation j has a size S_{ij} and contains τ seconds of playtime. After the initial delay T_0, the ith segment has to be downloaded and played out at time $D_i = T_0 + i\tau, \ i = 1, \ldots, n$ to avoid stalling.

According to the previous QoE results, the quality layer of each played out segment influences the overall perceived quality. Therefore, the quality value of a segment i in representation j, i.e., its contribution to the overall perceived quality, is set to $w_{ij} = j$ in order to optimize the average played out quality layer. Nevertheless, different options for expressing the value of a segment are possible [13]. The Boolean target variable x_{ij} indicates whether the client downloads segment i of representation j ($x_{ij} = 1$) or not ($x_{ij} = 0$). The optimal assignment x_{ij} describes the optimal segments to download in the investigated scenario and reaches an optimal quality value W_{opt}. This optimal adaptation is realizable under the given network and video characteristics, however, it does not indicate when to download which segment.

Table 2.5: Notations and variables used for the optimization problems.

Variable	Explanation
$V(t)$	Total amount of data received by a client during the time $[0, t]$
n	Number of segments
τ	Duration of a segment
D_i	Playback deadline for segment i
R	Available representations, i.e. $R = \{1, \ldots, r_{max}\}$
S_{ij}	Size of segment i from representation j
w_{ij}	Weighting factor indicating the QoE value of segment i for representation j
$x_{ij} \in \{0, 1\}$	Target variable indicating if client downloads segment i from representation j ($x_{ij} = 1$) or not ($x_{ij} = 0$)
W_{opt}	Optimal quality value of Optimization Problem 1

Using the notation, which was introduced above and which is summarized in Table 2.5, two optimization problems 1 and 2 can be formulated. The first optimization problem will maximize the average played out quality layer, as

representation quality and time on each layer have the most significant impact on the QoE of HAS. Although the number of quality switches has only little impact on the QoE, the second optimization problem will minimize the number of quality switches to avoid flickering, which negatively influences the perceived quality [138]. Thus, this two-step MILP will by design result in an optimal QoE without the need for a dedicated mapping of adaptation parameters to MOS.

Optimization Problem 1 (Maximize quality value)**.**

$$\textit{Maximize } W = \sum_{i=1}^{n} \sum_{j=1}^{r_{max}} w_{ij} x_{ij}, \tag{2.1}$$

$$\textit{subject to } \sum_{j=1}^{r_{max}} x_{kj} = 1 \qquad \forall k = 1, \dots, n, \tag{2.2}$$

$$\sum_{i=1}^{k} \sum_{j=1}^{r_{max}} S_{ij} x_{ij} \leq V(D_k) \quad \forall k = 1, \dots, n. \tag{2.3}$$

Optimization Problem 1 maximizes the sum of the quality values w_{ij} of all downloaded segments, which results in an optimization for the average quality layer. The first constraint (2.2) ensures that for each segment one representation is downloaded. Constraint (2.3) checks that all segments up to segment k can be downloaded with $V(D_k)$, i.e., the available download volume up to the playback deadline of segment k. This ensures that all segments can be downloaded until their respective playback deadline. The optimal quality value W resulting from Optimization Problem 1 will be denoted by W_{opt} and will be used as input to Optimization Problem 2.

Optimization Problem 2 (Minimize switches at given target quality W_{opt}).

$$\text{Minimize } \frac{1}{2} \sum_{i=1}^{n-1} \sum_{j=1}^{r_{max}} (x_{ij} - x_{i+1,j})^2, \tag{2.4}$$

$$\text{subject to } \sum_{j=1}^{r_{max}} x_{kj} = 1 \qquad \forall k = 1, \dots, n, \tag{2.5}$$

$$\sum_{i=1}^{k} \sum_{j=1}^{r_{max}} S_{ij} x_{ij} \leq V(D_k) \quad \forall k = 1, \dots, n, \tag{2.6}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{r_{max}} w_{ij} x_{ij} \geq W_{opt}. \tag{2.7}$$

Optimization Problem 2 minimizes the number of quality switches, which can be computed as sum of layer-wise differences between the indicator x_{ij} of adjacent segments. If no switch occurs after a segment, the differences will be 0 for all layers. In contrast, if a switch occurs, one layer will show a transition from $x_{ij} = 1$ to $x_{i+1,j} = 0$, and another layer from 0 to 1. The factor $\frac{1}{2}$ accounts for this double counting of each switch, and is only added here for the sake of completeness. Constraints (2.5) and (2.6) in problem 2 resemble constraints (2.2) and (2.3) in problem 1, and again ensure that one representation of each segment is downloaded and that it is downloaded before its playback deadline, respectively. The additional constraint (2.7) guarantees that the overall quality value is not decreased below the optimum W_{opt} by minimizing the number of quality switches.

Optimization Problem 1 is a variant of a Multiple-choice Nested Knapsack Problem (MCNKP) [161], while Problem 2 is a Quadratic MCNKP. The MCNKP is NP-hard, but pseudo-polynomial time algorithms exist. Therefore, the optimization software Gurobi[10] is utilized to solve the linear program. The resulting QoE-optimal adaptation can be used to evaluate the performance of HAS adap-

---

[10]http://www.gurobi.com/ [Online] – Accessed: 24.08.2017

tation strategies. Therefore, evaluation scenarios have to be defined, in which different adaptation logics are compared for different videos and network conditions. The presented linear program allows to compute an optimal adaptation strategy for each evaluation scenario, and thus, allows to quantify how good adaptation logics are with respect to the optimal QoE.

2.3.2 Evaluation Scenario

An exemplary evaluation is conducted for the complete "Tears of Steel"[78] short movie, which was split in $n = 367$ segments of $\tau = 2\,\text{s}$ playtime each. The movie was transcoded to H.264/SVC using QP 24, 24 fps, and three spatial quality layers, i.e., three different resolutions of 320x180 ($r = 1$), 640x360 ($r = 2$), and 1280x720 ($r = 3$). The encoded quality layers have average bit rates of 0.26 Mbps, 0.95 Mbps, and 2.67 Mbps and maximum bit rates of 1.28 Mbps, 3.37 Mbps, and 10.46 Mbps, respectively. Due to the usage of H.264/SVC, the segment size S_{ir} is the sum of the segment plus all required lower layer segments ($S_{ir} = \sum_{j=1}^{r} S_{ij}$). Table 2.6 summarizes the properties of each representation r. In this evaluation, no initial delay is considered, such that the video playback begins after the first segment has been downloaded.

Table 2.6: Characteristics of video contents and the segment sizes S_{ir} for the three quality layers, i.e., 320x180 ($r = 1$), 640x360 ($r = 2$), and 1280x720 ($r = 3$).

Representation	$r = 1$	$r = 2$	$r = 3$
Total volume [MB]	26.52	84.86	238.57
Mean segment size [kB]	75.77	242.47	681.64
Maximum segment size [kB]	301.17	789.66	2142.00
Minimum segment size [kB]	3.76	9.60	20.22
Standard deviation [kB]	37.14	127.09	419.74
Coefficient of variation	0.49	0.52	0.62
Lag-1 autocorrelation	0.76	0.82	0.87

The evaluation scenario is based on a realistic UMTS throughput pattern, which was recorded while driving on a highway around Klagenfurt, Austria [103]. The mean measured bandwidth was 2.81 Mbps. The throughput pattern was scaled to a mean adjusted bandwidth of 2.67 Mbps, such that after $n\tau$ seconds (i.e., the video duration) the video can be completely downloaded in its highest representation, i.e., $V(n\tau) = \sum_{i=1}^{n} S_{i3}$. For each run, a randomized starting point is selected and the network pattern is wrapped around. In this way, 30 different realistic bandwidth patterns with the same statistical characteristics were created and the optimization problem was solved for each pattern. The throughput patterns of the first three simulation runs are shown in Figure 2.9. It can be seen that the bandwidth fluctuates rapidly during each run between a minimum of 4.64 kbps and a maximum of 5.31 Mbps.
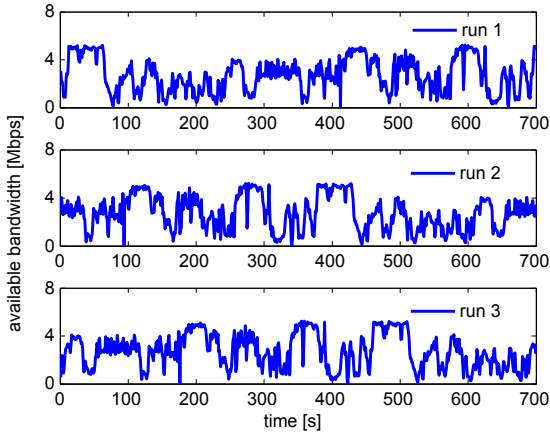
Figure 2.9: Network pattern, i.e., available bandwidth over time, of the first three evaluation runs. The measured traffic pattern was adjusted to the given video and wrapped around with a randomized starting point for each evaluation run.

2.3.3 Performance Evaluation Results

In a testbed, the four HAS adaptation algorithms introduced in Section 2.1.1 were implemented, namely, KLU [103], TUB [104], TRI [105], and BIEB [107]. Their adaptation was recorded for each of the 30 different bandwidth patterns. Additionally, the optimal strategy for each bandwidth pattern was computed by the linear program. It can be used as a reference to benchmark the different HAS adaptation logics with respect to the optimal QoE that can be achieved in the evaluation scenario.
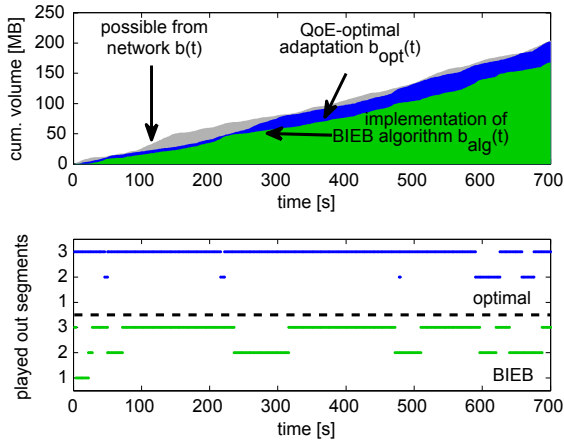
Figure 2.10: Single experiment comparing BIEB algorithm with theoretical QoE-optimal adaptation strategy for the network scenario sketched in Figure 2.9.

In Figure 2.10, the adaptation of the BIEB algorithm for the *run 1* network condition of in Figure 2.9 is shown. The upper plot shows the available cumulative download volume of the network ($b(t)$, gray), as well as the cumulative download volume utilized by BIEB ($b_{alg}(t)$, green) and the QoE-optimal adaptation ($b_{opt}(t)$, blue). The QoE-optimal playout strategy downloads the video

segments over time with bandwidth $b_{opt}(t) \leq b(t)$, but utilizes the available cumulative download volume almost completely. In contrast, a concrete implementation of a HAS adaptation logic, like the BIEB algorithm in Figure 2.10, utilizes the network less effective because a concrete adaptation logic does not have perfect knowledge about the current and future network conditions. Consequently, any HAS adaptation logic uses a bandwidth $b_{alg}(t) \leq b_{opt}(t) \leq b(t)$. The lower part of Figure 2.10 depicts the quality layers of the downloaded and played out segments. It can be seen that the QoE-optimal download strategy (blue) would enable to play out the highest quality layer for most of the time, while only occasionally switching to the intermediate layer for some segments. In contrast, the BIEB algorithm spends more time on the intermediate layer, and even uses the lowest quality layer at the start of the video, which would result in a lower, suboptimal QoE of the user.
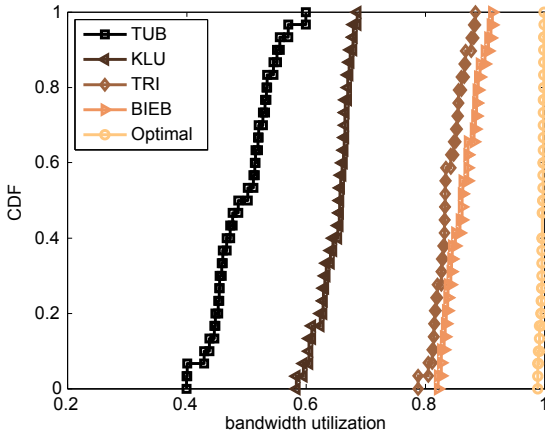
Figure 2.11: Bandwidth utilization of different adaptation logics.

Figure 2.11 shows the Cumulative Distribution Function (CDF) of the bandwidth utilization of each adaptation logic over all 30 bandwidth patterns. Note

that the bandwidth is adjusted, such that in order to download each segment of the video in the highest representation the whole available bandwidth has to be utilized, i.e., a bandwidth utilization of 100%. It can be seen that the QoE-optimal download strategy, as expected, almost always utilizes the entire available bandwidth. BIEB and TRI can utilize at least 82% and 78% of the available bandwidth, whereas KLU and TUB only utilize down to 58% and 40%, respectively. Also the maximum utilization of TUB (56%) and KLU (62%) is low compared to TRI and BIEB, which can come close to 90% bandwidth utilization.
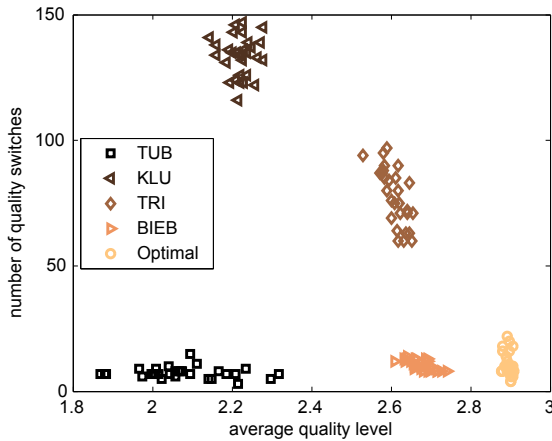
Figure 2.12: Scatter plot for 30 trial runs: average quality level vs. number of quality switches for the different adaptation algorithms and the optimal solution.

Still, the bandwidth utilization does not fully reflect the perceived quality of the algorithm. According to the QoE results above, the quality of a layer and the time on each quality layer have to be considered. Figure 2.12 depicts the achieved average quality level on the x-axis and the corresponding number of quality switches on the y-axis. Thus, the optimal combinations are located in

the bottom right corner of the plot. Each adaptation logic forms an own cluster, which indicates that the performance results presented in this evaluation are consistent over all runs. It can be seen that BIEB and TRI achieve the highest average quality levels with means of 2.67 and 2.61, which is close to the mean value of the QoE-optimal adaptation (2.90). TUB and KLU result in lower mean values of 2.07 and 2.22, respectively. Considering the number of quality switches, TUB and BIEB need few switches and adapt on average only every 92.00 s and 61.33 s, respectively. In contrast, KLU and TRI are more aggressive adaptation logics, which need much more quality switches and adapt the quality to the current network conditions more frequently every 5.15 s and 6.81 s, respectively.

The results of this exemplary benchmarking show that, in the given scenario, the BIEB adaptation logic has the best performance in terms of QoE. It reaches a high average quality level with a low number of quality switches, which brings it close to the QoE-optimal adaptation. TRI performed second in terms of playback quality, but the QoE might suffer from the high switching frequency. KLU achieved worse results both considering playback quality and switching frequency. TUB has the advantage of few quality switches, but reaches only an intermediate playback quality.

In summary, the presented linear program can be used to find the QoE-optimal adaptation for given network conditions and video characteristics. Therefore, it is a valuable tool for the performance evaluation of any HAS adaptation logic in different evaluation scenarios and allows for a more comprehensive assessment and benchmarking with respect to QoE.

2.4 Lessons Learned

In this chapter the impact of adaptation-related parameters on the Quality of Experience of HTTP Adaptive Video Streaming was investigated. This allowed to identify the most important QoE factors and quantify their impact on subjectively perceived quality of HAS services. These insights allow to improve HAS adaptation logics to mitigate the impact of adaptation on the QoE.

The investigation of the QoE of HAS started with a review of findings from previous studies. These works had concluded that initial delay, stalling, and quality adaptation had the biggest impact on the QoE. Nevertheless, the detailed results on the impact of adaptation on QoE were partly contradicting. Therefore, two crowdsourcing studies were conducted to collect subjective ratings for different adaptation patterns in order to investigate the impact of adaptation-related parameters.

The results of the study on adaptation-related parameters in Section 2.2.1 showed that only time on high layer and switching amplitude were main influence factors. In contrast, last quality level, recency time, and switching frequency had no significant impact. The results of related works on switching frequency might not be contradictory, when quality switches are correlated to time on each layer. The quality switches only call the user's attention to a perceivable degradation/improvement, for which again the duration matters. Similar findings were indicated in [142], which found that long low quality segments after high quality segments were perceived strongly negatively. Moreover, multiple quality switches were preferred over fewer switches, if the subject could watch a higher quality for a longer time. Thus, the impact of adaptation frequency is inferior to the impact of time on each layer. However, if the switching frequency is too high, such that the user perceives only flickering, the switching itself is the worst degradation, and the QoE might be even worse than for low video quality [138].

The results of the study on time on intermediate layer in Section 2.2.2 supported the findings of Section 2.2.1. It could be observed in both studies that the MOS values of the adaptation sequences were bounded by the MOS values of the constant high and low quality sequences. Moreover, it became evident that the quality of each layer (cf. switch amplitude in Section 2.2.1) and the time on each layer were the only adaptation-related parameters, which had a significant impact on the QoE of HAS. Based on these findings, QoE-based traffic management decisions could be taken in the network by monitoring the requested quality layers (e.g., by deep packet inspection or machine learning on

statistical packet features) and pooling the obtained metrics over time into QoE estimators. However, in order reach a high accuracy of estimation and to create a holistic QoE model for HAS systems, the results have to be extended and consolidated for different HAS systems (e.g., more quality layers, streaming on mobile devices) and different video contents (e.g., high motion sequences, longer sequence duration).

As it directly influences the adaptation of the played out video, the implementation of the HAS adaptation logic has the most impact on the QoE in a HAS system. The results showed that it should aim to maximize the played out video quality first before reducing the number of quality switches, for which no significant QoE degradation was found. In a step towards improved HAS adaptation logics on the client side, these QoE findings were incorporated in a linear program, which can be used to compute the QoE-optimal adaptation for given network conditions and video characteristics. This allows for a more comprehensive assessment and benchmarking of HAS adaptation logics with respect to QoE. The applicability of the benchmark was demonstrated in an exemplary performance evaluation of four adaptation strategies. Thus, the linear program is a valuable tool for the performance evaluation of existing and design of future HAS adaptation logics, which use an improved segment download strategy to maximize the played out video quality and reach a higher QoE for HAS systems.

3 QoE-aware Traffic Management for HTTP Adaptive Video Streaming

Traffic management is employed by network operators to meet Service-level Agreements (SLAs) for data transmissions and to reduce their costs by sophisticated utilization of the network resources. In early stages, pure network traffic management mainly focused on the efficient transmission of packets and flows. Nowadays, QoE-aware traffic management additionally aims to improve the QoE of networked services to reach a high end user satisfaction. This includes cross-layer traffic management, which utilizes information from different layers (e.g., application-layer information) for the traffic management process, and collaborative traffic management, which is based on the communication and information exchange between different stakeholders (e.g., exposure of client-side information to network) to manage the interplay of services and network.

The traffic management process can typically be described by a management cycle as depicted in Figure 3.1. It shows a networked service (orange) and the corresponding QoE-aware traffic management (blue) seen from different layers (vertical separation) and stakeholders (horizontal separation). The solid lines indicate classical network traffic management and the dashed lines show possible extensions by cross-layer and collaborative traffic management. First, the current situation in the network is monitored, e.g., in terms of QoS parameters measured on the network nodes or QoE parameters signaled by the client. The monitored data are collected, processed, and aggregated to performance met-

rics, and compared to target values. These target values can be predefined by the network operators, derived from the SLAs, or (dynamically) specified by the services. If the performance metrics and the targets diverge, a traffic management action has to be decided. Such actions include network mechanisms (e.g., routing, prioritization, bandwidth shaping, offloading, caching), but can also include sending requests to services to change their network demands. The network actions are put into effect by changing the settings of network nodes specifying how to handle the respective flows. Afterwards, the cycle restarts and the monitoring of the performance metrics continues. The extensions of the classical network traffic management cycle by cross-layer and collaborative approaches allow to also monitor QoE on application layer, e.g., within the client application, and user layer, e.g., through quality feedback within or after a session. Moreover, service characteristics can be considered for traffic management decisions or even altered by traffic management actions.
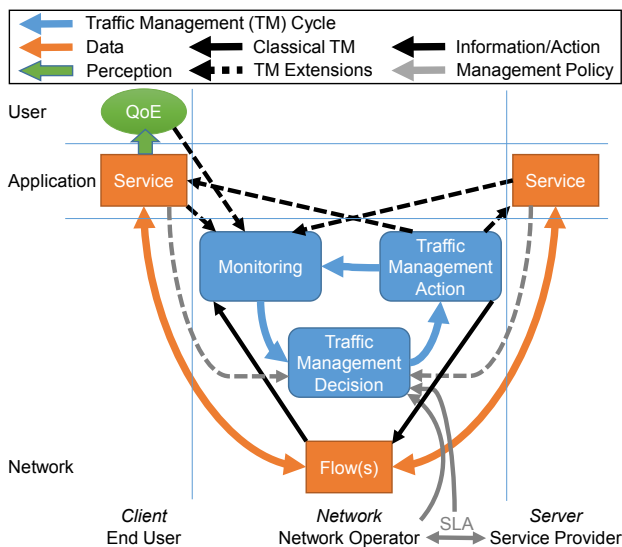
Figure 3.1: QoE-aware traffic management cycle.

Based on the identified major QoE influence factors of HAS, this chapter will investigate approaches to monitor the QoE of HAS on different layers. In particular, in-network measurements of objective per-frame and per-segment video quality metrics, which indicate the quality of each layer, are temporally pooled with different methods to account for the impact of the time on each layer. An Android app is presented, which not only allows to monitor application-layer QoE parameters, but also network and context parameters, and subjective feedback of end users. Due to the huge amount of monitored data on different layers, the app is well suited to further research the QoE of HAS on mobile devices. Moreover, the utilization of monitored network parameters to directly estimate QoE without considering application-layer QoE factors is investigated. As the QoE monitoring enables QoE-aware traffic management of HAS flows, a performance evaluation of different bandwidth allocation strategies for video flows on a shared bottleneck link is conducted. Thereby, the benefits of different levels of monitoring information, i.e., from flow type to video characteristics to real-time buffer fill information, are investigated.

Section 3.1 outlines the background and related work on QoE monitoring and QoE-aware traffic management for HAS. New approaches for monitoring QoE parameters for HAS are presented in Section 3.2. These include both monitoring methods in the network, which are based on [61, 80], and on application layer, which are based on [34, 45, 48]. On the basis of [12], Section 3.3 conducts an analytical and simulative performance evaluation of different QoE-aware traffic management approaches, which utilize bandwidth shaping and prioritization of video flows on a bottleneck link. This is a common situation in access networks where multiple users and applications often share a single access link. The findings of this chapter are summarized in Section 3.4.

3.1 Background and Related Work

This section presents background and related work on QoE-aware traffic management for HAS. The section is divided into two subsections. The first sub-

section focuses on QoE monitoring of HAS, and thus, corresponds directly to the monitoring component of the QoE-aware traffic management cycle depicted in Figure 3.1. It includes monitoring approaches, which estimate QoE or QoE factors from network parameters, and trending machine learning approaches. Other works monitor QoE factors directly in the application, which might then be mapped to a QoE prediction using QoE models. The second subsection presents traffic management mechanisms for improving the QoE of video streaming sessions. These include pure network traffic management mechanisms as well as cross-layer and collaborative traffic management approaches. Depending on the specific mechanism, they can be classified as standalone components or complete solutions, which comprise all the components of the traffic management cycle. The structure of this section is summarized in Table 3.1.

Table 3.1: Overview of related works for QoE-aware traffic management for HAS.

QoE monitoring	
Surveys	[162–167]
Monitoring in the network	
QoE estimation	[168–173]
QoE factor estimation	[16, 27, 63, 174–182]
Machine learning	[183–189]
Monitoring in the application	
QoE factors	[176, 190–196]
QoE models	[132, 148, 197–199]
QoE-aware traffic management mechanisms	
Surveys	[200, 201]
Network traffic management	[202–211]
Cross-layer and collaborative traffic management	[17, 176, 212–239]

3.1.1 QoE Monitoring of Video Streaming

Network providers rely on QoE monitoring to assess the satisfaction of end users and use the monitored information for traffic management decisions. Typically, they distribute hardware middleboxes in their networks, which investigate the traffic, extract relevant features, and estimate the QoE. Moreover, following the recent trend towards virtualization (Software-defined Networking (SDN), Network Function Virtualization (NFV)), also Virtual Network Functions (VNFs) for QoE monitoring are deployed in the network (e.g., virtual Deep Packet Inspection (DPI) functions [240, 241]).

Many approaches can be taken to monitor the QoE of HAS, which have different levels of complexity and accuracy. Most related works are already covered in recent survey works, which will be shortly outlined: [162] reviews methodologies to evaluate the QoE of video streaming. Similarly, [163] gives a tutorial on popular QoE assessment approaches for video streaming. Reference [164] surveys metrics, tools, and measurement methodologies to predict the QoE of video streaming, and [165, 166] not only focus on QoE factors and assessment, but also take QoE management into account. The current standardized quality assessment models are described in [167] differentiated by the level of information obtained from the media stream. In the following, details on some relevant monitoring approaches are given for both monitoring in the network and monitoring in the application.

Monitoring in the Network

A prerequisite to in-network monitoring of video flows is the detection and classification of such flows. Related works on this aspect (e.g., [168, 169]) will not be tackled in this section, but all presented related works require that video flows were already identified. Reference [170] used QoS parameters, namely, packet loss, packet loss burstiness, and jitter, as input for a fuzzy expert system to predict the MOS of video flows. Reference [171] monitored network parameters of an UMTS network (block error rate, mean burst length). For the prediction of

the MOS, additional application-layer information (content type, sender bit rate) were utilized. Reference [172] described a QoE measurement architecture using mobile Android devices. The authors created a QoE model for non-adaptive YouTube streaming of very low bit rate videos. Therefore, linear regression models were used to map objective network parameters to subjective QoE factors, whose weighted average predicted the overall QoE score. Recently, a parametric bit stream-based quality assessment model for HAS was standardized [173], which predicted the MOS from stream inspection and supported four different modes of input information.

Other works focused on estimating application-layer parameters, which can be mapped to QoE. Reference [16, 174, 175] transferred the approach of [176] and proposed an in-network system based on DPI to extract downloaded playtimes. They could be used to estimate the buffered playtime at the client, and thus, the corresponding stalling events. The mapping to QoE is done with the model presented in [63]. Reference [177] followed a similar approach but supported more video encodings and container formats. Also [178] followed a similar approach to predict stalling in LTE networks. Reference [179] not only monitored stalling but also the quality level from an intermediate network element. Reference [180] used a decision tree to detect stalling events from HTTP segment requests in the network. Reference [181] estimated stalling events based on the ratio of playback time and download time, but needed the total size of the video for real time estimation of stalling. Reference [27] studied the accuracy of a VNF-based buffer and stalling estimation. [182] used traffic profiling to estimate the video bit rate and streaming phase.

Due to the trend towards end-to-end encryption, the DPI-based approaches cannot be applied anymore. Therefore, machine learning methods are utilized to estimate the QoE based on statistical features of the video traffic. In [183], passive in-network measurements were collected and machine learning techniques were applied to obtain mappings between QoS and QoE of mobile video applications. Reference [184] predicted video abandonment from radio network statistics and flow records with a decision tree. Reference [185] used packet

level measurements to estimate the video bit rate based on decision trees. Reference [186] analyzed the predictive power of both network and application QoE features. Reference [187] classified the video quality of segments with a K-means++ algorithm using the download bit rate as feature. Reference [188] used random forests to detect stalling, average video quality, and quality variations from network features, such as round-trip time, packet loss, and segment size. Reference [189] directly classified the QoE from network parameters with different machine learning approaches. Again, random forest resulted in the best prediction accuracy.

Monitoring in the Application

The monitoring of QoE can also be directly integrated into the application. The advantage of application monitoring is that most QoE information can be accessed directly and accurately, and does not need to be estimated. However, in order to share the QoE information with network providers and use it for traffic management, additional signaling is required.

Reference [176] conducted a client-side DPI to detect video flows and extract video playtimes contained in the downloaded packets. Together with a browser plug-in, which monitored the current state of the player, an accurate calculation of the buffer and detection of stalling events was possible. Reference [190] used a customized flash player to monitor initial delay, stalling frequency, and mean stalling duration from the client. These parameters are then mapped to QoE with a linear function. In [191], the explanatory power of the model of [190] could be increased by adding user-viewing activities. Based on [190], [192] implemented an evaluation tool for YouTube QoE in Android mobile devices. However, this application did not take adaptive video streaming into account. In [193], a mobile application was presented to actively measure and analyze mobile app QoE on network and application layer. Reference [194–196] proposed a browser plug-in that passively monitors initial delay, stalling, and adaptation, and predicted the abandonment rate.

In case the QoE parameters are extracted from the client player, all QoE models from Section 2.1 can be utilized to estimate the QoE. For example, [132] input the quantization parameter, number of stalling events, and average and maximum duration of stalling events to a random neural network in order to estimate the QoE. Reference [197] predicted user engagement based on application-level QoE parameters. Reference [198] continued this work and applied different decision trees for the prediction of user engagement based on confounding factors, like device or video content. Reference [199] used only 15 s of stream state data, namely, number of stalling events, frame rate, bandwidth, and received packets per second, to predict the QoE using a K-nearest-neighbor algorithm with a dynamic time warping metric. Another automated QoE evaluation based on a non-linear function was described in [148] using metrics for objective content quality, stalling, and adaptation.

3.1.2 QoE-aware Traffic Management Mechanisms for Video Streaming

The goals of QoE-aware traffic management are to efficiently utilize the available network resources and to improve the QoE of end users, which typically includes to establish QoE fairness [242] between different applications and users. Reference [200] surveyed various approaches to QoE-aware traffic management approaches in wireless networks, whereas [201] only focused on mechanisms for improving the QoE of video streaming. Selected traffic management approaches are presented in chronological order and separated into pure network traffic management solutions, and mechanisms, which use information from applications and/or signal applications to change their network demands.

Network Traffic Management

Reference [202] proposed to limit the TCP congestion window on the streaming server to control the burstiness of adaptive video traffic, which resulted in reduced retransmissions and round trip times. Reference [203] used DPI to detect

video flows and redirected them to less congested links using SDN to maintain a high level of QoE. A similar approach is presented in [204], in which different layers of H.264/SVC video stream can be routed over different paths to reduce the stalling. References [205, 206] proposed to rewrite segment requests and to shape the TCP throughput of each client in an LTE network with a proxy server according to QoE parameters extracted from the index file and the streamed content. Reference [207] estimated the buffered playtime of video streams and used it to assign resources in LTE networks such that stalling could be avoided. Reference [208] considered dynamic offloading to improve the QoE of YouTube videos when multiple access networks were available. Reference [209] used a network proxy to guarantee the video delivery of H.264/SVC-based live streaming using Differentiated Services (DiffServ). Reference [210] detected and monitored the buffer of video flows on home gateways. Stalling of all flows could be minimized by assigning flows to different priority queues. Reference [211] proposed a system that uses a neural network to detect and estimate the QoE of video streams from flow parameters. The traffic management is done by bandwidth shaping, considering also premium users of HAS services.

Cross-layer and Collaborative Traffic Management

First cross-layer traffic management solutions for mobile video streaming were proposed in [212–215], which used information about the packet content (e.g., type of frame) to prioritize the most important frames (e.g., I-frames of MPEG-compliant video encodings) or transmit them over channels with the best transmission capabilities in order to reach a high video quality. References [17, 216] used the tool presented in [176] to detect and monitor video flows, and prioritized or routed them based on both network and application information.

Reference [217] proposed to utilize a proxy server to monitor the streaming and assign quality levels to clients in order to manage the QoE. Reference [218] used a measurement proxy server to detect the highest quality level currently supported by the network and accordingly instruct the client. Reference [219] used a cloud-based framework to transcode the video according to network re-

quirements. Reference [220] considered a collaborative traffic management for HAS including caching of segments on a proxy server. Reference [221] proposed an OpenFlow-based framework that monitored the status of all HAS applications and dynamically allocated network resources to each devices. Moreover, the system instructed clients which quality levels to request in order to achieve a high fairness among all users. Reference [222] used a proxy to share monitored information of all streaming sessions with the clients, such that a fair distribution of network resources was achieved.

Quality information about the layered content is used in [223] to adapt videos in order to improve the average QoE for a given total bandwidth, and in [224] to prioritize H.264/SVC layers in mobile networks according to their QoE contribution. Reference [225] used network and application information of multiple applications for resource allocation in LTE based on particle swarm optimization. References [226, 227] proposed server-side traffic shaping to improve the QoE of HAS sessions. Reference [228] described an SDN-based framework, which changes streaming servers and routing paths based on QoE feedback from the client. A similar approach using ant colony optimization to compute the best paths is presented in [229]. Reference [230] used an OpenFlow controller for prioritized delivery of HAS segments to avoid imminent stalling. Reference [231] estimated the QoE based on encoding, stalling, and video duration, and recommended quality adaptations. Reference [232] presented an SDN-based architecture, which dynamically allocated the network resources for each client based on its expected QoE. Reference [233] leveraged the edge computing paradigm to improve the QoE of HAS by modifying the HAS index files in the network. Reference [234] proposed bandwidth reservation for HAS flows and signaled clients which quality levels to request. Reference [235] proposed a system that recommends video adaptations to mobile video streaming clients based on video characteristics and not only based on the bit rate. Reference [236] described an adaptation logic, which used information about network conditions to achieve a fair QoE for multiple clients. [237] presented a collaborative traffic management system, which considered QoE fairness in case of encrypted HAS. Recently

Server and Network Assisted DASH (SAND) standardization efforts were started to exchange information measured at servers and network elements (parameters enhancing delivery) and also signal quality-related assisting information to clients (parameters enhancing reception) [238, 239].

3.2 QoE Monitoring of HAS

According to the QoE studies presented in Section 2.2, not only initial delay and stalling, but also adaptation and the corresponding time on each quality layer have a major impact on the QoE of HAS. Thus, these factors have to be considered for QoE monitoring. This section presents two approaches for estimating the application-layer QoE factors of HAS. The first implements in-network monitoring and is based on [61, 80], the other implements monitoring in the application based on the works in [45, 48]. Finally, an approach to estimating the QoE from simple network parameters is described [34], whose results encourage the usage of application-layer parameters for QoE monitoring.

3.2.1 Monitoring of QoE Parameters in the Network

Network operators favor QoE monitoring in the network as an input to their traffic management decisions. Having their own monitoring, they do not have to base the traffic management on information provided by other stakeholders, in which they might have limited trust. The monitoring is typically implemented on middleboxes or VNFs, which analyze the traffic and output QoE estimates on a per session base. For the estimation of stalling, several methods were already proposed in related work (cf. Section 3.1), and even deployed in operational mobile networks [16]. With the increasing popularity of HAS, for which also adaptation parameters influence the QoE, additional methods are required.

The following method complements the approach in [16], and is also based on video information, which can be obtained via DPI. More specifically, this method is based on objective video quality metrics, which indicate the image quality of

individual frames or segments. When the adaptation logic changes the quality of a HAS stream, the bit rate of the video, and thus, the image quality changes, which can be measured by these objective video quality metrics. As also the time on each quality level is an influence factor of the QoE of HAS, temporal pooling will be used to aggregate these periodical measures of the video sequence over time into one measure for the whole sequence.

The quality level of the segment can be obtained by monitoring the segment requests of the client. The network operator can intercept the index file, which contains the URLs of each segment and the corresponding segment quality (e.g., bit rate). As the client requests segments individually, the network operator can inspect the HTTP request, extract the URL, and thus, obtain the quality level of the requested segment. Moreover, the network operator can inspect all downlink packets and extract and decode the actual video file, which is streamed to the client. This allows to compute objective per-frame metrics. Note that additional information might be needed, e.g., in case of full reference image quality metrics, the reference video file in highest quality has to be additionally requested by the monitoring function. In this case, an in-band signaling of this additional information, e.g., in the index file, could be beneficial to reduce the (computational) overhead of QoE monitoring. The recent trend towards end-to-end encryption of HAS sessions will prevent DPI-based approaches. Still, end-to-end encrypted flows can be terminated by a man-in-the-middle proxy, which could be implemented, for example, by the home router in a residential environment, or the needed quality information can be estimated by or signaled to the monitoring function, which is out of scope of this work.

Instead, the accuracy of ten different temporal pooling methods for estimating the QoE of HAS will be studied. Therefore, three objective video quality metrics will be considered, namely, the ordinal quality layer of the segment and the segment bit rate, which both can be estimated from encrypted traffic [182, 185, 187, 188], and the per-frame full reference Structural Similarity (SSIM) metric [243], which was calculated for the luma component of the two considered video clips, and represents the situation when more information (unencrypted

Table 3.2: Quality layers and target bit rate settings.

Quality layer	1	2	3	4	5	6
Target bit rate [kbps]	128	210	350	545	876	1410

flow and additional reference video clips) is available. Further results for other objective metrics are given in [61, 80].

Study Description

The study was conducted within the project Service Quality Definition and Measurement (NGMN P-SERQU) of the Next Generation Mobile Networks (NGMN) Alliance [80] and used two 100 s videos from the open source Sintel[1] project, which had animated content with a high motion intensity and high degree of detail. Only one content class was used in order to focus on the relationship between different quality levels and QoE, and not different contents. The clips consisted of 20 segments of 5 s each, which were encoded with different QP to reach six target bit rates as described in Table 3.2. These bit rates were spread below 1.5 Mbps, which is the maximum bit rate supported by all generations of the considered end user devices, i.e., iPhone and iPod. Moreover, it was checked that the different quality layers could be visually distinguished. An HLS-based HAS system was set up in a wireless testbed, which emulated mobile LTE conditions (fading, interference/noise, latency, jitter, competing traffic), and the resulting adaptation patterns were recorded. Too similar patterns were removed and several artificial patterns were added, i.e., patterns, which were not recorded while using the HLS client, but could occur with other HAS adaptation logics. The 90 resulting patterns contained up to 20 quality changes but no stalling.

To collect a sufficiently high number of subjective ratings from diverse participants, a crowdsourcing study was designed. To overcome the reliability problems of an anonymous crowd, only voluntary participants were recruited from

---

[1]`https://durian.blender.org/` [Online] – Accessed: 24.08.2017

the NGMN P-SERQU partners. The tested patterns were prepared offline and offered as Podcasts consisting of five different video patterns each. The participants downloaded a Podcast on their iPhone or iPad, and accessed a study questionnaire with their browser. After downloading the Podcast and completing a personal data questionnaire, which included consistency questions to check the reliability, the users sequentially watched all five video clips. After each video, the participants accessed the study web page and rated their experience on a 5-point ACR scale. The study was completed by 494 users during the five month test phase. The following results are based on the subjective ratings of those 297 users, who watched the videos on an iPad.

Temporal Pooling

The basis of the temporal pooling analysis is formed by periodic values of an objective metric OM. In this work, the pooling is done on a per-frame base, i.e., for each frame $t, t = 1, \ldots, T$, an objective video metric $OM(t)$ is considered. Note that for segment metrics (i.e., number of quality layer, bit rate), each frame of the segment will have the same value, and the pooling on a per-segment base would give the same results because of the fixed segment length.

A literature review found several different temporal pooling methods, which are investigated in this study. The most intuitive approaches described in [244] are mean pooling (*Mean*, $\frac{1}{T} \cdot \sum_{t=1}^{T} OM(t)$) and last frames mean pooling (*MeanLastFrames*, $\frac{1}{F} \cdot \sum_{t=T-F}^{T} OM(t)$). These methods simply average the objective metrics over all frames, or the F most recent frames, respectively. In contrast, the influence of highest quality frames is emphasized in *Minkowski* summation (*Minkowski*, $[\frac{1}{T} \cdot \sum_{t=1}^{T} OM^p(t)]^{1/p}$) and exponentially-weighted Minkowski summation (*ExpMinkowski*, $[\frac{1}{T} \cdot \sum_{t=1}^{T} \exp(\frac{t-T}{\tau}) \cdot OM^p(t)]^{1/p}$) by tuning the parameter p. The additional parameter τ in *ExpMinkowski* controls the exponential weighting, and thus, the recency effect. Moreover, [244] described the histogram pooling (*Histogram*), which is computed as the k-th percentile of the cumulative histogram values. Thus, a low k value can take the impact of lowest quality frames on viewers into account. The local minimum of mean values of N

successive frames (*LocalMinimum*, $\min[\frac{1}{N}\sum_{i=1}^{N} OM(t+i)]$) also accounts for the poorest quality section. A related method, which was not described in [244], computes the mean of the p percent of overall frames with lowest quality (*Percentile*).

Reference [245] described the hysteresis effect, which combined an element that indicates the quality of the current frame with the minimum of quality scores over the last τ seconds. The final score was computed as the mean of these values over all frames (*Hysteresis*, see detailed description in [245]). Reference [246] divided the frames of lower and higher quality according to a percentile. When combining the scores, a scaling factor was multiplied to the scores of the lower quality frames (*SequenceLevel*, see detailed description in [246]). Finally, in [247] the mean of objective metrics and the differences between successive frames were considered. The score emphasized quality degradations down to a saturation threshold (*VQA*, see detailed description in [247] but slightly modified).

Some of the presented temporal pooling methods can be tuned by adjusting the parameters, e.g., to emphasize low or high quality frames or to account for recency effects. Thus, the optimal parameter values for each method have to be found, which maximize the correlation of the temporal pooling output with the MOS values of each pattern. Due to the limited data set, a cross-validation approach was implemented using the leave-one-out algorithm in order to achieve a good performance independent of the training and test sets. For every pattern, the leave-one-out algorithm computes the output of the pooling method after training on all remaining patterns. Eventually, there will be one cross-validated output for each sample in the data set. To evaluate the overall performance of the temporal pooling methods, the cross-validated outputs are correlated with the MOS values of the crowdsourcing study in terms of Pearson Linear Correlation Coefficient (PLCC) and SROCC.

Results

The results of the different temporal pooling methods after cross-validation are shown in Table 3.3. Note that some special cases exist, in which a parametrization of a temporal pooling method was identical to another method, i.e., *Minkowski* can be identical to *Mean* for $p = 1$, and *LocalMinimum* can be identical to *MeanLastFrames* when the minimum of the mean of all successive frames is the mean of the last frames. To keep the characteristics of the methods visible, the best result, which is not a special case is shown in the corresponding row. The three considered objective metrics are the ordinal quality layer, i.e., the number of the quality layer from 1 (lowest) to 6 (highest), the actual bit rate of each segment after encoding, which might differ from the target bit rate, and the per-frame SSIM computed with the highest quality representation of each video as reference.

Table 3.3: PLCC and SROCC between the pooled objective metrics (segment quality layer, segment bit rate, per-frame SSIM) and subjective MOS after cross-validation (sorted by PLCC of quality layer pooling).

Pooling method	Quality layer		Bit rate		SSIM	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
VQA	0.867	0.860	0.751	0.750	0.820	0.767
SequenceLevel	0.861	0.851	0.797	0.829	0.790	0.815
Mean	0.856	0.837	0.751	0.730	0.870	0.866
LocalMinimum	0.831	0.803	0.671	0.615	0.506	0.511
MeanLastFrames	0.829	0.805	0.721	0.693	0.437	0.445
Percentile	0.824	0.806	0.715	0.781	0.854	0.864
ExpMinkowski	0.822	0.802	0.730	0.711	0.572	0.513
Minkowski	0.820	0.800	0.718	0.678	0.815	0.842
Histogram	0.785	0.789	0.666	0.683	-0.004	-0.150
Hysteresis	0.766	0.726	0.664	0.662	0.867	0.859

The best performing methods for the segment quality layer metric, both in terms of PLCC and SROCC, are *VQA*, *SequenceLevel*, and *Mean*. The top results of *VQA* are a PLCC of 0.867 and SROCC of 0.860, which indicate a high correlation

to the MOS values. Nevertheless, most pooling methods reach high correlations above 0.8. For the sequence bit rate metric, the same pooling methods perform best. Additionally, *Percentile* reaches the second highest SROCC. However, the performance of this metric is generally lower than the segment quality layer, resulting in lower correlations for every pooling method. *Mean* pooling of the per-frame SSIM metric, gives the highest correlations (PLCC = 0.870, SROCC = 0.866). Also *Hysteresis* and *Percentile* perform well on the SSIM metric. Nevertheless, several pooling methods perform bad and only reach correlations below 0.6.

The results show that temporal pooling is generally well suited to account for the temporal dimension of quality adaptation in a HAS scenario. High correlations can be achieved for all levels of information taking into account that the methods were intentionally not developed for longer video sequences as used within this analysis. The most sophisticated metric, which takes the quality of every frame into account, reaches the top correlation, but does not outperform the other metrics. In contrast, for several pooling methods, the results were much worse. Surprisingly, the most trivial metric, considering only the number of the quality layer, gives a decent performance for most pooling methods. It reached higher correlations than the segment bit rate, which might be due to the complex relationship between video content, compression, and resulting video quality.

This means that a very low complexity algorithm, such as the mean of segment quality layer, gives already a decent prediction performance for the QoE of HAS, although it includes almost no information about the underlying content. Thus, this approach might be also applied in a scenario with end-to-end encryption, for which several works already tried to estimate the quality layer from network traffic features, e.g., [187, 188].

3.2.2 Monitoring of QoE Parameters on Application Layer

A different approach is monitoring QoE information on application layer. The advantage is that the QoE factors, which are perceived by the end user, can be directly obtained within the application without the need for network-layer DPI and/or estimation. This means, application-layer monitoring will always give accurate information, also in case of end-to-end encryption. The disadvantages, however, are that the monitoring has to be implemented at the client side, and that the monitored information has to be signaled to the network operator in order to utilize it for traffic management.

This section presents a methodology for monitoring application-layer QoE factors of YouTube on mobile devices with an Android app, which is called YoMoApp[2] (YouTube Monitoring App) [48]. As the main QoE influence parameters of HAS are initial delay, stalling, and adaptation, the app unobtrusively monitors the state and buffer of the video player, and the resolution of the played out videos.

In the following, the functionality of the app is described. The app uses an Android WebView[3] browser element to access the mobile YouTube website[4]. The mobile YouTube website includes Hypertext Markup Language Version 5 (HTML5) video playback using DASH technology. Similar to the browser plugin of YoMo [176], JavaScript functions are injected to the website, which detect the video element in the Hypertext Markup Language (HTML) Document Object Model (DOM) tree and ultimately perform the monitoring of the video playback. These functions add event listeners to the HTML5 video element to monitor changes of the player state (e.g., playing, paused, buffering, ended), and the height and width of the video element, which indicate the video resolution. The current playback time and the buffered playtime are polled periodically every 1 s. Some of the monitored parameters are visualized in Figure 3.2. Moreover,

---

[2]`https://play.google.com/store/apps/details?id=de.yomoapp` [Online] – Accessed: 24.08.2017

[3]`https://developer.android.com/reference/android/webkit/WebView.html` [Online] – Accessed: 24.08.2017

[4]`https://m.youtube.com/` [Online] – Accessed: 24.08.2017

the YouTube ID, title, duration of the watched video, and statistics about the precedent advertisement clip are retrieved. The monitored data are sent to the Android app, where they are processed and logged. Postprocessing is required because the usage of JavaScript can introduce inconsistencies and errors, e.g., missing/incorrect values or non-equidistant data queries.

(a) Playback of an exemplary EDGE session.

(b) Playback of an exemplary LTE session.
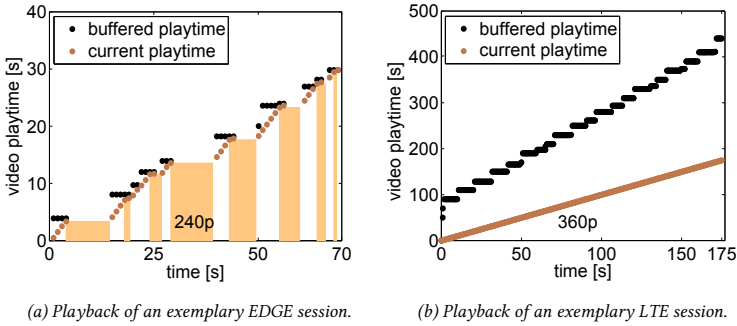
Figure 3.2: *Illustration of some of the monitored application-layer parameters for two streaming sessions: current video playtime (orange) and buffered video playtime (black). Stalling events are displayed as yellow boxes.*

In addition to the monitoring of the playback, the native Android part of YoMoApp also logs network and context parameters. It logs several device characteristics and listens to their changes, namely, screen size, screen orientation, volume, player size, and player mode (normal/full screen). The app monitors its network usage (both mobile and Wi-Fi networks) as shown in Figure 3.3, and thus, periodically logs the download and upload volume, as well as changes of operator, Radio Access Technology (RAT), cell ID, signal strength, and Global Positioning System (GPS) position. For each video session, i.e., the playback of a new video, separate log files are created. After each session (minimum session length: 20 s), the user is asked to rate the QoE of the session on a continuous 5-point ACR MOS scale ranging from 1 (bad) to 5 (excellent) [248]. Note that the user is not required to rate but can also close the rating dialogue. The log

files are locally cached and transmitted to an external database when closing the app, triggered manually by the user, or at fixed time intervals. Note that the most recent version of YoMoApp features a discrete 5-point ACR MOS scale for rating the video liking, video quality, streaming quality, and acceptability of the streaming. However, the results presented in the following were gathered with the previous version of YoMoApp.

(a) Network usage of an exemplary EDGE session.   (b) Network usage of an exemplary LTE session.
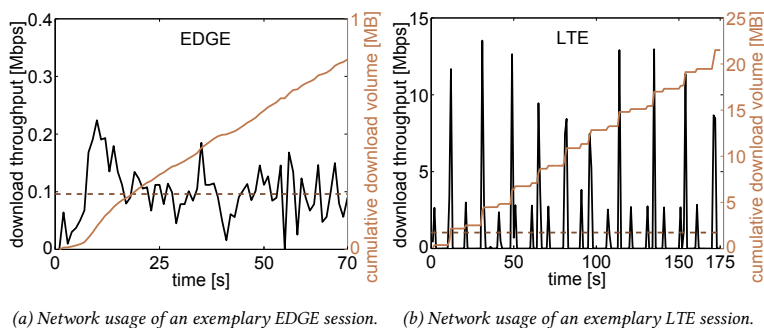
Figure 3.3: Illustration of some of the monitored parameters on network layer: download throughput (black) and cumulative download volume (orange). The depicted sessions correspond to the sessions of Figure 3.2.

To provide incentives for using the app, the user can access several aggregate statistics for each session as well as a visualization of the streaming similar to Figures 3.2 and 3.3. Moreover, a map view is included, which displays all subjective ratings and the corresponding network operators. The users can filter by operators, and thus, can compare with a simple heat map which operator performs best in which locations. Finally, users, especially researchers, which use YoMoApp can access the log files of the streaming sessions also on the YoMoApp web portal[5] for further evaluations.

Its rich functionality makes YoMoApp is not only an accurate application-layer QoE monitor, but the app is also a valuable tool for researching the QoE of

---

[5]`http://yomoapp.de/dashboard` [Online] – Accessed: 24.08.2017

HAS. Therefore, the applicability of YoMoApp is demonstrated in a subjective lab study [45], which investigates the impact of bandwidth throttling on the QoE. The participants of the study watched YouTube videos with YoMoApp and rated their subjective experience.

Study Description

The study was conducted in a dedicated lab for QoE tests at Forschungszentrum Telekommunikation Wien (FTW) in Vienna, Austria, which is compliant to the standards for subjective studies [84, 248, 249]. YoMoApp was installed on Android smartphones (Samsung Galaxy S4, Android 4.4 KitKat), which were connected to the Internet via Wi-Fi. To throttle the bandwidth, the downlink traffic was routed through a modified NetEm network emulator [250]. Three constant downlink bandwidths of 1 Mbps, 2 Mbps, and 4 Mbps were tested. Moreover, a variable bandwidth condition was included, for which the bandwidth was set to 1 Mbps, but increased to 3 Mbps for 5 s three times per minute (average bandwidth 1.5 Mbps). Finally, an outage condition was tested, for which the bandwidth was set to 4 Mbps with outages of 10 s twice per minute (average bandwidth 2.7 Mbps).

52 people (29 female, 23 male) participated and were compensated with vouchers, which asserted reliable and thoughtful study execution. They had to watch five two minute long YouTube videos, which were linked to a specific bandwidth condition each, according to Table 3.4. The videos were available as 4K ultra-HD videos (i.e., 2160p), and included a movie trailer and four nature-themed clips. After the playback, the participants rated the overall experience, initial delay, stalling, and video image quality on continuous 5-point ACR MOS scales [248]

Results

The evaluation of the results will only focus on stalling and adaptation, which are the most important QoE factors of HAS. Figure 3.4a shows the CDF of the

Table 3.4: Video content and download throughputs as measured in the lab experiment.

YouTube video ID	Avg. throughput (condition)	Avg. throughput (free)
6pxRHBw-k8M	2.8 Mbps (4 Mbps)	4.6 Mbps
iNJdPyoqt8U	1.0 Mbps (1 Mbps)	5.8 Mbps
kObNpTFPV5c	1.8 Mbps (2 Mbps)	5.0 Mbps
QS7lN7giXXc	2.3 Mbps (out)	5.0 Mbps
suWsd372pQE	1.3 Mbps (var)	3.9 Mbps

overall stalling duration of each bandwidth condition. As expected, almost no stalling occurs for the constant bandwidth conditions because the adaptation logic of YouTube can easily adapt the video quality. For 1 Mbps, 85.00% of the streaming sessions do not include stalling. For the 2 Mbps condition, 95.12% of the sessions, and for 4 Mbps, 93.48% of the sessions have a smooth playback without stalling. Also the remaining sessions have very small stalling times in the order of a few seconds, except for one outlier with 22 s stalling at 1 Mbps. In variable bandwidth condition ("var"), still 85.37% of the sessions show no stalling, but the stalling times of the remaining sessions are higher and reach a total stalling length of up to 34 s. The most stalling can be observed for the outage condition ("out") with an average total stalling length of 25 s. Only 21.43% of the sessions have a smooth playback. The remaining 78.57% of the sessions show stalling up to the maximum total stalling length of 41 s.

Figure 3.4b shows the corresponding subjective ratings in terms of MOS and 95% confidence intervals. After each video, the participants were asked to which extend the interruptions caused by stalling were perceived disturbing, ranging from 1 (very disturbing) to 5 (not disturbing at all). For the constant and the variable bandwidth conditions, which have a MOS of at least 4.41, the interruptions were not perceived disturbing. This corresponds to the monitored total stalling time, which showed very short stalling times or no stalling for these conditions. Stalling was only disturbing in the outage condition, which is indi-

(a) Distribution of total stalling length.

(b) MOS of stalling disturbance.
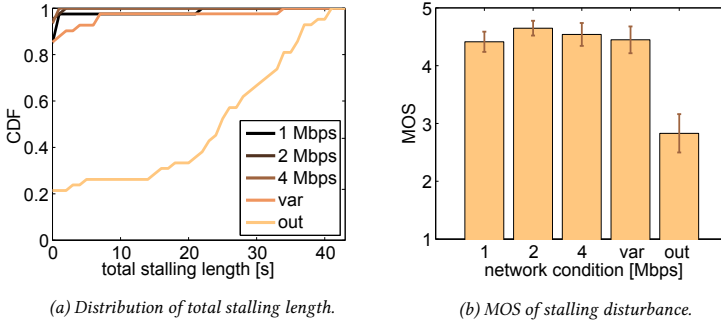
Figure 3.4: Monitoring of stalling and its impact on user experience. In (b), partic-
ipants rated the disturbance of stalling on a continuous scale ranging
from 1 (very disturbing) to 5 (not disturbing at all).

cated by a MOS of 2.83. However, this again corresponds to the high stalling times monitored by YoMoApp.

In Figure 3.5a, the time on each quality layer is analyzed. It shows the percentage of time, for which each resolution was played out during the streaming. Note that the highest played out resolution is 720p, which might be due to the limited device capabilities. In the 1 Mbps condition, mainly 360p (92.89%) is requested by the adaptation logic, while the other resolutions are only played out for a very short amount of time. For the constant conditions with higher bandwidth, the share of higher resolutions increases, which is very intuitive. In the 2 Mbps condition 480p (88.84%) is the dominant resolution, while in the 4 Mbps condition the most time is spent on 720p (67.15%) quality. Not surprisingly, the outage condition, which is a 4 Mbps on/off pattern, shows very a similar quality distribution to the 4 Mbps condition. However, the variable bandwidth condition plays out the lowest resolution for most of the time (240p, 80.86%), which indicates that the adaptation logic of YouTube cannot well adapt to fluctuating conditions, and shows a very conservative behavior.

(a) Distribution of time on each quality layer.

(b) MOS of image quality.
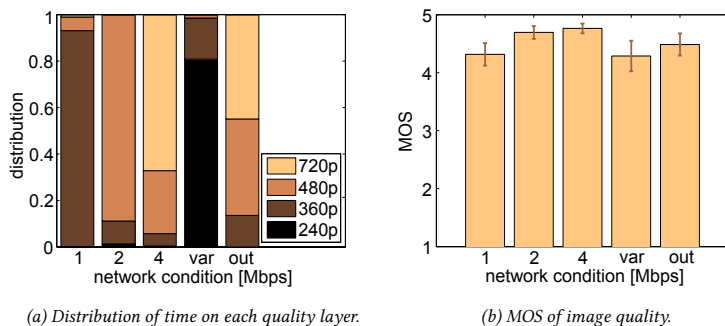
*Figure 3.5: Monitoring of time on each quality and its impact on user experience. In
(b), participants rated the image quality on a continuous scale ranging
from 1 (bad) to 5 (excellent).*

The corresponding subjective ratings on the image quality are presented in
Figure 3.5b. Again the MOS and 95% confidence intervals are displayed for each
condition. The most striking result is that the image quality is rated good for all
conditions having a MOS of 4.17 at least. This means that resolution adaptation
did not have a big impact on the QoE, which could be due to the used devices'
small screen size. This confirms the findings that the usage of HAS is especially
beneficial in mobile scenarios. Nevertheless, comparing the MOS values with
the time shares of Figure 3.5a, the general trend is visible that longer times on
higher resolution lead to higher QoE. This confirms the previous findings that
the time on each quality layer influences the QoE of HAS.

To sum up, YoMoApp can be considered a valuable tool for accurately and un-
obtrusively monitoring the QoE of HAS on application layer. As the monitoring
extends also to network and context parameters, and the app is able to collect
subjective feedback on video sessions, YoMoApp is also useful for researchers.
Another research application of YoMoApp to collect a data set for estimating
QoE from network parameters is described in the next section.

3.2.3 Monitoring of Network Parameters to Estimate QoE

The above presented study was continued as a field study to quantify the impact of realistic network conditions on the QoE of mobile YouTube streaming [34]. Moreover, the goal of the study was to estimate the QoE from monitoring the network parameters of the HAS session.

Study Description

The field study was conducted in Vienna, Austria, during two weeks in January 2015. 30 participants installed YoMoApp to their own smartphones and streamed videos through their own mobile ISP connection. Although YoMoApp was used to monitor the streaming, the used version was not yet able to monitor also network conditions and collect subjective ratings. Thus, another Android-based passive monitoring tool had to be installed to capture the network usage (flow start time, flow direction, flow duration, flow volume, flow throughput, mobile operator, RAT, cell ID, signal strength), and periodically upload the logs to an external server. The participants were asked to access a web-based questionnaire immediately after streaming a video. They rated the overall quality on a discrete 5-point ACR scale ranging from 1 (bad) to 5 (excellent). Moreover, they rated whether the session quality was acceptable on a binary scale (acceptable/not acceptable). The ratings were stored on the server for later analysis. Similar to the lab study, the participants were compensated with vouchers to assert reliable study involvement.

During the study, 85 videos were watched and the log files from the three sources (YoMoApp, network measurements, QoE ratings) were collected. As the timestamps were available, the identification of overlapping logs seemed a straightforward approach. However, only 30 YoMoApp logs could be mapped to network logs because for the remaining sessions the network monitoring app was not actively running on the participants' devices or a Wi-Fi network was used. Subjective ratings were only accepted when they were submitted within 15 minutes after the video session, and were filtered according to typical reliabil-

ity checks. This also resulted in only 30 YoMoApp logs, which could be mapped to subjective ratings. Eventually, only 10 YoMoApp logs could be mapped to both network logs and subjective ratings. These problems actually led to the improvement of YoMoApp by implementing capabilities for network monitoring and collection of subjective feedback.

Results

Figure 3.6 shows the correlation of the network measurements to the participants' subjective ratings. The correlations of the parameters with the MOS are shown in terms of SROCC (black), the correlation with the dichotomous acceptability in terms of Point-biserial Correlation Coefficient (PBCC) (yellow). For both throughput parameters small correlations to both MOS and acceptability can be observed. This means, a higher throughput is not directly related to a higher QoE. Flow duration, flow volume, and signal strength show negative correlations. This is surprising, especially for signal strength, because a higher signal strength seems to reduce the QoE. Note that this evaluation is based only on the 10 completely logged sessions, such that a generalization of the results is not possible. However, still there is evidence that network measurements are not sufficient for an accurate QoE estimation. In the following, the impact of network parameters on the streaming will be investigated in more detail.

Figure 3.7 illustrates the SROCC between different streaming parameters and the average flow throughput (dark brown) and signal strength (light brown). The data set consists of the 30 video sessions for which YoMoApp logs and network logs could be matched. The correlations are computed for initial delay, stalling parameters, adaptation parameters, and playback time. From the adaptation parameters, recency time refers to the time after the last quality change, and weighted time on layer is computed as the mean pooling of the ordinal quality layers, cf. Section 3.2.1. The average throughput has a fairly high correlation of 0.64 to both the end quality and the weighted time on layer, which confirms the findings of the lab study that a high throughput generally improves the streamed resolution of the video. Note that correlations are generally lower in this study
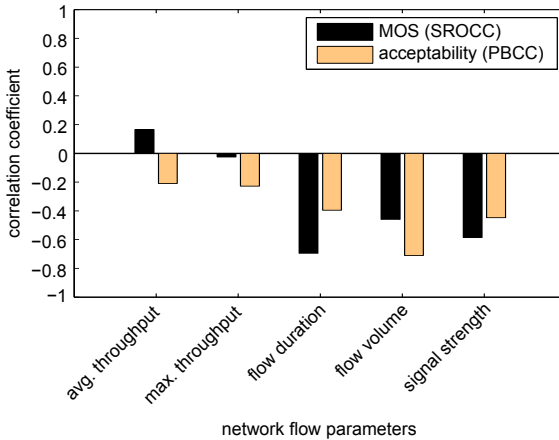
Figure 3.6: Correlations between network parameters and QoE ratings.

due to the very heterogeneous test environment. Fairly high negative correlations are visible for all stalling parameters, which also confirms that a higher throughput reduces the stalling. Similar trends can be observed for maximum flow throughput (similar SROCCs) and flow volume (lower SROCCs). An inverse trend can be observed for flow duration, which can be explained by the fact that longer flows are caused by worse network conditions. Again, signal strength only shows small SROCCs, which confirms that it cannot be directly used to indicate streaming parameters. These results shows that some network parameters, especially average or maximum flow throughput, are linked to streaming parameters and the resulting QoE. However, as observed above, the resulting QoE cannot be directly deduced from network parameters.

In contrast, the correlations of streaming parameters and the QoE ratings are investigated in Figure 3.8 for the data set of 30 video sessions with both YoMoApp logs and subjective feedback. Again, the SROCCs between streaming parameters and MOS are depicted as black bars, and the PBCCs of streaming
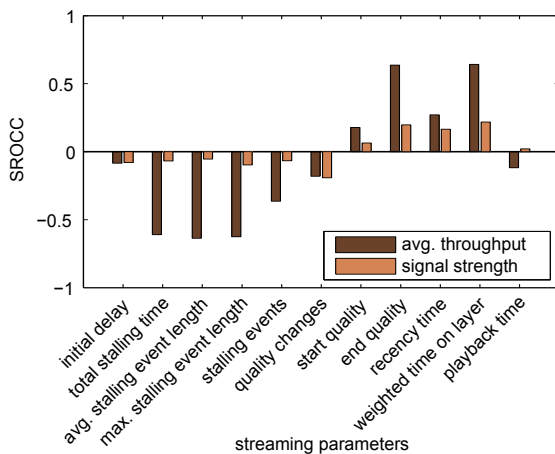
Figure 3.7: Correlations between network parameters and streaming parameters.

parameters and acceptability are shown in yellow. Initial delay does not show negative correlations, which might be due to very small initial delays during the field study, such that they were not an issue for the participants. Stalling, on the other hand, shows clear negative correlations to MOS, which confirms that stalling is the worst quality degradation of HAS [14]. The video quality parameters have fairly high correlations to the MOS, especially weighted time on quality layer (0.58), start quality (0.48), and end quality (0.48). In contrast, no correlation is visible for the number of quality switches (0.04). Still the fairly high correlation of recency time to MOS (0.52) could indicate that a stable streaming quality is valued by the users. Unlike MOS, the correlation coefficients of the streaming parameters and acceptability are very low, which could indicate that the acceptability is not influenced by a single streaming parameter but rather a more complex combination of them. Still, MOS and acceptability show a fairly high PBCC of 0.54 in the collected subjective ratings.
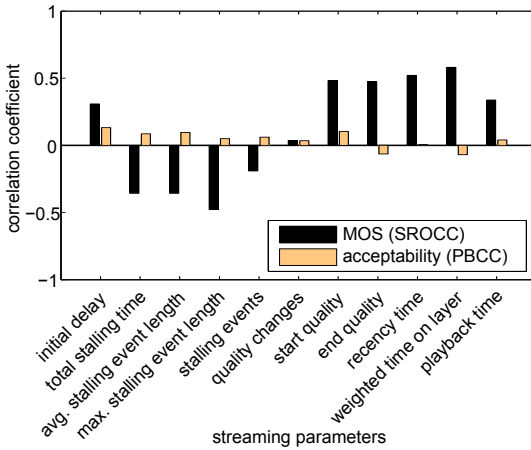
Figure 3.8: Correlations between streaming parameters and QoE ratings.

All in all, the field study shows that application-layer streaming parameters give better insights into the subjective experience of users than simple network parameters. Another big advantage is that application-layer parameters can be directly used as feedback in the QoE-aware traffic management cycle (cross-layer traffic management). The presented network monitoring approach shows that it is possible to extract QoE factors of HAS from unencrypted network traffic by DPI. Moreover, tools like YoMoApp can be utilized for application-layer QoE monitoring, which also work in scenarios with end-to-end encryption, but rely on signaling of QoE parameters for traffic management (collaborative traffic management). Finally, network-only monitoring and traffic management solutions might also be developed for encrypted network traffic by applying more complex machine learning approaches. Still, the estimation of application-layer parameters from the encrypted network traffic might be more beneficial and meaningful for QoE-aware traffic management than a direct estimation of the QoE from a complex, incomprehensible machine learning model.

3.3 QoE-aware Resource Allocation of Video Flows on Bottleneck Links

This section evaluates the performance of different QoE-aware traffic management approaches for a shared bottleneck link based on [12]. This is a common scenario in access networks, for example, when multiple users and applications share the same backhaul link. Traffic management will be applied by allocating the bandwidth resources of the bottleneck link to the users' video streaming and web browsing flows to achieve a high QoE. The study will confirm that a cross-layer approach utilizing application-layer QoE information is feasible and more dynamic than pure network-based traffic management. Section 3.3.1 formally describes the scenario and the resource allocation problem. The investigated traffic management strategies are presented in Section 3.3.2. Section 3.3.3 outlines the analytical and simulative performance evaluation methodology. Finally, the results of the performance evaluation are presented in Section 3.3.4.

3.3.1 Resource Allocation Problem on Bottleneck Links

The system of interest is a single shared bottleneck link, which is used by multiple users to stream videos or browse websites. All other links in the network are assumed to have a sufficiently high capacity. This means, the bottleneck link limits the throughput of each flow, and consequently, the throughput of each flow on the bottleneck link is the throughput of the flow in the whole network. To study the pure impact of the traffic management strategies, this scenario will only consider classical, i.e., non-adaptive, video streaming because the control loop of the adaptation logic would interfere with the control loop of the traffic management. Thus, an extreme and very critical HAS scenario is investigated, in which only the lowest quality layer is available. In order to achieve a high QoE for these video sessions, the goal of the traffic management is to avoid or reduce stalling [63]. With respect to web browsing, the page load time has the

most dominant impact on QoE [119], and thus, should be reduced by the traffic management strategies.

Table 3.5: Summary of used symbols for resource allocation problem formulation, resource allocation algorithms, and analytical performance evaluation model.

Symbol	Definition
$\mathcal{A}$	The set of active application flows
$a \in \{v, w\}$	An active video or web flow
$S \subset \mathcal{A}$	A subset of active flows
S_v, S_w, S_p	The set of active video/web/prioritized flows
C	Link capacity
q	Allocated bandwidth
q_a, q_S, q_p	Bandwidth allocated to flow/set of flows/set of prioritized flows
$\overline{b}_v$	Average video bit rate
σ_b	Standard deviation of video bit rate
s	Flow size
λ	Arrival rate
μ	Service rate
ρ	Utilization

The shared bottleneck link is assumed to have limited downlink capacity C. The uplink traffic will not be considered. The link is used to stream videos (identified by v) or browse websites (identified by w). Each application request triggers a download of content of size s_v or s_w, respectively. Thereby, each video has a certain video length t_v and an average bit rate $\overline{b}_v$, such that $s_v = t_v \cdot \overline{b}_v$. Video streaming and web browsing requests are modeled as Poisson processes, i.e., their inter-arrival times are exponentially distributed with rate λ_v or λ_w, respectively.

All currently active application flows are contained in the set $\mathcal{A} = \{a_1, ..., a_N\}$ and each flow may occupy a share of the link capacity C. In the typical case of transmission via TCP, all active flows share the link capacity equally. However, traffic management can influence the resource allocation for each flow $a_i \in \mathcal{A}$, or for each disjoint set of flows $S_j \subset \mathcal{A}$. Thus, the traf-

fic management strategy can assign a bandwidth q_{a_i} to each flow a_i, such that $\sum_{i=1}^{N} q_{a_i} \leq C$. Similarly, it can assign a bandwidth q_{S_j} to each set of flows S_j, such that $\sum_{S_j \subset \mathcal{A}} q_{S_j} \leq C$. The bandwidth allocated to a set of flows S_j will be shared equally among each flow $a_j \in S_j$, such that $q_{a_j} = \frac{q_{S_j}}{|S_j|}$.

When assigning bandwidth to the flows, the QoE-aware traffic management strategies have to consider different application characteristics, usage behaviors, and QoE factors. This problem can be formulated as follows: **Find a mapping** $\mathcal{A} \rightarrow \mathbb{R}, a_i \mapsto q_{a_i}$**, such that** $\sum_{i=1}^{N} q_{a_i} \leq C$ **holds, and the average QoE of all active flows** a_i**,** $i \in \{1, 2, \ldots, N\}$ **is maximized.** Note that the concept of QoE fairness [242] is not fully integrated, but it is assumed that improving the average QoE in the system will increase the number of users with an acceptable QoE. Nevertheless, the problem statement includes that the resource allocation takes different application types into account and the respective QoE factors. In this scenario, this means, the traffic management is supposed to achieve low stalling for video streams and short download times for web pages. In case of video streaming, therefore, the buffering ratio will be investigated, i.e., the total stalling time t_s divided by the total video length t_v. In case of web browsing, the download time of a web flow is considered.

3.3.2 QoE-aware Traffic Management Strategies

To target the objective, network operators can customize the bandwidth allocation according to different resource allocation strategies. These strategies may differ in needed input, complexity, and costs, and might result in different performance depending on different traffic distributions and application mixes. In the following, four algorithms for QoE-aware resource allocation on a shared bottleneck link are presented, which consider two applications, i.e., video streaming and web browsing. Network operators can implement these strategies, for example, by SDN [251, 252]. This emerging technology provides a unified interface to all networking elements, and thus makes the network more manageable and

controllable. Thus, it can be utilized to adjust the resource allocation on a per flow base according to the desired traffic management strategy [203, 253].

a) Fixed Bandwidth Allocation for All Videos Flows (FBV)

As video streaming is the more demanding application, the Fixed Bandwidth Allocation for All Videos Flows (FBV) strategy will allocate a fixed bandwidth for video traffic. Thereby, a fixed capacity q_{S_v} will be reserved for video traffic, which is equally shared by all video flows. The parameter q_{S_v} can be adjusted depending on the actual share of video traffic. To avoid starvation of the web browsing flows, q_{S_v} has to be smaller than C. Technically, this algorithm relies on a detection of the application type only (e.g., classification based on network traffic characteristics [168]).

Algorithm   The active flows are separated by type into the set of video streaming flows S_v and the set of web browsing flows S_w. All flows in S_v equally share the fixed allocated bandwidth q_{S_v}, such that each video flow receives $q_v = \frac{q_{S_v}}{|S_v|}$. The web flows share $q_{S_w} = C - q_{S_v}$ equally, such that each web browsing flow gets a bandwidth of $q_w = \frac{q_{S_w}}{|S_w|}$. If one set is empty, the other set can utilize the whole capacity C. In this case, $q_v = \frac{C}{|S_v|}$, if S_w is empty, and $q_w = \frac{C}{|S_w|}$, if S_v is empty.

The resource allocation is illustrated in Figure 3.9a. The two video flows v_1 and v_2 share q_{S_v} equally, and the web flow can utilize the remaining bandwidth $C - q_{S_v}$. Figure 3.9b shows the situation when S_w is empty and the video flows can share the whole capacity C equally.

b) Fixed Bandwidth Allocation for Each Video Flow (FBF)

The Fixed Bandwidth Allocation for Each Video Flow (FBF) strategy reserves a fixed bandwidth for each individual video flow, and thus, can take flow characteristics into account. For example, a Constant Bit Rate (CBR) video can be allocated its bit rate ($q'_{v_i} = \overline{b_{v_i}}$), or a Variable Bit Rate (VBR) video can be allo-
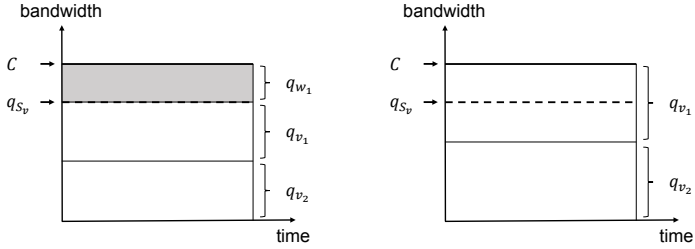
(a) FBV: Video flows share reserved data rate.　　(b) FBV: Video flows share link capacity.

Figure 3.9: Illustration of FBV behavior.

cated its average video bit rate plus its standard deviation ($q'_{v_i} = \overline{b_{v_i}} + \sigma_{b_i}$) to account for bit rate variations. Note that any flow characteristic, which is available to the traffic management decision entity can be used. To prevent starvation, a minimum bandwidth for web browsing q_{S_w} is required, which can be adjusted according to the traffic mix. If the video flows request more bandwidth than available, the allocated bandwidth has to be scaled.

Algorithm　The active flows are separated by type into the set of video streaming flows S_v and the set of web browsing flows S_w. Each active video streaming flow requests a bandwidth q'_{v_i} according to its characteristics. If all requests can be fulfilled, i.e., $q_r = C - q_{S_w} - \sum_{a_j \in S_v} q'_{v_j} \geq 0$, each video flow is allocated its requested bandwidth $q_{v_i} = q'_{v_i}$ and the remaining capacity q_r is added to q_{S_w}. Thus, each web browsing flows receives $q_w = \frac{q_{S_w} + q_r}{|S_w|}$. If more bandwidth is requested than available, i.e., $q_r < 0$, the requests of the video flows have to be scaled. Thus, each video flow receives a bandwidth of $q_{v_i} = (C - q_{S_w}) \cdot \frac{q'_{v_i}}{\sum_{a_j \in S_v} q'_{v_j}}$, and each web browsing flow is allocated $q_w = \frac{q_{S_w}}{|S_w|}$.

Figure 3.10a shows the situation in which the requested video rates $q'_{v_1} + q'_{v_2} \leq C - q_{S_w}$. The video flows are both allocated their requested bandwidth $q_{v_1} = q'_{v_1}$, $q_{v_2} = q'_{v_1}$, and the web flow receives the remaining capacity $q_{w_1} = q_{S_w} + q_r$. In Figure 3.10b, the requests exceed the available bandwidth, i.e., $q'_{v_1} + q'_{v_2} > C - q_{S_w}$. Thus, the requests are scaled according to their requested rates.
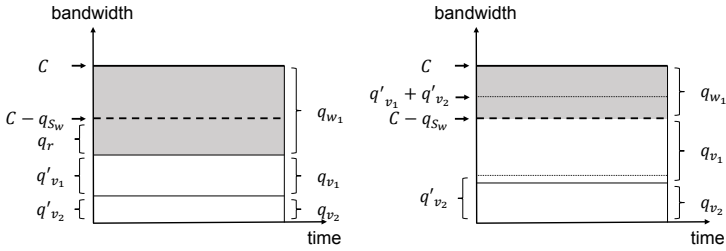
(a) FBF: Requested video rate does not exceed available bandwidth for video flows. (b) FBF: Requested video rate exceeds available bandwidth for video flows.

Figure 3.10: Illustration of FBF behavior.

c) Weighted Bandwidth Allocation for All Videos Flows (WBV)

The Weighted Bandwidth Allocation for All Videos Flows (WBV) strategy utilizes a fixed service ratio between the two application based on the idea that video streaming requires a multiple of the bandwidth, which is needed for web browsing. Thus, it uses a weight parameter w_V, and allocates each video flow a w_V times larger bandwidth than a web browsing flow. Within each application class, the allocated bandwidths for each flow are equal. The advantage is that the allocation scales with the number of active flows. Moreover, no fixed separation between the resources of each application is needed.

Algorithm The active flows are separated by type into the set of video streaming flows S_v and the set of web browsing flows S_w. Each video flow is assigned

$q_v = C \cdot \frac{w_V}{w_V \cdot |S_v| + |S_w|}$, which is w_V times the bandwidth allocated to web browsing flows $q_w = C \cdot \frac{1}{w_V \cdot |S_v| + |S_w|}$.

The behavior of the WBV algorithm with weight $w_V = 2$ is illustrated in Figure 3.11. In Figure 3.11a, one video flow v_1 and one web flow w_1 are on the link, and the ratio between $q_{v_1} = \frac{2}{3}C$ and $q_{w_1} = \frac{1}{3}C$ is w_V. In Figure 3.11b, an additional video streaming flow is added to the link. Now each video flow receives $\frac{2}{5}C$ and the web flow gets $\frac{1}{5}C$, such that the ratio between the bandwidths of a video and a web flow remains w_V.
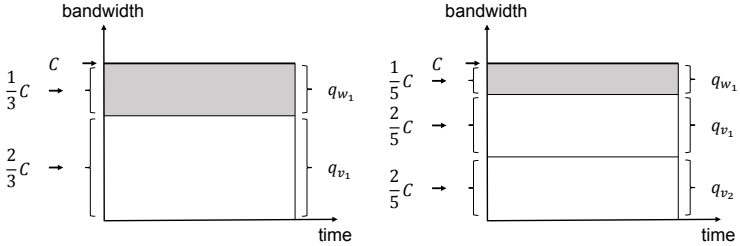
(a) WBV: Link share in case of two flows (one web flow, one video flow).

(b) WBV: Link share in case of three flows (one web flow, two video flows).

Figure 3.11: Illustration of WBV behavior ($w_V = 2$).

d) Dynamic Bandwidth Allocation for Each Video Flow (DBF)

The Dynamic Bandwidth Allocation for Each Video Flow (DBF) strategy is based on current application–layer information from video streaming flows, and will allocate bandwidth based on the buffer level. Generally, all flows will share the capacity equally. However, if the buffer level of a video flow is below a threshold t_l, the video flow will be prioritized. Prioritized flows can utilize a reserved bandwidth fraction q_p to fill the buffer until the buffer level exceeds a threshold t_h. Thereby, the traffic management can leverage the short term flexibility

introduced by the video buffer and shift bandwidth to struggling flows in order to avoid imminent stalling, which is the worst degradation of QoE. Note that other application-layer information could be used as well to reach different traffic management goals. Technically, this strategy requires that application-layer information is monitored periodically, such that the traffic management decision entity can react on a short timescale.

Algorithm   In the beginning, the active flows are added to a set S_u, and the set of prioritized flows S_p is empty. If the playout buffer of a video stream decreases below the lower threshold t_l, the flow is moved from S_u to S_p. If the buffer of a video flow in S_p increases above t_h, it is put back to S_u. All flows in S_p share q_p equally, i.e., $q_{v_p} = \frac{q_p}{|S_p|}$. The web browsing flows and non-prioritized video streaming flows share the remaining bandwidth equally, i.e., $q_{w/v_u} = \frac{C-q_p}{|S_u|}$, or $q_{w/v_u} = \frac{C}{|S_u|}$, if S_p is empty.

Figure 3.12a shows one web browsing flow and two video streaming flows, which both have a buffer level higher than t_l. Thus, all flows share the capacity C equally. In Figure 3.12b, the buffer of v_2 has fallen below the threshold t_l. Thus, v_2 is prioritized and receives $q_{v_2} = q_p$, while the other flows w_1 and v_1 share the remaining bandwidth $C - q_p$ equally.

3.3.3 Evaluation Methodologies and Scenario

This section outlines the analytical approach that is used for analyzing the download times of video streaming and web browsing flows without traffic management, as well as for the FBV and WBV strategies. As the applicability of the analytical models is limited, a simulation is needed, which resumes the performance evaluation for the download times of the more complex strategies FBF and DBF, and also provides stalling results. The simulation is described in detail, and the evaluation scenario is presented in the end of this section.
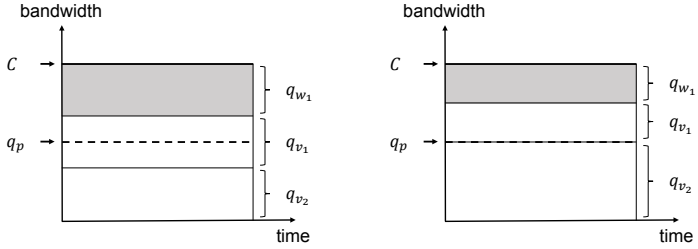
(a) DBF: All video flows have a buffer level higher (b) DBF: One video flow has a low buffer level and is than t_l.   *prioritized.*

Figure 3.12: Illustration of DBF behavior.

Analytical Approach

The shared bottleneck link is modeled as a birth-death process with a single server. Flows arrive individually and independently at the system, are served with a certain downlink bandwidth, and leave the system when the download is completed. For mathematical tractability, the arrivals are exponentially distributed with rate λ and the job size (i.e., video or website size) is exponentially distributed with mean s, which gives the service rate $\mu = \frac{C}{s}$. Thus, a system, which shares the capacity equally among all active flows, i.e., the normal TCP behavior, is described by an M/M/1 Queue with Processor Sharing Policy (M/M/1-PS).

Considering the two application types video streaming and web browsing, different arrival rates λ_v, λ_w and mean job sizes s_v, s_w have to be distinguished. The needed two class M/M/1-PS model can be derived from the M/M/1 Queue with Discriminatory Processor Sharing Policy (M/M/1-DPS) model described in [254]. According to [255], the conditional average response times $V(\tau)$ can be obtained for each class depending on the job size τ. Thus, the average download time can be computed for any given website size or video size. If additionally the

video length (depending on video size τ and video bit rate) is known, the average stalling time should be definable. However, not single flows, but the situation in the whole system is evaluated. Thus, the unconditional average response times V are considered:

$$
\begin{aligned}
V_v &= \frac{1}{\mu_v(1 - \rho)}, \\
V_w &= \frac{1}{\mu_w(1 - \rho)},
\end{aligned}
\tag{3.1}
$$

where $\rho = \rho_v + \rho_w = \frac{\lambda_v}{\mu_v} + \frac{\lambda_w}{\mu_w}$. Note that average stalling times cannot be computed from V_v, as video bit rate distributions have to be considered. Thus, the simulation approach described below will complement the analytical results and provide stalling times.

In the general M/M/1-DPS model, each class is assigned a weight to control the sharing of the processor capacity, which can be directly applied to model the WBV strategy. Thus, the generalized results from [255] provide results for the average download times of WBV when videos are assigned a w_V times larger bandwidth than web pages:

$$
\begin{aligned}
V_v &= \frac{1}{\mu_v(1 - \rho)}\left(1 + \frac{\mu_v \rho_w(1 - w_V)}{w_V \mu_v(1 - \rho_v) + \mu_w(1 - \rho_w)}\right), \\
V_w &= \frac{1}{\mu_w(1 - \rho)}\left(1 + \frac{\mu_w \rho_v(w_V - 1)}{w_V \mu_v(1 - \rho_v) + \mu_w(1 - \rho_w)}\right).
\end{aligned}
\tag{3.2}
$$

The M/M/1 Queue with Generalized Processor Sharing Policy (M/M/1-GPS) is a further generalization, in which each request or class of requests has an arbitrary service rate. Thus, to model FBV, an M/M/1-GPS with one class for video and one class for web pages has to be considered. The stationary joint distributions can be obtained from the work of [256], but no closed form of the conditional response time is given. Still, by solving the steady state equations of the FBV model, the mean number of jobs in the system can be computed from

the state probabilities. Little's Theorem is applied to obtain the mean delay in the system, which is equal to the mean service time in a processor sharing system.

In a similar way, FBF can be modeled by an M/M/1-GPS with an own class for each video streaming flow and one class for all web browsing flows. However, the video flows would have to be limited to a tractable number of classes, and the resulting model would be overly enlarged and complicated, such that no analytical performance evaluation was conducted. Also for DBF, which is a dynamic strategy prioritizing video flows depending on current buffer levels, no analytical model can be presented. Thus, the analysis of FBF and DBF strategies is solely based on the simulation, which is described in the following.

Simulation

The system is simulated with a Java Discrete Event Simulation (DES) to verify the analytical results for the average download times, to evaluate the strategies for which no analytical model is present, and to obtain detailed results on stalling. The simulation is based on the analytical model, and thus, the arrivals of both video streaming flows and web browsing flows are exponentially distributed with rates λ_v and λ_w, respectively. Also the size of the videos s_v and web pages s_w are exponentially distributed. The flows share a single link with a fixed capacity C according to TCP-like best effort sharing (no traffic management), or according to one of the four implemented traffic management strategies (FBV, FBF, WBV, DBF). The download times of each flow can be obtained. For web browsing flows, the download time relates to the page load time, which is a major QoE factor.

The simulation of video streaming considers VBR videos with an exponentially distributed average bit rate $\overline{b_v}$. The video is split in segments of 2 s playtime. The bit rate of each segment varies, but a constant bit rate is assumed within a segment. The segment bit rates change according to a simple autoregressive process, which is able to accurately model videos without scene changes [257]. In practice, if the video content consists of multiple scenes, higher

variations may occur, such that the presented results might be negatively affected.

The playout buffer of each video streaming session is accurately simulated and a simple buffering mechanism is implemented. After the first segment is completely downloaded, the playback starts. Thereby, the initial delay is minimal and does not significantly deteriorate the QoE [118]. When the video buffer runs empty during the playback, the stalling begins, and the playback resumes when the buffer is filled with 5 s of playtime. Thus, the stalling time of each video session and the buffering ratio, i.e., the ratio of stalling time and playback time, can be obtained, which indicates the QoE of the video streaming sessions.

Evaluation Scenario

The evaluation scenario is based on the findings of [125] for mobile YouTube video streaming. The default quality on smartphones is 360p with an average bit rate of $\overline{b_v} = 0.5\,\text{Mbps}$. 50 randomly chosen videos were analyzed to obtain the average video length, which resulted in $\overline{t_v} = 110\,\text{s}$. Thus, the average video size $s_v = \overline{b_v} \cdot \overline{t_v} = 6.875\,\text{MB}$. The download of 50 randomly chosen websites resulted in a mean size $s_w = 1.3\,\text{MB}$. To align the simulation with the analytical model, the size of each video or web flow is exponentially distributed with mean s_v or s_w, respectively. The videos are split into segments of 2 s playtime. For each segment i a random size $s_v^r(i)$ is computed, which is exponentially distributed with mean $\overline{b_v} \cdot 2\,\text{s} = 0.125\,\text{MB}$. The actual segment size $s_v(i)$ is determined according to the simple auto-regressive process $s_v(i) = 0.7 \cdot s_v(i-1) + 0.3 \cdot s_v^r(i)$. Note that the last segment has to be cropped to reach the desired size of the whole video. The capacity C of the link was set to 5 Mbps, 10 Mbps, 20 Mbps, and 100 Mbps, which defines $\mu_v = \frac{C}{s_v}$ and $\mu_w = \frac{C}{s_w}$ in the corresponding analytical models. The arrivals of video flows λ_v and web flows λ_w were varied to reach different loads $\rho = \rho_v + \rho_w = \frac{\lambda_v}{\mu_v} + \frac{\lambda_w}{\mu_w}$ and traffic mixes $\frac{\lambda_v}{\lambda_w}$. Per load, traffic mix, and resource management strategy, ten simulation runs were conducted with a simulation time of $100000\,\text{s} = 27.78\,\text{h}$ each.

3.3.4 Performance Evaluation Results

In this section, the performance evaluation of the traffic management strategies is presented. First, the situation in current best effort networks is described to investigate for which level of network usage resource allocation might improve the QoE of video streaming and web browsing. Based on these findings, the four traffic management strategies are evaluated analytically and simulatively for different application traffic mixes. Finally, the results for all algorithms are summarized, and trade-offs between the QoE of video streaming and the QoE of web browsing are discussed.

Situation in Current Networks Without Resource Management

To understand why QoE-aware traffic management is needed, the performance of flows in best effort networks without resource management is investigated. In best effort networks with TCP, all flows share the link capacity equally, such that every active flow receives the same bandwidth. Figure 3.13 shows the average download time of 1 MB depending on the link capacity and the load computed both with the analytical model (solid, M/M/1-PS model in Equation 3.1) and the simulation (dashed, average over ten runs including 95% confidence intervals). The normalized average download time only depends on the average download bandwidth, which is investigated below. Note that the analytical model is within the 95% confidence intervals of the simulation, which verifies the simulation implementation. For a given network load, it can be seen that a larger link capacity results in lower average download times. In contrast, the download time grows hyperbolically when the network load increases towards 1. This results in unacceptable download times, and consequently, for example, a bad web browsing QoE, which most network operators typically avoid by over-provisioning the network resources.

As different applications have different network requirements, the change of network QoS parameters, such as download bandwidth, impacts the QoE of those applications differently. Figure 3.14a investigates the average download
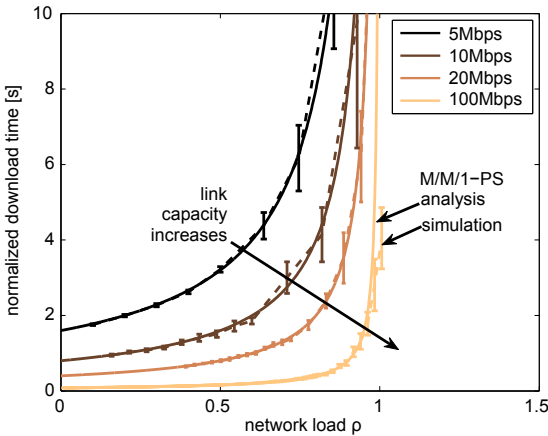
Figure 3.13: Normalized download time of 1 MB in best effort networks.

bandwidth for different capacities and loads. The average download bandwidth decreases linearly when the network load increases. To guarantee a smooth streaming without stalling in case of CBR videos, the average download bandwidth has to be at least as high as the average bit rate. Considering video streaming with an average bit rate $\overline{b_v} = 0.5\,\mathrm{Mbps}$, the figure shows that a 5 Mbps link can support such download bandwidths up to a load of 0.9, while higher capacities can support even higher loads. However, currently deployed streaming technologies use VBR videos and employ buffering strategies when downloading video content, which results in a different performance. For example, the usage of VBR videos leads to fluctuating network requirements, but the use of the playout buffer can overcome certain load peaks. Figure 3.14b depicts the average buffering ratio of the simulated VBR videos depending on the link capacity and network load. Due to the bit rate fluctuations, stalling of the videos can already occur with loads of 0.7 and above. This behavior could not be de-

rived from pure QoS parameters and shows the necessity of QoE-aware traffic management especially in high load scenarios.

In the following, if not stated otherwise, a link capacity of 10 Mbps and a network load, which is equivalent to 0.9 in the best effort network, will be considered. As the results of the investigated QoE-aware traffic management strategies also depend on the popularity of the applications in the network, different traffic mixes ($\frac{\lambda_v}{\lambda_w}$) will be investigated. For example, the traffic mix $\frac{\lambda_v}{\lambda_w} = 10$ means that on average ten times more video streaming flows arrive on the shared link than web browsing flows.

Results of Fixed Bandwidth Allocation for All Videos Flows

The FBV strategy reserves a fixed share of the link capacity to video streaming flows. However, if too much bandwidth is reserved for video flows, the page load time of web browsing might increase. On the other hand, if the allocated bandwidth is too small, stalling might increase. Therefore, a range of allocations from $q_{S_v} = 0.3C$ up to $q_{S_v} = 0.9C$ is investigated. Figure 3.15a shows the average download times of web pages depending on the reserved bandwidth q_{S_v} and the traffic mix $\frac{\lambda_v}{\lambda_w}$, which is depicted on the x-axis. Again both analytical results (solid, M/M/1-GPS model) and simulation results (dashed, averages over ten runs including 95% confidence intervals) are shown. The average download time of a web page in a best effort network is 9 s (highlighted in red, cf. Figure 3.13), which does not depend on the traffic mix because all application flows are treated equally in a best effort network. On the left side of the vertical, dashed line, the traffic mix is less than 1, i.e., more web browsing flows arrive on the shared link. If the video reservation q_{S_v} increases (lighter colors), the download times increase because less resources are left for the web flows. When $\frac{\lambda_v}{\lambda_w}$ increases towards $10^0 = 1$ and beyond, less web flows are on the link. The bandwidth reservations become better aligned to the traffic mix, which mitigates this effect. Towards the rightmost part of the plot, very high traffic mixes indicate that only few web flows are on the link. These flows can use the remaining, not
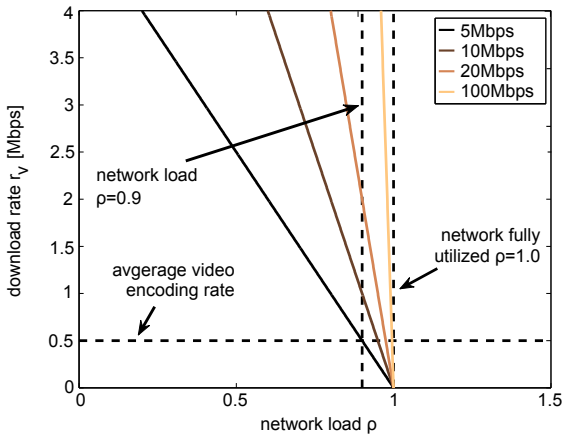
(a) Network load vs. download rate video.
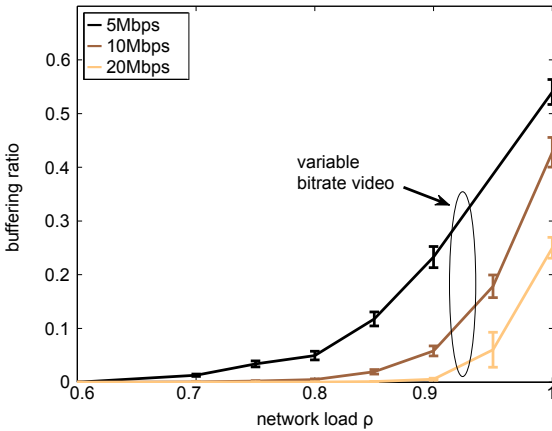
(b) Stalling ratio.

Figure 3.14: Video application performance in best effort networks.

reserved bandwidth $C-q_{S_v}$ almost exclusively, which results in short download times.

The corresponding results for the video streaming flows are depicted in Figure 3.15b, which shows the average buffering ratio depending on reservation q_{S_v} and traffic mix $\frac{\lambda_v}{\lambda_w}$. For high traffic mixes, the buffering ratio converges towards the best effort performance (red, cf. Figure 3.14b) for any reservation because web flows are rarely in the system and for short times only. Thus, the video flows compete among themselves most of the time, which results in a best effort behavior. In the left part of the figure, video flows are less frequently in the system but they can clearly benefit from the bandwidth reservation. Thereby, a higher q_{S_v} leads to smaller buffering ratios. Only if $q_{S_v} = 0.3C$, the allocated bandwidth is too small, which leads to higher buffering ratios than in the best effort case. Consequently, the reservation q_{S_v} has to be carefully selected to avoid negative effects for the QoE. The trade-offs for the optimization of both applications will be discussed below.

Results of Fixed Bandwidth Allocation for Each Video Flow

In contrast to FBV, the FBF strategy allocates bandwidth to individual video flows based on flow characteristics. Two allocations are investigated, namely, the allocation of the average bit rate, and the allocation of average bit rate plus standard deviation to account for bit rate fluctuations. Figure 3.16a depicts the average download times of web browsing flows for both allocations. The results represent the averages of ten simulation runs and the corresponding 95% confidence intervals. To compare the performance of FBF, selected FBV results ($q_{S_v} = [0.3\,0.6\,0.9]C$) are added in gray. It can be seen that the download times increase when the ratio of video flows increases. This is due to the fact that an increasing number of video flows increases the allocated bandwidth, which becomes unavailable for web browsing.

The buffering ratio of the video flows in the same scenario is depicted in Figure 3.16b. It can be seen that the allocation of the average bit rate performs badly because it cannot avoid the stalling of the simulated VBR videos. In contrast, the
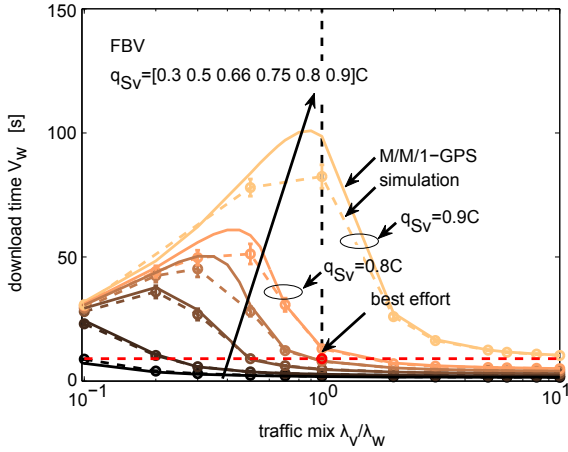
(a) Download time of web pages.
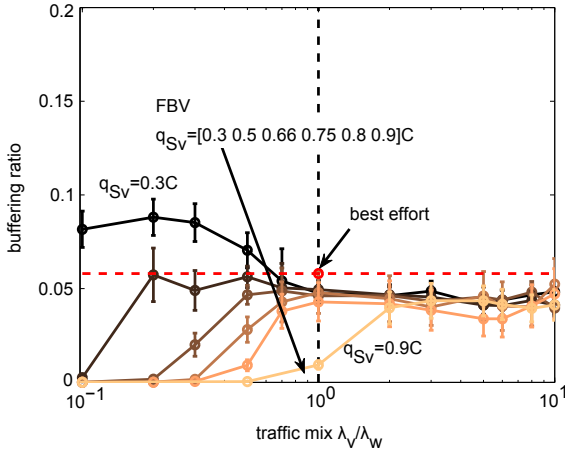
(b) Stalling ratio of video flows.

Figure 3.15: *FBV: Fixed bandwidth reservation q_{S_v} for all video traffic at a load $\rho = 0.9$.*

allocation of average bit rate plus standard deviation is sufficient to account for the bit rate fluctuation of the VBR content, and almost completely avoids stalling for traffic mixes below 1. For increasing traffic mixes, the requested allocations exceed the available capacity, such that the allocations have to be scaled down. This results in increasing buffering ratios and a worse video streaming QoE. Still, FBF performs better than the best effort case or the FBV strategy, which also shares the capacity equally if too many video flows are on the link, because the proportional sharing of FBF is better aligned to the demands of the video streams.

Results of Weighted Bandwidth Allocation for All Videos Flows

With the WBV strategy, each video streaming flow is allocated w_V times the bandwidth of a web browsing flow. Figure 3.17a shows the average download times of web pages obtained with from the analytical model (solid, M/M/1-DPS model in Equation 3.2) and the simulation (dashed, averages over ten runs and 95% confidence intervals). Different weights $w_V = [0.02\ 0.5\ 2\ 5\ 50]$ are considered, including weights that favor web browsing, i.e., $w_V < 1$. The results of the analytical model and the simulation are well aligned, especially for smaller w_V. For larger w_V the analytical model overestimates the page download time. The larger w_V the larger the download time because web browsing flows get a smaller share of the capacity. For large weights $w_V \gg 1$ and large traffic mixes, the bandwidth allocated to web flows becomes arbitrarily small, such that the average download load time quickly increases.

Figure 3.17b shows the average buffering ratio in the same scenario. The lines represent smooth exponential fittings of the simulation results, which are depicted by the markers. If the traffic mix is below 1, the video streaming flows benefit from WBV with $w_V > 1$. If the ratio of video streaming flows increases, this effect diminishes and the buffering ratio converges towards the best effort case (red), i.e., equal sharing. The results show that WBV with an appropriate weight w_V is able to align the bandwidth allocation to the flow sizes for a given traffic mix. The bandwidth allocation scales with the number of flows and signif-
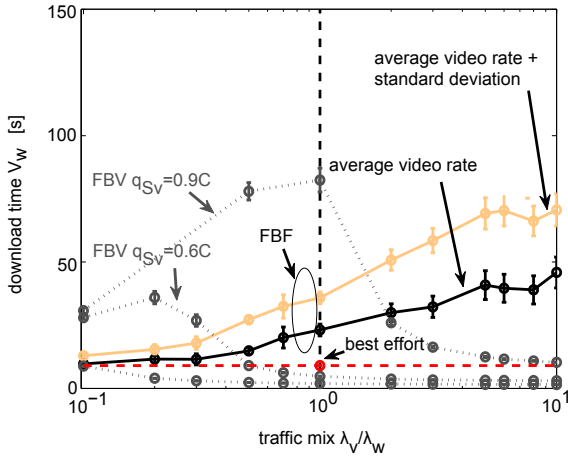
(a) Download time of web pages.
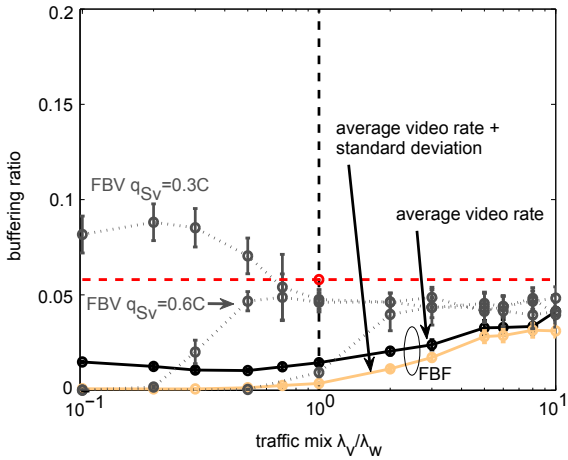
(b) Stalling ratio of video flows.

Figure 3.16: FBF: Fixed bandwidth reservation for each flow with either average bit rate or average bit rate plus standard deviation at link load $\rho = 0.9$.

icantly decreases the buffering ratio, while only slightly increasing the average download time of web pages. Further trade-offs between the two applications are discussed below.

Results of Dynamic Bandwidth Allocation for Each Video Flow

The results above showed the performance of traffic management strategies, in which the allocation of each flow was predefined at the flow start and only changed depending on the number of currently active flows. DBF, however, is a dynamic strategy, which utilizes monitored application-layer QoE information to update the bandwidth allocation. The buffer of each video flow is monitored, and if the buffer drops below a low threshold t_l, the flow is prioritized to avoid imminent stalling and keep a high QoE. When the buffer level rises above a threshold t_h, the flow is not prioritized anymore but is treated normally again. For the performance evaluation presented below, the prioritization bandwidth, which is allocated exclusively to prioritized video flows if there are any, is $q_p = 0.9C$, and the buffer thresholds are $t_l = 10\,\text{s}$ and $t_h = 20\,\text{s}$. In Figure 3.18a the average download times of web pages are compared for the DBF (dashed) and the WBV strategy, which showed a decent performance for both applications. The DBF strategy results in the highest download times for a balanced traffic mix $\frac{\lambda_v}{\lambda_w} = 1$. The more imbalanced the traffic mix to either more videos streaming flows or more web browsing flows, the shorter the average download times, which come close to the best effort situation.

Figure 3.18b presents the results for video flows, which face a significant performance improvement by DBF. For traffic mixes $\frac{\lambda_v}{\lambda_w} < 1$, stalling can be avoided almost completely in the investigated scenario. When the ratio of video flows increases, the average buffering ratio also increases and converges to the best effort situation. Nevertheless, DBF is able to support video flows even better than WBV with a high $w_V > 5$, but does not increase the page load times too much. Thus, it shows a balanced and fair behavior, which is a desired characteristic of a QoE-aware traffic management strategy. Note that the impact of the parameters q_p, t_l, and t_h was not investigated as the chosen parameters already
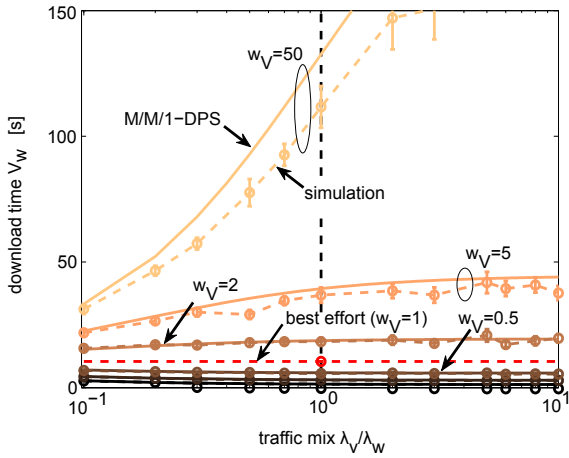
(a) Download time of web pages.
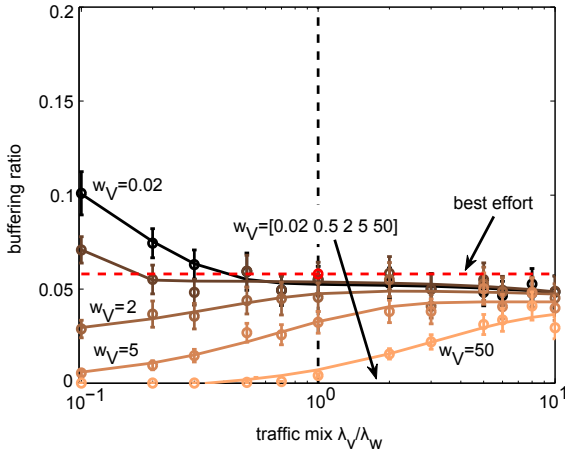
(b) Stalling ratio of video flows.

Figure 3.17: WBV: Weighted bandwidth reservation for all video traffic with weight w_V at link load $\rho = 0.9$.

proved a decent performance for avoiding stalling of video streaming sessions, and demonstrated the advantages of using application-layer QoE parameters for traffic management decisions.

Comparison of the Different Strategies

Finally, the different traffic management strategies are compared, and trade-offs between the QoE of the two applications are discussed. As the two QoE performance indicators, i.e., average download times of web pages and average buffering ratio of video streaming sessions, are conflicting, a Pareto analysis is presented in Figure 3.19 with the two objectives being the axes of the plot. The best effort performance (average web page download time of 9.04 s, average buffering ratio 0.058) serves as reference and is marked by a red cross. The Pareto-optimal performances of the traffic management strategies for a given traffic mix are shown in the plot. Five traffic mixes $\frac{\lambda_v}{\lambda_w} = [0.1\ 0.5\ 1\ 2\ 5]$ are considered, and the respective Pareto-optimal performances for each traffic mix are connected by colored lines. Note that the colored lines do not indicate the location of other Pareto-optimal performances (i.e., the Pareto frontier), but were added only for visualization purposes. Only the marked points are Pareto-optimal, and their shape represents the strategy, which achieved the respective performance.

It can be seen that Pareto-optimal performances, i.e., performances, for which the QoE of one application cannot be improved without deteriorating the QoE of the other application, are either left or below of the best effort performance. This means that, for any traffic mix, the usage of traffic management can improve at least the QoE of one application. The plot indicates that WBV algorithms are especially well suited to achieve short web page download times, while FBF and DBF strategies can be used to minimize the buffering ratio. The desired region for QoE-aware traffic management is the gray box in the bottom left part of the plot, which indicates a QoE improvement for both applications compared to the situation without traffic management (best effort). For almost all traffic mixes except $\frac{\lambda_v}{\lambda_w} = 0.1$, i.e., ten times more web flows than video flows, such performances are possible. They are achieved by WBV and FBV strategies if they
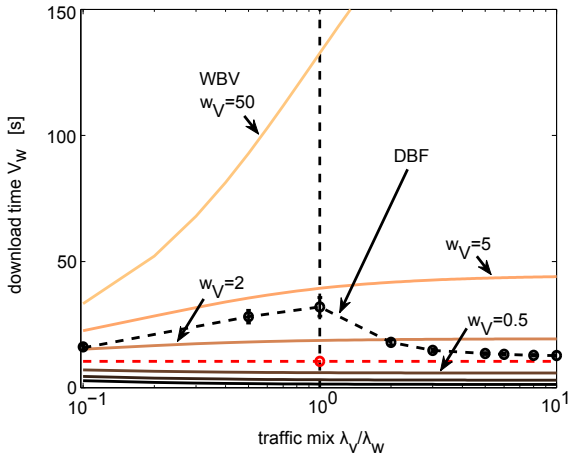
(a) Download time of web pages.
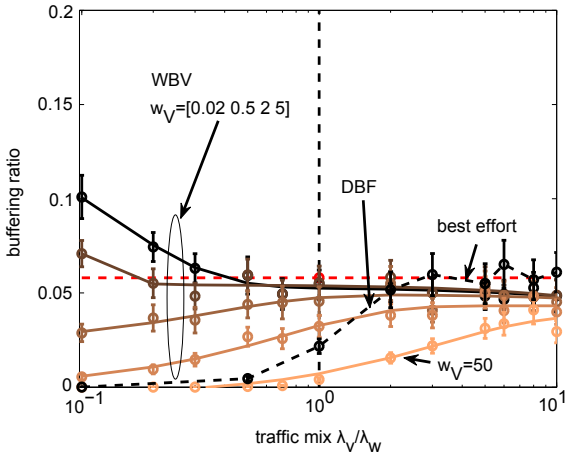
(b) Stalling ratio of video flows.

Figure 3.18: DBF: Dynamic bandwidth allocation for each video flow based on buffer fill at link load $\rho = 0.9$.

use the right parameter setting for the given traffic mix. Note that the specific parameters of the Pareto-optimal performances are not shown here but can be obtained from the individual result plots of each strategy above.
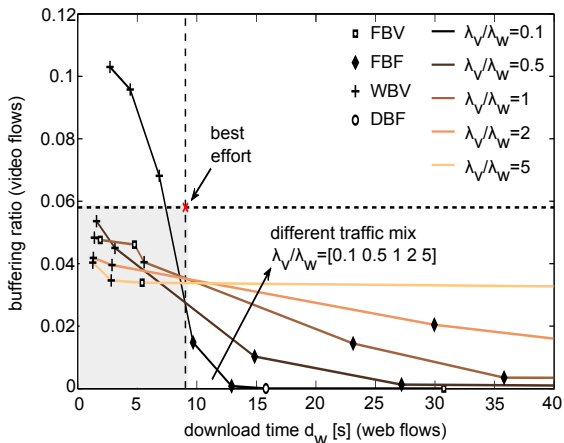
Figure 3.19: Overview of Pareto-optimal performances for different traffic mixes.

The performance evaluation of different QoE-aware traffic management strategies shows that it is possible to improve the QoE of applications, which compete on a shared bottleneck link, by changing the bandwidth allocation of the active flows. However, depending on the traffic mix and the preferences of the network operator, the right strategy and the right parameter settings have to be chosen. For managing the QoE of video streaming, the direct usage of QoE feedback in terms of application-layer QoE factors is feasible and can significantly reduce stalling, which is the most important QoE degradation. Although also the pure network-based strategies can achieve a decent performance, the advantage of using QoE feedback is that a dynamic reaction to imminent QoE degradations is possible.

3.4 Lessons Learned

This chapter investigated QoE-aware traffic management for HAS in access networks. The requirements are QoE monitoring, traffic management decisions, and traffic management actions. Two approaches were presented to monitor QoE within the network and within the application. Moreover, a direct estimation of QoE from monitored network parameters was studied. Eventually, bandwidth allocation strategies for video flows on a shared bottleneck link, such as a residential broadband connection in a typical home environment or a backhaul link of a mobile base station, were evaluated and the benefits of different levels of monitoring information were discussed.

To monitor the QoE of HAS, two approaches were presented. The first approach complemented the work in [16] and was an example of network-based monitoring. By DPI, objective video quality metrics were extracted from the network traffic. These included the requested video quality level of the HAS stream and the video bit rate, which are objective metrics of each segment, and the SSIM metric on a per-frame base. As temporal pooling is well aligned with the QoE findings of Chapter 2, it was applied to combine these objective metrics into a single measure for the whole streaming session. The decent performance of the QoE estimators confirms that temporal pooling is a suitable approach to combine periodical objective quality metrics when monitoring HAS. Thus, its application is also promising for the increasingly used end-to-end encrypted HAS, in which objective metrics cannot be extracted by DPI, but have to be estimated from the network traffic.

The second approach monitored the QoE unobtrusively within the video streaming application, and could thus provide accurate information on the QoE factors just as they are perceived by the end user. Therefore, JavaScript-based monitoring functions were injected into the streaming website, which monitored the state of the video player, the buffer, and the video quality level. The information about initial delay, stalling, and adaptation, which are the most important QoE factors of HAS, could be computed and might be signaled to the

network operator as meaningful QoE feedback for QoE-aware traffic management. The concept was implemented in an Android app, which could be used for researching QoE models for HAS. Moreover, it was used in a field study on QoE estimation from monitored network parameters. The results showed that application-layer information gave better insights into the QoE of HAS than simple network parameters.

Finally, four QoE-aware traffic management strategies based on bandwidth allocation were presented and their performance was evaluated in a shared bottleneck link scenario. Two applications, video streaming and web browsing, were considered and different ratios between the application flows were evaluated. The strategies had different complexity, dynamics, and used different levels of information ranging from pure flow type information to real-time application-layer information about the video buffer. Analytical models based on processor sharing queues were used to obtain average download times for some of the strategies. Additionally, all algorithms were assessed with a Java DES. Thereby, also the stalling of the video streaming sessions could be evaluated. The results showed that a joint QoE improvement was possible for both applications compared to the best effort scenario without traffic management. However, operators have to select the right strategy and the right parameter settings depending on the traffic mix in their networks and their objectives. Moreover, dynamic traffic management based on application-layer QoE factors proved to be feasible to react to imminent QoE degradations. Thereby, stalling could be significantly reduced, which is the most important QoE degradation of HAS.

4 Socially-aware Traffic Management for HTTP Adaptive Video Streaming based on Wi-Fi Offloading

Users of Internet applications provide much information not only about themselves but also about their interests, their activities, or their relations. Such social signals might be created unwittingly or voluntarily, but they can be collected, for example, from sensors (e.g., location, co-location with others, mobility), applications (e.g., usage behavior, usage preferences, co-usage with friends), or Online Social Networks (OSNs) (e.g., friendships, interests, opinions). Social awareness harvests these ubiquitous signals to extract and exploit useful information in order to improve a service. Social awareness is also beneficial for network operators and the traffic management of Internet services, such as HAS. Thereby, Socially-aware Traffic Management (SATM) mechanisms enhance the performance of services by utilizing social information in order to perform efficient network management, service placement, and traffic optimization.

The research field of SATM introduces promising approaches for enhanced service delivery on the Internet but also new interdisciplinary challenges. This is especially due to the fact that multiple stakeholders are involved. Therefore, incentive-compatible network management mechanisms are needed, which satisfy the goals of all stakeholders, e.g., in terms of high QoE or efficient resource

utilization. The realized SATM solutions have to be highly collaborative and utilize traffic management information and actions across different layers.

In this chapter, SATM is applied to improve the QoE of HAS. From the vast possibilities for SATM, a single traffic management action is considered, namely, Wi-Fi offloading. Thereby, mobile connections are offloaded to fixed networks using Wi-Fi, which comprises technology for Wireless Local Area Networks (WLANs) based on the Institute of Electrical and Electronics Engineers (IEEE) 802.11 standards[1][2]. In the last years, Wi-Fi offloading received a lot of attention from industry to handle the mobile data growth and reduce the load on stressed mobile networks [258]. Additionally, there is not only a widely distributed public Wi-Fi infrastructure, but Wi-Fi is also prevalent in residential environments of end users [259]. Thus, Wi-Fi offloading has a huge potential to be utilized for network-based SATM. A simple hotspot location model is presented, which can be used for the performance evaluation of public Wi-Fi offloading systems. Moreover, a home router platform is proposed to enable private Wi-Fi offloading and the socially-aware distribution of video content. A simulative performance evaluation study quantifies the impact of public and private Wi-Fi offloading on the QoE and energy consumption of HAS sessions, and confirms the applicability of the simple hotspot location model.

Section 4.1 summarizes the background on SATM based on [1], and outlines background and related works on Wi-Fi offloading. Section 4.2 presents a simple model to describe the locations of public Wi-Fi infrastructure in cities based on [10]. Moreover, it describes a collaborative SATM platform for HAS on residential Wi-Fi routers based on [42, 56], which supports private Wi-Fi offloading, as well as the distribution and caching of niche content. A performance evaluation of the gain of Wi-Fi offloading for HAS is conducted in Section 4.3 based on [29, 44, 47], and the applicability of the simple hotspot location model is investigated. Finally, Section 4.4 integrates the findings of this chapter.

---

[1]`http://standards.ieee.org/about/get/802/802.11.html` [Online] – Accessed: 24.08.2017

[2]`http://grouper.ieee.org/groups/802/11/` [Online] – Accessed: 24.08.2017

4.1 Background and Related Work

This section presents the background on SATM in general, as well as on its particular application to Wi-Fi offloading of HAS. The first part of this section follows [1] and defines SATM and the used terminology. Moreover, the involved stakeholders are introduced, and their interests and potential benefits are described. Background and related works on Wi-Fi offloading are presented in Section 4.1.2.

4.1.1 Socially-aware Traffic Management

The concept of SATM requires the utilization of social information and the collaboration of stakeholders. In the following, a definition of SATM is presented and the needed terms are introduced. These definitions were elaborated within the Framework Programme 7 (FP7) project Socially-aware Management of New Overlay Application Traffic combined with Energy Efficiency in the Internet (EU FP7 SmartenIT) funded by the European Union (EU).

Terminology and Definition

Any signal, which is emitted by a person, is a social signal. However, in the context of Internet services, a **social signal** is considered to be a signal, which is emitted on the Internet by a user of an Internet application. This also includes any interaction of an end user with an Internet service. Social signals do not contain any information, but they can be evaluated in particular contexts to create information out of them. In the context of OSNs, these signals are, for example, logins, postings about external events, indications of interest or liking, or friendship requests and confirmations. Other examples are sensor data of mobile devices, which are transmitted when using Internet services, e.g., GPS locations.

Social information comprises information about one or more persons, or their relationships. When social signals are brought into an appropriate context, social information can be deduced, which brings new insights about users or relationships. For example, evaluating the social signals of a user's friendship request in an OSN and another user's instant confirmation of that request, generates the social information that both users are friends. In a different context, it generates the social information that both users used the OSN service during the time of the signal emissions. As another example, analyzing the location signals of a user might create the social information that the user is at the same site every second Saturday. Adding the external information that there is a football stadium, will create the social information that the user is interested in football and a supporter of a certain team. These examples show that the created social information depends on the used social signals, external information, and the context of evaluation.

The term **social awareness** describes the utilization of such social information to improve an Internet service. It can include the collection and evaluation of social signals but also the collaboration with a social information provider. Social awareness exploits the created or provided social information to increase the QoE of end users and/or to provide the service more efficiently. **Socially-aware Traffic Management (SATM)** is an application of social awareness, in which the traffic management on the Internet is improved by social information on different layers. The goal of SATM is to efficiently transport data across the networks, thereby delivering services with a high QoE, avoid congestion, and save energy, resources, and costs. It can be employed by the services themselves, e.g., by scheduling of transmissions or service quality adaptation, or by the network operator, e.g., in terms of routing, bandwidth shaping, prioritization, caching, or offloading. Also collaborative traffic management solutions are possible. In the following, the stakeholders, their goals, and possible benefits of incorporating SATM are described.

Stakeholders, Goals, and Benefits

With SATM, five stakeholders have to be considered. Each of them has his own goals but also individual benefits from participating in SATM. Note that each actor can be an individual stakeholder, but actors can also have multiple roles.

The **cloud service provider** or **application provider** offers an Internet service to end users, which is running on the infrastructure of a cloud operator or on own infrastructure. He is interested in monetizing the service, which includes reducing the costs of consumed ISP and cloud infrastructure. Moreover, he wants to satisfy the end users to maintain or increase the number of customers. Therefore, the QoS and QoE requirements should be fulfilled [160]. By utilizing social information, QoS/QoE parameters may be improved and new services may be developed. If social information is utilized to increase the resource utilization, also infrastructure costs can be reduced.

The **cloud provider** or **data center operator** offers storage and computation resources to the application provider. Moreover, he needs connectivity between his sites and to the Internet, which he buys from an ISP. The cloud provider wants to monetize his infrastructure by guaranteeing satisfactory QoS parameters, and thus, fulfilling SLAs with the application provider. Moreover, he wants to reduce his costs by efficiently utilizing hardware in terms of resources and energy consumption. Similarly, he will seek the best SLA with an ISP for network access. To reach these goals, social information can be utilized, especially for the prediction of service demand and resource utilization.

Internet Service Providers (ISPs) operate communication network infrastructures, which they want to monetize. Therefore, they aim to provide high quality network services to reach a high satisfaction of cloud operators and end users. By employing social information, an ISP can enrich his network traffic management and support new services for application providers, which simultaneously makes him more competitive towards cloud providers and end users. These new services can also reduce the costs by more efficiently using own resources, and keeping traffic on transit links low.

End users are concerned with their QoE [160], their costs for services and network access, and their energy consumption [260]. Note that costs incur not only in terms of money but also by being exposed to advertisements. Still, end users are mainly customers of ISPs and application providers, and thus, not directly involved in the interactions of other stakeholders. End users provide social signals, which are utilized for SATM, and benefit in terms of personalized services and increased QoE.

Finally, **social information providers**, such as OSNs, have to be considered for SATM. They gather social signals and evaluate them to create social information, which they want to monetize. This information can be offered to application providers, cloud providers, or ISPs to support the operation of SATM mechanisms.

Figure 4.1 presents the stakeholders for SATM in a typical HAS scenario. The HAS service is offered by the cloud service provider. He pays the cloud provider to host the streaming service and store the video content. The clouds or data centers (orange) are connected to the networks of ISPs. Upon request, the video content is delivered to the end users and is handled by one or multiple ISPs in different Autonomous Systems (ASs) (blue). After the video consumption, the end user can share the video content with his friends, e.g., in an OSN (red), such that they will also request it. Social information can be collected and provided by any of the stakeholders in the delivery chain, i.e., by cloud service providers, cloud operators, and ISPs. Moreover, they can receive social information directly from end users or can buy social information from dedicated social information providers, which evaluate external sources of social information, such as OSNs.

Examples for Socially-aware Traffic Management of HTTP Adaptive Video Streaming

To illustrate its potential, some possible applications of SATM in the context of HAS are outlined. For example, social information about user location and mobility can be exploited to prepare for proximate handovers on both network and application layer [261]. Also the new trends of edge and fog computing, i.e.,
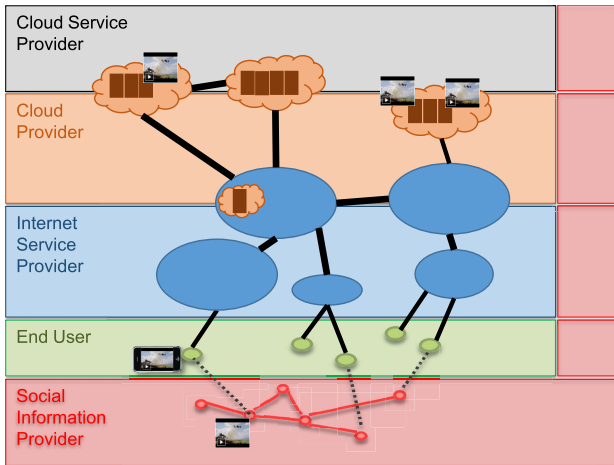
Figure 4.1: Stakeholders of SATM in a typical HAS scenario.

the instantiation of services and placement of data at the network edge close to
end users, benefits from this information. Based on the location of the user, a
different streaming server or content location can be chosen [33], or the routing
can be altered [262, 263]. Additionally, the social cascades of content and the
interests of users can be taken into account for content placement and caching
decisions, i.e., which video file to store on which server or cache [264–269]. This
is especially relevant for user-generated and niche contents, which are not glob-
ally popular and well covered by the widely dispersed CDNs, but still might be
frequently accessed by a single user (group). Such individually popular videos
could be cached, pre-fetched, or distributed on devices close to the user or di-
rectly on user devices [219, 270–278]. Finally, social information on explicit or
implicit trust relationships allows to share resources for trusted friends, which
was already implied in some of the above mentioned content delivery mech-
anisms. Furthermore, HAS sessions can not only benefit from shared comput-

ing or storage resources but also from a shared Internet access of a trusting friend [279–281], e.g., by Wi-Fi offloading.

To sum up, SATM is an opportunity to leverage ubiquitous social signals for improved Internet services. Many actors can be involved in collaborative and cross-layer SATM with various benefits for each of them [1]. Thus, even when considering only HAS, there is a huge potential for applications of SATM, which cannot be covered comprehensively. Therefore, in the remainder of this chapter, the focus is on evaluating the applicability of network-based SATM for improving the QoE of HAS, considering only Wi-Fi offloading as a traffic management action.

4.1.2 Wi-Fi Offloading

Wi-Fi offloading provides a complementary Internet access over a fixed network. It allows ISPs to handle the traffic in well-dimensioned fixed networks, which saves costs and reduces the load on stressed mobile networks. End users, on the other hand, can benefit from higher access bandwidths and avoid exceeding their data plans, especially when consuming demanding applications like mobile HAS. Wi-Fi infrastructure is growing, especially in urban environments, and free public Wi-Fi hotspots (e.g., provided by cafes, shops, libraries) are listed in hotspot databases (e.g., WiGLE[3]). Moreover, there are specialized Wi-Fi-sharing communities (e.g., Fon[4]), commercial infrastructures of big telecommunication operators (e.g., BT[5]), civic projects of municipalities (e.g., New York City[6]), as well as a countless number of residential private Wi-Fi networks. Wi-Fi offloading can leverage social information on location and mobility of users to select appropriate hotspots, as well as trust information to share private Wi-Fi networks. Thus, there is a huge possibility for investigating SATM for Wi-Fi offloading.

---

[3] `https://wigle.net/` [Online] – Accessed: 24.08.2017
[4] `https://fon.com` [Online] – Accessed: 24.08.2017
[5] `http://www.btwifi.co.uk/` [Online] – Accessed: 24.08.2017
[6] `https://link.nyc/` [Online] – Accessed: 24.08.2017

In [282], currently available features for offloading mobile traffic were presented, such as multipath TCP [283–285]. Reference [286] analyzed and modeled offloading in heterogeneous networks. Reference [287] outlined approaches for enabling mobility and multihoming. Reference [288] studied incentives and algorithms for sharing Wi-Fi infrastructure with nomadic users. Reference [289] found that Wi-Fi offloading can save between 75% and 90% of the energy for network transmissions compared to Third Generation Mobile Telecommunications (3G) connectivity only. References [290, 291] presented systems for sharing private Wi-Fi credentials among trusted friends with an OSN app. Reference [292] analyzed the offloading efficiency based on Wi-Fi density. Reference [293] computed the minimum required number of Wi-Fi access points to provide the same average throughput as a cellular network. Reference [294] proposed analytical models for the Wi-Fi offloading efficiency.

References [295–297] investigated offloading to public Wi-Fi infrastructure in urban environments. Reference [298] scanned available governmental and private Wi-Fi hotspots in Paris, France, and found a high potential for Wi-Fi offloading. Reference [299] used user mobility, interaction, and traffic demands to find a deployment of hotspots in metropolitan areas, which enables a higher offloading ratio. Similarly, [300, 301] studied hotspot deployment, but additionally considered also battery levels. A tracking method was applied by [302] to measure the spatial distribution of Wi-Fi hotspots, and a high density of hotspots was observed in residential areas. This relates to [303], which found that base stations of different operators are clustered according to population density, and indicates that population density models [304–306] could be used to model the distribution of Wi-Fi hotspots. Also different stochastic geometry models were used to model the density of cellular networks [307, 308]. However, these models lack to generate hotspot distributions for cities of irregular shape, e.g., cities with a natural coastline. Thus, they are not sufficient for the performance evaluation of public Wi-Fi offloading solutions, and better models are needed.

Traffic management for mobile services, such as Wi-Fi offloading, can benefit from social information, such as the location of users. The user location

can be gathered from the mobile network operator, from services (e.g., [309]), or directly from end users. They often share their location with applications, or as explicit postings in OSNs or specialized services like Foursquare[7]. Reference [310] exploited mobility prediction and prefetching to enhance mobile data offloading. Reference [311] used number of users and load in the network to decide on offloading. Reference [312] proposed a system that selected the offloading of video and download flows between LTE and Wi-Fi networks based on QoE and network information. Similarly, [313] evaluated a LTE/Wi-Fi system and presented a rate redistribution algorithm to optimize the QoE of video streams.

4.2 Public and Private Wi-Fi Offloading

Wi-Fi offloading is a highly relevant traffic management action due to the wide distribution of Wi-Fi infrastructure, both in public places and in private homes. The performance of Wi-Fi offloading mechanisms is largely depending on the coverage of hotspots and the strength of the received signal. To design and evaluate such mechanisms, thus, a model for the locations of Wi-Fi hotspots is required. Based on [10], Section 4.2.1 analyzes the distribution of public Wi-Fi hotspots in large cities and develops a simple model, which can be used to generate realistic hotspot locations for arbitrary cities. Wi-Fi is also prevalent in home environments to provide Internet access for mobile devices. Thus, there is a huge potential to incorporate also this infrastructure in traffic management. However, private Wi-Fi networks should not be publicly shared, but only to trusted friends. Therefore, Section 4.2.2 presents a framework for Home Router Sharing based on Trust (HORST) following [42, 56]. The framework enables private Wi-Fi offloading, and additionally supports content delivery, caching, and prefetching mechanisms for HAS.

---

[7]`https://foursquare.com` [Online] – Accessed: 24.08.2017

4.2.1 Simple Hotspot Location Model for Public Wi-Fi Offloading

The potential for Wi-Fi offloading is influenced by the distribution of the Wi-Fi hotspots and their transmission range, i.e., their coverage. If flows are offloaded, the received throughput depends on the backhaul link, the signal strength, and the number of simultaneous users in the Wi-Fi network. Thus, the locations of users and Wi-Fi hotspots have to be considered for the performance evaluation of traffic management mechanisms utilizing Wi-Fi offloading. A generic model would allow to generate hotspot distributions for cities of different size, population density, shape, and number of hotspots. These hypothetical distributions can be used to design novel mechanisms, which rely on offloading to Wi-Fi infrastructure, and evaluate their performance and scalability.

To develop a model for the distribution of Wi-Fi hotspots, the hotspot locations of ten large cities were collected from a public database and their properties are analyzed. A simple, intuitive, and generic model is presented, which shows a sufficient accuracy for all cities. Note that also more complex distributions or higher order models can be used to model the hotspot locations, which might better reproduce the characteristics of particular cities. However, these models are less generic and must be fitted for each city individually. Moreover, in case individual fitting is necessary, the actual hotspot locations can be taken from public databases, which give a higher accuracy and render the fitted model useless.

Methodology

Ten large cities, two in Europe and eight in the United States, with a large number of Wi-Fi hotspots and different layouts (e.g., grid-based or ring-based layout) were selected to develop a general model for the geographic distribution of hotspots. Table 4.1 lists the investigated cities and some of their characteristics, namely, the number of hotspots, the area, and the population. The addresses of public hotspots in these cities were manually searched on the web-

based database OpenWiFiSpots[8] and parsed from the search results. Note that some public hotspots might not be listed in the database, and thus, the collected hotspot addresses are only a sample of a possibly larger number of Wi-Fi hotspots. By using the MapQuest[9] geocoding service, the obtained addresses were transformed into an ordered pair of geographic coordinates (φ, λ), i.e., latitude φ and longitude λ, which uniquely identify each location on the surface of the earth. To illustrate this process, Figure 4.2 depicts a map extract of London (map source: OpenStreetMap[10]), and visualizes the hotspots in this area.

Table 4.1: General information about investigated cities.

City	Number of hotspots	Total investigated area [km^2]	Population [in thousands]
Austin	220	220	843
Berlin	110	250	3502
Boston	193	173	637
Brooklyn (NYC)	454	419	2566
Houston	307	306	2161
Los Angeles	199	165	3858
London	668	367	8308
Portland	419	465	603
San Francisco	214	241	826
Seattle	296	202	635

To find general characteristics for all cities, the hotspot distributions are analyzed relative to the city center. The city center (φ_c, λ_c) is calculated by using the k-means algorithm to compute the centroid of the hotspot locations. Then, the geographic coordinates of the Wi-Fi hotspots are mapped into a polar coordinate system, which has (φ_c, λ_c) as origin and north as reference direction:

---

[8]http://openwifispots.com/ [Online] – Accessed: 24.08.2017
[9]https://developer.mapquest.com/ [Online] – Accessed: 24.08.2017
[10]https://www.openstreetmap.org/ [Online] – Accessed: 24.08.2017
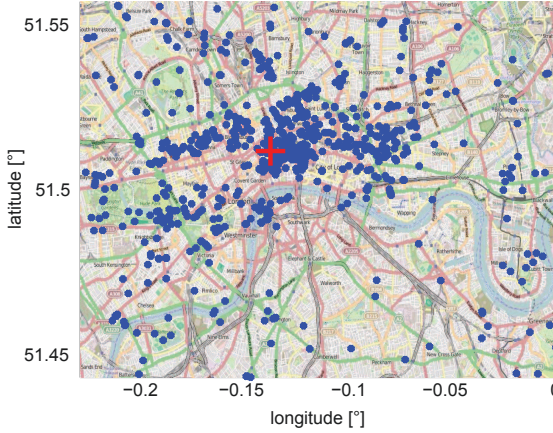
Figure 4.2: Extract of hotspot locations (blue) of London and computed center (red cross) at (51.5126N, 0.1405W). Map source: OpenStreetMap.

$$a = \sin^2\left(\frac{\varphi - \varphi_c}{2}\right) + \cos\varphi \cdot \cos\varphi_c \cdot \sin^2\left(\frac{\lambda - \lambda_c}{2}\right), \tag{4.1}$$

$$d = 2 \cdot r_E \cdot atan2(\sqrt{a}, \sqrt{1-a}), \tag{4.2}$$

$$\theta = atan2(\sin(\lambda - \lambda_c) \cdot \cos\varphi,$$

$$\cos\varphi_c \cdot \sin\varphi - \sin\varphi_c \cdot \cos\varphi \cdot \cos(\lambda - \lambda_c)). \tag{4.3}$$

Equations 4.2 and 4.3 transform the geographic coordinates (φ, λ) (in radians) of each hotspot location into polar coordinates (d, θ) with the spherical distance d from the city center and angle θ towards north[11]. The equations use the haversine formula (Equation 4.1), the mean radius of the earth $r_E \approx 6371$ km,

---

[11] http://www.movable-type.co.uk/scripts/latlong.html [Online] – Accessed: 24.08.2017

and the two-argument $atan2$ function[12], which is a four-quadrant arctangent function implemented by many programming languages and has a codomain of $[-\pi, \pi]$. Note that positive angles of θ point clockwise from north, whereas negative angles point counterclockwise from north.

Analysis of Hotspot Distributions

(a) Angular distribution. *(b) Distance distribution.*
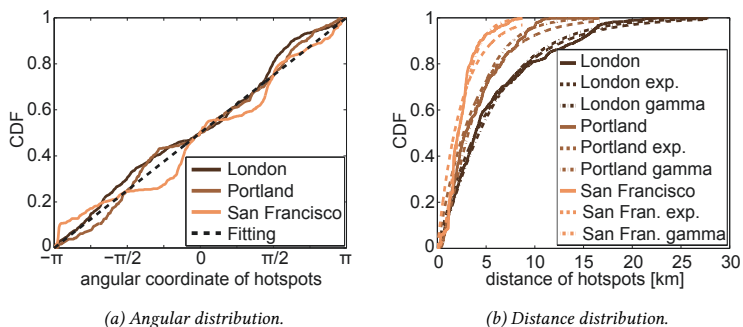
Figure 4.3: Angular and distance distributions of London, Portland, and San Francisco hotspots with respect to city center.

In the following, the distributions of distance and angle of the polar coordinates are analyzed. Figure 4.3 visualizes these distributions for the three diverse cities of London, Portland, and San Francisco. Figure 4.3a compares the CDFs of the angles of the hotspots' polar coordinates (solid) to a uniform distribution $F(x) = \frac{x+\pi}{2\pi}, x \in [-\pi, \pi)$ (black dashed). It can be observed that the angular distributions are not perfectly uniform because of city-specific geographic peculiarities like parks or water areas, which cause less hotspots at the corresponding angles. Nevertheless, still the uniform distributions show a good approximation. To quantitatively assess the goodness of fit, two standard methods for compar-

---

[12]`http://www.mathworks.com/help/matlab/ref/atan2.html` [Online] – Accessed: 24.08.2017

ing distributions are applied and the results for all ten cities are shown in Table 4.2. The used metrics are the maximum absolute error, i.e., the Kolmogorov-Smirnov statistic D, and the mean absolute error (mae)[13], which indicate how far the model is from the real distribution at most (D) and on average (mae). Confirming the impressions from Figure 4.3a, the uniform distributions have a rather high D for all cities due to the particular characteristics of each city. For example, the highest D value of the city of Austin is caused by the city shape, which contributed to a slightly elliptic hotspot distribution. However, the low mae values indicate that the uniform distribution nevertheless well approximates the angular distribution of hotspots, and achieves a sufficient accuracy for practical applications (see below).

Table 4.2: *Maximum (D) and mean (mae) absolute error for uniform fittings of angular distribution.*

City	D	mae
Austin	0.1619	0.0677
Berlin	0.0841	0.0401
Boston	0.0809	0.0274
Brooklyn (NYC)	0.1142	0.0537
Houston	0.1023	0.0412
Los Angeles	0.0757	0.0288
London	0.0855	0.0308
Portland	0.0713	0.0244
San Francisco	0.1036	0.0370
Seattle	0.0887	0.0353

Figure 4.3b depicts the CDFs of the hotspots' distances from the city center (solid). The CDFs show a high similarity to an exponential distribution $F(x, \mu) = 1 - \exp(-\frac{x}{\mu}), x \geq 0$ (dashed) with mean μ. By estimating the mean μ of the exponential distribution from the hotspot distances in a maximum like-

---

[13]http://www.mathworks.com/matlabcentral/fileexchange/
22020-goodness-of-fit--modified-/content/gfit2.m [Online] – Accessed: 24.08.2017

lihood sense (cf. second column of Table 4.3), a good fitting is obtained. As the coefficients of variation c_v (cf. third column of Table 4.3) indicate that the distributions are not perfectly exponential ($c_v \approx 1$), they are also compared to a more general gamma distribution $F(x, \alpha, \beta) = \int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} u^{\alpha-1} \exp(-\beta u) \mathrm{d}u = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}, x \geq 0$ (dash-dot lines), where γ is the lower incomplete gamma function. Note that the parameters α and β can be estimated from μ and c_v. The goodness of the fittings is shown in the last columns of Table 4.3 using again the maximum (D) and mean (mae) absolute error metrics. As observed in Figure 4.3b, the D_E values confirm that the distance distributions are not perfectly exponential due to city-specific peculiarities. The highest D value is observed in San Francisco, for example, because there is a high hotspot density along the northeast waterfront that cannot be accurately covered by the exponential distribution ($c_v \ll 1$). Yet again, the generally low mae values show that a decent approximation can be reached. Using the more general gamma distribution with two parameters results in a better fitting in terms of D and mae for most cities. The smaller the c_v values, the better the goodness of the gamma fitting compared to the exponential fitting. Thus, the additional parameter of the gamma distribution helps to decrease the D and mae metrics especially for cities with low c_V, such as Austin or San Francisco, and improves the approximation of the real distance distributions. In contrast, the closer c_v to 1, e.g., Berlin or London, the smaller the gain of using a gamma distribution.

These results show that the transformation of the geographic coordinates of the hotspot locations into polar coordinates with respect to the city center allowed to find general characteristics of the hotspot distributions among diverse cities. In particular, it could be observed that the uniform distribution can provide a decent approximation of the angular distribution, while the distance distributions can be fitted with exponential or gamma distributions. These characteristics can now be used to create artificial hotspot distributions for generic cities.

Table 4.3: Mean (μ) and coefficient of variation (c_v) of distribution of hotspot distances, and D and mae for exponential (E) and gamma (G) fittings.

City	μ	c_v	D_E	mae_E	D_G	mae_G
Austin	3.2041	0.6886	0.1262	0.0545	0.0814	0.0319
Berlin	5.0306	0.9335	0.0661	0.0270	0.0767	0.0295
Boston	2.5224	0.7201	0.1884	0.0476	0.1048	0.0195
Brooklyn (NYC)	5.5942	0.7948	0.2057	0.0383	0.1674	0.0466
Houston	6.7299	0.7294	0.1645	0.0335	0.0608	0.0124
Los Angeles	2.7116	0.7239	0.1079	0.0542	0.1027	0.0418
London	5.7011	0.8826	0.0488	0.0185	0.0609	0.0177
Portland	3.8773	0.7394	0.1186	0.0339	0.0640	0.0199
San Francisco	2.5045	0.6294	0.2732	0.0710	0.0943	0.0179
Seattle	2.9365	0.7968	0.1136	0.0464	0.1304	0.0456

Generation of a Hotspot Distribution for a Generic City

The generation of a hotspot distribution will facilitate the observed characteristics of the hotspots' polar coordinates. Thus, the coordinates of the city center (φ_c, λ_c) (latitude/longitude) have to be defined first. Then, a random Wi-Fi hotspot location can be computed by generating a uniformly distributed angle θ and a distance d, which follows the desired exponential or gamma distribution, for example, by utilizing inverse transform sampling. The polar coordinates (d, θ) are then transformed to latitude/longitude coordinates (φ, λ) (in radians) taking into account the city center (φ_c, λ_c) and the radius of the earth r_E as described in Equations 4.4 and 4.5[14]:

$$\varphi = \arcsin(\sin\varphi_c \cdot \cos\frac{d}{r_E} + \cos\varphi_c \cdot \sin\frac{d}{r_E} \cdot \cos\theta), \qquad (4.4)$$

$$\lambda = \lambda_c + atan2(\sin\theta \cdot \sin\frac{d}{r_E} \cdot \cos\varphi_c, \cos\frac{d}{r_E} - \sin\varphi_c \cdot \sin\varphi). \qquad (4.5)$$

---

[14]`http://www.movable-type.co.uk/scripts/latlong.html` [Online] – Accessed: 24.08.2017

The limitation of this naive method is that it distributes the Wi-Fi hotspots over a circular and possibly unlimited area. To distribute hotspots only within a city with a given shape, additionally an accept-reject method has to be applied, rejecting all hotspots outside the city limits. However, the resulting truncated distribution of the accept-reject method will have different characteristics than the modeled distribution. Figure 4.4 shows this effect for the city of San Francisco. The first subplot visualizes the Wi-Fi coverage (blue) taking into account the real hotspot locations and a transmission range of 100 m. Additionally, the city limits, which were computed as the convex hull of the hotspot locations, are shown in black. The subplots below depict artificial hotspot distributions with the same number of hotspots generated according to the fittings presented in Tables 4.2 and 4.3. It can be seen that the naive exponential and gamma models place many Wi-Fi hotspots outside the city limits (red). By accepting only locations inside the city limits, the truncated exponential and truncated gamma models result.
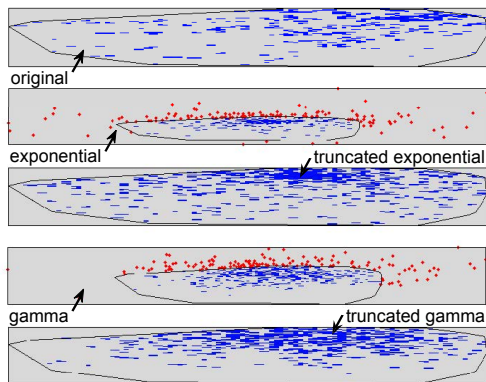
Figure 4.4: Original and generated Wi-Fi coverage (single run) of San Francisco.

Comparison of Model and Hotspot Characteristics

In the following, the characteristics of the original and the generated hotspot distributions are compared in detail for the exemplary city of San Francisco. Note that the accuracy of the models for this city is rather low (compared to the other cities), such that the presented results can be almost considered a worst case analysis. Figure 4.5a shows the CDFs of the distance of the hotspots to the city center for all four generation approaches. It can be seen that the truncation causes smaller distances compared to the original distribution, which means that the model parameters have to be adjusted to account for the truncation. From a practical point of view, the distance of any point to the closest Wi-Fi hotspot is relevant because it determines the most important offloading parameters, such as coverage, interference, or signal strength. Figure 4.5b depicts the CDF of this distance for the original distribution (black), and the mean distance over 50 generated distributions using all four generation approaches. Although the impact of the truncation is visible again, a high similarity can be reached. It can be observed that the exponential models can better replicate this characteristic due to the higher variance of the distances. As this effect can be seen for all ten cities, only the truncated exponential model will be investigated in the following.

The presented simple model implicitly includes the independence of the two dimensions of the polar coordinates, which might not hold for the realistic hotspot distributions. Therefore, the hotspot densities of the original and the truncated exponential distributions of San Francisco are spatially investigated. Figure 4.6 investigates the clustering of hotspot locations by applying the density-based clustering algorithm DBSCAN [314] to find clusters of at least 10 hotspots with two different density values $\epsilon = 0.003$ and $\epsilon = 0.005$. Note that a larger ϵ allows larger distances between hotspots in the same cluster. The algorithm detects five clusters in the original hotspot locations in Figure 4.6a, a large one and four small, very dense, and spatially separated clusters. The corresponding analysis of the generated hotspot distribution in Figure 4.6b also results in one large and four small clusters. However, the small clusters are rather concentrated around the center and less dense than the clusters in the original

(a) Distance distribution.

(b) Distance of random point to closest hotspot.
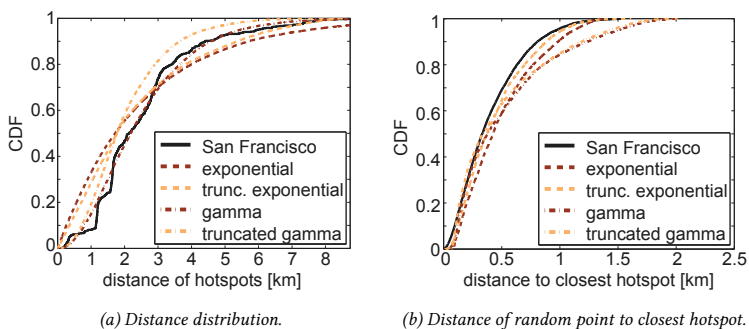
Figure 4.5: Impact of models (exponential, gamma) and truncation to city limits (convex hull of hotspots) on generated hotspot distribution in San Francisco.

hotspot locations. The differences are even more striking if the larger ϵ is used in Figures 4.6c and 4.6d. The original hotspots show four irregularly shaped and distributed clusters. In contrast, the generated hotspot distribution has a large, centered cluster, and thus, does not reproduce the spatial clustering of the original hotspots.

To quantify this analysis and generalize it to all cities, the spatial autocorrelation of the number of hotspots in range is investigated in terms of Moran's I [315], which measures the global spatial autocorrelation, and Geary's C [316], which is more sensitive to local spatial autocorrelation. Moran's I ranges from -1 (regular dispersion) to 0 (random pattern) to 1 (high clustering), while Geary's C takes values from 0 (high clustering) to 1 (random pattern) to 2 (regular dispersion). Figure 4.7 presents Moran's I for different Wi-Fi ranges in San Francisco. It can be seen that Moran's I approaches 1 when the transmission range increases because the coverage areas grow and the hotspots are more clustered. Still, for all ranges, the I values of the model (yellow) are similar to the original data (blue). Table 4.4 lists Moran's I and Geary's C for all ten cities and a Wi-Fi range of 50 m, and shows that the spatial autocorrelations generally have less

(a) Clustering of original hotspot locations with $\epsilon = 0.003$.

(b) Clustering of generated hotspot locations with $\epsilon = 0.003$.

(c) Clustering of original hotspot locations with $\epsilon = 0.005$.

(d) Clustering of generated hotspot locations with $\epsilon = 0.005$
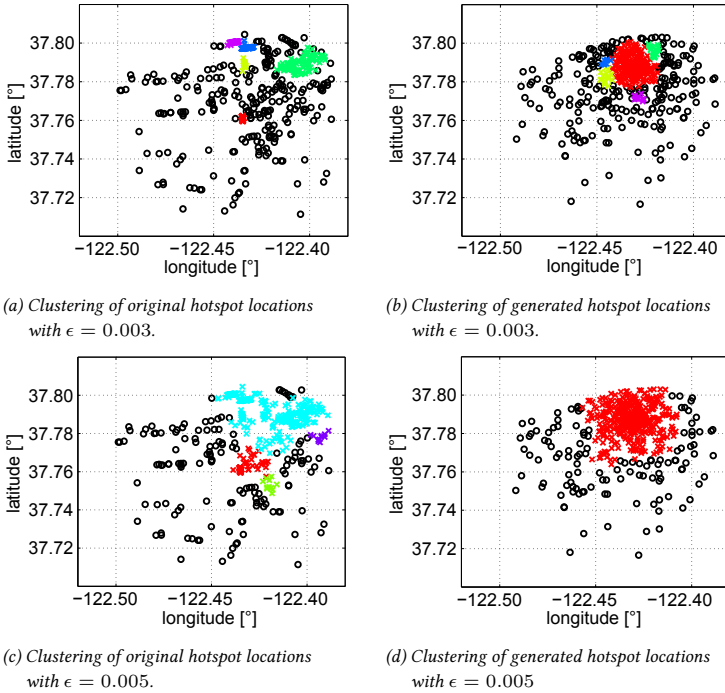
Figure 4.6: Clustering with density-based clustering algorithm DBSCAN.

discrepancy between the original data and the model than San Francisco, and resemble the original distributions even better.
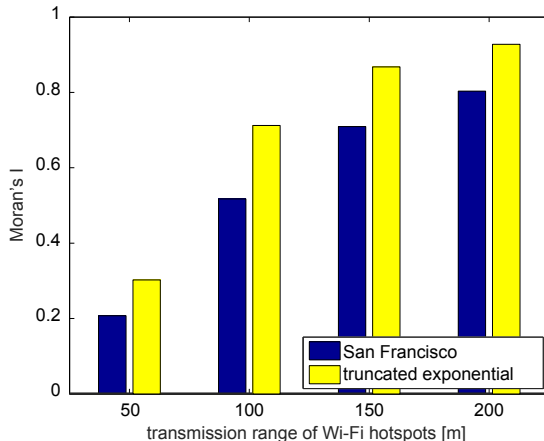
Figure 4.7: Spatial autocorrelation of number of hotspots in range measured by Moran's I for different Wi-Fi ranges.

The results show that the simple truncated exponential model cannot generate the spatial patterns of the original hotspot distributions in all details, e.g., clustering of hotspots. However, the model can replicate some characteristics with a high accuracy, such as the distance to the next hotspot or the spatial autocorrelation of the hotspot coverage. The presented results for San Francisco suffer from the poor fitting of the exponential model for this city compared to other cities (cf. Table 4.3), and it can be seen that the simple model performs better for the other cities.

Table 4.4: Spatial autocorrelation of hotspot coverage for Wi-Fi range of 50 m.

City	I (orig.)	I (model)	C (orig.)	C (model)
Austin	0.0730	0.0783	0.9295	0.9368
Berlin	0.0201	-0.0012	0.9317	1.0164
Boston	0.0918	0.1432	0.9080	0.8718
Brooklyn (NYC)	0.0333	0.0762	0.9732	0.9389
Houston	0.0042	0.0386	1.0110	0.9766
Los Angeles	0.0180	0.0138	0.9680	1.0013
London	0.1420	0.0511	0.8731	0.9640
Portland	0.0720	0.1574	0.9345	0.8577
San Francisco	0.2073	0.3344	0.8066	0.6804
Seattle	0.1301	0.1741	0.8721	0.8409

Applicability of the Model

To investigate the applicability of the model for practical performance evaluations, three use cases are selected. First, the Wi-Fi coverage within a city is investigated. Then, the number of hotspots in range is analyzed, and finally, an application to Wi-Fi mesh networks is considered.

An accurate estimation of Wi-Fi coverage is the basis for evaluations of offloading potential, signal strength, and their corresponding applications. The distance to the closest hotspot (cf. Figure 4.5b) and the Wi-Fi transmission range determine the coverage of a certain location. Figure 4.8 shows the coverage of the three examples Berlin, London, and San Francisco depending on the Wi-Fi radius. It depicts the mean coverage percentage (dashed) and 95% confidence intervals for ten generated distributions and compares to the original coverage (solid). A high similarity between the generated and original hotspot distribution is visible, which follows from the similar distance of arbitrary points to the closest hotspots presented in Figure 4.5b. This high accuracy also holds for the other investigated cities and confirms that the simple truncated exponential model can be used to investigate the coverage, for example, to quantify how well traffic management using Wi-Fi offloading will perform.
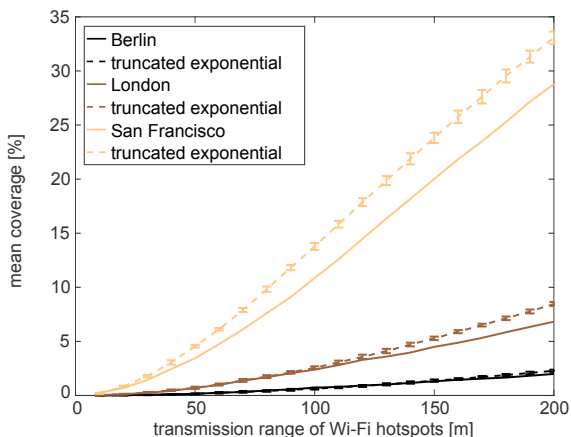
Figure 4.8: Mean coverage for different Wi-Fi ranges.

The number of hotspots in range are essential for analyses of handovers, interference, bandwidth resource sharing, and their corresponding applications. Figure 4.9a presents the average number of hotspots in range and 95% confidence intervals depending on the Wi-Fi range. Again ten runs were conducted for each of the three example cities. The plot shows that the average number of hotspots can be accurately replicated by the model. In Figure 4.9b, the standard deviation of the number of hotspots in range has a high similarity for Berlin and London, but there are large differences for the city of San Francisco. This confirms the above findings, and it follows that the application of the model suffers if the spatial clustering of hotspots cannot be accurately replicated.

Wi-Fi mesh networks rely on direct communication among the hotspots to save or aggregate backhaul traffic volume (e.g., BeWi-Fi[15] [8, 26]) or for privacy reasons (e.g., Freifunk[16]). Therefore, the mesh of the hotspots in a city is rep-

---

[15]`http://www.tid.es/research/areas/bewifi` [Online] – Accessed: 24.08.2017
[16]`https://freifunk.net/en/` [Online] – Accessed: 24.08.2017

(a) *Mean of number of hotspots in range.*

(b) *Standard deviation of number of hotspots in range.*
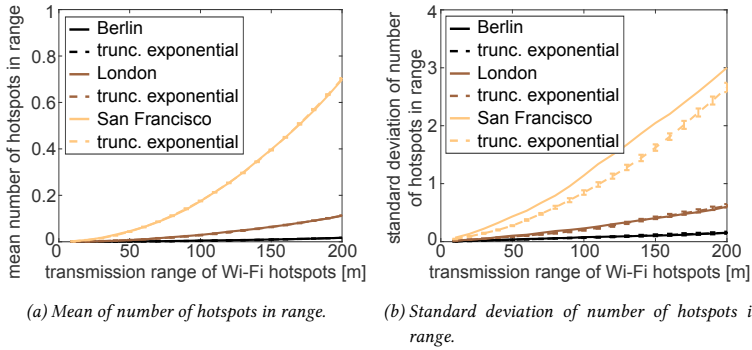
Figure 4.9: Comparison of number of hotspots in range.

resented as a graph with hotspots as nodes, which are connected by an edge if they are within Wi-Fi range, and thus, can communicate. Figure 4.10a depicts the number of connected components for the mesh graph of the original hotspots of San Francisco (blue) and the generated hotspots (yellow) for different Wi-Fi ranges. It can be seen that the number of connected components decreases if the transmission range increases. The model has a higher number of connected components especially for small Wi-Fi ranges. Figure 4.10b presents the CDF of the betweenness centrality in San Francisco for a transmission range of 50 m. A higher betweenness centrality means that a hotspot is more important for the mesh because more flows would be routed through this node. It can be seen that the number of isolated hotspots, which have a betweenness centrality of 0, is higher for the generated distribution. Again the different clustering of both distributions in the case of San Francisco is evident. The original hotspots are more locally centered with more locally important nodes with a high betweenness centrality and less isolated nodes. In contrast, the generated distribution is highly centralized (cf. Figure 4.6), which results in a few very highly ranked hotspots. Thus, again for this city, the application of the model to investigate Wi-Fi mesh networks is not recommended.

(a) Number of connected components.   *(b) Betweenness centrality for Wi-Fi range of 50 m.*
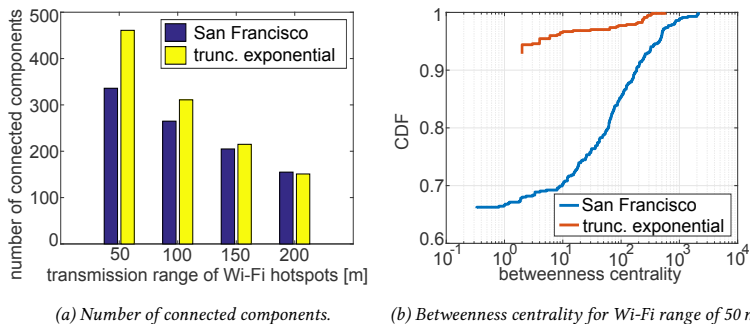
Figure 4.10: Metrics for Wi-Fi mesh network graph representation of San Francisco.

To sum up, the limitation of the simple truncated exponential model to replicate the spatial clustering of the hotspots for some cities propagate to some desired applications, which require proper hotspot collocation, such as handovers, interference, bandwidth sharing, or mesh networks. Still, for many applications, such as offloading potential, coverage, or signal strength, the simple model provides a sufficiently high accuracy. Thus, the model can be utilized for the performance evaluation of traffic management mechanisms, which rely on coverage and throughput of public Wi-Fi infrastructure in cities.

4.2.2 Home Router Sharing based on Trust

Next to public Wi-Fi hotspots, also the resources of private Wi-Fi infrastructure in home environments have to be considered for SATM. Social awareness allows to predict social cascades of the user and his friends, i.e., the propagation of content in OSNs, as well as their interests and mobility. This allows to predict where, when, and by whom content will be requested, and thus, to utilize the Internet access and storage capacities of private home routers accordingly. However, as the access to a private Wi-Fi network could be abused by malicious users, it should not be open to any user, but only to trusted users. Again, social informa-

tion can be used to infer trust relationships among users, e.g., from friendship structures and interactions in OSNs. In the following, the Home Router Sharing based on Trust (HORST) mechanism is presented, which enables SATM for HAS on private home routers by offering Wi-Fi offloading, caching, prefetching, and content delivery. This mechanism was developed and evaluated in the context of the EU FP7 SmartenIT project. First, the basic functionality and use cases for HORST are outlined. Then, the components of HORST, namely, a firmware for home routers, an OSN app, and a mobile device app, are introduced. Finally, the prototype implementation of the mechanism is described.

a) Basic Functionality and Use Cases

The main goal of HORST is to enable Wi-Fi offloading for all participating users based on trust relationships. Therefore, two Wi-Fi networks (Service Set Identifiers (SSIDs)) are established, a private and a shared network. The shared Wi-Fi is open and can be used to negotiate access to the private Wi-Fi. The access information of the private Wi-Fi network is uploaded to an OSN, where it can be shared with other trusted users upon request. The location information available in the OSN can be used to recommend near Wi-Fi networks to users, such that the user can request access and connect for Wi-Fi offloading. Additionally, the home router establishes a User-owned Nano Data Center (UNaDa) [270, 271] for the management of relevant video contents. Based on the friendship structures, the activities, and the interests of the users, social cascades can be computed in the OSN to predict when which videos will be requested by which user. If the future requests will be at a user's or a friend's home, this allows to cache or prefetch the relevant videos to the corresponding home router, such that they can be delivered when the user is connected to the Wi-Fi network. Eventually, the UNaDas can be federated to form a CDN overlay to efficiently distribute the cached and prefetched videos among friends. Thereby, the social cascades of the content in the OSN can be mimicked for the content delivery, which will save resources of service providers and cloud operators, and can even reduce the

expensive inter-AS traffic of ISPs if users are in the same AS. The basic function-
ality is also shown in Figure 4.11.
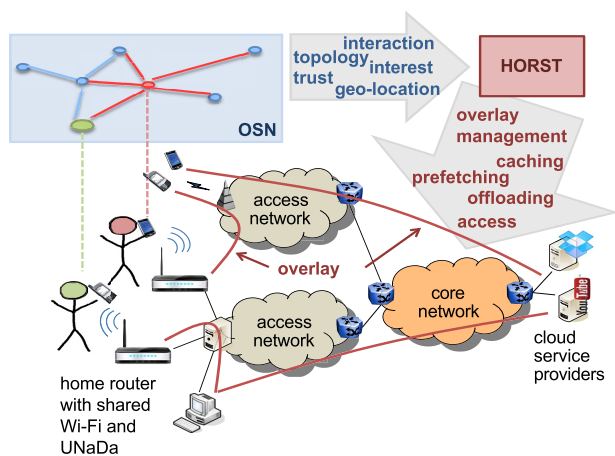
*Figure 4.11: Basic HORST functionality. HORST uses information from the OSN to
provide end users access to shared, private Wi-Fi access points based
on their trust scores, to enable content caching and prefetching, and to
control an efficient content delivery overlay.*

Wi-Fi Offloading   As mentioned above, offloading flows from mobile to fixed
networks via Wi-Fi reduces the load on mobile networks and saves costs for
mobile network operators. End users benefit from a higher QoE due to higher
bandwidths in residential Wi-Fi networks and lower consumption of mobile data
volume and energy on the mobile devices. The participation of more trusting
friends in HORST will extend the Wi-Fi coverage, and thus, the offloading poten-
tial. Further incentives could be given to users, which share their home routers,
but are out of scope of this work. Based on the location of private hotspots and
the mobility of users, HORST can guide users to other Wi-Fi networks and man-

age the access request. Additionally, the transmission of delay-tolerant content can be scheduled and delayed until the user is connected to a Wi-Fi network. Note that not only trust but also other incentive mechanisms could be possible to regulate the sharing of Wi-Fi access. These mechanisms should also consider users who have no home router to share, but also want to participate in HORST for improved QoE. The strict separation of private and shared Wi-Fi networks will prevent improper access to private network devices, and traffic management can be applied to the shared access link to make a compromise over the resource allocation for private flows and flows of visiting users.

Content Caching and Prefetching   The UNaDa on the home router can provide local storage for relevant videos, which can be filled proactively (prefetching, i.e., download content and place it in storage before first request) or reactively (caching, i.e., keep downloaded content in storage for future requests). The temporal and spatial popularity of video content can be predicted from social cascades in OSNs or the access history of UNaDas. This allows to select content for prefetching and improve the local caching on the home router. Thus, users who are interested in content, which is shared via OSN or frequently requested by friends, will often find that the UNaDa already stores the desired content. This locally stored content can be requested in the Wi-Fi network with much less delay and a higher bandwidth, which generally results in a higher QoE for HAS. HORST also allows to specify that content will be accessed at a later time, e.g., when users are back at home or when they have shared Wi-Fi access. In this case, HORST prefetches the video to the specified UNaDa in the indicated period, which allows to schedule the transmission, such that peak loads in the network can be avoided. Thus, also ISPs benefit from HORST's caching and prefetching capabilities, which reduce the load on the network and shift delay-tolerant traffic to off-peak periods.

Content Delivery   The social cascades of video content in OSNs, in which video links propagate from users to their friends, can also be reproduced for

content delivery. Therefore, HORST federates the UNaDas on home routers and establishes an overlay CDN. Following the social cascades, also the video contents, which are shared in the OSNs, can be delivered from users to their friends. Again, the video distribution can be scheduled to non-peak hours in case of delay-tolerant content. For each content request, HORST decides from which resource to download the content depending on the SATM preferences. If the content is available is the UNaDas' overlay CDN, the content can be requested from another UNaDa to reduce the load on the service provider and cloud operator. UNaDas in the same AS can be preferred, which will save inter-AS traffic for the ISP. Note that the content delivery via the UNaDas' overlay CDN might not provide the same performance than downloading content from the service provider due to the prevalent asymmetric Internet access of home routers (e.g., Asymmetric Digital Subscriber Line (ADSL)) with a reduced upload bandwidth. However, if the expected QoE is not sufficient or the video cannot be requested from a UNaDa, it is available on the server of the cloud service provider at last.

b) Components of Home Router Sharing based on Trust

Three components are needed to implement the HORST mechanism. These are a home router firmware, an OSN application, and an application running on mobile devices.

Home Router Firmware   The home router firmware needs to be modified to bring the HORST functionality to home routers. Due to legal and privacy issues, the home routers have to provision at least two Wi-Fi networks with different SSIDs. The shared Wi-Fi network is open to all users but only provides minimal bandwidth to negotiate access to the private network. Other traffic in the shared network is blocked. If the negotiation is successful, i.e., if the requesting user is trustworthy, the protected private Wi-Fi network can be accessed. Therefore, appropriate Wi-Fi hardware is required as well as the support of Virtual Access Points (VAPs). Several projects implement VAP support, such

as DD-WRT[17], OpenWrt[18], and Freetz[19]. To provide connectivity for all devices, the firmware has to distinguish the Basic Service Set Identifier (BSSID), which uniquely identifies a specific access point interface mostly with its Media Access Control (MAC) address. Note that problems might occur in end user devices if multiple SSIDs share the same MAC address.

The hardware requirements for home routers to serve as cache and contribute to the CDN overlay are memory and storage. The available memory on current home routers is sufficient for basic operations, and external Universal Serial Bus (USB) drives can increase the storage. Then, the firmware has to set up the UNaDa and run an overlay management software. Content has to be pushed or pulled from other nodes in the overlay network (i.e., other HORST routers) or the original content provider. Content requests have to be intercepted to serve cached content, or redirect the request to another node in the overlay. Thereby, additional load balancing and traffic management can be applied to assure a high service quality.

Online Social Network Application   The OSN application is the major innovation of HORST and provides input for all SATM decisions. It allows to utilize the convenient user management of the OSN for HORST, such that users can simply log in to the OSN and grant permissions to the app in order to participate in HORST. The app requires permissions to retrieve personal data, position data, and communication data. Additionally, users have to submit information about their own home router, i.e., home router position, Internet Protocol (IP) address, Wi-Fi SSIDs, and passwords for Wi-Fi networks and UNaDa access.

The OSN app computes trust scores from social information in the OSN. Explicit trust scores can be obtained by directly asking users to indicate whether they trust other users. Nevertheless, also implicit trust scores can be computed based on personal data, communication data, and OSN topology. Trusted users

---

[17]`http://www.dd-wrt.com/` [Online] – Accessed: 24.08.2017
[18]`https://openwrt.org/` [Online] – Accessed: 24.08.2017
[19]`http://freetz.org/` [Online] – Accessed: 24.08.2017

are then determined by comparing their implicit trust score to a threshold defined by the user. Eventually, a combination of both mechanisms is possible, e.g., a recommendation of trustworthy users based on their implicit trust scores, which have to be confirmed by the user explicitly. If users want to access another Wi-Fi network, they have to send a request to the owner via the OSN app. The shared Wi-Fi provides a minimal bandwidth to access the OSN for submitting the request. Also the credentials for the private Wi-Fi can be received in the shared Wi-Fi if the requesting user is considered trustworthy, i.e., after implicit or explicit trust scores were obtained.

The OSN app can further collect and analyze social information for traffic management about users' positions, interests, and preferences, as well as content popularity or social cascades. Based on the position data, the app can recommend and request access to near Wi-Fi networks. Content access patterns can be analyzed to detect and distribute popular content over the UNaDas' overlay CDN. Thereby, the transmission of delay-tolerant content can be scheduled to non-peak hours to more efficiently utilize network resources. The same mechanisms can be applied to prefetch or cache content for individual users on the home router to which he already is or soon will be connected. Again, delay-tolerant content can be predicted, or the user himself can explicitly indicate the desired content, the time of consumption, and the home router on which the content has to be stored, thereby, allowing for an efficient transmission scheduling.

Mobile Device Application   The mobile application brings HORST directly to mobile devices. It manages the request of Wi-Fi credentials from the OSN app, and stores them on the device for automatic access to the Wi-Fi network. Moreover, it automatically interacts with the OSN app without the user having to be constantly engaged manually and upload information. For example, position data or delayed content requests can be uploaded. On the other hand, the mobile app can receive information from the OSN app, e.g., about positions of near Wi-Fi networks, cached video contents, or delayed requests. Furthermore,

the mobile app provides interfaces to configure the settings of HORST, such as trust thresholds or prefetching/caching preferences.

c) Design and Prototype Implementation

For the design of an actual HORST implementation, the functionality of mechanisms for Wi-Fi offloading, caching, prefetching, and content delivery has to be specified. The specification of these mechanisms is part of this work and will be outlined in the following. In contrast, the actual architecture and implementation of the prototype are not part of this work, and thus, only some implementation details are mentioned here to better understand the specification of the mechanisms. The entire design and implementation of the prototype is described in detail in [42] and its functionality was demonstrated in [65].

For the prototype implementation, the Raspberry Pi[20] Single-board Computer (SBC) with Raspbian operating system was selected as the home router. It was equipped with a USB-based Wi-Fi dongle supporting two SSIDs (private and shared) and an Secure Digital (SD) card for storage. The HORST OSN app was implemented for the OSN Facebook[21], and thus, to participate in the HORST platform, users have to log in to the app on Facebook and accept the required app permissions. The Facebook app is used for user management, collection of social information, and computation of trust relationships. For the prototype implementation, users are considered to trust each other if they are friends on Facebook. A mobile Android application, which is installed on the user's mobile device, handles the user authentication, receives the credentials, installs them in the mobile device, and connects to the private Wi-Fi network without entering a password or using Wi-Fi Protected Setup (WPS).

Caching mitigates the load on networks by avoiding repeated downloads of the same content. In contrast, prefetching shifts video downloads to off-peak periods, e.g., during night. However, the prefetched video contents have to be

---

[20]`https://www.raspberrypi.org/` [Online] – Accessed: 24.08.2017
[21]`https://www.facebook.com/` [Online] – Accessed: 24.08.2017

selected carefully to avoid unnecessary downloads. Therefore, the content consumption of clients is predicted in two separate processes. The overlay-based prediction uses information about cached and watched videos on other home routers in the overlay. The social prediction, on the other hand, takes into account content shared in the OSN by friends, the location of the user, the age and global popularity of videos, and the user's previous requests of the content. The social information is collected from the Facebook app via the Facebook Graph Application Programming Interface (API) and the video content provider, which is the Vimeo[22] video sharing portal. The exact specifications of the prediction algorithms implemented in the prototype can be found in [42]. Both algorithms output rankings of relevant video contents. The top contents of each ranking are downloaded until the cache is full. Periodically, new rankings are computed and irrelevant content is replaced according to the Least Recently Used (LRU) strategy. The used software to run the prediction algorithms and manage the cache is a Java web application, which is deployed in a Jetty[23] web server. The web server also provides a user interface, which can be accessed by users with any browser to manage their local cache and access point, and a Representational State Transfer (REST) API to allow the Android app to register users in remote home routers. Additionally, the mitmproxy[24] software employs a proxy server to intercept content requests, cache videos, and rewrite content requests, e.g., if the content is cached locally or available in the overlay.

The prototype incorporates a Peer-to-peer (P2P)-based overlay following the RB-Tracker approach [317] to interconnect the participating home routers using the TomP2P[25] library. The RB-Tracker mechanism uses a Distributed Hash Table (DHT) for the first look up of nodes, which store the content, and afterwards exchanges direct messages to find the closest content provider. The advantage of the distributed approach is that sensitive data, e.g., social information collected from Facebook, are always stored on the own device only, which reduces

---

[22]`https://vimeo.com/` [Online] – Accessed: 24.08.2017
[23]`http://www.eclipse.org/jetty/` [Online] – Accessed: 24.08.2017
[24]`https://mitmproxy.org/` [Online] – Accessed: 24.08.2017
[25]`https://tomp2p.net/` [Online] – Accessed: 24.08.2017

the risk of data leaks. Only the list of cached content is exchanged with other home routers. Thereby, the home routers are identified by IP addresses only, from which the identity of individual users cannot be deduced. When rewriting content requests to the home router overlay, the distance of home routers in terms of AS hops is considered, which is computed by using traceroute in combination with an AS directory. If the closest home router providing the content is two hops or more away, the content is requested from the data center or CDN of the service provider instead. Thus, expensive transit traffic of ISPs is avoided, since it would imply at least two AS hops.

These specifications allowed to implement, evaluate, and demonstrate a prototype of the HORST framework [42, 65]. Using the HORST Android application, users were able to request access at a Facebook friend's HORST home router, receive the credentials for the private Wi-Fi, connect to it, and access the Internet. This proves the feasibility of the presented framework to increase the Wi-Fi offloading potential by sharing private Wi-Fi infrastructure in home environments among trusted friends. Additionally, HORST supports other SATM actions for HAS, namely, caching, prefetching, and content delivery, to reduce the load on networks and reduce costs for ISPs, which were evaluated in [42]. In this work, only the performance of Wi-Fi offloading for HAS is analyzed, and the applied methodology and results are presented in the following section.

4.3 Wi-Fi Offloading of Mobile HAS Sessions

This section describes the performance evaluation of Wi-Fi offloading for HAS based on [29, 44, 47]. Thereby, mobile video streaming sessions are simulated in a city, which can be offloaded if shared Wi-Fi access points are in range, otherwise they are served in a mobile network. The video streaming sessions are assessed in terms of QoE factors and smartphone energy consumption depending on different RATs and Wi-Fi sharing percentages, i.e., the percentage that an existing Wi-Fi hotspot can be accessed. The used data sets for the evaluation scenario are presented in Section 4.3.1. Section 4.3.2 describes the simulation

framework, and Section 4.3.3 shows the performance evaluation results for the Wi-Fi offloading of HAS flows.

4.3.1 Data Sets for Performance Evaluation

To simulate mobile HAS on smartphones, several data sets are required, which are presented in this section. These data sets include video characteristics of mobile requests, coverage and throughput for different RATs, access point locations, way points of mobile users, as well as an energy model for smartphones.

Characteristics of mobile YouTube video streaming were found in [318]. The authors streamed 2000 YouTube videos in mobile networks, and analyzed the video formats, bit rates, and durations. They found that Android and iOS devices selected the *itag*36 format (240p, 25 fps, H.263, Advanced Audio Coding (AAC), 3GPP File Format (3GP)) in more than 80% of the sessions. Most videos had a bit rate between 220 and 250 kbps. In this performance evaluation, the bit rate distribution of the *itag*36 format and the distribution of the video durations, which were obtained from [318] and are shown in Figure 4.12, are used to generate realistic mobile video requests.

(a) Bit rate distribution.　　　　(b) Video duration distribution.
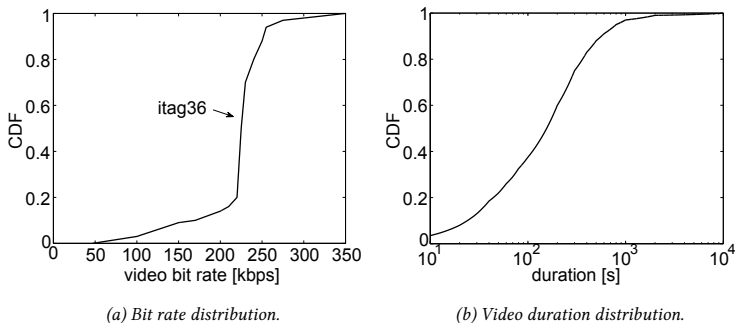
Figure 4.12: Bit rate and duration of mobile YouTube videos in itag36 format [318].

A large measurement study on the coverage and performance of cellular and public Wi-Fi networks was conducted in and around the city of Darmstadt, Germany, representing a medium sized urban area (population: ~150000, area: ~122.2 km^2, two universities, southern part of Frankfurt Rhine-Main Metropolitan Region, 30 km south of Frankfurt am Main, Germany). Since 2013, data has been collected using the NetworkCoverage[26] [319] app in both dedicated and voluntary measurement studies. The collected data include location, time, and measurement device, as well as network information like RAT, throughput, Round Trip Time (RTT), and signal strength. For this performance evaluation, only the RAT coverage and the active measurements of the mobile and Wi-Fi networks are used. The mobile measurements were filtered to only one large mobile network operator to exclude the impact of different backbone networks on the measurements, which results in 4436 Fourth Generation Mobile Telecommunications (4G), 1043 3G, and 23 Second Generation Mobile Telecommunications (2G) throughput measurements. 173 Wi-Fi measurements were conducted in a variety of different public networks. To investigate the performance of private Wi-Fi offloading, the average connection speeds of fixed broadband connections in Germany during the time of the mobile measurements were obtained from [320]. Figure 4.13 shows the CDF of the throughputs for each RAT. It can be seen that 2G (General Packet Radio Service (GPRS), Enhanced Data Rates for GSM Evolution (EDGE)) has the worst download rates. The best performance is reached by 4G (LTE), although some measurements seem to be limited to 20 Mbps. 3G (UMTS, High Speed Packet Access (HSPA)) data rates are lower than 4G, but similar to fixed broadband throughputs, while public Wi-Fi only shows a performance between 2G and 3G.

The signals of Wi-Fi access points in the inner city of Darmstadt (area: ~1.5 km^2) were measured by [321]. The locations of the hotspots can be interpolated from the observed Wi-Fi beacons at street level. In total, 1971 hotspot locations were obtained. Additionally, way points in the same area can be re-

---

[26]`https://play.google.com/store/apps/details?id=de.tudarmstadt.`
`networkcoverage` [Online] – Accessed: 24.08.2017
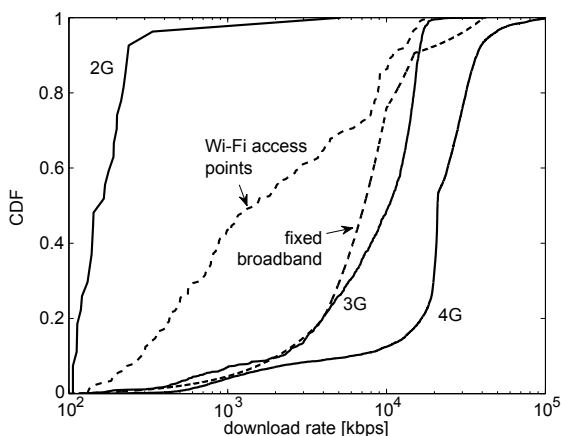
Figure 4.13: Throughput of mobile connections for different RATs in Darmstadt and residential fixed broadband connections.

trieved from OpenStreetMap[27], where users indicated intersections of streets, buildings, facilities, local businesses, or sights. These way points provide a good approximation for the locations of mobile users in this area. The access point and way point locations are visualized in Figure 4.14.

The power consumption of the Google Nexus 5 smartphone was modeled in [322]. This model was based on indoor measurements with stable device configuration (e.g., display brightness, active components), good network availability, good signal strength, and different data rates, using the voltage and current sensors of the device. The model can be used in this study to evaluate the energy consumption of individual connections, which is composed of the idle power consumption of the network interface, the cost for network connection and disconnection, and the energy expense per transferred byte. Thereby, the energy

---

[27]http://www.openstreetmap.org/export#map=14/49.8788/8.6628 [Online] – Accessed: 24.08.2017
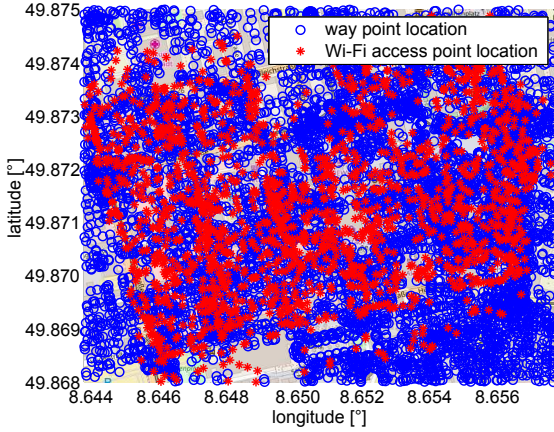
Figure 4.14: Way point and Wi-Fi access point locations in the inner city of Darmstadt. Map source: OpenStreetMap.

consumption is expected to be similar for indoor and outdoor communication, as the power consumption of commonly used power amplifiers does not depend on the output power.

Figure 4.15 shows the power consumption of an exemplary data transfer using 3G connectivity. At time 0, the device connects to the network (ramp state). When the connection is set up (second vertical line), the data transmission begins. Afterwards, the device disconnects from the network (tail state, third and fourth vertical line), and returns to idle state. The measurements indicated that ramp and tail durations and energies can be considered as constants for each RAT. Thus, the energy consumption of a continuous data burst i_j with volume v_{i_j} can be computed as

$$E_{i_j} = E_{\text{ramp}} + \frac{v_{i_j}}{\rho_{i_j}} \cdot P(\rho_{i_j}) + E_{\text{tail}}, \tag{4.6}$$
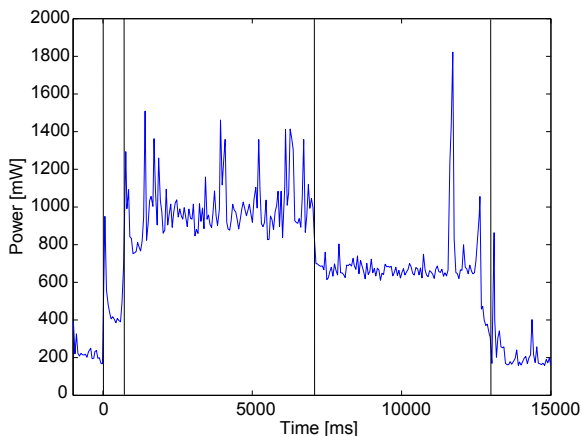
Figure 4.15: Exemplary data transfer using the Google Nexus 5 on 3G. The vertical lines indicate the start and end of the ramp and tail state.

depending on the interface power consumption $P(\rho_{i_j})$ during a data transmission with rate ρ_{i_j}. Figure 4.16 shows the transmission costs per data rate for each RAT based on the power models of the Google Nexus 5. It can be seen that Wi-Fi connections have the lowest power consumption, followed by the 3G and 4G. Note that no higher data rates than 900 kBps could be achieved with 3G. The energy consumption of an individual connection i can be computed by adding the energy consumption of all data bursts. If the interval between two bursts i_j and i_{j+1} is smaller than the ramp and tail durations, the bursts can be combined, such that the ramp and tail energies are only added once. This corresponds to the bearer release timeout in cellular networks.
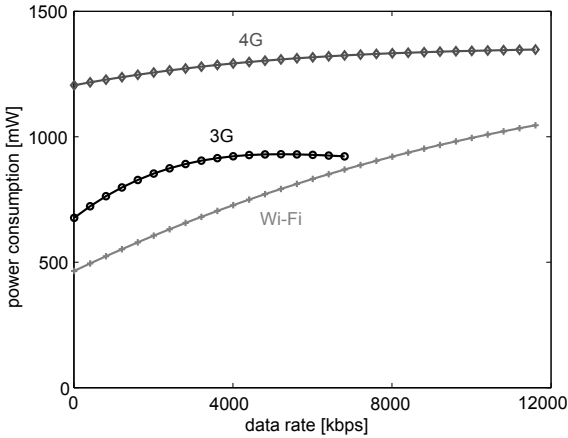
Figure 4.16: Power consumption of the interfaces of the Google Nexus 5 at different data rates.

4.3.2 Simulation Framework

The simulation considers the set of access point locations A and way point locations W. The indicator function $\chi_r : W \times A \mapsto \{0, 1\}$ returns 1 if way point $w \in W$ is in transmission range of an access point $a \in A$, depending on the global transmission range r. Additionally, for each access point, a Bernoulli trial with probability p (global sharing probability) determines whether the access point is shared, i.e., mobile connections can be offloaded to this access point. The shared access points are added to the subset $A_s \subset A$.

In each simulation run, $n = 10000$ mobile video requests are generated according to a Poisson process with rate $\lambda = 1$, i.e., one request per second. The mean baseline bit rate b_i^b and the video duration d_i are determined according to the measured characteristics of mobile YouTube videos (*itag*36 format, i.e., 240p). To account for HAS, a bit rate factor f is introduced, which indicates the

ratio of the current resolution and the 240p baseline (e.g., $\frac{1920x1080}{320x240} = 27$ for 1080p). The simulation considers the typical YouTube resolutions with the following bit rate factors: 240p: 1, 360p: 2.25, 480p: 4, 720p: 12, and 1080p: 27. Thus, the current video bit rate b_i can change according to the selected quality, such that $b_i = f \cdot b_i^b$.

For each request $1 \leq i \leq n$, a random way point $w_i \in W$ is drawn. The request can be offloaded if a shared Wi-Fi access point is in range, i.e., $\exists a \in A_s : \chi_r(w_i, a) = 1$, otherwise it will be served by the cellular network with 2G, 3G, or 4G access according to the measured RAT probabilities. The average throughput ρ_i of each request depends on the RAT and is determined randomly from the measured throughput distributions. To simulate fluctuating network conditions, the actual bandwidth is periodically modified and follows an exponential distribution with mean ρ_i.

YouTube currently employs a streaming behavior based on thresholds of the playback buffer [5]. The main idea is to download data only until the buffered playtime exceeds an upper threshold β to minimize the amount of unnecessarily downloaded video data if the streaming is aborted. If the buffered playtime drops below a lower threshold α, the download of video data resumes to avoid imminent stalling. In this evaluation, the used thresholds are $\alpha = 30\,\mathrm{s}$ and $\beta = 100\,\mathrm{s}$. The framework also has to take into account the playback and quality adaptation. Therefore, the playback stops if the buffer is empty, and it starts or resumes (after the initial delay or a stalling event) when the buffer exceeds the lower threshold α. To simulate HAS, a simple adaptation logic has to be implemented to adapt the current bit rate $b_i(t) = f \cdot b_i^b$ to the network conditions. Every time the bandwidth $\rho_i(t)$ changes, the implemented conservative adaptation logic sets the bit rate factor to the highest factor f, such that $\rho_i(t) > b_i^b \cdot f$ (i.e., the buffer increases), or $f = 1$ else. The remaining download volume of the video is updated accordingly.

The framework consists of a DES implemented in MATLAB. It utilizes four streaming-related events $(\overline{dl_i}, \overline{play_i})$, $(dl_i, \overline{play_i})$, $(dl_i, play_i)$, $(\overline{dl_i}, play_i)$, which are shown in Figure 4.17, and a periodical $(change\_\rho_i)$ event. The video

request i starts at t_i with the $(\overline{dl_i}, \overline{play_i})$ event, in which the baseline video bit rate b_i^b, video duration d_i, way point, RAT, and throughput ρ_i are determined. The events calls a $(dl_i, \overline{play_i})$ event at the same time t_i, at which the download of the first traffic burst i_1 immediately starts. The $(dl_i, \overline{play_i})$ event indicates that video data are downloaded, but the video is not played back. Four cases have to be considered to compute the times of the subsequent streaming events:
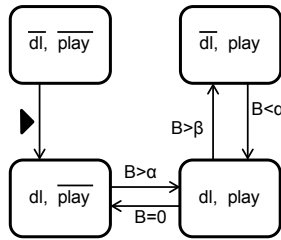
Figure 4.17: *State diagram of the video streaming process (i.e., downloading dl and playback $play$) depending on buffered playtime B and the burst thresholds α and β.*

a) Case $\rho_{i_j}(t) > b_i(t)$:

If $\rho_{i_j}(t) > b_i(t)$, the download rate is larger than the bit rate and the buffer increases, such that the upper threshold β will be reached. Then, the burst download can be divided into three phases Δt, $\Delta t'$, and $\Delta t''$, which are depicted in Figure 4.18. The figure shows the buffered playtime B and the two thresholds on the y-axis. The time is plotted on the x-axis.

At the start of the first phase, the video buffer is empty (initial delay or stalling), and the video is downloaded with rate $\frac{d}{dt} = \rho_{i_j}(t)$. The video playback starts when the buffered playtime B exceeds the threshold α
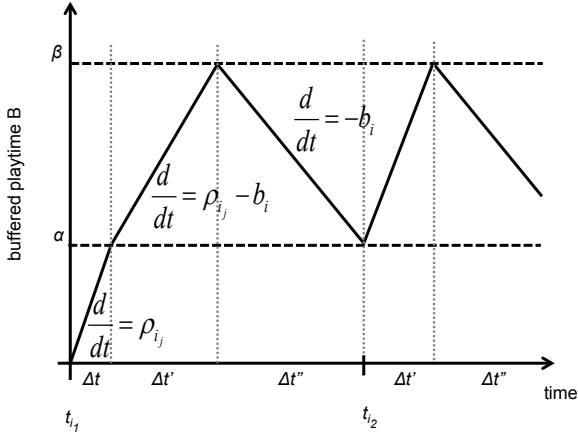
Figure 4.18: Playback buffer in case of $\rho_{i_j}(t) > b_i(t)$.

or the video is completely downloaded. The playback start is initiated by the $(dl_i, play_i)$ event after time Δt, i.e., $\alpha \cdot b_i(t) = \rho_{i_j}(t) \cdot \Delta t$, and

$$\Delta t = \min\left(\frac{\alpha \cdot b_i(t)}{\rho_{i_j}(t)}, \frac{v_i(t) - \sum_{k=1}^{j-1} v_{i_k}}{\rho_{i_j}(t)}\right). \tag{4.7}$$

To minimize the negative effect of initial delay on the QoE of the video streaming session, the start quality is set to the highest bit rate factor f, such that, for the current bandwidth $\rho_{i_1}(t)$ and the selected bit rate $b_i(t) = f \cdot b_i^b$, the initial delay is expected to be below 5 s, or $f = 1$ else.

The burst continues until the upper threshold β is reached or the video is completely downloaded, which is triggered by a $(\overline{dl_i}, play_i)$ event. In this phase, the gradient of the buffer is $\frac{d}{dt} = \rho_{i_j}(t) - b_i(t)$, and the

downloaded data volume is either $(\beta - \alpha) \cdot b_i(t)$, or the remaining video data. Thus,

$$\Delta t' = \min(\frac{(\beta - \alpha)) \cdot b_i(t)}{\rho_{i_j}(t) - b_i(t)}, \frac{v_i(t) - (\alpha \cdot b_i(t))}{\rho_{i_j}(t)}, \frac{v_i(t) - \sum_{k=1}^{j-1} v_{i_k}}{\rho_{i_j}(t)}).$$
(4.8)

If the video is not completely downloaded, a third phase has to be considered, in which no data is downloaded until the buffer drops below α. The playback decreases the buffer with gradient $\frac{d}{dt} = -b_i(t)$, thus, $(\alpha - \beta) \cdot b_i(t) = -b_i(t) \cdot \Delta t''$, and

$$\Delta t'' = \frac{(\alpha - \beta) \cdot b_i(t)}{-b_i(t)} = \beta - \alpha.$$
(4.9)

This phase determines the time of the next $(dl_i, play_i)$ event, i.e., the start of the next burst i_{j+1}. Afterwards, the $\Delta t'$ and $\Delta t''$ phases alternate until the video is completely downloaded.

b) Case $\rho_{i_j}(t) \leq b_i^b$:

If $\rho_{i_j}(t) \leq b_i^b$, and thus, $f = 1$, the received data rate cannot increase the playback buffer. Hence, the upper threshold β will not be reached, and the download of the video data does not pause. Whenever the playback of the video starts, the buffer of the video will eventually deplete, which results in periodical stalling. In this case, there are two different download phases, Δt and $\Delta t'$, which are depicted in Figure 4.19.

At the start of the first phase, the video buffer is empty (initial delay or stalling), and subsequently, phases Δt and $\Delta t'$ alternate until the video is completely downloaded. The bursts in each phase have to be considered as individual bursts for computational reasons, but have to be combined later. During the initial delay or stalling, the video is not played back until
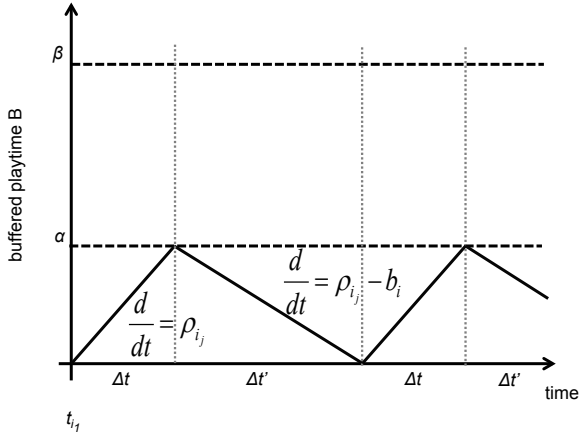
Figure 4.19: Playback buffer in case of $\rho_{i_j}(t) > b_i(t)$.

the buffer is filled up to the threshold α, for which a $(dl_i, play_i)$ event is scheduled after time Δt:

$$\Delta t = \min\left(\frac{\alpha \cdot b_i^b}{\rho_{i_j}(t)}, \frac{v_i(t) - \sum_{k=1}^{j-1} v_{i_k}}{\rho_{i_j}(t)}\right). \tag{4.10}$$

Then, the playback resumes, which will lead to further stalling, i.e., $(dl_i, \overline{play_i})$, after time $\Delta t'$:

$$\Delta t' = \min\left(\frac{\alpha \cdot b_i^b}{b_i^b - \rho_{i_j}(t)}, \frac{v_i(t) - \sum_{k=1}^{j-1} v_{i_k}}{\rho_{i_j}(t)}\right). \tag{4.11}$$

As the download of video data continues throughout the two phases, the back-to-back bursts eventually have to be combined into a single burst.

c) Bandwidth Changes:

Apart from the described streaming events, the $(change\_\rho_i)$ event is scheduled periodically every second for every request. When the event is processed, the new current bandwidth of request i, $\rho_i(t)$, is determined randomly according to an exponential distribution with mean ρ_i. The adaptation logic sets f to the highest value, such that $\rho_i(t) > b_i^b \cdot f$ (i.e., the buffer increases), or $f = 1$ otherwise. The current bit rate $b_i(t)$, the volume of the video $v_i(t)$, the volume of the current burst i_j, and the computed time of appearance of the subsequent streaming event are updated. Moreover, the quality adaptation is monitored in the simulation framework. If the bandwidth changes within one of the basic cases, i.e., Case a) or b), the respective equations can be slightly modified to compute the next event. However, if the new bandwidth changes the cases either from Case a) to b), or vice versa, a transition phase Δt^* might have to be included.

d) Transition Phases:

If the bandwidth $\rho_{i_j}(t)$ drops below b_i^b, f is set to 1 and the transition from Case a) to b) has to be considered. If the transition occurs in phase Δt, the updated time of appearance of the $(dl_i, play_i)$ event can be computed from the corresponding phase equation in Case b). If it occurs in phase $\Delta t''$, the $(dl_i, play_i)$ event does not need to be updated. In both situations, the phase $\Delta t'$ follows. In case the transition occurs in phase $\Delta t'$, the ongoing burst from Case a) is continued in a transition phase with time Δt^* and in subsequent phases. In the transition phase, the buffer will completely deplete, which results in a stalling event $(dl_i, \overline{play_i})$, or the completion of the video download.

$$\Delta t^* = \min\left(\frac{B \cdot b_i^b}{b_i^b - \rho_{i_j}(t)}, \frac{v_i(t) - \sum_{k=1}^{j-1} v_{i_k}}{\rho_{i_j}(t)}\right). \qquad (4.12)$$

155

In the opposite case of a transition from Case b) to a), i.e., a bandwidth increase above b_i^b, two situations have to be distinguished. If the transitions occurs in phase Δt, again the updated $(dl_i, play_i)$ event can be computed from the corresponding phase equation of Case a). In case it occurs in phase $\Delta t'$, the burst continues for Δt^* until the upper threshold β is reached and phase $\Delta t''$ starts triggered by a $(\overline{dl_i}, play_i)$ event, or the download is completed.

$$\Delta t^* = \min(\frac{(\beta - B)) \cdot b_i(t)}{\rho_{i_j}(t) - b_i(t)}, \frac{v_i(t) - \sum_{k=1}^{j-1} v_{i_k}}{\rho_{i_j}(t)}). \qquad (4.13)$$

The implemented events and the discussed phases allow to obtain all QoE factors of HAS for a video session i (i.e., initial delay, stalling, quality adaptation), as well as the energy consumption based on the arrival times t_{i_j} and volumes v_{i_j} of each burst i_j.

4.3.3 Performance Evaluation Results

The simulation framework was first applied to evaluate Wi-Fi offloading for classical video streaming, i.e., video streaming without adaptation. The results, which are described in [44, 47], showed that the QoE and energy consumption were slightly worse if the flow was offloaded to public Wi-Fi hotspots due to lower throughput compared to 3G and 4G. The following evaluation uses the same approach and simulation framework, but investigates the performance of Wi-Fi offloading for HAS, thereby considering both offloading to public and private Wi-Fi hotspots. The presented mean results are based on five simulation runs per test condition, which proved to be sufficient to obtain small confidence intervals. The results for initial delay, stalling, and adaptation are presented, but no mapping to QoE is shown, as no holistic QoE model for HAS has been widely accepted, which considers all these QoE factors.

First, the baseline results for the QoE of HAS are shown for all RATs (2G, 3G, 4G, public Wi-Fi) according to the measured throughput distributions. Fig-

ure 4.20a depicts the mean and 95% confidence interval of the initial delays of HAS sessions for all access technologies. The highest mean initial delay of around 40 s occurs for 2G access, which has the lowest throughput. The other RATs have lower mean initial delays below 10 s. These delays were considered negligible QoE degradations because people are accustomed to some delays [118]. Figure 4.20b shows the mean and 95% confidence interval of the number of stalling events per streaming session. For public Wi-Fi, 3G, and 4G, HAS is able to almost completely avoid stalling, which is the worst QoE degradation [63]. Only for 2G, due to the too low throughput, the mean number of stalling events is significantly higher around 2.3.

HAS trades off stalling by switching to a reduced image quality. The CDF of the number of quality changes per second is shown in Figure 4.20c. 4G access results a constant video quality for around 75% of the videos. However, an almost linear increase until the maximum switching frequency of 0.6 switches per second can be observed, which is a big QoE degradation for long videos [138]. 3G and public Wi-Fi show a similar behavior but the ratio of constant quality HAS sessions is lower at around 40% and 20%, respectively. Also around 20% of 2G sessions have no quality switches, but the CDF increases faster, which indicates that adaptation occurs more rarely. The corresponding time percentages on each layer are shown in Figure 4.20d. With 2G access, video streams have a resolution of 240p most of the time, but no higher resolutions for most videos, such that the QoE of users is poor. The other RATs allow that High Definition (HD) resolution (720p and 1080p) is streamed for a substantial percentage of time. Thereby, the HD percentage is around 20% for Wi-Fi, around 50% for 3G, and around 85% for 4G. This shows that these access technologies better support video streaming, as they provide a higher quality layer for a longer time, which results in higher QoE. Note that the video resolutions of Wi-Fi are generally smaller than 3G or 4G due to lower throughput.

The performance of Wi-Fi offloading for HAS will be evaluated depending on the transmission range of hotspots and the Wi-Fi sharing probability, i.e., the probability that a Wi-Fi hotspot can be accessed. The sharing probability can

(a) Initial delay.

(b) Number of stalling events.

(c) Frequency of quality switches.

(d) Time on quality layer (video resolution).
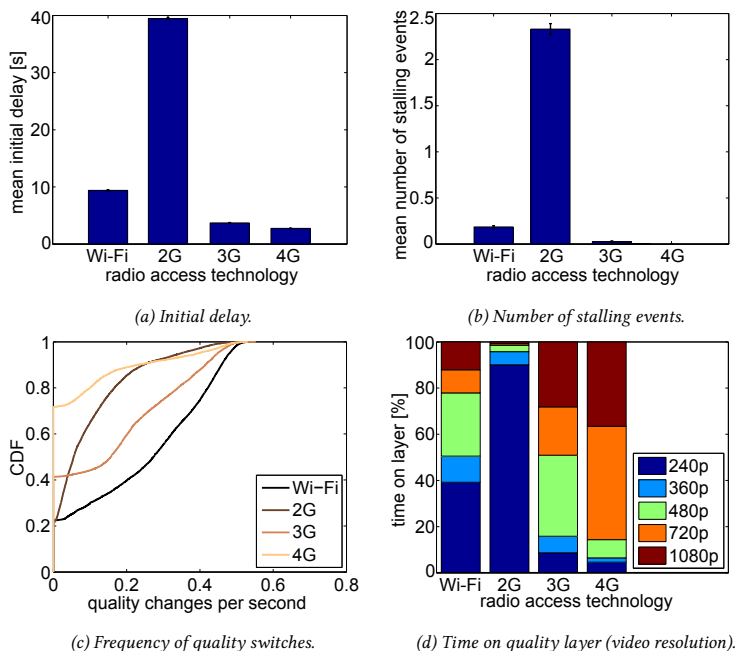
Figure 4.20: Key performance indicators for QoE of mobile HAS depending on network interface.

account for different Wi-Fi access, such as open access to public infrastructure, contracted access to hotspots of a single provider, or private sharing of Wi-Fi networks. In terms of the Wi-Fi backhaul links, not only the measured throughput distribution of public Wi-Fi infrastructure will be considered but also the throughput of residential broadband connections to also account for offloading to private hotspots (e.g., HORST, Wi-Fi sharing community). As 2G access was rarely observed in the coverage measurements and it shows a significantly

worse performance than the other cellular RATs, the corresponding results will not be presented.

(a) Number of stalling events.

(b) Percentage of time on highest quality (1080p).
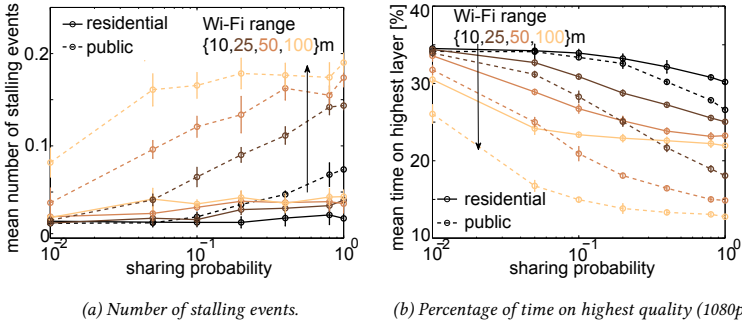
Figure 4.21: *QoE indicators of HAS for public and residential Wi-Fi offloading scenarios.*

First, the mean number of stalling events over all HAS sessions in a run is evaluated. Figure 4.21a depicts the mean and 95% confidence intervals over all simulation runs. The x-axis shows the different Wi-Fi sharing probabilities. The different colors indicate different Wi-Fi transmission ranges from 10 m (black) to 100 m (yellow). The dashed lines connect the results for public Wi-Fi offloading, i.e., using the measured throughput distribution of the Darmstadt Wi-Fi networks, while the solid lines connect the results for residential Wi-Fi offloading, i.e., using the throughput distribution of fixed broadband connections in Germany. It can be seen for public Wi-Fi offloading that the mean number of stalling events increases when the sharing probability is larger and more sessions can be offloaded to a hotspot. The same effect can be observed for increasing Wi-Fi range. The reason for both effects is that the throughput of public Wi-Fi networks is worse compared to the mobile throughput of 3G and 4G. Thus, when the sharing probability or Wi-Fi range increases, more connections can be offloaded, but thereby, receive less bandwidth, such that stalling is more likely. For residential Wi-Fi offloading, the throughput distribution is similar to 3G.

Because of the higher throughput, stalling can almost completely be avoided when offloading to residential hotspots, and the QoE of HAS is not deteriorated.

Figure 4.21b shows the mean percentage of time on the highest quality layer, i.e., 1080p, and 95% confidence intervals. For public Wi-Fi offloading, less time is spent on the highest quality layer, when the sharing probability or Wi-Fi range increase. Also the percentages of time on all other quality layers approach the Wi-Fi baseline (cf. Figure 4.20d), and thus, the QoE of HAS decreases. Considering residential Wi-Fi offloading, the time on highest layer also decreases when more HAS sessions are offloaded. However, due to the higher throughput, a higher percentage can be achieve, which results in a better QoE than public Wi-Fi offloading.

(a) Impact of Wi-Fi sharing probability.　　(b) Impact of Wi-Fi hotspot range.
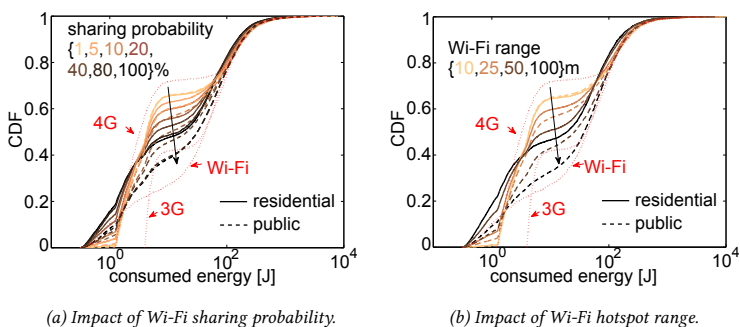
Figure 4.22: Energy consumption of HAS for public and residential Wi-Fi offloading scenarios.

The power consumption of the smartphone is analyzed in Figure 4.22a. It shows the CDFs of the baseline power consumption of HAS sessions with public Wi-Fi, 3G, and 4G in red. It can be seen that the consumed energy of public Wi-Fi is lower than 4G in 15%, and lower than 3G in 25% of the HAS sessions. For the offloading scenarios, again, public (dashed) and residential (solid) Wi-Fi offloading is distinguished. The sharing probability varies from 1% (yellow) to 100% (black), while the Wi-Fi range is set to 25 m. When the sharing proba-

bility increases, the energy consumption of public Wi-Fi offloading approaches the Wi-Fi baseline. This means, the number of sessions with very little energy consumption increases, but for most sessions more energy is consumed. The reason is again the lower throughput, and thus, longer transmissions from the public Wi-Fi hotspots. Note that for a sharing probability of 100%, the CDF does not overlap the baseline CDF because not all sessions can be offloaded, e.g., if no Wi-Fi hotspot is in range. In case of residential Wi-Fi offloading, the energy consumption decreases due to higher throughput than public Wi-Fi offloading. Moreover, it can be seen that less energy is consumed compared to the baseline 3G access although both have comparable access speeds. Figure 4.22b depicts the power consumption for different Wi-Fi ranges from 10 m (yellow) to 100 m (black) for a fixed sharing probability of 10%. A similar behavior is visible, i.e., the energy consumption generally deteriorates when the Wi-Fi range increases, and thus, more sessions are offloaded.

To sum up, the results showed that HAS is well suited for mobile usage. For 3G and 4G, almost all stalling could be avoided and the videos could be watched at least 50% of the time in HD resolution (720p or 1080p). Thereby, 4G proved to be the best RAT for HAS due to the highest throughput, which results in longer times on high quality layers. Due to the high throughput, also the energy consumption is lowest because shorter transmission bursts are needed to download the video with 4G access.

The utilization of Wi-Fi offloading deteriorated the QoE and energy consumption of end users due to lower throughput at public Wi-Fi hotspots. Thus, although mobile network operators benefit from Wi-Fi offloading, end users currently do not. The evaluation of Wi-Fi offloading to residential hotspots showed that the QoE and energy consumption were better compared to public Wi-Fi offloading, and also to 3G access, which had a comparable throughput distribution. Thus, Wi-Fi offloading still can be attractive for end users, however, operators have to ensure that similar or higher throughputs can be achieved in Wi-Fi networks than in cellular networks.

4.3.4 Applicability of Simple Hotspot Model for the Performance Evaluation of Public Wi-Fi Offloading

In Section 4.2.1, a simple model for generating hotspot distributions in cities was presented. In this section, its applicability is revisited for the performance evaluation of public Wi-Fi offloading. Therefore, the simulation runs for public Wi-Fi offloading in Darmstadt are repeated with a hotspot distribution, which was generated for Darmstadt according to the simple hotspot location model.
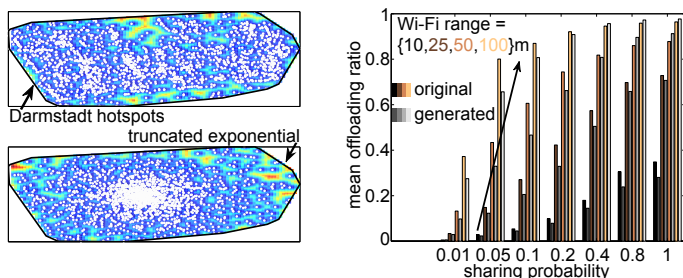
(a) *Original and generated hotspot distribution for Darmstadt.*

(b) *Ratio of sessions offloaded to Wi-Fi for original and generated hotspot distribution.*

Figure 4.23: *Evaluation of applicability of Wi-Fi hotspot location model to Wi-Fi offloading.*

Figure 4.23a visualizes the original (top) and a generated (bottom) hotspot distributions for Darmstadt. White dots indicate the locations of hotspots, and the colors indicate the hotspot proximity from close (blue) to far (red). Generated hotspot locations outside the convex hull of the original hotspot locations (black) were rejected, and the resulting hotspot distribution has the same number of hotspots as the original distribution, i.e., 1971 hotspots. At first sight, both distributions do not show the same spatial characteristics, which confirms the findings from Section 4.2.1. Figure 4.23b compares the offloading ratio, i.e., the mean ratio of sessions offloaded to Wi-Fi hotspots. The x-axis shows different

Wi-Fi sharing probabilities. The bars indicate the offloading ratio for different Wi-Fi ranges, while colored bars represent the original hotspot distribution and gray bars represent the generated hotspot distribution. It can be seen that the corresponding offloading ratios for the original and the generated distribution are very close especially for high sharing probability and high Wi-Fi range. For low sharing probabilities, i.e., in situations when HAS sessions can rarely be offloaded, the locations of the few available hotspots have a big impact on the offloading ratio, which can cause some differences. Nevertheless, the observed differences are not larger than 10%.

(a) Mean number of stalling events.

(b) Mean percentage of time on highest layer.
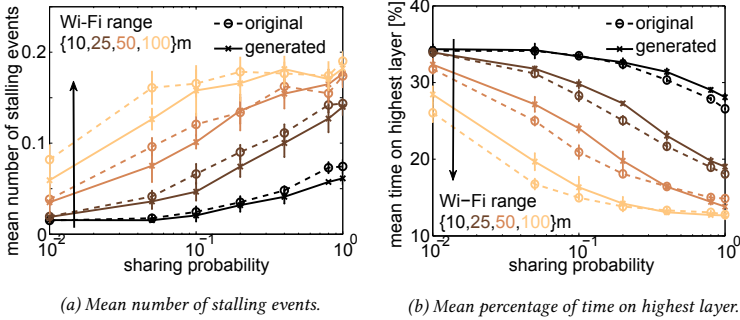
Figure 4.24: Comparison of QoE results for original and generated hotspot distribution.

Figure 4.24 compares the QoE results for both hotspot distributions. Thereby, it resembles Figure 4.21. In Figure 4.24a, the dashed lines connect the mean number of stalling events for the original hotspot distribution, and the solid lines connect the corresponding results for the generated hotspot distribution. It can be seen that the results for both hotspot distributions are very close, and the 95% confidence intervals overlap in most cases. Also the results for the time on highest layer, which are presented in Figure 4.24b confirm the high similarity between the original and the generated hotspot distribution. Although the con-

fidence intervals rarely overlap, the mean time percentages on highest layer do not differ more than 5%.

The application of the generated hotspot distribution for the performance evaluation of public Wi-Fi offloading confirms the practical value of the simple hotspot location model. The obtained results about offloading ratio and the QoE of HAS sessions only showed small deviations from the original hotspot distribution. Nevertheless, if the actual locations of Wi-Fi hotspots are available, they should be used for the performance evaluation. However, the model is general and can be applied to arbitrary cities to generate hotspot distributions, e.g., if the actual Wi-Fi hotspot locations are not available, or a future hotspot distribution should be evaluated. In these scenarios, the usage of generated hotspot distributions for the performance evaluation of public Wi-Fi offloading will provide a meaningful approximation to the results, which could be obtained with the real, unknown hotspot distributions.

4.4 Lessons Learned

This chapter investigated SATM for improving the QoE of HAS. SATM utilizes ubiquitous social signals for improved traffic management of Internet services. It is a highly collaborative and cross-layer approach involving many stakeholders, and thus, has a huge potential for applications. As not all applications can be covered comprehensively, the focus of this chapter was on Wi-Fi offloading. Wi-Fi offloading provides a complementary Internet access over a fixed network to reduce the load on mobile networks. The applicability of Wi-Fi offloading can benefit from social information, such as information about location and mobility of users to select appropriate public hotspots, or information about trust between users to share private Wi-Fi networks. A simple hotspot location model for the performance evaluation of public Wi-Fi offloading was designed and its applicability was confirmed. A socially-aware home router platform for private Wi-Fi offloading and video content distribution was presented. Finally, the im-

pact of public and private Wi-Fi offloading on the QoE and energy consumption of HAS sessions was investigated in a simulation study.

The performance of Wi-Fi offloading in cities depends on the locations and coverage of Wi-Fi hotspots. To find a general model for such hotspot distributions, the characteristics of public Wi-Fi infrastructure in cities were analyzed. After a transformation to polar coordinates with respect to the city center, a uniform distribution of the angle and an exponential distribution of the distance was fitted. A simple model for generating hotspot distributions for arbitrary cities could be derived. Although the generated hotspot distributions cannot accurately recreate the spatial patterns of real hotspot locations, they could replicate the offloading potential, coverage, or signal strength in a city. Thus, the simple model can be used to generate hotspot distributions for the performance evaluation of Wi-Fi offloading in scenarios, for which real hotspot locations are not available. A performance evaluation study for public Wi-Fi offloading with a generated hotspot distribution showed similar results to the study with the original hotspot distribution, and thus, confirmed the applicability of the simple hotspot location model.

Apart from to public hotspots, also private Wi-Fi infrastructure can be utilized for SATM. However, the sharing of private Wi-Fi networks requires a trust relationship to the owner to avoid abuse by malicious users. To easily enable private Wi-Fi offloading among trusted users, the HORST mechanism was developed, which additionally supports caching, prefetching, and content delivery on the home router. Therefore, HORST consists of a firmware for home routers, an OSN app, and a mobile device app. The firmware hosts SATM mechanisms on the private home router, which leverage social information from the OSN app and the mobile app to improve the QoE of HAS. A HORST prototype was implemented, evaluated, and demonstrated, which allowed users to request access at a friend's HORST home router, receive the credentials for the private Wi-Fi, connect to it, and access the Internet. This proved the feasibility of HORST to increase the Wi-Fi offloading potential. The utilization of other SATM mecha-

nisms, i.e., caching, prefetching, and content delivery, for improved QoE of HAS provided additional incentives to offload with HORST.

The performance evaluation of public Wi-Fi offloading for HAS in the city of Darmstadt, Germany, showed that QoE and smartphone energy consumption of offloaded HAS sessions deteriorated. The reason is the lower throughput of public Wi-Fi hotspots compared to 3G and 4G access. The analysis of offloading to residential Wi-Fi hotspots, which had a fixed broadband connection and a throughput distribution similar to 3G, indicated it is possible to improve the energy consumption compared to public Wi-Fi offloading, and also to 3G access. Thereby, the results came closer to 4G access, which allowed for the best HAS performance in terms of QoE and energy consumption. The results showed that Wi-Fi offloading of HAS sessions is only beneficial for end users if the received bandwidth in the Wi-Fi network is not lower than in the mobile network. This means, the throughput and coverage of public Wi-Fi infrastructure has to keep up with the increasing 4G coverage if improved QoE and energy consumption shall be incentives to encourage offloading. This has to be taken into account by operators, which offer public Wi-Fi as an alternative Internet access. For the moment, private Wi-Fi offloading, e.g., using HORST, is a more promising approach to improve the QoE and energy consumption of HAS, and to reduce the load on cellular networks, especially taking the increasing speeds of residential fixed broadband connections into account [323].

5 Conclusion

This thesis focused on QoE and access network traffic management of HAS. In the following, the thesis is summarized, especially considering the research questions of Section 1.1, which are highlighted in bold. The major findings for each research question are presented and conclusions are discussed. Finally, an outlook to future research on the basis of this thesis completes this monograph.

5.1 Summary and Contributions

Video streaming is one of the most popular and demanding applications of today's Internet. The high data volumes, bandwidth requirements, and delay constraints of video traffic pose a lot of challenges to ISPs, which want to deliver the traffic as efficiently as possible, while maintaining a high subjectively perceived service quality. To measure the satisfaction of end users with a networked service, the concept of Quality of Experience (QoE) has been established. QoE research aims to identify QoE factors on different layers, which have a high correlation to the subjectively perceived quality. Ideally, all QoE factors of a service can be integrated into a holistic QoE model, which allows to quantify the satisfaction of end users based on the values of the identified QoE factors.

The current HTTP Adaptive Video Streaming (HAS) technology allows to align the streaming demands to the network conditions by adapting the video bit rate. However, the impact of quality adaptation on the QoE has not been comprehensively investigated yet. Therefore, this thesis reviewed the previous QoE results and **investigated the impact of adaptation on the QoE of HAS** in subjective crowdsourced QoE studies. It was found that only the time on each

layer and its respective image quality impacted the subjectively perceived quality, while the other adaptation-related parameters, such as number of quality changes, did not show a significant effect. Based on these results better adaptation logics can be designed, which reach a higher QoE by maximizing the time on high quality layers first, instead of focusing on low switching frequency or conservative up-switching behavior. These results were also used to formulate an optimization problem as a linear program, which can be solved to compute the QoE-optimal adaptation strategy for given throughput conditions. This framework allows to benchmark and compare HAS adaptation logics in terms of distance to the optimally achievable QoE. Video streaming service providers can use this benchmark to find the best adaptation logics for realistic network conditions, in which their service is typically consumed.

The identification of the most important QoE factors of HAS allowed to **develop monitoring approaches for these factors on network and adaptation layer**. Temporal pooling of objective per-frame or per-segment metrics, which were collected by DPI, reached high correlations to subjective MOS values. ISPs can deploy such monitoring in their networks to keep track of the QoE of their customers in real-time. As only little information about the underlying content is used, the concept of temporal pooling might also be applied in the context of end-to-end encrypted video traffic. Irrespective of end-to-end encryption, application-layer monitoring can be implemented as it monitors the QoE factors directly at the client application. An Android app was developed to monitor the video streaming by injecting JavaScript monitoring functions into the website, which contains the video player element. Additionally, the app allows to log network and device statistics, which can also be used to estimate the QoE. Finally, the app cannot only be used for QoE monitoring, but it can also collect subjective quality ratings of users. Such an unobtrusive monitoring app can be provided by video services or network operators to their customers to gain a full understanding of all QoE factors perceived by the end user. Additionally, subjective QoE studies can be conducted with the app, which makes it a valuable tool for researching the QoE of HAS.

The monitored information is beneficial as input to QoE-aware traffic management solutions. This thesis **investigated different resource allocation strategies for video streaming and web browsing flows on a shared bottleneck link**. This scenario relates to access networks where multiple users and applications typically share a single backhaul link. Analytical and simulative performance evaluations were conducted and identified the benefits of QoE-aware traffic management and the trade-offs between the different resource allocation strategies. As a joint QoE improvement was possible for both applications compared to the best effort scenario without traffic management, ISPs have to invest into identifying the right traffic management strategy for the specific application mix in their networks. Moreover, it was found that dynamically considering application-layer information for traffic management has advantages over pure network-based strategies. This further confirms the potential of collaborative and cross-layer approaches for QoE-aware traffic management, such as application-aware networks and network-aware applications. Given these findings, network operators and service providers have to agree on standard interfaces to exchange application and network information, which allows for an improved QoE-aware traffic management.

Considering the end user, his shared resources, and ubiquitous social information about users, their interests, or their interactions with other users can further enhance QoE-aware traffic management. The concept of Socially-aware Traffic Management (SATM) was introduced and the most important stakeholders were identified. These stakeholders can collaborate to deliver a personalized service to the end user to reach a high satisfaction with the service. Therefore, all members of the video service delivery chain should identify and bring together the relevant stakeholders for implementing SATM. In particular, they have to investigate what social signals they can harvest and how they can benefit from social information. Then again, standard interfaces have to be defined to exchange social information among the different stakeholders for improved services and an improved traffic management.

Due to increasing user mobility, which puts huge loads on cellular networks, growing public Wi-Fi infrastructure, and the prevalence of residential Wi-Fi networks in home environments, **Wi-Fi offloading was considered as a SATM mechanism to improve the QoE of HAS**. To evaluate the potential for public Wi-Fi offloading, the distribution of Wi-Fi hotspot locations was modeled. The developed simple model reached a high accuracy in terms of offloading potential, coverage, or signal strength. It could be successfully applied in a real performance evaluation study, in which the generated hotspot distribution gave similar results to the original hotspot distribution. Based on this simple model, Wi-Fi hotspot location distributions can be generated in arbitrary cities, for which real hotspot locations are not available. This allows to design and evaluate the performance of new SATM mechanisms based on Wi-Fi offloading. Nevertheless, the simple model had limited applicability in case the spatial collocation of hotspots has to be reconstructed. This thesis also presented a trust-based SATM framework for home routers, which allows to improve the QoE of HAS by offering Wi-Fi offloading, caching, prefetching, and content delivery. As home routers are typically provided by ISPs to their customers, this SATM platform can be easily deployed to share private Wi-Fi networks, improve traffic management, and reach a higher QoE. Finally, the performance of Wi-Fi offloading of mobile HAS sessions was investigated in terms of QoE and energy efficiency. It was found that the utilization of Wi-Fi offloading deteriorated the QoE factors and energy consumption of end users due to lower throughput at public Wi-Fi hotspots. However, if a comparable throughput distribution is available at the Wi-Fi hotspots, e.g., in residential environments with fixed broadband connections, both subjectively perceived metrics can be improved. Based on these findings, in order to encourage end users to offload in terms of QoE and energy consumption, the capacity of the available public Wi-Fi infrastructure has to be increased to provide comparable throughput like current mobile networks. Moreover, the sharing of private Wi-Fi infrastructure has to be facilitated and incentivized, e.g., by utilizing the proposed SATM framework for home routers.

5.2 Outlook

Continuing the work of this thesis, future research should strive to even better understand the QoE of HAS. The previous QoE results and the findings of this thesis have to be combined to eventually develop a holistic QoE model for HAS. Therefore, more subjective studies will be needed to investigate HAS systems with more or different quality layers, video lengths, and contents. Also the impact of non-technical influence factors of QoE, such as human and context influence factors [82], has to be addressed. Based on the findings of these studies, new monitoring concepts have to be developed to cover all relevant QoE factors. The trend towards end-to-end encryption will restrain the usage of current network monitoring mechanisms. It has to be investigated if established methods, like temporal pooling, can be transferred to encrypted traffic. The performance of recent trends, such as the estimation of QoE from statistical features of the encrypted traffic with machine learning, or QoE monitoring with VNFs, have to be evaluated and improved.

QoE-aware traffic management has to leverage new networking technologies, such as SDN, which allows to flexibly program the network behavior. This leads to more possibilities for QoE monitoring and traffic management in the network. As this thesis identified their gain, the implementation of cross-layer and collaborative approaches with new networking technologies has to be tackled to benefit from additional information and take enhanced traffic management decision. Moreover, the trend towards edge and fog computing has to be integrated into traffic management solutions. Edge resources of ISPs or end users can be utilized to host data and services in the proximity of end users, thereby reducing delays and mitigating load on core networks and data centers. Therefore, the concept of the trust-based SATM platform on home routers, which was presented in this thesis, can be extended to include also other edge and fog devices. The mobility of end users can be supported when virtualized services or service components can be dynamically migrated or instantiated. The orchestration and consolidation of such services and service chains on edge resources

has to target a high cost and energy efficiency, and also a high QoE of end users. As many traffic management solutions currently only focus on the QoE of single users, new mechanisms have to be developed, which also consider the QoE of multiple users. This requires more research on QoE fairness and coordinated traffic management actions.

More research is also needed in the field of SATM. This thesis showed, for the example of utilizing private and public Wi-Fi infrastructure, that end users were able to participate in SATM and that both ISPs and end users could benefit. Thus, in future works, Wi-Fi offloading has to be leveraged and also other mechanisms need to be investigated. Ultimately, different approaches and mechanisms should be combined to create synergies, which result in improved traffic management on all layers. Therefore, it is necessary to define and implement interfaces for the collaboration of stakeholders in terms of exchange of information and coordination of traffic management actions. With respect to social information, it has to be researched how relevant social information can be identified and best exploited. Due to the vast amount of data, e.g., in OSNs, sampling techniques, big data analytics, or deep learning might be utilized. Furthermore, models for social interactions are needed, which could be used for improved traffic management decisions. First attempts, e.g., to model the social cascades of video sharing in OSNs, have to be refined, and new models have to be developed. Finally, privacy issues have to be considered to prevent the abuse of sensitive personal data, which might require additional legal regulations. Nevertheless, SATM foreshadows a huge potential for improving service delivery and QoE, which still has to be uncovered and popularized.

The above presented outlook touches only the obvious future work, for which this monograph can be used as a starting point to launch new research activities. It is intentionally modest, as disruptive technology trends and groundbreaking research milestones might change the challenges, goals, and prospects of research at any time. At the time of writing this thesis, the presented findings on the relevant QoE factors of HAS, and the conducted studies on QoE-aware and socially-aware traffic management approaches help to enhance the satisfaction

of end users with video streaming services in many ways. This thesis allows service providers to design and benchmark improved HAS adaptation logics. Moreover, it offers ISPs concepts for QoE monitoring on different layers and for QoE-aware traffic management on bottleneck links, such as in access networks. Finally, it shows how end users can participate and benefit from SATM on the example of offloading to Wi-Fi infrastructure, which is prevalent in homes and public spaces. Beyond that, the general concepts and methodologies, which were discussed throughout this thesis, are not limited to specific technologies and services, but can still be applied when the QoE of other or future networked services is to be investigated and improved.

Acronyms

2G Second Generation Mobile Telecommunications.

3G Third Generation Mobile Telecommunications.

3GP 3GPP File Format.

3GPP Third Generation Partnership Project.

4G Fourth Generation Mobile Telecommunications.

AAC Advanced Audio Coding.

ACR Absolute Category Rating.

ADSL Asymmetric Digital Subscriber Line.

ANOVA Analysis of Variance.

API Application Programming Interface.

AS Autonomous System.

BIEB Bandwidth Independent Efficient Buffering Adaptation Logic.

BSSID Basic Service Set Identifier.

CBR Constant Bit Rate.

CDF Cumulative Distribution Function.

CDN Content Delivery Network.

COST European Cooperation in Science and Technology.

COST Qualinet European Network on Quality of Experience in Multimedia Systems and Services.

DASH Dynamic Adaptive Streaming over HTTP.

DBF Dynamic Bandwidth Allocation for Each Video Flow.

DCR Degradation Category Rating.

DES Discrete Event Simulation.

DFG Deutsche Forschungsgemeinschaft.

DFG Crowdsourcing Design und Bewertung neuer Mechanismen für Crowdsourcing als neue Form der Arbeitsorganisation im Internet.

DFG ÖkoNet Entwurf und Bewertung neuer Mechanismen für das Internet der Zukunft - Neue Paradigmen und ökonomische Aspekte.

DFG QoE-DZ Analyse und Optimierung des Trade-offs zwischen QoE und Energieeffizienz in Datenzentren.

DHT Distributed Hash Table.

DiffServ Differentiated Services.

DOM Document Object Model.

DPI Deep Packet Inspection.

EDGE Enhanced Data Rates for GSM Evolution.

EU European Union.

EU FP7 SmartenIT Socially-aware Management of New Overlay Application Traffic combined with Energy Efficiency in the Internet.

EU H2020 INPUT In-network Programmability for Next-generation Personal Cloud Service Support.

EU H2020 MONROE Measuring Mobile Broadband Networks in Europe.

EU H2020 MONROE/Mobi-QoE Monitoring and Analysis of Quality of Experience in Mobile Broadband Networks.

FBF Fixed Bandwidth Allocation for Each Video Flow.

FBV Fixed Bandwidth Allocation for All Videos Flows.

FP7 Framework Programme 7.

fps Frames per Second.

FTW Forschungszentrum Telekommunikation Wien.

GPRS General Packet Radio Service.

GPS Global Positioning System.

H.264/AVC H.264 Advanced Video Coding.

H.264/SVC H.264 Scalable Video Coding.

H.265/HEVC H.265 High Efficiency Video Coding.

H2020 Horizon 2020.

HAS HTTP Adaptive Video Streaming.

HD High Definition.

HDS Adobe HTTP Dynamic Streaming.

HLS Apple HTTP Live Streaming.

HORST Home Router Sharing based on Trust.

HSPA High Speed Packet Access.

HTML Hypertext Markup Language.

HTML5 Hypertext Markup Language Version 5.

HTTP Hypertext Transfer Protocol.

IEEE Institute of Electrical and Electronics Engineers.

IP Internet Protocol.

IPTV Internet Protocol Television.

IQX Exponential Interdependency of QoE and QoS.

ISP Internet Service Provider.

KLU Klagenfurt University Adaptation Logic.

LRU Least Recently Used.

LTE Long Term Evolution.

M/M/1-DPS M/M/1 Queue with Discriminatory Processor Sharing Policy.

M/M/1-GPS M/M/1 Queue with Generalized Processor Sharing Policy.

M/M/1-PS M/M/1 Queue with Processor Sharing Policy.

MAC Media Access Control.

MCNKP Multiple-choice Nested Knapsack Problem.

MILP Mixed Integer Linear Program.

MOS Mean Opinion Score.

MPEG Moving Picture Experts Group.

MSS Microsoft Silverlight Smooth Streaming.

NAS Network-attached Storage.

NFV Network Function Virtualization.

NGMN Next Generation Mobile Networks.

NGMN P-SERQU Service Quality Definition and Measurement.

OSI Open Systems Interconnection.

OSN Online Social Network.

P2P Peer-to-peer.

PBCC Point-biserial Correlation Coefficient.

PLCC Pearson Linear Correlation Coefficient.

QoE Quality of Experience.

QoS Quality of Service.

QP Quantization Parameter.

RAT Radio Access Technology.

REST Representational State Transfer.

RTT Round Trip Time.

SAND Server and Network Assisted DASH.

SATM Socially-aware Traffic Management.

SBC Single-board Computer.

SD Secure Digital.

SDN Software-defined Networking.

SLA Service-level Agreement.

SROCC Spearman Rank Order Correlation Coefficient.

SSID Service Set Identifier.

SSIM Structural Similarity.

STB Set-top Box.

TCP Transmission Control Protocol.

TRI Tribler-based Adaptation Logic.

TUB Technical University Berlin Adaptation Logic.

UMTS Universal Mobile Telecommunications System.

UNaDa User-owned Nano Data Center.

URL Uniform Resource Locator.

USB Universal Serial Bus.

VAP Virtual Access Point.

VBR Variable Bit Rate.

VCEG Video Coding Experts Group.

VNF Virtual Network Function.

WBV Weighted Bandwidth Allocation for All Videos Flows.

WLAN Wireless Local Area Network.

WPS Wi-Fi Protected Setup.

Bibliography and References

Bibliography of the Author

Book Chapters

[1] M. Seufert, G. Darzanos, I. Papafili, R. Łapacz, V. Burger, and T. Hoßfeld, "Socially-Aware Traffic Management", in *Socioinformatics – The Social Impact of Interactions between Humans and IT*, K. Zweig, W. Neuser, V. Pipek, M. Rohde, and I. Scholtes, Eds., Springer International Publishing, 2014, pp. 25–43.

[2] V. Burger, D. Hock, I. Scholtes, T. Hoßfeld, D. Garcia, and M. Seufert, "Social Network Analysis in the Enterprise: Challenges and Opportunities", in *Socioinformatics – The Social Impact of Interactions between Humans and IT*, K. Zweig, W. Neuser, V. Pipek, M. Rohde, and I. Scholtes, Eds., Springer International Publishing, 2014, pp. 95–120.

Journal Papers

[3] C. Metter, M. Seufert, F. Wamser, T. Zinner, and P. Tran-Gia, "Analytical Model for SDN Signaling Traffic and Flow Table Occupancy and its Application for Various Types of Traffic", *IEEE Transactions on Network and Service Management*, 2017, in press.

[4] T. Zinner, S. Geissler, S. Lange, S. Gebert, M. Seufert, and P. Tran-Gia, "A Discrete-Time Model for Optimizing the Processing Time of Virtualized Network Functions", *Computer Networks*, 2017, in press.

[5]   F. Wamser, P. Casas, M. Seufert, C. Moldovan, P. Tran-Gia, and T. Hoßfeld, "Modeling the YouTube Stack: from Packets to Quality of Experience", *Computer Networks*, vol. 109, no. 2, pp. 211–224, 2016.

[6]   M. Seufert, V. Burger, K. Lorey, A. Seith, F. Loh, and P. Tran-Gia, "Assessment of Subjective Influence and Trust with an Online Social Network Game", *Computers in Human Behavior*, vol. 64, pp. 233–246, 2016.

[7]   S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnström, and N. García, "Perceptual Quality of HTTP Adaptive Streaming Strategies: Cross-Experimental Analysis of Multi-Laboratory and Crowdsourced Subjective Studies", *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 8, pp. 2141–2153, 2016.

[8]   V. Burger, M. Seufert, T. Hoßfeld, and P. Tran-Gia, "Performance Evaluation of Backhaul Bandwidth Aggregation Using a Partial Sharing Scheme", *Physical Communication*, vol. 19, pp. 135–144, 2016.

[9]   P. Casas, M. Seufert, F. Wamser, B. Gardlo, A. Sackl, and R. Schatz, "Next to You: Monitoring Quality of Experience in Cellular Networks from the End-devices", *IEEE Transactions on Network and Service Management*, vol. 13, no. 2, pp. 181–196, 2016.

[10] M. Seufert, T. Griepentrog, V. Burger, and T. Hoßfeld, "A Simple WiFi Hotspot Model for Cities", *IEEE Communications Letters*, vol. 20, no. 2, pp. 384–387, 2016.

[11] M. Seufert, S. Lange, and T. Hoßfeld, "More than Topology: Joint Topology and Attribute Sampling and Generation of Social Network Graphs", *Computer Communications*, vol. 73, no. B, pp. 176–187, 2016.

[12] F. Wamser, A. Blenk, M. Seufert, T. Zinner, W. Kellerer, and P. Tran-Gia, "Modeling and Performance Analysis of Application-Aware Resource Management", *International Journal of Network Management*, vol. 25, no. 4, pp. 223–241, 2015.

[13]   T. Hoßfeld, M. Seufert, C. Sieber, T. Zinner, and P. Tran-Gia, "Identifying QoE Optimal Adaptation of HTTP Adaptive Streaming Based on Subjective Studies", *Computer Networks*, vol. 81, pp. 320–332, 2015.

[14]   M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A Survey on Quality of Experience of HTTP Adaptive Streaming", *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.

[15]   T. Hoßfeld, M. Seufert, C. Sieber, T. Zinner, and P. Tran-Gia, "Close to Optimum? User-centric Evaluation of Adaptation Logics for HTTP Adaptive Streaming", *PIK - Praxis der Informationsverarbeitung und Kommunikation*, vol. 37, pp. 275–285, 2014.

[16]   P. Casas, M. Seufert, and R. Schatz, "YOUQMON: A System for On-line Monitoring of YouTube QoE in Operational 3G Networks", *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 2, pp. 44–46, 2013.

[17]   F. Wamser, D. Hock, M. Seufert, B. Staehle, R. Pries, and P. Tran-Gia, "Using Buffered Playtime for QoE-Oriented Resource Management of YouTube Video Streaming", *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 3, pp. 288–302, 2013.

[18]   D. Hock, F. Wamser, M. Seufert, R. Pries, and P. Tran-Gia, "OC^2E^2AN: Optimized Control Center for Experience Enhancements in Access Networks", *PIK - Praxis der Informationsverarbeitung und Kommunikation*, vol. 36, p. 40, 2013.

Conference Papers

[19]   M. Seufert, S. Lange, and M. Meixner, "Automated Decision Making based on Pareto Frontiers in the Context of Service Placement in Networks", in *Proceedings of the 29th International Teletraffic Congress (ITC)*, Genoa, Italy, 2017.

[20] F. Wamser, S. Höfner, M. Seufert, and P. Tran-Gia, "Client-side Dynamic Server Selection for MPEG DASH Video Streaming", in *Proceedings of the ACM SIGCOMM Workshop on QoE-based Analysis and Management of Data Communication Networks (Internet-QoE)*, Los Angeles, CA, USA, 2017.

[21] M. Seufert, B. Kamneng Kwam, F. Wamser, and P. Tran-Gia, "EdgeNetworkCloudSim: Placement of Service Chains in Edge Clouds Using NetworkCloudSim", in *Proceedings of the 1st IEEE International Workshop on Network Programmability – From the Data Center to the Ground (NetFoG)*, Bologna, Italy, 2017.

[22] A. Schwind, M. Seufert, Ö. Alay, P. Casas, P. Tran-Gia, and F. Wamser, "Concept and Implementation of Video QoE Measurements in a Mobile Broadband Testbed", in *Proceedings of the IEEE/IFIP Workshop on Mobile Network Measurement (MNM)*, Dublin, Ireland, 2017.

[23] M. Seufert, O. Zach, M. Slanina, and P. Tran-Gia, "Unperturbed Video Streaming QoE Under Web Page Related Context Factors", in *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, Germany, 2017.

[24] M. Seufert, N. Wehner, F. Wamser, P. Casas, A. D'Alconzo, and P. Tran-Gia, "Unsupervised QoE Field Study for Mobile YouTube Video Streaming with YoMoApp", in *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, Germany, 2017.

[25] P. Casas, A. D'Alconzo, F. Wamser, M. Seufert, B. Gardlo, A. Schwind, P. Tran-Gia, and R. Schatz, "Predicting QoE in Cellular Networks using Machine Learning and in-Smartphone Measurements", in *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, Germany, 2017.

[26] V. Burger, M. Seufert, T. Zinner, and P. Tran-Gia, "An Approximation of the Backhaul Bandwidth Aggregation Potential Using a Partial Sharing

Scheme", in *Proceedings of the 15th IFIP/IEEE International Symposium on Integrated Network Management (IM)*, Lisbon, Portugal, 2017.

[27] L. Dinh-Xuan, M. Seufert, F. Wamser, and P. Tran-Gia, "Study on the Accuracy of QoE Monitoring for HTTP Adaptive Video Streaming Using VNF", in *Proceedings of the 1st IFIP/IEEE International Workshop on Quality of Experience Management (QoE-Management)*, Lisbon, Portugal, 2017.

[28] O. Zach, M. Seufert, M. Hirth, M. Slanina, and P. Tran-Gia, "On Use of Crowdsourcing for H.264/AVC and H.265/HEVC Video Quality Evaluation", in *Proceedings of Radioelektronika*, Brno, Czech Republic, 2017.

[29] M. Seufert, V. Burger, and F. Kaup, "Evaluating the Impact of WiFi Offloading on Mobile Users of HTTP Adaptive Video Streaming", in *Proceedings of the 5th IEEE International Workshop on Quality of Experience for Multimedia Communications (QoEMC)*, Washington, DC, USA, 2016.

[30] C. Metter, M. Seufert, F. Wamser, T. Zinner, and P. Tran-Gia, "Analytic Model for SDN Controller Traffic and Switch Table Occupancy", in *Proceedings of the 12th International Conference on Network and Service Management (CNSM)*, Best Paper Award, Montreal, Canada, 2016.

[31] M. Seufert, S. Lange, and M. Meixner, "Automated Decision Making Methods for the Multi-objective Optimization Task of Cloud Service Placement", in *Proceedings of the 1st International Workshop on Programmability for Cloud Networks and Applications (PROCON)*, Würzburg, Germany, 2016.

[32] M. Seufert and T. Hoßfeld, "One Shot Crowdtesting: Approaching the Extremes of Crowdsourced Subjective Quality Testing", in *Proceedings of the 5th ISCA/DEGA Workshop on Perceptual Quality of Systems (PQS)*, Berlin, Germany, 2016.

[33]   F. Wamser, M. Seufert, S. Höfner, and P. Tran-Gia, "Concept for Client-initiated Selection of Cloud Instances for Improving QoE of Distributed Cloud Services", in *Proceedings of the ACM SIGCOMM Workshop on QoE-based Analysis and Management of Data Communication Networks (Internet-QoE)*, Florianópolis, Brazil, 2016.

[34]   M. Seufert, P. Casas, F. Wamser, N. Wehner, R. Schatz, and P. Tran-Gia, "Application-Layer Monitoring of QoE Parameters for Mobile YouTube Video Streaming in the Field", in *Proceedings of the 6th IEEE International Conference on Communications and Electronics (ICCE)*, Ha Long, Vietnam, 2016.

[35]   L. Dinh-Xuan, M. Seufert, F. Wamser, and P. Tran-Gia, "QoE Aware Placement of Content in Edge Networks on the Example of a Photo Album Cloud Service", in *Proceedings of the 6th IEEE International Conference on Communications and Electronics (ICCE)*, Ha Long, Vietnam, 2016.

[36]   M. Seufert, O. Zach, T. Hoßfeld, M. Slanina, and P. Tran-Gia, "Impact of Test Condition Selection in Adaptive Crowdsourcing Studies on Subjective Quality", in *Proceedings of the 8th International Conference on Quality of Multimedia Experience (QoMEX)*, Lisbon, Portugal, 2016.

[37]   M. Seufert, T. Hoßfeld, A. Schwind, V. Burger, and P. Tran-Gia, "Group-based Communication in WhatsApp", in *Proceedings of the 1st IFIP Internet of People Workshop (IoP)*, Vienna, Austria, 2016.

[38]   V. Burger, J. F. Pajo, O. R. Sanchez, M. Seufert, C. Schwartz, F. Wamser, F. Davoli, and P. Tran-Gia, "Load Dynamics of a Multiplayer Online Battle Area and Simulative Assessment of Edge Server Placements", in *Proceedings of the ACM Multimedia Systems Conference (MMSys)*, Klagenfurt, Austria, 2016.

[39]   M. Seufert, T. Hoßfeld, and C. Sieber, "Impact of Intermediate Layer on Quality of Experience of HTTP Adaptive Streaming", in *Proceedings of*

the 11th International Conference on Network and Service Management (CNSM), Barcelona, Spain, 2015.

[40]   P. Casas, B. Gardlo, M. Seufert, F. Wamser, and R. Schatz, "Taming QoE in Cellular Networks: from Subjective Lab Studies to Measurements in the Field", in *Proceedings of the 11th International Conference on Network and Service Management (CNSM)*, Barcelona, Spain, 2015.

[41]   V. Burger, G. Darzanos, I. Papafili, and M. Seufert, "Trade-Off between QoE and Operational Cost in Edge Resource Supported Video Streaming", in *Proceedings of the 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, Krakow, Poland, 2015.

[42]   A. Lareida, G. Petropoulos, V. Burger, M. Seufert, S. Soursos, and B. Stiller, "Augmenting Home Routers for Socially-Aware Traffic Management", in *Proceedings of the 40th Annual IEEE Conference on Local Computer Networks (LCN)*, Clearwater Beach, FL, USA, 2015.

[43]   M. Seufert, A. Schwind, T. Hoßfeld, and P. Tran-Gia, "Analysis of Group-based Communication in WhatsApp", in *Proceedings of the 7th EAI International Conference on Mobile Networks and Management (MONAMI)*, Santander, Spain, 2015.

[44]   V. Burger, F. Kaup, M. Seufert, M. Wichtlhuber, D. Hausheer, and P. Tran-Gia, "Energy Considerations for WiFi Offloading of Video Streaming", in *Proceedings of the 7th EAI International Conference on Mobile Networks and Management (MONAMI)*, Santander, Spain, 2015.

[45]   M. Seufert, F. Wamser, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "YouTube QoE on Mobile Devices: Subjective Analysis of Classical vs. Adaptive Video Streaming", in *Proceedings of the 6th International Workshop on Traffic Analysis and Characterization (TRAC)*, Dubrovnik, Croatia, 2015.

[46]   P. Casas, R. Schatz, F. Wamser, M. Seufert, and R. Irmer, "Exploring QoE in Cellular Networks: How Much Bandwidth do you Need for Popular Smartphone Apps?", in *Proceedings of the 5th ACM SIGCOMM Workshop on All Things Cellular: Operations, Applications and Challenges (ATC)*, London, UK, 2015.

[47]   V. Burger, M. Seufert, F. Kaup, M. Wichtlhuber, D. Hausheer, and P. Tran-Gia, "Impact of WiFi Offloading on Video Streaming QoE in Urban Environments", in *Proceedings of the IEEE Workshop on Quality of Experience-based Management for Future Internet Applications and Services (QoE-FI)*, London, UK, 2015.

[48]   F. Wamser, M. Seufert, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "YoMoApp: a Tool for Analyzing QoE of YouTube HTTP Adaptive Streaming in Mobile Networks", in *Proceedings of the European Conference on Networks and Communications (EuCNC)*, Paris, France, 2015.

[49]   M. Seufert, V. Burger, F. Wamser, P. Tran-Gia, C. Moldovan, and T. Hoßfeld, "Utilizing Home Router Caches to Augment CDNs toward Information-Centric Networking", in *Proceedings of the European Conference on Networks and Communications (EuCNC)*, Paris, France, 2015.

[50]   S. Egger, B. Gardlo, M. Seufert, and R. Schatz, "The Impact of Adaptation Strategies on Perceived Quality of HTTP Adaptive Streaming", in *Proceedings of the 1st Workshop on Design, Quality and Deployment of Adaptive Video Streaming (VideoNext)*, Sydney, Australia, 2014.

[51]   K. Wajda, R. Stankiewicz, Z. Dulinski, T. Hoßfeld, M. Seufert, D. Hausheer, M. Wichtlhuber, I. Papafili, M. Dramitinos, P. Cruschelli, S. Soursos, R. Lapacz, and B. Stiller, "Socially-aware Management of New Overlay Applications Traffic - The Optimization Potentials of the SmartenIT Approach", in *Proceedings of the 6th International Conference on Mobile Networks and Management (MONAMI)*, Würzburg, Germany, 2014.

[52] T. Hoßfeld, M. Seufert, C. Sieber, and T. Zinner, "Assessing Effect Sizes of Influence Factors Towards a QoE Model for HTTP Adaptive Streaming", in *Proceedings of the 6th International Workshop on Quality of Multimedia Experience (QoMEX)*, Singapore, 2014.

[53] I. Papafili, K. Wajda, R. Lapacz, A. Predieri, T. Bocek, and M. Seufert, "An Overview of Application Traffic Management Approaches: Challenges and Potential Extensions", in *Proceedings of the 8th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, Birmingham, UK, 2014.

[54] B. Gardlo, S. Egger, M. Seufert, and R. Schatz, "Crowdsourcing 2.0: Enhancing Execution Speed and Reliability of Web-based QoE Testing", in *Proceedings of the International Conference on Communications (ICC)*, Sydney, Australia, 2014.

[55] M. Seufert, K. Lorey, M. Hirth, and T. Hoßfeld, "Gamification Framework for Personalized Surveys on Relationships in Online Social Networks", in *Proceedings of the 1st International Workshop on Crowdsourcing and Gamification in the Cloud (CGCloud)*, Dresden, Germany, 2013.

[56] M. Seufert, V. Burger, and T. Hoßfeld, "HORST - Home Router Sharing based on Trust", in *Proceedings of the Workshop on Social-aware Economic Traffic Management for Overlay and Cloud Applications (SETM)*, Zurich, Switzerland, 2013.

[57] M. Seufert, G. Darzanos, V. Burger, I. Papafili, and T. Hoßfeld, "Socially-Aware Traffic Management", in *Proceedings of the Workshop Sozioinformatik*, Koblenz, Germany, 2013.

[58] V. Burger, T. Hoßfeld, D. Garcia, M. Seufert, I. Scholtes, and D. Hock, "Resilience in Enterprise Social Networks", in *Proceedings of the Workshop Sozioinformatik*, Koblenz, Germany, 2013.

[59]   A. Sackl, M. Seufert, and T. Hoßfeld, "Asking Costs Little? The Impact of Tasks in Video QoE Studies on User Behavior and User Ratings", in *Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS)*, Vienna, Austria, 2013.

[60]   P. Casas, M. Seufert, and R. Schatz, "YOUQMON: A System for On-line Monitoring of YouTube QoE in Operational 3G Networks", in *Proceedings of the 31st International Symposium on Computer Performance, Modeling, Measurements and Evaluation (IFIP Performance)*, Vienna, Austria, 2013.

[61]   M. Seufert, M. Slanina, S. Egger, and M. Kottkamp, "To Pool or not to Pool: A Comparison of Temporal Pooling Methods for HTTP Adaptive Video Streaming", in *Proceedings of the 5th International Workshop on Quality of Multimedia Experience (QoMEX)*, Klagenfurt, Austria, 2013.

[62]   P. Casas, M. Seufert, S. Egger, and R. Schatz, "Quality of Experience in Remote Virtual Desktop Services", in *Proceedings of the IFIP/IEEE International Workshop on Quality of Experience Centric Management (QCMan)*, Ghent, Belgium, 2013.

[63]   T. Hoßfeld, R. Schatz, M. Seufert, M. Hirth, T. Zinner, and P. Tran-Gia, "Quantification of YouTube QoE via Crowdsourcing", in *Proceedings of the International Workshop on Multimedia Quality of Experience - Modeling, Evaluation, and Directions (MQoE)*, Dana Point, CA, USA, 2011.

Software Demonstrations

[64]   F. Wamser, F. Loh, R. Bruschi, P. Lago, M. Seufert, and P. Tran-Gia, *Dynamic Cloud Service Placement for Live Video Streaming with a Remote-Controlled Drone*, 15th IFIP/IEEE International Symposium on Integrated Network Management (IM), Best Demonstration Award, Lisbon, Portugal, 2017.

[65]   G. Petropoulos, A. Lareida, S. Soursos, M. Seufert, V. Burger, and B. Stiller, *WiFi Offloading and Socially Aware Prefetching on Augmented Home Routers*, 40th Annual IEEE Conference on Local Computer Networks (LCN), Best Demonstration Award, Clearwater Beach, FL, USA, 2015.

[66]   M. Seufert, F. Wamser, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, *On the Monitoring of YouTube QoE in Cellular Networks from End-devices*, 7th Wireless of the Students, by the Students, for the Students Workshop (S3), Paris, France, 2015.

[67]   F. Wamser, M. Seufert, P. Tran-Gia, R. Irmer, P. Casas, and R. Schatz, *YoMoApp - Performance Monitoring of Mobile Networks*, European Conference on Networks and Communications (EuCNC), Paris, France, 2015.

[68]   F. Wamser, D. Hock, M. Seufert, T. Zinner, and P. Tran-Gia, *Demonstrating the Benefit of Joint Application and Network Control Within a Wireless Access Network*, IEEE Conference on Computer Communications (INFOCOM), Turin, Italy, 2013.

[69]   D. Hock, F. Wamser, M. Seufert, R. Pries, and P. Tran-Gia, *OC^2E^2AN: Optimized Control Center for Experience Enhancements in Access Networks*, Conference on Networked Systems (NetSys), Stuttgart, Germany, 2013.

[70]   F. Wamser, D. Hock, M. Seufert, R. Pries, and P. Tran-Gia, *Performance Optimization in Access Networks Using a Combined Control Strategy*, 12th Würzburg Workshop on IP: ITG Workshop 'Visions of Future Generation Networks' (Euroview), Würzburg, Germany, 2012.

Poster Presentations

[71]   M. Seufert, T. Hoßfeld, and C. Sieber, *Impact of Intermediate Layer on Quality of Experience of HTTP Adaptive Streaming.* 11th International

Conference on Network and Service Management (CNSM), Barcelona, Spain, 2015.

[72]   M. Seufert, F. Wamser, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, *On the Monitoring of YouTube QoE in Cellular Networks from End-devices*, 7th Wireless of the Students, by the Students, for the Students Workshop (S3), Paris, France, 2015.

[73]   F. Wamser, M. Seufert, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, *Understanding YouTube QoE in Cellular Networks with YoMoApp – a QoE Monitoring Tool for YouTube Mobile*, 21st Annual International Conference on Mobile Computing and Networking (MobiCom), 2015.

[74]   M. Seufert, V. Burger, and T. Hoßfeld, *HORST – Home Router Sharing based on Trust*, Workshop on Social-aware Economic Traffic Management for Overlay and Cloud Applications (SETM), Zurich, Switzerland, 2013.

[75]   F. Wamser, D. Hock, M. Seufert, R. Pries, and P. Tran-Gia, *Application and Quality of Experience Aware Resource Management in Wireless Access Networks*, EuroNF Summer School Modeling and Analysis of Novel Mechanisms in Future Internet Applications, Würzburg, Germany, 2012.

Technical Reports

[76]   M. Seufert, T. Hoßfeld, G. Sperb Machado, T. Bocek, M. Biancani, P. Cruschelli, R. Lapacz, G. Darzanos, I. Papafili, and K. Wajda, "A Survey of Cloud Services and Potential Applications of Social Awareness", University of Würzburg, Tech. Rep. 496, 2015.

[77]   T. Hoßfeld, M. Hirth, J. Redi, F. Mazza, P. Korshunov, B. Naderi, M. Seufert, B. Gardlo, S. Egger, and C. Keimel, "Best Practices and Recommendations for Crowdsourced QoE - Lessons learned from the Qualinet Task Force Crowdsourcing", COST Action IC1003 European Network on

Quality of Experience in Multimedia Systems and Services (QUALINET), Tech. Rep., 2014.

[78]   T. Hoßfeld, M. Seufert, C. Sieber, and T. Zinner, "Crowdsourced Subjective User Study Results on QoE Influence Factors of HTTP Adaptive Streaming", University of Würzburg, Tech. Rep. 491, 2014.

[79]   M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A Survey on Quality of Experience of HTTP Adaptive Streaming", University of Würzburg, Tech. Rep. 490, 2014.

[80]   M. Wennesheimer, D. Robinson, J. De Vriendt, D. De Vleeschauwer, C. Bahr, K. Heise, S. Argyropoulos, B. Feiten, M. Bilgic, T. Raiha, S. Egger, M. Seufert, M. Slanina, M. Kottkamp, H. Gsödl, H. Ibl, and R. García Pérez, "Service Quality Definition and Measurement", Next Generation Mobile Networks Alliance (NGMN), Tech. Rep., 2013, Version 1.0.4.

[81]   T. Hoßfeld, T. Zinner, R. Schatz, M. Seufert, and P. Tran-Gia, "Transport Protocol Influences on YouTube QoE", University of Würzburg, Tech. Rep. 482, 2011.

General References

[82]   P. Le Callet, S. Möller, and A. Perkis (eds), "Qualinet White Paper on Definitions of Quality of Experience", European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Tech. Rep., 2013, Version 1.2.

[83]   International Telecommunication Union, *ITU-T Recommendation E.800: Definitions of Terms Related to Quality of Service*, 2008.

[84]   ——, *ITU-T Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications*, 2008.

[85]   D. F. Brueck and M. B. Hurst, "Apparatus, System, and Method for Multi-bitrate Content Streaming", US7818444 B2, 2010.

[86]   A. Zambelli, "Smooth Streaming Technical Overview", Microsoft Corporation, Tech. Rep., 2009.

[87]   Adobe Systems Inc., "HTTP Dynamic Streaming on the Adobe Flash Platform", Adobe Systems Inc., Tech. Rep., 2010.

[88]   European Telecommunications Standard Institute (ETSI), *Universal Mobile Telecommunication System (UMTS); LTE; Transparent End-to-end Packet-switched Streaming Service (PSS); Protocols and Codecs (3GPP TS 26.234 version 9.1.0 Release 9)*, 2009.

[89]   ——, *Universal Mobile Telecommunication System (UMTS); LTE; Transparent End-to-end Packet-swiched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH) (3GPP TS 26.247 version 1.0.0 Release 10)*, 2010.

[90]   International Standards Organization/International Electrotechnical Commission (ISO/IEC), *23009-1:2012 Information Technology – Dynamic Adaptive Streaming over HTTP (DASH) – Part 1: Media Presentation Description and Segment Formats*, 2012.

[91]   ——, *14496-10:2012 Information Technology – Coding of Audio-visual Objects – Part 10: Advanced Video Coding*, 2012.

[92]   ——, *23008-2:2013 Information Technology – High Efficiency Coding and Media Delivery in Heterogeneous Environments – Part 2: High Efficiency Video Coding*, 2013.

[93]   H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[94]   M. Wien, H. Schwarz, and T. Oelbaum, "Performance Analysis of SVC", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1194–1203, 2007.

[95]   R. Gupta, A. Pulipaka, P. Seeling, L. Karam, and M. Reisslein, "H.264 Coarse Grain Scalable (CGS) and Medium Grain Scalable (MGS) Encoded Video: A Trace Based Traffic and Quality Evaluation", *IEEE Transactions on Broadcasting*, vol. 58, no. 3, pp. 428–439, 2012.

[96]   M. Grafl, C. Timmerer, H. Hellwagner, D. Negru, W. Cherif, and S. Battista, "Scalable Video Coding Guidelines and Performance Evaluations for Adaptive Media Delivery of High Definition Content", in *Proceedings of the IEEE Symposium on Computers and Communications (ISCC)*, Split, Croatia, 2013.

[97]   M. Grafl, C. Timmerer, H. Hellwagner, W. Cherif, and A. Ksentini, "Evaluation of Hybrid Scalable Video Coding for HTTP-based Adaptive Media Streaming with High-definition Content", in *Proceedings of the 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Madrid, Spain, 2013.

[98]   J. Famaey, S. Latre, N. Bouten, W. Van de Meerssche, B. De Vleeschauwer, W. Van Leekwijck, and F. De Turck, "On the Merits of SVC-based HTTP Adaptive Streaming", in *Proceedings of the 13th IFIP/IEEE International Symposium on Integrated Network Management (IM)*, Ghent, Belgium, 2013.

[99]   T. Arsan, "Review of Bandwidth Estimation Tools and Application to Bandwidth Adaptive Video Streaming", in *Proceedings of the 9th International Conference on High-Capacity Optical Networks and Emerging/Enabling Technologies (HONET)*, Istanbul, Turkey, 2012.

[100]   D. Yun, K. Chung, and J. Hong, "Efficient Bandwidth Estimation for HTTP Adaptive Streaming", in *Proceedings of the International Conference on Information Networking (ICOIN)*, Phuket, Thailand, 2014.

[101]   S. S. Chaudhari and R. C. Biradar, "Survey of Bandwidth Estimation Techniques in Communication Networks", *Wireless Personal Communications*, vol. 83, no. 2, pp. 1425–1476, 2015.

[102]   K. Lazic, I. Basicevic, and J. Kovacevic, "Bandwidth Estimation in Adaptive Video Streaming over HTTP", in *Proceedings of the 5th IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 2015.

[103]   C. Müller, S. Lederer, and C. Timmerer, "An Evaluation of Dynamic Adaptive Streaming over HTTP in Vehicular Environments", in *Proceedings of the 4th Workshop on Mobile Video (MoVID)*, Chapel Hill, NC, USA, 2012.

[104]   K. Miller, E. Quacchio, G. Gennari, and A. Wolisz, "Adaptation Algorithm for Adaptive Streaming over HTTP", in *Proceedings of the 19th International Packet Video Workshop (PV)*, Munich, Germany, 2012.

[105]   S. Oechsner, T. Zinner, J. Prokopetz, and T. Hoßfeld, "Supporting Scalable Video Codecs in a P2P Video-on-Demand Streaming System", in *Proceedings of the 21th ITC Specialist Seminar on Multimedia Applications – Traffic, Performance and QoE (ITC-SS21)*, Miyazaki, Japan, 2010.

[106]   J. A. Pouwelse, P. Garbacki, J. Wang, A. Bakker, J. Yang, A. Iosup, D. H. J. Epema, M. Reinders, M. R. van Steen, and H. J. Sips, "Tribler: A Social-based Peer-to-peer System", *Concurrency and Computation: Practice and Experience*, vol. 20, no. 2, pp. 127–138, 2008.

[107]   C. Sieber, T. Hoßfeld, T. Zinner, P. Tran-Gia, and C. Timmerer, "Implementation and User-centric Comparison of a Novel Adaptation Logic for DASH with SVC", in *Proceedings of the IFIP/IEEE International Workshop on Quality of Experience Centric Management (QCMan)*, Ghent, Belgium, 2013.

[108]   O. Oyman and S. Singh, "Quality of Experience for HTTP Adaptive Streaming Services", *IEEE Communications Magazine*, vol. 50, no. 4, pp. 20–27, 2012.

[109]   G. Tian and Y. Liu, "Towards Agile and Smooth Video Adaptation in Dynamic HTTP Streaming", in *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*, Nice, France, 2012.

[110]   T. Huang, R. Johari, and N. McKeown, "Downton Abbey Without the Hiccups", in *Proceedings of the ACM SIGCOMM Workshop on Future Human-centric Multimedia Networking (FhMN)*, Hong Kong, China, 2013.

[111]   L. De Cicco, V. Caldaralo, V. Palmisano, and S. Mascolo, "Elastic: A Client-side Controller for Dynamic Adaptive Streaming over HTTP (DASH)", in *Proceedings of the 20th International Packet Video Workshop (PV)*, San Jose, CA, USA, 2013.

[112]   Z. Li, X. Zhu, J. Gahm, R. Pan, H. Hu, A. C. Begen, and D. Oran, "Probe and Adapt: Rate Adaptation for HTTP Video Streaming at Scale", *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 4, pp. 719–733, 2014.

[113]   Z. Li, A. C. Begen, J. Gahm, Y. Shan, B. Osler, and D. Oran, "Streaming Video over HTTP with Consistent Quality", in *Proceedings of the 5th ACM Multimedia Systems Conference (MMSys)*, Singapore, 2014.

[114]   T.-Y. Huang, R. Johari, N. McKeown, M. Trunnell, and M. Watson, "A Buffer-based Approach to Rate Adaptation: Evidence from a Large Video Streaming Service", *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 187–198, 2015.

[115]   X. Yin, A. Jindal, V. Sekar, and B. Sinopoli, "A Control-theoretic Approach for Dynamic Adaptive Video Streaming over HTTP", *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4, pp. 325–338, 2015.

[116]   P. Juluri, V. Tamarapalli, and D. Medhi, "QoE Management in DASH Systems using the Segment Aware Rate Adaptation Algorithm", in *Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS)*, Istanbul, Turkey, 2016.

[117]   G. Cofano, L. De Cicco, and S. Mascolo, "Modeling and Design of Adaptive Video Streaming Control Systems", *IEEE Transactions on Control of Network Systems*, 2016, in press.

[118]   T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial Delay vs. Interruptions: Between the Devil and the Deep Blue Sea", in *Proceedings of the 4th International Workshop on Quality of Multimedia Experience (QoMEX)*, Yarra Valley, Australia, 2012.

[119]   S. Egger, P. Reichl, T. Hoßfeld, and R. Schatz, "Time is Bandwidth? Narrowing the Gap between Subjective Time Perception and Quality of Experience", in *Proceedings of the IEEE International Conference on Communications (ICC)*, Ottawa, Canada, 2012.

[120]   R. E. Kooij, A. Kamal, and K. Brunnström, "Perceived Quality of Channel Zapping", in *Proceedings of the Communication Systems and Networks (CSN)*, Palma de Mallorca, Spain, 2006.

[121]   S. Egger, T. Hoßfeld, R. Schatz, and M. Fiedler, "Waiting Times in Quality of Experience for Web Based Services", in *Proceedings of the 4th International Workshop on Quality of Multimedia Experience (QoMEX)*, Yarra Valley, Australia, 2012.

[122]   A. Sackl, S. Egger, and R. Schatz, "Where's the Music? Comparing the QoE Impact of Temporal Impairments Between Music and Video Streaming", in *Proceedings of the 5th International Workshop on Quality of Multimedia Experience (QoMEX)*, Klagenfurt, Austria, 2013.

[123]   T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, and L. Martens, "Quantifying the Influence of Rebuffering Interruptions on the User's

Quality of Experience During Mobile Video Watching", *IEEE Transactions on Broadcasting*, vol. 59, no. 1, pp. 47–61, 2013.

[124] M.-N. Garcia, D. Dytko, and A. Raake, "Quality Impact Due to Initial Loading, Stalling, and Video Bitrate in Progressive Download Video Services", in *Proceedings of the 6th International Workshop on Quality of Multimedia Experience (QoMEX)*, IEEE, Singapore, 2014.

[125] A. Finamore, M. Mellia, M. M. Munafò, R. Torres, and S. G. Rao, "YouTube Everywhere: Impact of Device and Infrastructure Synergies on User Experience", in *Proceedings of the Internet Measurement Conference (IMC)*, Berlin, Germany, 2011.

[126] L. Chen, Y. Zhou, and D. M. Chiu, "Video Browsing - A Study of User Behavior in Online VoD Services", in *Proceedings of the 22nd International Conference on Computer Communications and Networks (ICCCN)*, Nassau, Bahamas, 2013.

[127] S. S. Krishnan and R. K. Sitaraman, "Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-experimental Designs", *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 2001–2014, 2013.

[128] A. Raake and S. Egger, "Quality and Quality of Experience", in *Quality of Experience: Advanced Concepts, Applications and Methods*, S. Möller and A. Raake, Eds., Springer, 2014.

[129] Y. Qi and M. Dai, "The Effect of Frame Freezing and Frame Skipping on Video Quality", in *Proceedings of the 2nd International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Pasadena, CA, USA, 2006.

[130] T. N. Minhas and M. Fiedler, "Impact of Disturbance Locations on Video Quality of Experience", in *Proceedings of the 2nd Workshop of Quality of Experience for Multimedia Content Sharing (QoEMCS)*, Lisbon, Portugal, 2011.

[131]   Q. Huynh-Thu and M. Ghanbari, "Temporal Aspect of Perceived Quality in Mobile Video Broadcasting", *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 641–651, 2008.

[132]   K. D. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of Experience Estimation for Adaptive HTTP/TCP Video Streaming Using H. 264/AVC", in *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, USA, 2012.

[133]   D. Ghadiyaram, J. Pan, and A. C. Bovik, "A Time-varying Subjective Quality Model for Mobile Streaming Videos with Stalling Events", in *Proceedings of SPIE Applications of Digital Image Processing XXXVIII*, San Diego, CA, USA, 2015.

[134]   K. Zeng, H. Yeganeh, and Z. Wang, "Quality-of-experience of Streaming Video: Interactions between Presentation Quality and Playback Stalling", in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016.

[135]   J. Yao, S. S. Kanhere, I. Hossain, and M. Hassan, "Empirical Evaluation of HTTP Adaptive Streaming Under Vehicular Mobility", in *Proceedings of the 10th International IFIP TC 6 Networking Conference: Networking*, Valencia, Spain, 2011.

[136]   T. Zinner, T. Hoßfeld, T. N. Minash, and M. Fiedler, "Controlled vs. Uncontrolled Degradations of QoE – The Provisioning-Delivery Hysteresis in Case of Video", in *Proceedings of the 1st Workshop of Quality of Experience for Multimedia Content Sharing (QoEMCS)*, Tampere, Finland, 2010.

[137]   B. Lewcio, B. Belmudez, A. Mehmood, M. Wältermann, and S. Möller, "Video Quality in Next Generation Mobile Networks – Perception of Time-varying Transmission", in *Proceedings of the IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR)*, Naples, FL, USA, 2011.

[138]   P. Ni, R. Eg, A. Eichhorn, C. Griwodz, and P. Halvorsen, "Flicker Effects in Adaptive Video Streaming to Handheld Devices", in *Proceedings of the 19th ACM International Conference on Multimedia (MM)*, Scottsdale, AZ, USA, 2011.

[139]   M. Grafl and C. Timmerer, "Representation Switch Smoothing for Adaptive HTTP Streaming", in *Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS)*, Vienna, Austria, 2013.

[140]   M. Zink, J. Schmitt, and R. Steinmetz, "Layer-encoded Video in Scalable Adaptive Streaming", *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 75–84, 2005.

[141]   Y. Shen, Y. Liu, Q. Liu, and D. Yang, "A Method of QoE Evaluation for Adaptive Streaming based on Bitrate Distribution", in *Proceedings of the Workshop on QoE-centric Network and Application Management (QoE-NAM)*, Sydney, Australia, 2014.

[142]   A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. De Veciana, "Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies", *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 652–671, 2012.

[143]   J. De Vriendt, D. De Vleeschauwer, and D. Robinson, "Model for Estimating QoE of Video Delivered Using HTTP Adaptive Streaming", in *Proceedings of the 13th IFIP/IEEE International Symposium on Integrated Network Management (IM)*, Ghent, Belgium, 2013.

[144]   F. Wang, Z. Fei, J. Wang, Y. Liu, and Z. Wu, "HAS QoE Prediction Based on Dynamic Video Features with Data Mining in LTE Network", *Science China Information Sciences*, vol. 60, no. 4, p. 042 404, 2017.

[145]   D. Z. Rodríguez, Z. Wang, R. L. Rosa, and G. Bressan, "The Impact of Video-quality-level Switching on User Quality of Experience in Dynamic Adaptive Streaming over HTTP", *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 216, pp. 1–15, 2014.

[146]   H. T. Tran, T. Vu, N. P. Ngoc, and T. C. Thang, "A Novel Quality Model for HTTP Adaptive Streaming", in *Proceedings of the 6th IEEE International Conference on Communications and Electronics (ICCE)*, Ha Long, Vietnam, 2016.

[147]   M.-N. Garcia, F. De Simone, S. Tavakoli, N. Staelens, S. Egger, K. Brunnström, and A. Raake, "Quality of Experience and HTTP Adaptive Streaming: A Review of Subjective Studies", in *Proceedings of the 6th International Workshop on Quality of Multimedia Experience (QoMEX)*, Singapore, 2014.

[148]   C. Alberti, D. Renzi, C. Timmerer, C. Mueller, S. Lederer, S. Battista, and M. Mattavelli, "Automated QoE Evaluation of Dynamic Adaptive Streaming over HTTP", in *Proceedings of the 5th International Workshop on Quality of Multimedia Experience (QoMEX)*, Klagenfurt, Austria, 2013.

[149]   J. Xue, D.-Q. Zhang, H. Yu, and C. W. Chen, "Assessing Quality of Experience for Adaptive HTTP Video Streaming", in *Proceedings of the International Conference on Multimedia and Expo Workshops (ICMEW)*, Chengdu, China, 2014.

[150]   Y. Liu, S. Dey, F. Ulupinar, M. Luby, and Y. Mao, "Deriving and Validating User Experience Model for DASH Video Streaming", *IEEE Transactions on Broadcasting*, vol. 61, no. 4, pp. 651–665, 2015.

[151]   J. A. Redi, T. Hoßfeld, P. Korshunov, F. Mazza, I. Povoa, and C. Keimel, "Crowdsourcing-based Multimedia Subjective Evaluations: A Case Study on Image Recognizability and Aesthetic Appeal", in *Proceedings of the 2nd International Workshop on Crowdsourcing for Multimedia (CrowdMM)*, Barcelona, Spain, 2013.

[152]   M. Varela, T. Mäki, L. Skorin-Kapov, and T. Hoßfeld, "Increasing Payments in Crowdsourcing: Don't Look a Gift Horse in the Mouth", in *Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS)*, Vienna, Austria, 2013.

[153] M. Hirth, S. Scheuring, T. Hoßfeld, C. Schwartz, and P. Tran-Gia, "Predicting Result Quality in Crowdsourcing Using Application Layer Monitoring", in *Proceedings of the 5th International Conference on Communications and Electronics (ICCE)*, Da Nang, Vietnam, 2014.

[154] T. Hoßfeld, M. Hirth, P. Korshunov, P. Hanhart, B. Gardlo, C. Keimel, and C. Timmerer, "Survey of Web-based Crowdsourcing Frameworks for Subjective Quality Assessment", in *Proceedings of the 16th International Workshop on Multimedia Signal Processing (MMSP)*, Jakarta, Indonesia, 2014.

[155] T. Hoßfeld and C. Keimel, "Crowdsourcing in QoE Evaluation", in *Quality of Experience*, S. Möller and A. Raake, Eds., Springer, 2014, pp. 315–327.

[156] C. Keimel, J. Habigt, C. Horch, and K. Diepold, "QualityCrowd: A Framework for Crowd-based Quality Evaluation", in *Proceedings of the Picture Coding Symposium (PCS)*, Krakow, Poland, 2012.

[157] T. Zinner, O. Hohlfeld, O. Abboud, and T. Hoßfeld, "Impact of Frame Rate and Resolution on Objective QoE Metrics", in *Proceedings of the 2nd International Workshop on Quality of Multimedia Experience (QoMEX)*, Trondheim, Norway, 2010.

[158] L. Janowski and P. Romaniak, "QoE as a Function of Frame Rate and Resolution Changes", in *Proceedings of the 3rd International Workshop on Future Multimedia Networking (FMN)*, Krakow, Poland, 2010.

[159] A. Sackl, P. Zwickl, and P. Reichl, "The Trouble with Choice: An Empirical Study to Investigate the Influence of Charging Strategies and Content Selection on QoE", in *Proceedings of the 9th International Conference on Network and Service Management (CNSM)*, Zurich, Switzerland, 2013.

[160] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A Generic Quantitative Relationship Between Quality of Experience and Quality of Service", *IEEE Network*, vol. 24, no. 2, pp. 36–41, 2010.

[161]   E. Y.-H. Lin, "A Biblographical Survey on Some Well-known Non-Standard Knapsack Problems", *INFOR: Information Systems and Operational Research*, vol. 36, no. 4, pp. 274–317, 1998.

[162]   O. B. Maia, H. C. Yehia, and L. de Errico, "A Concise Review of the Quality of Experience Assessment for Video Streaming", *Computer Communications*, vol. 57, pp. 1–12, 2015.

[163]   Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: a Tutorial on Video Quality Assessment", *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.

[164]   P. Juluri, V. Tamarapalli, and D. Medhi, "Measurement of Quality of Experience of Video-on-Demand Services: A Survey", *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 401–418, 2016.

[165]   M. G. Martini, C. T. Hewage, M. M. Nasrall, and O. Ognenoski, "QoE Control, Monitoring, and Management Strategies", in *Multimedia Quality of Experience (QoE): Current Status and Future Requirements*, C. W. Chen, P. Chatzimisios, T. Dagiuklas, and L. Atzori, Eds., John Wiley & Sons, 2015, pp. 149–168.

[166]   T. Zhao, Q. Liu, and C. W. Chen, "QoE in Video Transmission: A User Experience-Driven Strategy", *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 285–302, 2017.

[167]   A. Raake, J. Gustafsson, S. Argyropoulos, M.-N. Garcia, D. Lindegren, G. Heikkilä, M. Pettersson, P. List, and B. Feiten, "IP-based Mobile and Fixed Network Audiovisual Media Services", *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 68–79, 2011.

[168]   T. T. Nguyen and G. Armitage, "A Survey of Techniques for Internet Traffic Classification Using Machine Learning", *IEEE Communications Surveys & Tutorials*, vol. 10, no. 4, pp. 56–76, 2008.

[169] N. Namdev, S. Agrawal, and S. Silkari, "Recent Advancement in Machine Learning Based Internet Traffic Classification", *Procedia Computer Science*, vol. 60, pp. 784–791, 2015.

[170] J. Pokhrel, B. Wehbi, A. Morais, A. Cavalli, and E. Allilaire, "Estimation of QoE of Video Traffic Using a Fuzzy Expert System", in *Proceedings of the 10th IEEE Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, USA, 2013.

[171] A. Khan, L. Sun, and E. Ifeachor, "QoE Prediction Model and its Application in Video Quality Adaptation over UMTS Networks", *IEEE Transactions on Multimedia*, vol. 14, no. 2, pp. 431–442, 2012.

[172] I. Ketykó, K. De Moor, T. De Pessemier, A. J. Verdejo, K. Vanhecke, W. Joseph, L. Martens, and L. De Marez, "QoE Measurement of Mobile YouTube Video Streaming", in *Proceedings of the 3rd Workshop on Mobile Video Delivery (MoViD)*, Florence, Italy, 2010.

[173] International Telecommunication Union, *ITU-T Recommendation P.1203: Parametric Bitstream-based Quality Assessment of Progressive Download and Adaptive Audiovisual Streaming Services over Reliable Transport*, 2016.

[174] R. Schatz, T. Hoßfeld, and P. Casas, "Passive YouTube QoE Monitoring for ISPs", in *Proceedings of the 2nd International Workshop on Future Internet and Next Generation Networks (FINGNet)*, Palermo, Italy, 2012.

[175] P. Casas, R. Schatz, and T. Hoßfeld, "Monitoring YouTube QoE: Is Your Mobile Network Delivering the Right Experience to Your Customers?", in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, 2013.

[176] B. Staehle, M. Hirth, R. Pries, F. Wamser, and D. Staehle, "YoMo: A YouTube Application Comfort Monitoring Tool", in *Proceedings of the 1st Workshop of Quality of Experience for Multimedia Content Sharing (QoEMCS)*, Tampere, Finland, 2010.

[177] M. Eckert, T. M. Knoll, and F. Schlegel, "Advanced MOS Calculation for Network Based QoE Estimation of TCP Streamed Video Services", in *Proceedings of the 7th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Gold Coast, Australia, 2013.

[178] P. Ameigeiras, A. Azcona-Rivas, J. Navarro-Ortiz, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A Simple Model for Predicting the Number and Duration of Rebuffering Events for YouTube Flows", *IEEE Communications Letters*, vol. 16, no. 2, pp. 278–280, 2012.

[179] R. Huysegems, B. De Vleeschauwer, K. De Schepper, C. Hawinkel, T. Wu, K. Laevens, and W. Van Leekwijck, "Session Reconstruction for HTTP Adaptive Streaming: Laying the Foundation for Network-based QoE Monitoring", in *Proceedings of the 20th IEEE International Workshop on Quality of Service (IWQoS)*, Coimbra, Portugal, 2012.

[180] T. Wu, S. Petrangeli, R. Huysegems, T. Bostoen, and F. De Turck, "Network-based Video Freeze Detection and Prediction in HTTP Adaptive Streaming", *Computer Communications*, vol. 99, pp. 37–47, 2017.

[181] P. Szilágyi and C. Vulkán, "Network side Lightweight and Scalable YouTube QoE Estimation", in *Proceedings of the IEEE International Conference on Communications (ICC)*, London, UK, 2015.

[182] D. Tsilimantos, T. Karagkioules, A. Nogales-Gómez, and S. Valentin, "Traffic Profiling for Mobile Video Streaming", in *Proceedings of the IEEE International Conference on Communications (ICC)*, Paris, France, 2017.

[183] V. Aggarwal, E. Halepovic, J. Pang, S. Venkataraman, and H. Yan, "Prometheus: Toward Quality-of-Experience Estimation for Mobile Apps from Passive Network Measurements", in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications (HotMobile)*, Santa Barbara, CA, USA, 2014.

[184] M. Z. Shafiq, J. Erman, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Understanding the Impact of Network Dynamics on Mobile Video User Engagement", in *Proceedings of the ACM SIGMETRICS*, Austin, TX, USA, 2014.

[185] W. Pan, G. Cheng, H. Wu, and Y. Tang, "Towards QoE Assessment of Encrypted YouTube Adaptive Video Streaming in Mobile Networks", in *Proceedings of the 24th IEEE/ACM International Symposium on Quality of Service (IWQoS)*, Vilanova i la Geltrú, Spain, 2016.

[186] M. Katsarakis, R. Teixeira, M. Papadopouli, and V. Christophides, "Towards a Causal Analysis of Video QoE from Network and Application QoS", in *Proceedings of the ACM SIGCOMM Workshop on QoE-based Analysis and Management of Data Communication Networks (Internet-QoE)*, Florianópolis, Brazil, 2016.

[187] R. Dubin, O. Hadar, I. Richman, O. Trabelsi, A. Dvir, and O. Pele, "Video Quality Representation Classification of Safari Encrypted DASH Streams", in *Proceedings of the Digital Media Industry & Academic Forum (DMIAF)*, Santorini, Greece, 2016.

[188] G. Dimopoulos, I. Leontiadis, P. Barlet-Ros, and K. Papagiannaki, "Measuring Video QoE from Encrypted Traffic", in *Proceedings of the ACM Internet Measurement Conference (IMC)*, Santa Monica, CA, USA, 2016.

[189] I. Orsolic, D. Pevec, M. Suznjevic, and L. Skorin-Kapov, "YouTube QoE Estimation Based on the Analysis of Encrypted Network Traffic Using Machine Learning", in *Proceedings of the 5th IEEE International Workshop on Quality of Experience for Multimedia Communications (QoEMC)*, Washington, DC, USA, 2016.

[190] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, "Measuring the Quality of Experience of HTTP Video Streaming", in *Proceedings of the 12th IFIP/IEEE International Symposium on Integrated Network Management (IM)*, Dublin, Ireland, 2011.

[191]   R. K. P. Mok, E. W. W. Chan, X. Luo, and R. K. C. Chan, "Inferring the QoE of HTTP Video Streaming from User-Viewing Activities", in *Proceedings of the 1st ACM SIGCOMM Workshop on Measurements Up the STack (W-MUST)*, Toronto, Canada, 2011.

[192]   G. Gómez, L. Hortigüela, Q. Pérez, J. Lorca, R. García, and M. C. Aguayo-Torres, "YouTube QoE Evaluation Tool for Android Wireless Terminals", *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 164, pp. 1–14, 2014.

[193]   Q. A. Chen, H. Luo, S. Rosen, Z. M. Mao, K. Iyer, J. Hui, K. Sontineni, and K. Lau, "QoE Doctor: Diagnosing Mobile App QoE with Automated UI Control and Cross-layer Analysis", in *Proceedings of the Internet Measurement Conference (IMC)*, Melbourne, Australia, 2014.

[194]   H. Nam, K.-H. Kim, D. Calin, and H. Schulzrinne, "YouSlow: A Performance Analysis Tool for Adaptive Bitrate Video Streaming", *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 111–112, 2014.

[195]   H. Nam, K.-H. Kim, and H. Schulzrinne, "QoE Matters More Than QoS: Why People Stop Watching Cat Videos", in *Proceedings of the IEEE INFOCOM*, San Francisco, CA, USA, 2016.

[196]   H. Nam and H. Schulzrinne, "YouSlow: What Influences User Abandonment Behavior for Internet Video?", Columbia University, Tech. Rep., 2016.

[197]   F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the Impact of Video Quality on User Engagement", in *Proceedings of the ACM SIGCOMM*, Toronto, Canada, 2011.

[198]   A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a Predictive Model of Quality of Experience for Internet Video", *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 339–350, 2013.

[199] A. C. Dalal, A. K. Bouchard, S. Cantor, Y. Guo, and A. Johnson, "Assessing QoE of On-demand TCP Video Streams in Real Time", in *Proceedings of the IEEE International Conference on Communications (ICC)*, Ottawa, Canada, 2012.

[200] S. Baraković and L. Skorin-Kapov, "Survey and Challenges of QoE Management Issues in Wireless Networks", *Journal of Computer Networks and Communications*, vol. 2013, no. 165146, pp. 1–28, 2013.

[201] Q. M. Qadir, A. A. Kist, and Z. Zhang, "Mechanisms for QoE Optimisation of Video Traffic: A Review Paper", *Australasian Journal of Information, Communication Technology and Applications*, vol. 1, no. 1, pp. 1–18, 2015.

[202] M. Ghobadi, Y. Cheng, A. Jain, and M. Mathis, "Trickle: Rate Limiting YouTube Video Streaming", in *Proceedings of the USENIX Annual Technical Conference (USENIX ATC)*, Boston, MA, USA, 2012.

[203] M. Jarschel, F. Wamser, T. Hohn, T. Zinner, and P. Tran-Gia, "SDN-Based Application-Aware Networking on the Example of YouTube Video Streaming", in *Proceedings of the 2nd European Workshop on Software Defined Networks (EWSDN)*, Berlin, Germany, 2013.

[204] S. Laga, T. Van Cleemput, F. Van Raemdonck, F. Vanhoutte, N. Bouten, M. Claeys, and F. De Turck, "Optimizing Scalable Video Delivery through OpenFlow Layer-based Routing", in *Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS)*, Krakow, Poland, 2014.

[205] A. El Essaili, D. Schroeder, D. Staehle, M. Shehada, W. Kellerer, and E. Steinbach, "Quality-of-Experience Driven Adaptive HTTP Media Delivery", in *Proceedings of the IEEE International Conference on Communications (ICC)*, Budapest, Hungary, 2013.

[206] A. El Essaili, D. Schroeder, E. Steinbach, D. Staehle, and M. Shehada, "QoE-based Traffic and Resource Management for Adaptive HTTP

Video Delivery in LTE", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 6, pp. 988–1001, 2015.

[207] J. Navarro-Ortiz, P. Ameigeiras, J. M. Lopez-Soler, J. Lorca-Hernando, Q. Perez-Tarrero, and R. Garcia-Perez, "A QoE-Aware Scheduler for HTTP Progressive Video in OFDMA Systems", *IEEE Communications Letters*, vol. 17, no. 4, pp. 677–680, 2013.

[208] F. Wamser, T. Zinner, P. Tran-Gia, and J. Zhu, "Dynamic Bandwidth Allocation for Multiple Network Connections: Improving User QoE and Network Usage of YouTube in Mobile Broadband", in *Proceedings of the ACM SIGCOMM Capacity Sharing Workshop (CSWS)*, Chicago, IL, USA, 2014.

[209] N. Bouten, M. Claeys, S. Latré, J. Famaey, W. Van Leekwijck, and F. De Turck, "Deadline-based Approach for Improving Delivery of SVC-based HTTP Adaptive Streaming Content", in *Proceedings of the IEEE Network Operations and Management Symposium (NOMS)*, Krakow, Poland, 2014.

[210] F. Wamser, L. Iffländer, T. Zinner, and P. Tran-Gia, "Implementing Application-Aware Resource Allocation on a Home Gateway for the Example of YouTube", in *Mobile Networks and Management: 6th International Conference, MONAMI 2014, Würzburg, Germany, September 22-26, 2014, Revised Selected Papers*, R. Agüero, T. Zinner, R. Goleva, A. Timm-Giel, and P. Tran-Gia, Eds., Springer, 2015, pp. 301–312.

[211] T. Phan-Xuan and E. Kamioka, "Efficiency of QoE-driven Network Management in Adaptive Streaming over HTTP", in *Proceedings of the 22nd Asia-Pacific Conference on Communications (APCC)*, Yogyakarta, Indonesia, 2016.

[212] J. Gross, J. Klaue, H. Karl, and A. Wolisz, "Cross-layer Optimization of OFDM Transmission Systems for MPEG-4 Video Streaming", *Computer Communications*, vol. 27, no. 11, pp. 1044–1055, 2004.

[213]   S. Khan, Y. Peng, E. Steinbach, M. Sgroi, and W. Kellerer, "Application-Driven Cross-Layer Optimization for Video Streaming over Wireless Networks", *IEEE Communications Magazine*, vol. 44, no. 1, pp. 122–130, 2006.

[214]   L. Superiori, M. Wrulich, P. Svoboda, M. Rupp, J. Fabini, W. Karner, and M. Steinbauer, "Content-aware Scheduling for Video Streaming over HSDPA Networks", in *Proceedings of the 2nd International Workshop on Cross Layer Design (IWCLD)*, Palma de Mallorca, Spain, 2009.

[215]   S. Thakolsri, S. Khan, E. Steinbach, and W. Kellerer, "QoE-driven Cross-layer Optimization for High Speed Downlink Packet Access", *Journal of Communications*, vol. 4, no. 9, pp. 669–680, 2009.

[216]   B. Staehle, M. Hirth, R. Pries, F. Wamser, and D. Staehle, "Aquarema in Action: Improving the YouTube QoE in Wireless Mesh Networks", in *Proceedings of the Baltic Congress on Future Internet Communications (BCFIC)*, Riga, Latvia, 2011.

[217]   N. Bouten, J. Famaey, S. Latré, R. Huysegems, B. D. Vleeschauwer, W. V. Leekwijck, and F. D. Turck, "QoE Optimization Through In-network Quality Adaptation for HTTP Adaptive Streaming", in *Proceedings of the 8th International Conference on Network and Service Management (CNSM)*, Las Vegas, NV, USA, 2012.

[218]   R. K. Mok, X. Luo, E. W. Chan, and R. K. Chang, "QDASH: a QoE-aware DASH system", in *Proceedings of the 3rd ACM Multimedia Systems Conference (MMSys)*, Chapel Hill, NC, USA, 2012.

[219]   X. Wang, M. Chen, T. T. Kwon, L. Yang, and V. C. Leung, "AMES-Cloud: A Framework of Adaptive Mobile Video Streaming and Efficient Social Video Sharing in the Clouds", *IEEE Transactions on Multimedia*, vol. 15, no. 4, pp. 811–820, 2013.

[220] V. Krishnamoorthi, N. Carlsson, D. Eager, A. Mahanti, and N. Shah-mehri, "Helping Hand or Hidden Hurdle: Proxy-assisted HTTP-based Adaptive Streaming Performance", in *Proceedings of the 21st IEEE International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, San Francisco, CA, USA, 2013.

[221] P. Georgopoulos, Y. Elkhatib, M. Broadbent, M. Mu, and N. Race, "Towards Network-wide QoE Fairness using OpenFlow-assisted Adaptive Video Streaming", in *Proceedings of the ACM SIGCOMM Workshop on Future Human-centric Multimedia Networking (FhMN)*, Hong Kong, China, 2013.

[222] S. Petrangeli, M. Claeys, S. Latré, J. Famaey, and F. De Turck, "A Multi-agent Q-Learning-based Framework for Achieving Fairness in HTTP Adaptive Streaming", in *Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS)*, Krakow, Poland, 2014.

[223] B. Fu, G. Kunzmann, D. Corujo, M. Wetterwald, and R. Costa, "QoE-aware Traffic Management for Mobile Video Delivery", in *Proceedings of the IEEE International Conference on Communications Workshops (ICC)*, Budapest, Hungary, 2013.

[224] B. Fu, D. Staehle, G. Kunzmann, E. Steinbach, and W. Kellerer, "QoE-aware Priority Marking and Traffic Management for H.264/SVC-based Mobile Video Delivery", in *Proceedings of the 8th ACM Workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Networks (PM2HW2N)*, Barcelona, Spain, 2013.

[225] P. Tang, P. Wang, N. Wang, and V. Nguyen Ngoc, "QoE-Based Resource Allocation Algorithm for Multi-Applications in Downlink LTE Systems", in *Proceedings of the International Conference on Computer, Communications and Information Technology (CCIT)*, Beijing, China, 2014.

[226] S. Akhshabi, L. Anantakrishnan, C. Dovrolis, and A. C. Begen, "Server-based Traffic Shaping for Stabilizing Oscillating Adaptive Streaming Players", in *Proceedings of the 23rd ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, Oslo, Norway, 2013.

[227] X. Liu and A. Men, "QoE-aware Traffic Shaping for HTTP Adaptive Streaming", *International Journal of Multimedia and Ubiquitous Engineering*, vol. 9, no. 2, pp. 33–44, 2014.

[228] H. Nam, K.-H. Kim, J. Y. Kim, and H. Schulzrinne, "Towards QoE-aware Video Streaming using SDN", in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, Austin, TX, USA, 2014.

[229] O. Dobrijevic, M. Santl, and M. Matijasevic, "Ant Colony Optimization for QoE-centric Flow Routing in Software-defined Networks", in *Proceedings of the 11th International Conference on Network and Service Management (CNSM)*, Barcelona, Spain, 2015.

[230] S. Petrangeli, T. Wauters, R. Huysegems, T. Bostoen, and F. De Turck, "Network-based Dynamic Prioritization of HTTP Adaptive Streams to Avoid Video Freezes", in *Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management (IM)*, Ottawa, Canada, 2015.

[231] T. Kimura, M. Yokota, A. Matsumoto, K. Takeshita, T. Kawano, K. Sato, H. Yamamoto, T. Hayashi, K. Shiomoto, and K. Miyazaki, "QUVE: QoE Maximizing Framework for Video-Streaming", *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 138–153, 2017.

[232] A. Bentaleb, A. C. Begen, and R. Zimmermann, "SDNDASH: Improving QoE of HTTP Adaptive Streaming Using Software Defined Networking", in *Proceedings of the ACM Multimedia Conference (MM)*, Amsterdam, The Netherlands, 2016.

[233] Y. Li, P. A. Frangoudis, Y. Hadjadj-Aoul, and P. Bertin, "A Mobile Edge Computing-based Architecture for Improved Adaptive HTTP Video Delivery", in *Proceedings of the IEEE Conference on Standards for Communications and Networking (CSCN)*, Berlin, Germany, 2016.

[234] G. Cofano, L. De Cicco, T. Zinner, A. Nguyen-Ngoc, P. Tran-Gia, and S. Mascolo, "Design and Experimental Evaluation of Network-assisted Strategies for HTTP Adaptive Streaming", in *Proceedings of the 7th ACM International Conference on Multimedia Systems (MMSys)*, Klagenfurt, Austria, 2016.

[235] S. Wilk, D. Stohr, and W. Effelsberg, "A Content-Aware Video Adaptation Service to Support Mobile Video", *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 12, no. 5s, 82:1–82:23, 2016.

[236] S. Petrangeli, J. Famaey, M. Claeys, S. Latré, and F. De Turck, "QoE-driven Rate Adaptation Heuristic for Fair Adaptive Video Streaming", *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 12, no. 2, 28:1–28:24, 2016.

[237] J. Chen, M. Ammar, M. Fayed, and R. Fonseca, "Client-Driven Network-level QoE Fairness for Encrypted 'DASH-S'", in *Proceedings of the ACM SIGCOMM Workshop on QoE-based Analysis and Management of Data Communication Networks (Internet-QoE)*, Florianopolis, Brazil, 2016.

[238] E. Thomas, M. O. van Deventer, T. Stockhammer, A. C. Begen, M.-L. Champel, and O. Oyman, "Applications and Deployments of Server and Network Assisted DASH", in *Proceedings of the International Broadcasting Convention (IBC)*, Amsterdam, The Netherlands, 2016.

[239] International Standards Organization/International Electrotechnical Commission (ISO/IEC), *23009-5:2015 Information Technology – Dynamic Adaptive Streaming over HTTP (DASH) – Part 5: Server and network assisted DASH (SAND)*, 2016.

[240]   M. Bouet, J. Leguay, and V. Conan, "Cost-based Placement of Virtualized Deep Packet Inspection Functions in SDN", in *Proceedings of the IEEE Military Communications Conference (MILCOM)*, San Diego, CA, USA, 2013.

[241]   M. Bouet, J. Leguay, T. Combe, and V. Conan, "Cost-based Placement of vDPI Functions in NFV Infrastructures", *International Journal of Network Management*, vol. 25, no. 6, pp. 490–506, 2015.

[242]   T. Hoßfeld, L. Skorin-Kapov, P. Heegaard, and M. Varela, "Definition of QoE Fairness in Shared Systems", *IEEE Communications Letters*, vol. 21, no. 1, pp. 184–187, 2017.

[243]   Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[244]   S. Rimac-Drlje, M. Vranjes, and D. Zagar, "Influence of Temporal Pooling Method on the Objective Video Quality Evaluation", in *Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Bilbao, Spain, 2009.

[245]   K. Seshadrinathan and A. C. Bovik, "Temporal Hysteresis Model of Time Varying Subjective Video Quality", in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011.

[246]   K. Lee, J. Park, S. Lee, and A. C. Bovik, "Temporal Pooling of Video Quality Estimates Using Perceptual Motion Models", in *Proceedings of the 17th IEEE International Conference on Image Processing*, Hong Kong, China, 2010.

[247]   A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Considering Temporal Variations of Spatial Visual Distortions in Video Quality Assessment", *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 253–265, 2009.

[248] International Telecommunication Union, *ITU-T Recommendation P.800: Methods for Subjective Determination of Transmission Quality*, 1996.

[249] ——, *ITU-T Recommendation P.1501: Subjective Testing Methodology for Web Browsing*, 2013.

[250] S. Hemminger, "Network Emulation with NetEm", in *Proceedings of the 6th Australia's National Linux Conference (LCA)*, Canberra, Australia, 2005.

[251] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling Innovation in Campus Networks", *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, p. 69, 2008.

[252] M. Jarschel, T. Zinner, T. Hoßfeld, P. Tran-Gia, and W. Kellerer, "Interfaces, Attributes, and Use Cases: A Compass for SDN", *IEEE Communications Magazine*, vol. 52, no. 6, pp. 210–217, 2014.

[253] T. Zinner, M. Jarschel, A. Blenk, F. Wamser, and W. Kellerer, "Dynamic Application-Aware Resource Management Using Software-Defined Networking: Implementation Prospects and Challenges", in *Proceedings of the IFIP/IEEE International Workshop on Quality of Experience Centric Management (QCMan)*, Krakow, Poland, 2014.

[254] L. Kleinrock, "Time-shared systems: A theoretical treatment", *Journal of the ACM (JACM)*, vol. 14, no. 2, pp. 242–261, 1967.

[255] G. Fayolle, I. Mitrani, and R. Iasnogorodski, "Sharing a Processor Among Many Job Classes", *Journal of the ACM (JACM)*, vol. 27, no. 3, pp. 519–532, 1980.

[256] J. W. Cohen, "The multiple phase service network with generalized processor sharing", *Acta informatica*, vol. 12, no. 3, pp. 245–284, 1979.

[257] S. Tanwir and H. Perros, "A Survey of VBR Video Traffic Models", *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1778–1802, 2013.

[258] Wireless Broadband Alliance, "WBA Industry Report 2011: Global Developments in Public Wi-Fi", Tech. Rep., 2011.

[259] Strategy Analytics, "Global Broadband and WLAN (Wi-Fi) Networked Households Forecast 2009-2018", Tech. Rep., 2014.

[260] S. Ickin, K. Wac, M. Fiedler, L. Janowski, J.-H. Hong, and A. K. Dey, "Factors Influencing Quality of Experience of Commonly Used Mobile Applications", *IEEE Communications Magazine*, vol. 50, no. 4, pp. 48–56, 2012.

[261] F. Kaup and D. Hausheer, "Optimizing Energy Consumption and QoE on Mobile Devices", in *Proceedings of the IEEE International Conference on Network Protocols (ICNP)*, Göttingen, Germany, 2013.

[262] P. Costa, C. Mascolo, M. Musolesi, and G. P. Picco, "Socially-aware Routing for Publish-Subscribe in Delay-tolerant Mobile Ad Hoc Networks", *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 5, pp. 748–760, 2008.

[263] A. Roy, P. De, and N. Saxena, "Location-based Social Video Sharing over Next Generation Cellular Networks", *IEEE Communications Magazine*, vol. 53, no. 10, pp. 136–143, 2015.

[264] N. Sastry, E. Yoneki, and J. Crowcroft, "Buzztraq: Predicting Geographical Access Patterns of Social Cascades Using Social Networks", in *Proceedings of the 2nd ACM EuroSys Workshop on Social Network Systems (SocialNets)*, Nuremberg, Germany, 2009.

[265] S. Scellato, C. Mascolo, M. Musolesi, and J. Crowcroft, "Track Globally, Deliver Locally: Improving Content Delivery Networks by Tracking Geographic Social Cascades", in *Proceedings of the 20th International Conference on World Wide Web (WWW)*, Hyderabad, India, 2011.

[266]   Z. Wang, L. Sun, X. Chen, W. Zhu, J. Liu, M. Chen, and S. Yang, "Propagation-based Social-aware Replication for Social Video Contents", in *Proceedings of the 20th ACM International Conference on Multimedia*, Nara, Japan, 2012.

[267]   Y. Wu, C. Wu, B. Li, L. Zhang, Z. Li, and F. Lau, "Scaling Social Media Applications into Geo-distributed Clouds", *IEEE/ACM Transactions on Networking (TON)*, vol. 23, no. 3, pp. 689–702, 2015.

[268]   Z. Wang, J. Liu, and W. Zhu, *Social Video Content Delivery*, ser. Springer-Briefs in Electrical and Computer Engineering. Springer, 2016.

[269]   I. Kilanioti and G. A. Papadopoulos, "Content Delivery Simulations Supported by Social Network-awareness", *Simulation Modelling Practice and Theory*, vol. 71, pp. 114–133, 2017.

[270]   N. Laoutaris, P. Rodriguez, and L. Massoulie, "ECHOS: Edge Capacity Hosting Overlays of Nano Data Centers", *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 1, pp. 51–54, 2008.

[271]   V. Valancius, N. Laoutaris, L. Massoulié, C. Diot, and P. Rodriguez, "Greening the Internet with Nano Data Centers", in *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies (Co-NEXT)*, Rome, Italy, 2009.

[272]   J. He, A. Chaintreau, and C. Diot, "A Performance Evaluation of Scalable Live Video Streaming with Nano Data Centers", *Computer Networks*, vol. 53, no. 2, pp. 153–167, 2009.

[273]   Z. Li, H. Shen, H. Wang, G. Liu, and J. Li, "SocialTube: P2P-assisted Video Sharing in Online Social Networks", in *Proceedings of the IEEE INFO-COM*, Orlando, FL, USA, 2012.

[274]   S. Traverso, K. Huguenin, I. Triestan, V. Erramilli, N. Laoutaris, and K. Papagiannaki, "TailGate: Handling Long-Tail Content with a Little Help from Friends", in *Proceedings of the 21st International Conference on World Wide Web (WWW)*, Lyon, France, 2012.

[275]   F. Zhou, L. Zhang, E. Franco, A. Mislove, R. Revis, and R. Sundaram, "WebCloud: Recruiting Social Network Users to Assist in Content Distribution", in *Proceedings of the 11th IEEE International Symposium on Network Computing and Applications (NCA)*, Cambridge, MA, USA, 2012.

[276]   K. Huguenin, A.-M. Kermarrec, K. Kloudas, and F. Taïani, "Content and Geographical Locality in User-generated Content Sharing Systems", in *Proceedings of the 22nd International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, Toronto, Canada, 2012.

[277]   X. Wang, T. Kwon, Y. Choi, H. Wang, and J. Liu, "Cloud-assisted Adaptive Video Streaming and Social-aware Video Prefetching for Mobile Users", *IEEE Wireless Communications*, vol. 20, no. 3, pp. 72–79, 2013.

[278]   M. Ma, Z. Wang, K. Su, and L. Sun, "Understanding the Power of Smartrouter-based Peer CDN for Video Streaming", in *Proceedings of the 25th International Conference on Computer Communication and Networks (ICCCN)*, Waikoloa, HI, USA, 2016.

[279]   G. Nan, Z. Mao, M. Yu, M. Li, H. Wang, and Y. Zhang, "Stackelberg Game for Bandwidth Allocation in Cloud-based Wireless Live-streaming Social Networks", *IEEE Systems Journal*, vol. 8, no. 1, pp. 256–267, 2014.

[280]   L. Gao, M. Tang, H. Pang, J. Huang, and L. Sun, "Performance Bound Analysis for Crowdsourced Mobile Video Streaming", in *Proceedings of the Annual Conference on Information Science and Systems (CISS)*, Princeton, NJ, USA, 2016.

[281]   Y. Cao, T. Jiang, X. Chen, and J. Zhang, "Social-aware Video Multicast Based on Device-to-device Communications", *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1528–1539, 2016.

[282]   Y. Khadraoui, X. Lagrange, and A. Gravey, "A Survey of Available Features for Mobile Traffic Offload", in *Proceedings of the 20th European Wireless Conference*, Barcelona, Spain, 2014.

[283] M. Gonzalez, T. Higashino, and M. Okada, "Radio Access Considerations for Data Offloading with Multipath TCP in Cellular/WiFi Networks", in *Proceedings of the International Conference on Information Networking (ICOIN)*, Bangkok, Thailand, 2013.

[284] C. Paasch, G. Detal, F. Duchene, C. Raiciu, and O. Bonaventure, "Exploring Mobile/WiFi Handover with Multipath TCP", in *Proceedings of the ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design*, Helsinki, Finland, 2012.

[285] S. Chen, Z. Yuan, and G.-M. Muntean, "An Energy-aware Multipath-TCP-based Content Delivery Scheme in Heterogeneous Wireless Networks", in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, 2013.

[286] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights", *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2484–2497, 2013.

[287] A. Gladisch, R. Daher, and D. Tavangarian, "Survey on Mobility and Multihoming in Future Internet", *Wireless Personal Communications*, vol. 74, no. 1, pp. 45–81, 2014.

[288] L. Mamatas, I. Psaras, and G. Pavlou, "Incentives and Algorithms for Broadband Access Sharing", in *Proceedings of the ACM SIGCOMM Workshop on Home Networks*, New Delhi, India, 2010.

[289] N. Gautam, H. Petander, and J. Noel, "A Comparison of the Cost and Energy Efficiency of Prefetching and Streaming of Mobile Video", in *Proceedings of the 5th Workshop on Mobile Video (MoVid)*, Oslo, Norway, 2013.

[290] C. B. Lafuente, X. Titi, and J.-M. Seigneur, "Flexible Communication: A Secure and Trust-Based Free Wi-Fi Password Sharing Service", in *Proceedings of the 10th IEEE International Conference on Trust, Security*

and Privacy in Computing and Communications (TrustCom), Changsha, China, 2011.

[291]   L. J. Donelson and C. W. Sweet, *Method, Apparatus and System for Wireless Network Authentication Through Social Networking*, US Patent App. 13/287,931, 2012.

[292]   K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile Data Offloading: How Much Can WiFi Deliver?", *IEEE/ACM Transactions on Networking (TON)*, vol. 21, no. 2, pp. 536–550, 2013.

[293]   J. Kim, N.-O. Song, B. H. Jung, H. Leem, and D. K. Sung, "Placement of WiFi Access Points for Efficient WiFi Offloading in an Overlay Network", in *Proceedings of the 24th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, London, UK, 2013.

[294]   D. Suh, H. Ko, and S. Pack, "Efficiency Analysis of WiFi Offloading Techniques", *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3813–3817, 2016.

[295]   N. Sastry, J. Crowcroft, and K. Sollins, "Architecting Citywide Ubiquitous Wi-Fi Access", in *Proceedings of the 6th Workshop on Hot Topics in Networks (HotNets)*, Atlanta, GA, USA, 2007.

[296]   P. Vidales, A. Manecke, and M. Solarski, "Metropolitan Public WiFi Access Based on Broadband Sharing", in *Proceedings of the Mexican International Conference on Computer Science (ENC)*, Mexico City, Mexico, 2009.

[297]   S. Dimatteo, P. Hui, B. Han, and V. O. Li, "Cellular Traffic Offloading Through WiFi Networks", in *Proceedings of the 8th IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, Valencia, Spain, 2011.

[298]   V. F. Mota, D. F. Macedo, Y. Ghamri-Doudane, and J. M. S. Nogueira, "On the Feasibility of WiFi Offloading in Urban Areas: The Paris Case Study", in *Proceedings of the IFIP Wireless Days (WD)*, Valencia, Spain, 2013.

[299]   E. M. R. Oliveira and A. C. Viana, "From Routine to Network Deployment for Data Offloading in Metropolitan Areas", in *Proceedings of the 11th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, Singapore, 2014.

[300]   E. Bulut and B. K. Szymanski, "WiFi Access Point Deployment for Efficient Mobile Data Offloading", *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 17, no. 1, pp. 71–78, 2013.

[301]   ——, "Rethinking Offloading WiFi Access Point Deployment from User Perspective", in *Proceedings of the 12th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, New York, NY, USA, 2016.

[302]   P. Rouveyrol, P. Raveneau, and M. Cunche, "Large Scale Wi-Fi Tracking Using a Botnet of Wireless Routers", in *Proceedings of the Workshop on Surveillance & Technology*, Philadelphia, PA, USA, 2015.

[303]   M. Michalopoulou, J. Riihijarvi, and P. Mahonen, "Studying the Relationships between Spatial Structures of Wireless Networks and Population Densities", in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, Miami, FL, USA, 2010.

[304]   C. Clark, "Urban Population Densities", *Journal of the Royal Statistical Society. Series A (General)*, vol. 114, no. 4, pp. 490–496, 1951.

[305]   J. F. McDonald, "Econometric Studies of Urban Population Density: A Survey", *Journal of Urban Economics*, vol. 26, no. 3, pp. 361–385, 1989.

[306]   D. A. Griffith and D. W. Wong, "Modeling Population Density Across Major US Cities: A Polycentric Spatial Regression Approach", *Journal of Geographical Systems*, vol. 9, no. 1, pp. 53–75, 2007.

[307]   M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.

[308]   H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic Geometry for Modeling, Analysis, and Design of Multi-tier and Cognitive Cellular Wireless Networks: A Survey", *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 996–1019, 2013.

[309]   A. Ruhela, R. M. Tripathy, S. Triukose, S. Ardon, A. Bagchi, and A. Seth, "Towards the Use of Online Social Networks for Efficient Internet Content Distribution", in *Proceedings of the 5th IEEE International Conference on Advanced Networks and Telecommunication Systems (ANTS)*, Bangalore, India, 2011.

[310]   V. A. Siris and M. Anagnostopoulou, "Performance and Energy Efficiency of Mobile Data Offloading with Mobility Prediction and Prefetching", in *Proceedings of the 14th IEEE International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Madrid, Spain, 2013.

[311]   B. H. Jung, N.-O. Song, and D. K. Sung, "A Network-assisted User-centric WiFi-offloading Model for Maximizing Per-user Throughput in a Heterogeneous Network", *IEEE Transactions on Vehicular Technology*, vol. 63, no. 4, pp. 1940–1945, 2014.

[312]   R. Mahindra, H. Viswanathan, K. Sundaresan, M. Y. Arslan, and S. Rangarajan, "A Practical Traffic Management System for Integrated LTE-WiFi Networks", in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking (MobiCom)*, Maui, HI, USA, 2014.

[313]   M. Seyedebrahimi and X.-H. Peng, "Optimising QoE Distribution for Video Applications through LTE-WiFi Interworking", in *Proceedings of the 9th International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST)*, Cambridge, UK, 2015.

[314]   M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD)*, Portland, OR, USA, 1996.

[315]   P. A. P. Moran, "Notes on Continuous Stochastic Phenomena", *Biometrika*, vol. 37, no. 1/2, pp. 17–23, 1950.

[316]   R. C. Geary, "The Contiguity Ratio and Statistical Mapping", *The Incorporated Statistician*, vol. 5, no. 3, pp. 115–146, 1954.

[317]   A. Lareida, T. Bocek, M. Waldburger, and B. Stiller, "RB-Tracker: A Fully Distributed, Replicating, Network-, and Topology-aware P2P CDN", in *Proceedings of the 13th IFIP/IEEE International Symposium on Integrated Network Management (IM)*, Ghent, Belgium, 2013.

[318]   J. J. Ramos-Muñoz, J. Prados-Garzon, P. Ameigeiras, J. Navarro-Ortiz, and J. M. López-Soler, "Characteristics of Mobile YouTube Traffic", *IEEE Wireless Communications*, vol. 21, no. 1, pp. 18–25, 2014.

[319]   F. Kaup, F. Jomrich, and D. Hausheer, "Demonstration of NetworkCoverage – A Mobile Network Performance Measurement App", in *Proceedings of the International Conference on Networked Systems (NetSys)*, Cottbus, Germany, 2015.

[320]   Akamai, "Q4 2014 State of the Internet", Tech. Rep., 2015.

[321]   K. Panitzek, I. Schweizer, T. Bönning, G. Seipel, and M. Mühlhäuser, "First Responder Communication in Urban Environments", *International Journal of Mobile Network Design and Innovation*, vol. 4, no. 2, pp. 109–118, 2012.

[322]   F. Kaup, M. Wichtlhuber, S. Rado, and D. Hausheer, "Can Multipath TCP Save Energy? A Measuring and Modeling Study of MPTCP Energy Consumption", in *Proceedings of the 40th Annual IEEE Conference on Local Computer Networks (LCN)*, Clearwater Beach, FL, USA, 2015.

[323]   Akamai, "Q4 2016 State of the Internet", Tech. Rep., 2017.