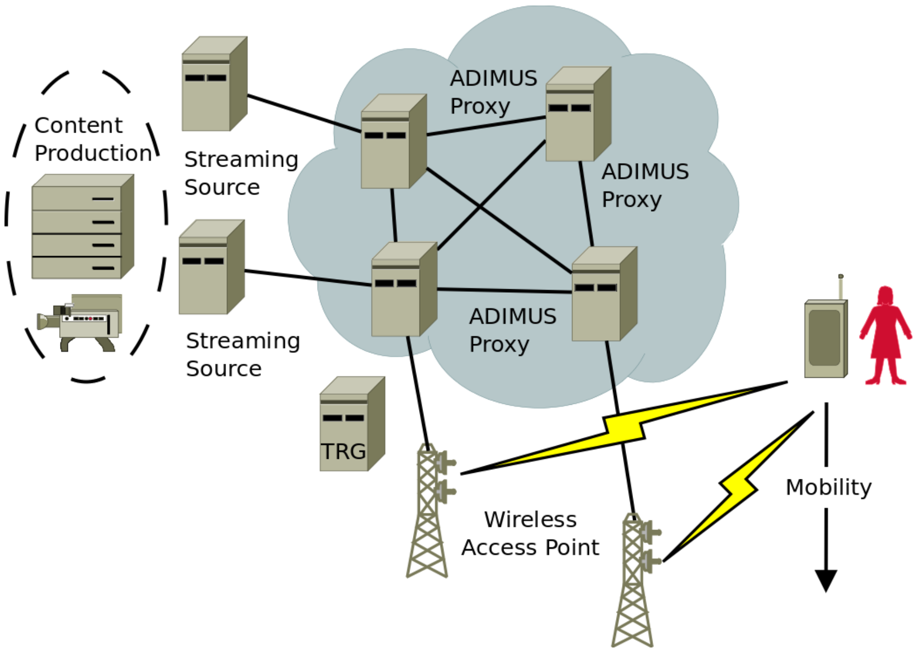
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Report

ADIMUS — Adaptive Internet Multimedia Streaming

Final Project Report



Report no 1026

Authors
Wolfgang Leister
Tiia Sutinen
Svetlana Boudko
Ian Marsh
Carsten Griwodz
Pål Halvorsen

Date September 1st, 2010

ISBN 978-82-539-0536-5

The authors

Wolfgang Leister, Dr. rer.nat., is a chief research scientist at Norsk Regnesentral, with research interests in multimedia, computer graphics, computer and sensor networks, health care applications, mobile systems, and free software.

Tiia Sutinen, M.Sc. (Tech.), is a researcher at VTT Technical Research Centre of Finland, and PhD student at the University of Oulu with research interests in cross-layer assisted adaptive video transmission in heterogeneous multi-access networks.

Svetlana Boudko, M.Sc., is a PhD student at Norsk Regnesentral, with research interests in distributed multimedia systems.

Ian Marsh, tekn.dr., is a researcher at SICS, and postdoc at the University of Porto. He has had research interests in real-time multimedia streaming. Now he is working with VANET communications.

Carsten Griwodz, Dr.Ing., is a Professor at the Department of Informatics, University of Oslo and a researcher at Simula Research Laboratory. His research focuses mainly on distributed multimedia systems.

Pål Halvorsen, Dr.Scient., is an Associate Professor at the Department of Informatics, University of Oslo and a researcher at Simula Research Laboratory. His research focuses mainly on distributed multimedia.

The work described in this document has been conducted as a part of the ADIMUS (Adaptive Internet Multimedia Streaming) project funded by the NORDUnet3 programme.

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modeling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Title	**ADIMUS** **Adaptive Internet Multimedia Streaming** **Final Project Report**
Authors	**Wolfgang Leister, Tiia Sutinen, Svetlana Boudko, Ian Marsh, Carsten Griwodz, Pål Halvorsen**
Quality assurance	Knut Holmqvist , Trenton Schulz
Date	September 1st, 2010
ISBN	978-82-539-0536-5
Publication number	1026

Abstract

This document is the final scientific report of the ADIMUS project which was funded by the NORDUnet3 programme. ADIMUS addresses the problem of enhancing the quality of multimedia streams at run-time based on the perceived quality. We present an architecture that addresses the different requirements from end-to-end for a mobile terminal. The architecture comprises an overlay network in the long-distance part, while the multi-access network employs cross-layer technology. Both parts interact, and use adaptation techniques shown in this report. We also present quality estimation techniques for both audio and video streams. This report is based on the scientifc papers published during the project, and sets these papers into the context of the ADIMUS project.

Keywords	multimedia streaming, overlay network, multi-access network, QoE, QoS, adaptation, video, audio
Target group	Researcher community, Internet and content providers
Availability	Open
Project	ADIMUS
Project number	803000
Research field	Multimedia, Networks
Number of pages	35
© Copyright	Norwegian Computing Center

Preface

This document is the final scientific report of the Adaptive Internet Multimedia Streaming (ADIMUS) project, funded by the NORDUnet3 programme, hosted by Norsk Regnesentral (NR), Norway, and supervised by Dr. Wolfgang Leister. Participating institutions are the Computer and Network Architecture Group at SICS, Sweden, VTT Technical Research Centre of Finland, and the Department of informatics at the University of Oslo (Ifi/UiO), Norway. The project grant period was from June 2006 to June 2010.

The project grant is used by researchers at the participating institutions, namely Svetlana Boudko at NR; Tiia Sutinen at VTT; and Dr. Ian Marsh at SICS to perform their PhD studies. Supervisors are Dr. Wolfgang Leister at NR, Dr. Carsten Griwodz and Dr. Pål Halvorsen at Ifi/UiO, and Dr. Marko Jurvansuu and Jyrki Huusko at VTT.

During the course of the project, the participating organisations collaborated scientifically by organising four project workshops, authoring joint articles to scientific conferences and journals, and supporting three scientists in their PhD studies.

In the ADIMUS project, we addressed the problem of enhancing the quality of multimedia streams from the service provider to the end user and trying to enhance the end user's quality of experience. For this we developed the ADIMUS architecture. The ADIMUS architecture divides the network into the parts: the "overlay network" and the "multi-access network". Within these parts, we concentrated our research on adaptation mechanisms.

The authors are grateful for the support of the NORDUnet3 programme for providing grants, and giving us the opportunity to work on this very interesting subject.

The authors want to thank Trenton Schulz and Knut Holmqvist for proofreading and discussions while preparing this report.

Contents

Executive Summary

This document is the final scientific report of the ADIMUS project that was funded by the NORDUnet3 programme. This report is based on the scientific papers published during the project, and puts these papers in the context of the ADIMUS project.

ADIMUS addresses the problem of enhancing the quality of multimedia streams at run-time based on the perceived quality. We present an architecture that addresses the different requirements from end-to-end for a mobile terminal. Due to the fact that different parts of a stream delivery system have different timing requirements, the architecture comprises an overlay network in the long-distance part, while the multi-access network employs cross-layer technology. Both parts interact and use adaptation techniques shown in this report.

We also present quality estimation techniques for both audio and video streams. We discuss quality metrics for video, estimation models, and present a quality degradation model adapted to ADIMUS. The work on quality analysis of audio presents techniques that can be adapted to ADIMUS.

For the multi-access network, we discuss the requirements for improving multimedia, and especially video stream delivery in heterogeneous multi-access network environments. We present the design of our proposed cross-layer signalling architecture that is needed to realize efficient handover management and video adaptation in heterogeneous networks during handovers. In addition to traditional mobility scenarios where the video stream is received using one access link at a time, we have also investigated the case of enabling the stream reception via two or more links concurrently. The solution presented here relies on the more flexible streaming capabilities of H.264 Scalable Video Coding (SVC). We also discuss the work done in the context of optimizing scalable video delivery in IEEE 802.11e networks, including selected simulation results.

For the overlay network we developed benchmark systems for multipath streaming that will be used to compare the adaptation algorithms in the ADIMUS proxies. Both unicast and multicast with caching are considered in the benchmark and the algorithms.

In the integration of the two parts of ADIMUS, we discuss the overlay routing reconfiguration triggered by changes in the multi-access network, and the signalling protocols among the Internet protocols that are suitable for ADIMUS.

1 Introduction

The ADIMUS project addresses the problem of enhancing the quality of multimedia streams at run-time. It develops methods to configure delivery systems in order to maximise the end-user's subjective experience, the perceived quality or *Quality of Experience* (QoE), and to adapt the system in response to changes in the resource availability.

The basis for such configuration and adaptation is a model for user-experienced service quality. As a main objective, ADIMUS developed adaptation technologies for streaming multimedia content based on subjective service quality and objectively measured *Quality of Service* (QoS) values. Special characteristics of the media streams, such as voice, audio, video, general application data, and their inter-dependencies are taken into account, also looking at emerging Internet structures, such as multi-access networks.

The goal of the ADIMUS project is to implement methods that adapt to changes in resource availability in order to maximise the end-user's subjective experience. We developed quality estimates that express the perceived end-user quality. ADIMUS considered applications that aim at achieving sufficient perceived quality with the least possible resource cost, and applications that aim at maximising perceived quality under present resource conditions, ADIMUS atomated system adaptation to reach these aims. This requires metrics for perceived quality. ADIMUS adopted such metrics, taking the best possible trade-off between resource use and quality improvements into account.

The methods developed within ADIMUS exploit the abilities of media encoders and transcoders to work within parameter ranges as well as the possibility of applications to dynamically choose encoding, or employ layered approaches such as H.264 SVC. The methods exploit the variability of end-to-end protocols that can employ buffering, retransmission and forward error correction that can achieve the desired trade-off between end-to-end latency, bandwidth variation and error rate. The mechanisms also exploit overlay routing and peer-to-peer approaches that can be used to overcome resource bottlenecks, improve availability and, in some cases, provide resource guarantees.

2 Architecture

The ADIMUS architecture, described by Leister et al. (2008), aims to support end-to-end video streaming to mobile terminals, and consider entertainment content with high media quality. We are therefore concerned with data delivery mechanisms that impact quality. Since the content is streamed from live feeds/servers to the mobile devices, the interplay of backbone and access technologies is considered. To be able to deliver streamed content with reasonable quality and startup time we look into the signalling in our architecture.

In our work, we consider a delivery system that consists of multiple service providers streaming video content via IP-based networks to multiple mobile terminals. These mo-

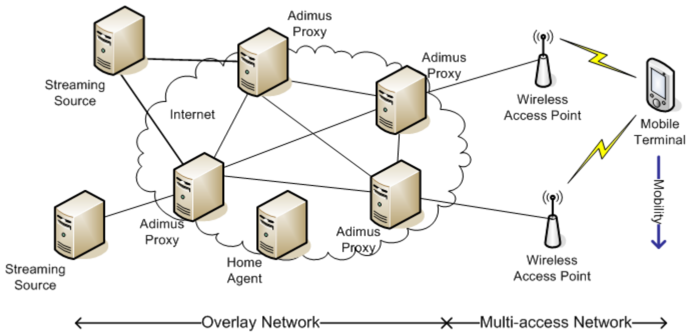


Figure 1. Overview of the ADIMUS architecture.

bile terminals may use diverse network technologies and different types of terminals to access and view the content. As the video is transmitted from a service provider to a mobile terminal, its quality is degraded by several factors that are specific to different parts of the network infrastructure. Thus, we study the degradations originating from the network and the adaptation mechanisms used as a remedy in two distinct parts: *the overlay network* in the backbone and *the multi-access network*.

The ADIMUS architecture, shown in Figure 1, comprises a delivery infrastructure based on an *overlay network*, which includes streaming source nodes at the service provider, ADIMUS proxies (AXs), and *multi-access networks* supporting mobile terminals. Thus, the ADIMUS architecture includes the following elements:

- In the backbone network, the data is routed through an *overlay network* which implements application-layer routing servers. To adapt to varying resource availability in the Internet, the overlay network monitors connections and makes application-layer forwarding decisions to change routes.
- Near the mobile terminal, a heterogeneous *multi-access network* provides application adaptation and handover mechanisms to maximise the Quality of Experience (QoE), and to support different types of mobility. A cross-layer signalling system is utilised to feed the different decision-points with continual information of the system status.
- The architecture contains QoS estimation mechanisms based on subjective (QoE) and objective (QoS) metrics from measured values.

The overlay network. The overlay network of the streaming infrastructure uses appropriate streaming protocols and source-driven mechanisms for applying quality-improving mechanisms. Streaming servers representing source nodes, and the AXs operated by service providers, are placed in the Internet to form an overlay network. Such overlays constitute fully meshed networks that allow overlay re-routing when IP-based routing cannot maintain the required QoE.

The AXs monitor network conditions using both passive and active network measurements, and they interchange information about observed network conditions. Statistical information about the observed conditions of the network is used to estimate trends in, e.g., bandwidth, latency or packet loss at a given link or path.

Terminals can initiate a session with the streaming server. Based on the overlay's monitor function of the network, the most appropriate overlay route is chosen. The last overlay node on the route always acts as a data source for the terminal. A re-invitation to the terminal is issued when route changes in the overlay involve a change in the last node. Such re-invitations are also used when route changes are triggered by reports from a mobile terminal.

Furthermore, ADIMUS supports multipath streaming to provide failure resistance and load balancing. Usually, the overlay nodes connected to the multi-access network use different access links for the different IP addresses of each of the mobile terminal's wireless devices. Aside from cross-layer information provided for faster reaction to changes at one of the wireless links, the multipath support is implemented entirely at the application level. This implies that multipath streaming is only possible to mobile terminals that run application-aware application layer software that handles buffering and reordering.

The multi-access network. The multimedia streaming end point at the terminal-side is the mobile terminal, which resides in the multi-access network consisting of different IP-based access networks of different technologies and managed by different parties. The mobile terminal is equipped with multiple network interfaces, and has support for IP mobility protocols, such as Mobile IP (MIP) (Perkins, 1998). The mobile terminal is thus capable of roaming between IP networks. In the case of MIP, route optimisation needs to be supported for the mobile terminal to be able to update its mobility information directly to the correspondent node, i.e., the AX.

The multi-access network environment allows for using handovers as a means for maintaining the QoE. Specifically, the mobile terminal is capable of selecting an alternative access when the current link does not meet the minimum QoS requirements of our video streaming service. To make an informed handover decision, the mobile terminal collects and utilises information related to the available access options' characteristics obtained through a cross-layer mechanism. Candidate cross-layer signalling frameworks for our work are the IEEE 802.21 Media Independent Handover (MIH) framework (IEEE, 2009) and the triggering framework presented by Makela and Pentikousis (2007).

In addition, to minimize the effects of transient congestion to multimedia transmission and QoE in a wireless access link, link level adaptation is used. In general, video streams, and especially scalable video streams, contain packets that have a differing impact on the decoded video quality. That is, the video streams can tolerate some packet loss (depending on the decoder), and loosing certain types of packets has a smaller impact on the decoded video quality than some other packet types (e.g., I frames vs. P and B frames). In the case of Scalable Video Coding (SVC) (Schwarz et al., 2006), the video can be decoded as long as the base layer is received correctly. Thus, it is possible to remove excess enhancement layers from the stream on the fly without affecting session continuity; the only effect is on the QoE. In the multi-access part of the ADIMUS architecture, link level adaptation is used to ensure that the terminal receives at least the most important frames (i.e., base layer) under poor link conditions.

Scalable Video Coding. For video streaming services, some of the optimizations supported by the ADIMUS system assume that the video has been encoded using H.264 Scalable Video Coding (SVC) standard (Schwarz et al., 2006). This is because SVC has been designed especially with heterogeneous networks and terminals in mind.

SVC provides the ability to include different representations of a video into a single encoded bit stream to serve a wide range of terminals with different requirements. Today's (mobile) users want to access the same services with diverse terminal devices ranging from high-definition TV sets to small handheld devices and over a wide range of access network technologies with differing characteristics. An SVC encoded bit stream has a layered structure that can be used to reconstruct the desired video presentation by simply discarding certain parts, referred to as enhancement layers, from the bit stream. The base layer that provides the basic video quality is always needed for decoding. The enhancement layers are used to enhance either the temporal resolution (frame rate), the spatial resolution (video dimensions), or the fidelity (SNR) of each picture of the video bit stream.

In SVC bit streams, each picture is divided into two or more units called Network Abstraction Layer Units (NALUs). The exact number of NALUs per picture depends on, among other things, the number of enhancement layers. SVC introduces a three-byte extension header to the H.264/AVC NALU header and this extension contains information about the SVC layer the NALU belongs to as well as the type of scalability used. In addition to picture NALUs, SVC bit stream includes NALUs that contain meta-information, namely Parameter Set (PS) NALUs and Supplemental Enhancement Information (SEI) NALUs.

The adaptation of SVC streams is achieved by filtering out enhancement layers to produce the desired video presentation. The filtering can be done in different parts of the end-to-end path, i.e., the server, a media-aware networ entity (MANE) or the terminal, with some restrictions. In addition to intra-layer prediction, each enhancement layer predicts from the layer of the same scalability dimension beneath it that needs to be taken into account. Moreover, to avoid missing reference NALUs, filtering of spatial or temporal enhancement layers can only start at the beginning of a GOP. However, filtering of fidelity enhancement layers can be started at every frame thanks to key frames, which are introduced into the fidelity enhancement layer(s) to provide a resynchronization point for the decoder in order to limit decoder drift. Therefore, SVC can support both coarse-grain scalability (CGS) and medium-grain scalability (MGS) for quality-scaled streams.

Finally, the SVC base layer is compatible with H.264/AVC, and is therefore playable by existing AVC players that just ignore the enhancement layers.

3 Quality of Experience

Mechanisms in the ADIMUS architecture (Leister et al., 2008) require the measurement or estimation of the Quality of Experience (QoE) for decision making in the adaptation processes. We take a closer look into such methods for video streaming, and for audio in real-time voice communication.

3.1 Quality Metrics for Video

Quality of Experience (QoE) is a subjective measure from the user's perspective of the overall quality of a service. QoE in the context of ADIMUS is concerned with users' experience of video delivery from servers through several networks to mobile devices up to and including its rendering on these devices. It depends on the technical performance of servers, networks, and devices involved in the streaming process. The QoE is also terminal-dependent, influenced by the environment in which content is displayed. The QoE is related to particular user expectations, the nature of the content and the user's intentions. Not all contributing factors to QoE can be taken into a account in a real system. ADIMUS can only make decisions based on:

A set of measurable, environmental parameters. In a practical system it is impossible to take users' intentions into account. It is to some degree possible to adapt to users' context, which provides a reasonable estimation of a user's intention. However, a simple example demonstrates that intention cannot be predicted in all cases: without additional systems such a eye-tracking equipment or brainwave scanners, ADIMUS cannot guess whether a football commentator is watching re-runs of a game with a focus on a particular player (where zooming in and high quality would increase QoE), or on the cooperation of a team (where zooming out and high framerate would increase QoE). Therefore, ADIMUS restricts itself to environmental parameters provided by network measurement and sensors on the user's device.

A finite, countable number of states. Experts who have created standards for video quality metrics are aware of the limitations of current standards and the limitations of the state of the art. In a recent interview by Mu (2009), these experts warn that existing metrics are known to be stable only if environmental conditions, encoding formats, display formats, video content, evaluation procedure and statistical method are kept unchanged. The stability of these procedures and their wider applicability is subject to ongoing research. Given the current state of the art, QoE could only be "predicted" by creating a database of contents and situations. This is, of course, infeasible. Therefore, ADIMUS must use estimation.

The model that was chosen for ADIMUS is based earlier work in the MOVIS project[1], refined by Leister et al. (2010). Concerning video transmission on the Internet, Leister et al. (2010) present a model for estimating the subjective quality from objective measurements at the transmission receivers and on the network. This model reflects the quality degrad-

1. For information on the MOVIS project we refer to `http://movis.nr.no`, last accessed August 19, 2010.

ation subject to parameters like packet loss ratio and bit rate and is calibrated using the results from subjective quality assessments. Besides the model and the calibration, the main achievement of this paper is the model's validation by implementation in a monitoring tool. It can be used by content and network providers to help swiftly localise the causes of a possibly poor quality of experience (QoE). It also can help content providers make decisions regarding the adjustment of vital parameters, such as bit rate and other error correction mechanisms.

3.1.1 Estimation Models

Existing methods for QoS measurement can be classified into network and application level measurements. Examples of widely used metrics for network level QoS include connectivity, one and two-way delay, jitter, throughput, as well as packet loss, as used in the initiatives of the IETF IPPM (IP Performance Metric) working group, such as RFC 2330 (Paxson et al., 1998). For measuring the QoS, there are several approaches, which can be classified based on whether they are subjective or objective, direct or indirect, in-service or out-of-service, real-time or deferred time, continuous or sampled, intrusive or non-intrusive, and single-ended or double-ended. Existing work that claims to derive QoE values directly from network QoS frequently applies peak-signal-to-noise ratio (PSNR) as its metric for translating QoS to QoE. There are very strong arguments against this practice, for example presented by Huynh-Thu and Ghanbari (2008) and by Ni et al. (2009). The picture appraisal rating for MPEG-2 (PAR) by Knee (2006) assesses the quality of indivual pictures instead of videos as well and suffers the same problems as PSNR. A very simple approximation in this general family of approaches was proposed by Koumaras et al. (2009). Rate-distortion models as introduced by Chiang and Zhang (1997) are also based on PSNR, which constitutes a problem, but are less affected by unstable quality because quality is adapted by changing quantization factors over time.

The just noticable differences (JND) provides an objective metric that tries to emulate the human visual apparatus in a way that considers the change of pictures over time, as do approaches like the Video Quality Metric (VQM) (Wolf and Pinson, 2002). Several experts explained the applicability and pitfalls of these techniques in an interview with Mu (2009).

3.1.2 Quality Degradation Model

The ADIMUS architecture is designed to deliver adaptive multimedia streams transferred from one or more servers to multiple terminals. In Figure 2, we show the architecture consisting of a delivery infrastructure based on an *overlay network*, which includes streaming source nodes at the service provider, ADIMUS proxies, and *multi-access networks* supporting mobile clients. We also show which parts of the network have an impact on the QoE for the end user, denoted by the M_i.

The ADIMUS architecture has similarities with the scenario used to develop the estimation model by Leister et al. (2010). Streamed data are transported from the content provider through the overlay network and the multiaccess network to the device where the

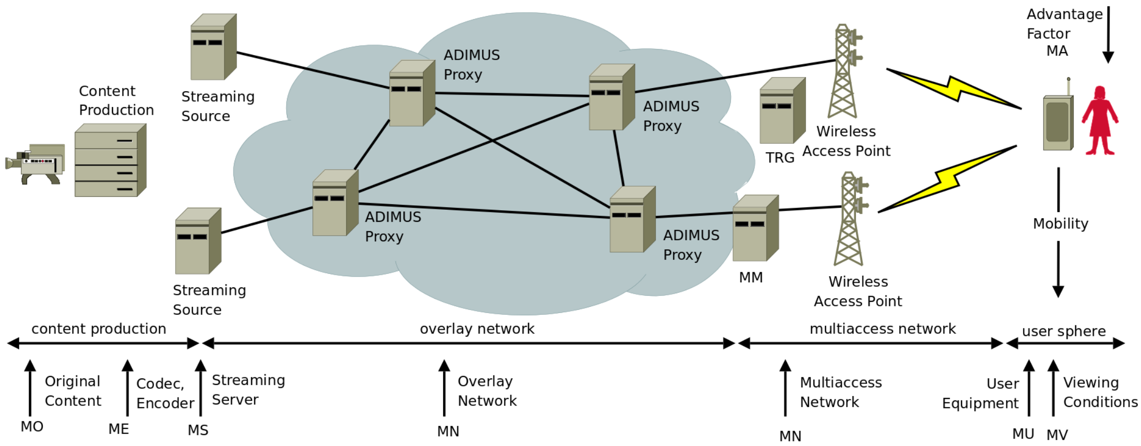


Figure 2. Transmission chain in the ADIMUS architecture. entities which have an influence on the consumer's QoE are identified.

stream is presented. Each entity along this chain can potentially mean a decrease in quality as compared to the original quality Q_{O}. Therefore, we apply the above mentioned estimation model for subjective QoE to ADIMUS, both for single consumers based on factors reducing the original quality of content, and for groups of consumers based on the APDEX measure described by Sevcik (2005).

QoE for one consumer. We estimate the quality perceived by the end user as a product of the original quality Q_{O}, and a number of influencing factors. Each factor is related to a certain entity of the delivery chain in our scenario and hereby represents the respective impact on the consumer quality. Thus, the estimated QoE for one consumer is defined as

$$\tilde{Q} = Q_{\text{O}} \cdot \prod_{i \in \{\text{E,S,N,U,V,A}\}} M_i,$$

where Q_{O} is the original quality measure, $M_{\text{A}} \geq 1$, and $0 < M_i \leq 1$ for $i \in \{\text{E,S,N,U,V}\}$. Setting $M_i = 1$ denotes the lack of influence, such as a transparent channel. In the following we describe each single factor:

M_{E}: Influence of the encoding on the delivered content. It depends on the codec and the content (fast/slow movements, colour, contrast, etc.).

M_{S}: Influence of the streaming server on the delivered content. It depends on the streaming protocol, the implementation of the streaming server and, to a certain extent, on the codec.

M_{N}: Influence of the network on the delivered content. The quality is determined by technical parameters like delay, jitter, congestion, packet loss, and mis-ordered packet arrival. It also depends on codecs and protocols. The parameter consists in ADIMUS of two distinct parts: the overlay network, and the multiaccess network.

M_{U}: Influence of the consumer's equipment on the delivered content. Hardware type and parameters (e.g., CPU speed, memory size), system and application software, and a system load parameters have an influence on M_{U}.

$M_{\mathbf{V}}$: Influence of the viewing conditions.

$M_{\mathbf{A}}$: Advantage factor from the use of the content, modelling cognitive effects like the acceptance of worse quality when consuming the content in a mobility situation. This increases the value of the content subjectively, even if the technical QoE is worse.

Note that in the general case some of the factors are not orthogonal, i.e., they depend on the impact factors of previous steps. For instance, a particular networking error can be visible in different manners for different codecs or bandwidth settings.

The concrete factors M_i must be derived from an objective measurement, here called assessment process, and mapped/scaled to the allowable range of values as defined above. Ideally, a model calibration process using regression analysis could be used to derive the scale factors for calculating $\tilde{Q}$. Since performing such an analysis requires a large data set from assessments we, instead, simplify the model and select only significant parameters.

Simplified model and parameter scaling. Parameters that are outside the control of ADIMUS are set to a constant value, e.g., are set to 1 if the parameter is supposed to be transparent. Parameters out of scope include M_{V}, and M_{A}, and partially M_{U}.

M_{E} and M_{S} depend on encoding and streaming parameters. Since ADIMUS, at the moment, does not perform transcoding in the ADIMUS proxies, we combine both factors to $M_{\mathrm{E,S}}$, which is determined by an assessment process shown by Leister et al. (2010) in Section IV.

M_{N} depends on the overlay network and the multiaccess network, which results in measurable delay, jitter, and packet loss. For the QoE of the consumer, the resulting packet loss is most important. Modern media players typically employ some form of buffers, such that jitter and delay eventually will result in packet loss due to these packets arriving too late for decoding. Therefore, we consider packet loss as the relevant parameter for M_{N}. We derive M_{N} from an assessment that evaluates the influence of packet loss on the QoE. Under varying network conditions, the influence of packet losses on the perceived QoE is measured, and the resulting score is scaled to fit the range (0,1].

QoE for groups of consumers. The estimation method by Leister et al. (2010) can be applied to ADIMUS also for groups of consumers. In this case, we calculate the value A_{M} at the ADIMUS proxies. We apply the APDEX model (Sevcik, 2005) as follows; we classify the consumers into three quality classes according to their current $\tilde{Q}$. Given the threshold values T_{S} and T_{U}, the consumer is in the set $M^{(\mathrm{S})}$ for $\tilde{Q} > T_{\mathrm{s}}$, in the set $M^{(\mathrm{T})}$ for $T_{\mathrm{s}} > \tilde{Q} > T_{\mathrm{u}}$, and in the set $M^{(\mathrm{U})}$ for all other cases. The threshold T_{S} is suggested to be at 60-80% of the maximum $\tilde{Q}$, and T_{U} at about 40%, depending on the expectations of the consumers.

We then apply the formula

$$A_{\mathrm{M}} = \frac{\left|M^{(\mathrm{S})}\right| + \left(\left|M^{(\mathrm{T})}\right|/2\right)}{\left|M^{(\mathrm{S})}\right| + \left|M^{(\mathrm{T})}\right| + \left|M^{(\mathrm{U})}\right|},$$

and rank A_{M} into U (unacceptable), P (poor), F (fair), G (good) and E (excellent) with the threshold values $0 \leq \{U\} \leq 0.5 < \{P\} \leq 0.7 < \{F\} \leq 0.85 < \{G\} \leq 0.94 < \{E\} \leq 1$. These values can be used for decision making in the ADIMUS proxies.

Since the ADIMUS proxies can employ multipath streaming to the multi-access network (which currently employs single-link streaming at the client) we need to extend the calculation of A_{M} by taking only the portion of the stream into account that is available at the ADIMUS proxy in question.

3.2 Quality Analysis of Audio

The work on real-time voice addressed topics within real-time voice communication. The focus was on the *quality aspects* of voice communication, since poor quality often leads to user dissatisfaction. The techniques presented in the dissertation by Marsh (2009) partially solved the research problems independent of ongoing network QoS efforts.

Each of the thesis's publications draws similar conclusions, that is, reasonable quality Internet telephony can be offered provided that the whole system is carefully engineered. This implies the introduction of mechanisms to preserve the subjective quality when impediments are, or are about to, occur. Some of the conclusions from our research include that the network load should be controlled for links that carry real-time voice. This means providing and dimensioning links with sufficient capacity, or alternatively, restricting the admission of voice calls to heavily loaded links. The monitoring of network conditions, in particular loss, should be used to signal potential quality problems on particular paths. We have presented a solution where the end system can do the monitoring where such network functionality is absent. Questions such as, should we require earlier indications of impending problems, tracking the network delay or jitter at the end system can be investigated. This technique has been used in our handover studies, where several network parameters have been combined to schedule a handover. Continuing in the wireless case, we have proposed mechanisms for maintaining quality by switching to lower data rates, or even switching to an alternative technology where available.

Since the scope of this work is broad, we have taken different cuts through IP telephony research by looking at the access and backbone networks, using modelling, simulation and experimental techniques; we have considered both fixed and wireless networks using subjective and objective quality tests to obtain the most appropriate solution for a particular problem. We have also looked at systems with and without background traffic, used real-time and off-line techniques, and finally applied cross layer approaches that combine normally separated layers of the protocol stack. The main contribution of this work is a near-complete system study concerning quality aspects of an Internet telephony system. We have looked at a number of different methods to enable adaptable solutions for maintaining acceptable quality. We have often found that relatively simple changes can lead to substantial user quality gains.

The tangible outcome of our research has been a number of software tools. These include an IP based voice measurement package, a handover algorithm for wireless terminals, a

VoIP traffic generator and a PESQ processing package.

The work presented in this thesis is complementary to the work presented in the rest of this report. That is, the voice work concentrates on the data stream, whilst some of the other work looks at establishing sessions with good quality. Many of the techniques presented in the thesis can extend to video. Voice and video need an architecture within which to operate that is described in the next section. Depending on where voice and video is sent and received, we will need to study the access and backbone networks. There are sections specifically on these technologies later in the report.

4 The Multiaccess Network

In this section, we discuss the work done with *Multi-access Support for Adaptive Applications*. The discussion is structured as a review of the publications produced in this work package during the project. These include six conference papers: Leister et al. (2008); Luoto and Sutinen (2008); Mäkelä et al. (2009); Piri et al. (2009); Sutinen and Frantti (2008); Sutinen and Huusko (2011), of which the EUMOB paper (Mäkelä et al., 2009) was extended to a journal article (Mäkelä et al., 2010); three related conference papers produced in the context of other projects: Backman et al. (2008); Dousson et al. (2007); Piri et al. (2010); as well as one demo paper (Sutinen et al., 2010) that will be published in the proceedings of the SVCVision workshop co-located with the Mobimedia'10 conference. These papers give an overview of the topics addressed in the project related to the multi-access network part of the ADIMUS system. These include the supported cross-layer signalling architecture and the network protocols used for the communications (i.e., stream delivery, mobility management, and signalling).

The theoretical work introduced in the papers has been backed up with real world implementations and simulations of the proposed solutions. The demonstrator building related to multi-access networks was conducted in co-operation with VTT's two other related projects: Tekes MERCoNe and Celtic SCALNET.[2] The OPNET simulator was used to model the multi-access network, related to scenarios discussed by Sutinen and Frantti (2008). Due to the vast amount of prior work published by other researchers in the field of optimizing multimedia delivery in multi-access networks, and especially during handovers, the main scientific contribution obtained from the OPNET simulation study is related to optimizing scalable video delivery in the medium access control (MAC) layer of a wireless network or link. Thus, we focus on presenting the results obtained from this study in the following subsections. We also mention briefly another link-level video adaptation related study that was performed using the OMNeT++ simulator in the context of the ICT-OPTIMIX project[3].

2. For the project Tekes MERCoNe we refer to `http://research.nokia.com/research/projects/mercone`, last accessed August 19, 2010; for the project Celtic SCALNET we refer to the project leaflet at `http://www.celtic-initiative.org/Projects/SCALNET/`, last accessed August 19, 2010.
3. Project web site: `http://www.ict-optimix.eu`, last accessed August 24, 2010.

The organisation of the rest of this section is as follows: First, we discuss the requirements that have been identified for improving multimedia, and especially video stream delivery in heterogeneous multi-access network environments. Then, we present the design of our proposed cross-layer signalling architecture that is used to realize efficient handover management and video adaptation in heterogeneous networks during handovers. In addition to traditional mobility scenarios where the video stream is received using one access link at a time, we have also investigated the case of enabling the stream reception via two or more links concurrently. The solution presented here relies on the more flexible streaming capabilities of H.264 Scalable Video Coding (SVC) (Schwarz et al., 2006). Finally, we discuss the work done in the context of optimizing scalable video delivery in IEEE 802.11e networks, including selected simulation results.

4.1 The design of a mobility-enhanced multimedia service according to proposed reference architecture approach

Heterogeneous multi-access networks pose new challenges for service development as supporting QoS-sensitive multimedia applications in such a dynamic environment requires additional mechanisms. Such mechanisms supporting session continuity need to be designed so that they respect the existing communication principles. Only this way they can be sustainable solutions. Sutinen and Frantti (2008) introduce a reference architecture and approach for multimedia services in heterogeneous multi-access networks. The proposed architecture encompasses the requirements set for the service by related domains, such as the Internet architecture. The reference architecture is validated with a session continuity example that follows a specific reference approach.

In heterogeneous networks, handovers can occur in different dimensions, e.g., within one access technology or between different technologies (horizontal and vertical handovers) and within one network domain or between different domains (intra- and inter-domain handovers). Also the client device may change in a handover. Handovers in heterogeneous networks introduce significant variance to the network connection and client device characteristics, including factors like the available bandwidth, delay, and display resolution. Thus QoS-sensitive multimedia applications require additional support in order to maintain their operation in the presence of heterogeneous networks and mobility.

For multimedia services that are to be used in today's Internet, Sutinen and Frantti (2008) identified four domains that need to be taken into account in the process: the administrative, user, application, and Internet architecture. For service design, each of the reference architecture domains poses a specific set of requirements or limitations that need to be satisfied. That is, there are a number of requirements demanded by administrative or regulative sources, like transmission power level and uninterrupted or continuous channel usage time. Applications itself set their unique requirements, e.g., application protocol interfaces to the communication protocols. Applications also define allowable delay, loss, and throughput limits. Users set a number of requirements in terms of usability, reliability, and security for applications/services. The QoE is solely seen from the user perspective. Finally, all the requirements or their implementations have to be designed to be in line with the existing networking (Internet) architecture. The purpose of the reference architecture is to enable overall improvement of the application/service implementation by avoiding to make proprietary solutions that may serve a valuable short-term purpose but significantly impair the long-term flexibility of the Internet.

Additionally Sutinen and Frantti (2008) derive the requirements for a mobility-enhanced video streaming service as a validation of our reference architecture. The analysis includes a narrative scenario description and five selected use cases. In addition, a sequence diagram is used to illustrate the time dynamics of the scenario. Based on the scenario and use cases we were able to collect a set of requirements and to discuss them against the existing Internet architecture. The final result of the analysis was a concise and unambiguous presentation of the requirements for the service being designed.

The requirements identified for a mobility-enhanced video streaming service were as follows: *a*) One or more access networks needs to be simultaneously available. *b*) The application needs to be able to adapt to varying network QoS. *c*) Video needs to be adapted to the terminal (e.g., screen size). *d*) Mobility management should be realized using Host Identity Protocol (HIP) or Mobile IPv6 (MIPv6). *e*) Multihoming and simultaneous multi-access support should be realized using HIP or MIPv6 extensions. *f*) Proactive connection deterioration detection needs to be enabled through extensive monitoring, including streaming performance (e.g., throughput and jitter) as well as access link status (e.g., signal strength). *g*) The access selection algorithm of the mobility manager needs to be both QoS- and application-aware. *h*) Cross-layer signalling and information collection through standard mechanisms, such as the ones provided by IEEE 802.21 Media Independent Handover Services (MIH). *i*) Unnecessary handovers should be avoided to reduce overhead. *j*) The consequences of a handover to applications (e.g., temporary connection outage or reduction in network bandwidth) should be estimated in advance, if possible, to avoid unnecessary handovers.

The derived requirements were utilised as guidelines when designing the multi-access network part of the ADIMUS system architecture described in Section 2, as well as by Leister et al. (2008).

4.2 Cross-layer signalling architecture

It is difficult for state-of-the-art multimedia applications to fully utilise the potential of heterogeneous multi-access networks as host mobility (and the necessary handover management that this leads to) undermine QoS. This is because major effects caused by the varying network QoS need to be handled in order to maintain a high-level user experience, for example, in video delivery. Currently, it is difficult for applications to take sufficient corrective measures (such as adaptation) to cope with handover-induced QoS changes as today's layer 2 and 3 (L2/L3) mobility management solutions aim at hiding the mobility effect completely from higher layers. Moreover, in many cases, the handover decision-making of current mobility protocols is suboptimal for multimedia applications as the decisions are solely based on lower layer information (e.g., RSS measurements) and do not take into account the applications' requirements for network access.

Cross-layer signalling (i.e., signalling between entities located on non-adjacent protocol layers) is needed to provide the required awareness to mobility managers and adaptive applications regarding the system status. In ADIMUS, the solutions that can be applied for cross-layer signalling have been discussed in several papers. The issue has also been studied in different viewpoints, taking into account the requirements of mobility management and adaptive applications. Backman et al. (2008); Dousson et al. (2007); Luoto and Sutinen (2008); Mäkelä et al. (2009); Mäkelä et al. (2010) mostly focus on the optimisation of mobility management and handover performance using cross-layer signalling, whereas Piri et al. (2009) focus on the multimedia adaptation viewpoint on the issue. For the ADIMUS architecture, both aspects are important as improving handover performance (i.e., reducing the handover delay and the resulting packet loss) through low-level triggers improves the user-perceived quality of the mobility-enhanced video streaming service. However, efficient adaptation of a video stream to changed network conditions after a vertical handover is crucial to ensure session continuity as well as optimized QoE to the user. This can only be done if the application is made aware of the handovers via a cross-layer signalling system.

The cross-layer signalling enhancements supported by the ADIMUS system build on top of a cross-layer signalling framework, like the one described by Piri et al. (2009). The signalling framework allows collecting and distributing information between system elements (local or remote). The solution makes use of the Triggering Engine (TRG) described by Makela and Pentikousis (2007) in distributing information between application, network and link layer entities in the mobile terminal as well as between the mobile terminal and remote network nodes (e.g., AXs). In specific, the Triggering Engine provides, for example, timely information exchange between the video streaming application and the mobility client, a.k.a. the Mobility Manager (MM), in order for them to support efficient video adaptation in the presence of handovers as well as application-aware handover decision-making.

The MM is responsible for the handover decisions that are then enforced by the Mobile IP protocol. As shown by Luoto and Sutinen (2008); Mäkelä et al. (2009); Mäkelä et al. (2010), the handover performance of standard Mobile IP can be improved significantly by utilising additional link status information from the link layer. This is because the handover decision-making supported by the standard is solely based on the router advertisements sent by the home agent (HA) and is thus incapable of detecting sudden changes in link conditions causing unnecessary handover delay. In the papers, proprietary low level triggers were used. As soon as the deployment of the IEEE 802.21 Media Independent Handover (MIH) services, published by the IEEE (2009), becomes more common, one can rely on standard MIH events to detect changes in link availability and conditions.

Finally, in order to support application-aware handover decisions, the MM needs to obtain information about the application requirements. One approach for this would be that the streaming client passes the information from the session description (SDP) it receives from the streaming server in the beginning of the session to the MM through TRG. The session description contains basic information of the video stream characteristics (e.g., the proposed bandwidth to be used by the session or media) that can be used when deciding which network interface to use. During the streaming, the MM can also make use of QoS and QoE measurement information regarding the video streaming performance to detect problems in the currently used connection.

4.3 Proof-of-concept implementation of multi-interface streaming

The fairly recent video coding standard H.264 SVC (Schwarz et al., 2006) has created interesting new possibilities for video service delivery in heterogeneous multi-access networks. The flexibility of SVC allows new communication scenarios as the video can be split into independently streamable layers. Since today's client devices often have multiple network interfaces, often with multihoming and mobility capabilities, they are able to utilise the available network connections simultaneously for accessing networked services. For streaming SVC-encoded video, an interesting new use case therefore is the ability to receive a same video stream concurrently over multiple network connections. Such multi-interface streaming capability would benefit situations where no single access network within the user's reach is capable of supporting the whole video stream or when the service provider wants to make only the base quality stream available to all users (e.g., via DVB-H) but offers a quality enhancement for paying customers via some other connection (e.g., WLAN). To demonstrate the potential of such a streaming system we have built a prototype implementation of multi-interface streaming for SVC presented by Sutinen et al. (2010).

The proof-of-concept demonstrator introduced by Sutinen et al. (2010) has two intercommunicating modules: an SVC streaming server which is implemented based on the Darwin Streaming Server (DSS) and a terminal which is a modified version of the MPlayer with integrated OpenSVCDecoder decoding tools for SVC support. The server binds to several IP addresses on the host on which it runs (possibly using virtual interfaces, if needed). Each SVC layer is then served from a different address. The mapping of SVC layers to server addresses advertised to the terminal via SDP (Handley et al., 2006). The video used in the demonstrator has two quality-scaled layers and the different layers are sent as separate RTP streams to the terminal. The terminal has two network interfaces (e.g., HSPA or DVB-H and WLAN), which are used concurrently to receive the video stream. The terminal orders the desired layers from server via RTSP and directing the requests through selected access networks. The simplest way to achieve the desired mapping of SVC layers to access networks in the client side is through altering the client's routing tables. During the streaming, the client synchronises the playback of the enhancement layer to the base layer, and the two sub-streams are combined and forwarded to the decoder as one continuous stream. To demonstrate dynamic dropping and adding of the enhancement layer, the client can execute a layer switch in the middle of the streaming by sending PAUSE and PLAY RTSP messages to the enhancement layer interface of the server. In the case of DVB-H, which is a unidirectional multicast link, the base layer stream setup is somewhat different. In this case, end-to-end signalling cannot take place in the session setup, but the client orders the base layer sub-stream by joining the corresponding multicast group.

With this demonstrator we showed how scalable video coding and the multi-session capability of RTP allow – with simple routing adjustments – for utilising multiple network interfaces in the stream reception simultaneously to maximise the QoE of the video service. As future work, we plan to test the prototype in various networking scenarios to verify its operation. In addition, we intend to investigate and develop more advanced mechanisms for synchronising the SVC sub-streams in the client side as well as for intelligent interface selection (using cross-layer information) and dynamic routing for the client.

4.4 Link level adaptation of scalable video streams

End-to-end congestion control algorithms performing bit rate adaptation on video streams, such as the TCP Friendly Rate Control (TFRC) by Handley et al. (2003), cannot always react fast enough to changes in access link conditions. Transmission characteristics of wireless access links may vary in response not only to handovers but also to factors like congestion, signal fading or multi-path propagation. The inability of TFRC to react to such fast changes is because it relies to estimates of the loss rate, round-trip time, and throughput as smoothed averages instead of the actual measured values. Additionally, sending the end-to-end feedback messages causes additional delays. Thus, it takes time for changes in the link conditions to take effect in the stream bit rate and TFRC therefore can induce transient congestion into a bottleneck link of a network.

In order to minimise the effects of transient congestion to video streaming and the user's QoE in the wireless access link, link level adaptation can be used. In general, video streams, and especially scalable video streams, contain packets that have a differing impact on the decoded video quality. That is, the video streams can tolerate some packet loss (depending on the decoder) and loosing certain types of packets has a smaller impact on the decoded video quality than some other packet types (e.g., I frames vs. P and B frames). In the case of H.264 SVC, the video can be decoded as long as the base layer is received correctly (some packet loss can be tolerated here as well, depending on the decoder). Thus, it is possible to remove excess enhancement layers from the stream on the fly without affecting session continuity; the change affects only the QoE and its impact depends on which type of scalability is being used (i.e., spatial, temporal or SNR). In the ADIMUS architecture, link level adaptation is used to ensure that the terminal receives at least the most important frames (i.e., base layer) under poor link conditions.

The adaptation in the link level is achieved via prioritising more important video packets over less important ones as well as applying selective packet drops in case of congestion. Link level adaptation can be implemented through a multiple queue system in the MAC level as presented by Piri et al. (2010). The use of multiple queues makes it possible to apply prioritised scheduling to video and other traffic. Although shown to be fair to other types of traffic than video, this solution is not fully compliant to existing MAC-level QoS architectures. To address this issue, Sutinen and Huusko (2011) have proposed another scheduling solution. Here, a generic hierarchical MAC-level QoS architecture is proposed that can be used to implement both inter- and intra-traffic class QoS needed for optimal and fair scalable video transmission.

5 The Overlay Network

In this section, we discuss the work done related to the overlay network employed in the ADIMUS architecture. The discussion is structured as a review of the publications produced, as well as further elaborations not yet published. The conference papers by Boudko et al. (2008) and Boudko et al. (2010) both show benchmarking systems that can be used to evaluate algorithms that are currently developed for adaptation and routing decisions in the overlay network.

For adaptations in the backbone delivery of streams, we consider a multipath streaming system that is built upon a network of overlay servers. The nodes of this overlay network are able to connect directly to each other using the Internet meaning that they form a fully meshed overlay network. Our scenario considers a situation where several multimedia providers use the overlay network to stream video content to end-user terminals.

The advantage of using an overlay network for multisource, multipath streaming under these conditions is that streams can be split up and re-routed to improve the total amount of bandwidth that is available for streaming from the servers to every single client, even when individual overlay links have insufficient bandwidth. Clients that do not suffer from bottlenecks in the access network are likely to compete for the bandwidth on at least some overlay links. To improve the quality of multimedia streams delivered to clients over the shared Internet infrastructure, algorithms have been proposed that exploit multipath delivery by splitting streams between different paths (Chu and Nahrstedt, 1997; Golubchik et al., 2002).

However, one cannot evaluate how good these strategies are in terms of optimal resource utilisation when several multimedia senders and receivers share the same delivery infrastructure and compete for the same network resources. Therefore, we propose a benchmarking system that provides us with the best possible distribution of the streams over available delivery paths given that the complete knowledge of the network, including its topology and resource availability, is obtainable. The benchmarking system can then be used to quantify the difference between the optimal solution and solutions provided by algorithms that operate in dynamically changing networking environments with partial knowledge of the network.

In our work, we consider unicast and multicast scenarios separately. The benchmarking systems and the multipath algorithms for these scenarios are presented below.

5.1 Unicast Benchmarks

To build our benchmarking system, we model a network that includes senders, receivers and overlay nodes as a graph $D = (V, A)$, where V is the set of vertices that represent the nodes of the network, and A is the set of arcs that represent the links between the nodes.

$$b : A \to R_0^+ \tag{1}$$

$$b(p) = \min\{b(a)\}, a \in p \tag{2}$$

$$\delta(a,p) = \begin{cases} 1, \text{if} & a \in p \\ 0, \text{if} & a \notin p \end{cases} \tag{3}$$

$$\max \sum_{p_{i,j}^k \in P} f(x_{i,j}^k) \tag{4}$$

$$\forall\{i,j\} : \sum_{k=0,K_{i,j}} x_{i,j}^k \cdot r_{i,j} \geq r_{i,j}^b \tag{5}$$

$$\forall\{i,j\} : \sum_{k=0,K_{i,j}} x_{i,j}^k \leq 1 \tag{6}$$

$$\forall\{a\} : \sum_{p_{i,j}^k \in P} x_{i,j}^k \cdot r_{i,j} \cdot \delta(a, p_{i,j}^k) \leq b(a) \tag{7}$$

$$\forall\{p\} : x_{i,j}^k \cdot r_{i,j} \leq b(p_{i,j}^k) \tag{8}$$

Table 1. Equations used in the benchmark model for the multipath unicast case.

The sets $V_s, V_o, V_r \in V$ are disjoint subsets of vertices representing respectively sender, overlay and receiver nodes. In the graph D, we define a set of all possible paths between the receivers and the senders. This includes the direct paths between the senders and the receivers and the paths that are constructed in the overlay plane using all possible permutations of the overlay nodes.

Then, $p_{i,j}^k$ denotes the k-th path in the set of paths connecting the sender i with the receiver j. $K_{i,j}$ denotes the number of paths from the sender i with the receiver j.

To model network resources and constraints, we have used the formulas defined in Table 1. First, we define the bandwidth function on the underlay arcs that expresses the available bandwidth of the arc a using Eq. 1. The same function is defined on the paths where the available bandwidth of a path p is defined as the lowest bandwidth among all arcs that belong to the path p (Eq. 2). Then, for each path p and each arc a, we define a function δ in Eq. 3.

Furthermore, the streaming requirements of the requested streams are defined by a matrix R, where an element $r_{i,j}$ represents the required bitrate at which the stream from the sender v_s^i is streamed to the receiver v_r^j. In addition, we define the matrix R^b which is similar to the above matrix R, but contains the bitrates for the base layers. The variable $x_{i,j}^k$ denotes the share of the multimedia stream that is sent from the content provider v_S^i to the receiver v_R^j through the path k, and with this we can define the objective function f in Eq. 4.

This function represents two utility functions. The first one is a linear function that maximises the bandwidth assigned to all streams in the system. The second one is a logarithmic utility function that is defined as a sum of video qualities experienced by all users of the system. To construct the logarithmic utility function, we used a very simple theoretical framework for end-to-end video quality prediction that was presented by Koumaras et al. (2009).

These objective functions are subject to the set of constraints given in Eqs. 5-8. First, in order to be able to play out the video, each session requires enough bandwidth for the base layer (Eq. 5). Second, the sum of sending rates along all paths from one sender to one receiver should not exceed the bitrate assigned to this stream (Eq. 6). Third, with respect to shared underlay links, the total sending rate must not exceed the available bandwidth of the shared link (Eq. 7). Finally, the bitrate assigned to the path should not exceed the bandwidth of the path (Eq. 8).

We applied the benchmarks to several network topologies and different clients requests. For each of these topologies, the bandwidth of the links has been degraded gradually several times, and the optimal solutions for each of these degraded trials have been computed. Our experiments show that the benchmark with the logarithmic utility function provides better bandwidth allocation in terms of achieved video quality over the whole set of users while providing the same optimal solution in terms of utilisation of available bandwidth.

When analysing and assessing resource allocation algorithms, such as rate control algorithms for multiple clients and multiple paths, an important issue to consider is fairness. Several authors have addressed this problem and defined several quantitative measures of fairness (Bertsekas and Gallager, 1987; Jain et al., 1984; Kelly, 1997; Kelly et al., 1998). From these measures we have chosen proportional fairness (Kelly, 1997; Kelly et al., 1998) to be implemented in the system. Proportional fairness is better in terms of utilisation of resources (Le Boudec, 2008), and it does not require control over all routers as the min-max fairness by Bertsekas and Gallager (1987) does. Hence, it is a better fit for the overlay system that relies on controls in the overlay nodes and cannot adjust the rate allocations of the streams at each router.

As shown by Kelly (1997), optimisation of a logarithmic function over a compact and convex region provides a proportionally fair solution. Hence, the logarithmic benchmark is also proportionally fair in allocating available bandwidth to different clients.

5.2 Unicast Multipath Algorithms

For the unicast multipath scheme, we have selected two algorithms from the literature and an algorithm designed by us. These algorithms are currently implemented in the OMNet++ environment (Pongor, 1993), and will be evaluated using the above benchmarks. Here, we give a short introduction to the selected algorithms.

The first algorithm is a simple TCP-based multipath streaming algorithm proposed by Wang et al. (2007), who assume that a multipath streaming architecture exists. There is only one pair server-receiver considered in the paper. The server opens several TCP connections to the receiver, one for each path, and stripes video content over these connections. However, the authors do not consider how the multipath infrastructure is built, and they do not study possible trade-offs between multiple streams.

The algorithm by Hefeeda et al. (2003) introduces a topology-aware peer-to-peer streaming system. Candidate peers are selected based on a proposed "goodness" metric, which is is calculated by using the packet loss and the available bandwidth between a peer and a client. The system is client-based meaning that all decisions about the peer selection are made at the client side and the interplay in decision making between different clients is not considered. The selfishness of the clients may lead to the situation where the clients that start the peer selection earlier consume the whole bandwidth.

We have also developed and implemented our own algorithm that is to be tested and compared with both the benchmarks and the two above algorithms. In this algorithm, the decision of rate allocation is done both in the overlay nodes and in the sender. The sender starts streaming over the direct path and adds an overlay path if the direct link experiences congestion. If the overlay node experiences a bandwidth reduction over its overlay links it informs the senders that the rate must be adjusted proportionally to the requested bandwidths. After receiving the notification from an overlay node the senders adjust the streaming rate over the congested overlay link to the suggested one, and to compensate the reduction include an overlay path that does not contain this link.

5.3 Multicast with Caching Benchmark

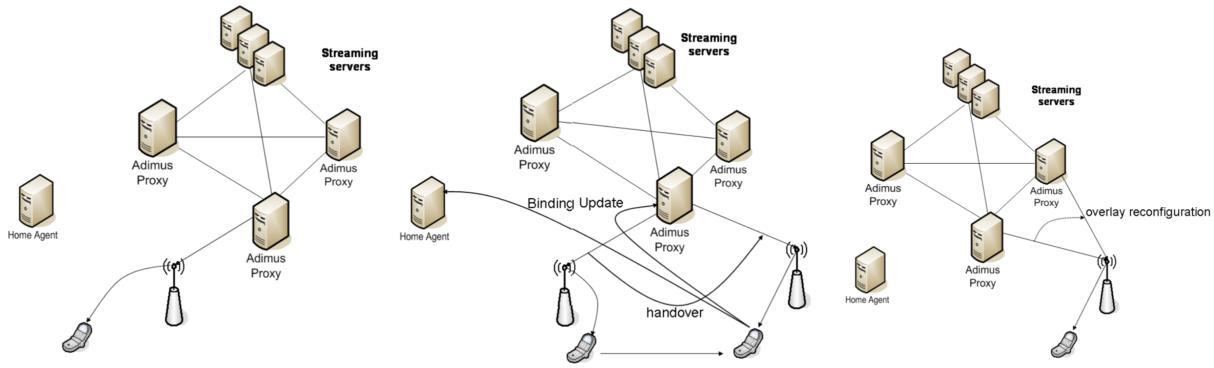
To proceed with our work, we intend to develop a benchmark for multicast multipath streaming with caching. The multicast part of this benchmark is similar to the unicast one and differs in introduction multiple trees from senders to groups of clients instead of multiple paths from senders to clients. The placement of caching should be included in the benchmark.

5.4 Multipath Algorithms for Multicast with Caching

We have studied algorithms from the literature and compared these with the benchmarks developed earlier. Ramesh et al. (2001) have proposed a multimedia streaming scheme (MCache) that combines multicast with caching. While requests are batched together to form a multicast group, the clients can start getting video from regional proxies. This implies that the requests that come later can still be served from the multicast group without any delay for this group of clients.

We considered several algorithms that use multicast without caching. SplitStream (Castro et al., 2003) is built upon an overlay network called cooperative environments and is used for multicast and content distribution. In SplitStream, multiple trees are built and used for streaming in order to balance the forwarding load. The peers are organised in trees in a way that each peer serves as an internal node in one tree and as a leaf node in other trees. This principle guarantees that the failure of one node will affect only one tree. The content is split into several stripes, which are then multicast using separate trees.

Outreach (Small et al., 2007) is another topology construction algorithm that is intended to optimise the peer-to-peer overlay construction. It can be considered as a peer-to-peer based multicast streaming scheme that maximises the utilisation of the peers' available upload bandwidth, in order to minimise the bandwidth requirements on the streaming server.



(a) Before a handover (b) Handover with route optimisation (c) Overlay reconfiguration

Figure 3. Overlay network reconfiguration when a mobile terminal roams from one access network to another.

We also considered the algorithms by Zhang et al. (2005), who present a data-driven overlay network for live media streaming (DONet). DONet adaptively forwards data over an overlay network using the proposed scheduling algorithm.

These algorithms are currently being implemented in the OMNet++ environment, and will be compared with the benchmarks developed earlier.

6 Integration

In order to present the ADIMUS architecture as a whole, the distinct parts need to be integrated by employing means for synchronisation and signalling. We discuss such mechanisms by showing how overlay routing is re-configurated when triggered by a handover in the multiaccess network. We also look into the need for signalling for ADIMUS in general.

6.1 Overlay Routing Reconfiguration

Leister et al. (2008) discuss the situation where the mobile terminal's mobility in the multi-access network requires dynamic changes in the overlay network to use resources in an efficient manner. In Figure 3 we show a scenario presenting part of the signalling taking place between the different entities during a handover-triggered overlay reconfiguration. There are two approaches for the reconfiguration sequence. The fast approach requires that the client application running on the mobile terminal is overlay-aware and can inform the border overlay node of the new network conditions when a handover is performed. The overlay network can then use static information as well as path probing to find more appropriate overlay nodes.

For overlay-unaware client applications, the handovers occur transparently for the overlay network, and each overlay node that sends a stream to the mobile terminal adapts to congestion information. When service quality changes significantly, the overlay initiates

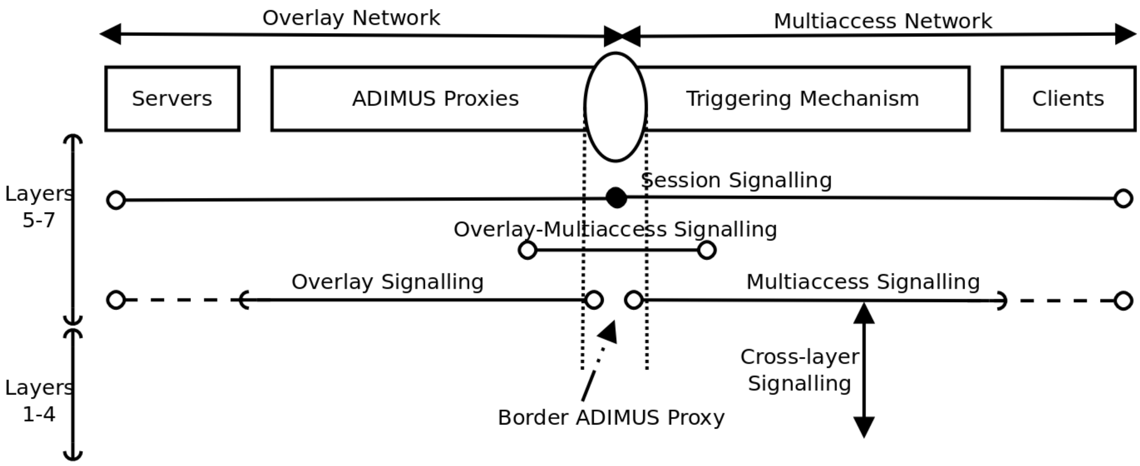


Figure 4. Signalling in the ADIMUS architecture.

path probing to discover a better suited overlay node. After a more appropriate overlay node is found, the overlay initiates a move of the session from the current to the new overlay node as shown in Fig. 3c, within a timescale of tenths of a second.

6.2 Signalling

Given the ADIMUS architecture the following four parts need to employ signalling as shown in Figure 4. The architecture requires both signalling on a command-level, as well as informing the different entities in the ADIMUS architecture about state, routing requests, support messages for decision making, network conditions, and measurements for, e.g., bandwidth, delay, or packet loss. The following types of signalling are envisaged: *a*) signalling within the multi-access network; including cross-layer signalling; *b*) signalling within the overlay network between the ADIMUS-Proxies (overlay nodes) which includes routing decisions and QoS/QoE information; *c*) signalling between overlay network and multi-access network at the borders of these networks; *d*) End-to-End signalling, e.g., start, or stop.

Protocols. The Internet protocol suite contains protocols suitable for signalling, addressing real-time multimedia. We review some of these protocols shortly.

The Real-Time Transport Protocol (RTP) by Schulzrinne et al. (2003) defines a standardised packet format for delivering multimedia over the Internet. RTP specifies the data transfer, including several profiles and payload formats, as well as the Real Time Control Protocol (RTCP). RTCP is used to specify QoS feedback and synchronisation between the media streams, in the form of out-of-band statistics and control information for an RTP flow. RTCP gathers statistics on quality aspects of the media distribution during a session, and transmits this data to the session media source and other session participatns, in order to adapt the media stream, and to detect transmission faults. RTCP also provides canonical end-point identifiers (CNAME) to all session participants.

The Real-Time Streaming Protocol (RTSP) by Schulzrinne et al. (1998) is a stateful network control protocol to establish and control media sessions between end points. The Session Description Protocol (SDP) by Handley et al. (2006) is intended for describing multimedia communication sessions for the purposes of session announcement, session invitation, and parameter negotiation. The set of properties and parameters are called a *session profile*. The Session Announcement Protocol (SAP) by Handley et al. (2000) is used for broadcasting multicast session information, typically using SDP as the format of the session description.

The Session Initiation Protocol (SIP) by Rosenberg et al. (2002) is a widely used signalling protocol for multimedia sessions. It can be used for creating, modifying, and terminating unicast or multicast sessions consisting of one or several media streams.

For mobility, the Mobile IP (MIP) by Perkins (1998) architecture can be employed.

These protocols together comprise the most important elements of the protocol suite that the IETF[4] promotes for streaming media communication in the Internet. ADIMUS makes use of all of the above protocols. Most of these protocols cover their dedicated portion of relevant tasks for enabling streaming applications that are not provided by one of the other protocols in the Internet protocol suite. One exception is the pair RTSP and SIP, both of which are application-layer signalling protocols that negotiate the establishment of media streams and carry information for controlling streams as time-dependent objects, but which don't consider issues of routing or quality. RTSP being defined for on-demand video streaming and SIP for conferences, they have overlapping functionality and can replace each other in most functions. ADIMUS has chosen RTSP because its fits the expected media streaming applications of ADIMUS better than SIP.

SAP played a role in the MBone (Multicast backbone) as a protocol for announcing media streams. ADIMUS terminals are supposed to discover the availability of media streams by other means (browsing or searching using HTTP), meaning that SAP has no place in the ADIMUS protocol suite.

None of the IETF protocol is meant for the internal management of an overlay network that requires the exchange of QoS management information and command functions for re-routing, caching and media transformation in an application-layer protocol. Although protocols such as SDP could play a role in this, these are only small contributions, and they have not been used in this manner in ADIMUS. Approaches that address multipath transport in overlay networks directly such as the multiparty transport overlay control protocol (MTOCP)[5] have not passed the experimental stage in the IETF either. These functions are therefore handled by private, non-standardised protocols.

4. See `www.ietf.org`, last accessed August 30, 2010.
5. See `http://tools.ietf.org/html/draft-kellil-sam-mtocp-01`, last accessed August 19, 2010. This is work in progress at the IETF.

7 Discussion and Conclusions

The goal of the ADIMUS project has been to develop methods that adapt to service quality changes to maximise the end-user's subjective experience. The ADIMUS architecture divides the network into two distinct parts, the overlay network and the multi-access network. The overlay network is responsible for the long-distance transport in the backbone network from the streaming server to a proxy close to the access network. The multi-access network allows to stream via the access network to the mobile terminal, selecting the possibility with the best QoE.

While defining the ADIMUS architecture, we recognised that a unified end-to-end adaptation solution is not viable since the challenges and requirements in the backbone network, and in the multi-access network, are too different. While decisions in the multi-access network need to be made very fast, involving cross-layer information, the overlay network cannot react that fast. Instead, decisions about lost and overloaded links and servers are treated in the overlay network, offering a resilient solution. The ADIMUS proxies offer a connection between these entirely different requirements with regard to the time-scale, so that ADIMUS as a whole can provide adaptation decisions end-to-end.

From this insight we developed both parts of the ADIMUS architecture separately, binding them together in the ADIMUS proxy nodes.

7.1 Research Challenges

The project set out to develop suitable metrics and mechanisms for quality measurement of multimedia streams to be used in adaptation processes so that the end user can receive optimal quality. The integration of the quality assessment in the system is to be designed in a way that the end user is not disturbed by the measurement activities, and we consider both end-user QoS and QoE.

The methods applied in ADIMUS need to choose the right adaptation method for video and audio, based on the right adaptation methods for these types of streamed data and the user preferences. Doing so, we need to define the right adaptation thresholds, e.g., delay, jitter levels, channel status, to support the algorithm operation when using different network technologies for access, including the means for decision-making in the vertical handover in the access networks.

7.2 Main achievements

ADIMUS is a step towards providing adaptive multimedia streaming on the Internet using multi-access technologies for the terminals. Challenges from diversity of the requirements of the different parts in the streaming chain are addressed by the ADIMUS architecture dividing the network into two distinct parts, namely the overlay network in the Internet backbone addressing long-term adaptation goals, and the multi-access network addressing mobility and short-term adaptation goals. Together, the parts of ADIMUS provide a framework on which streaming applications can build.

This work could also contribute to relations between perceived quality of experience, and measurements of QoS values, both for video and audio. We are on the way to apply these results to provide a basis for adaptation decisions.

For multimedia streaming in the backbone Internet, we work on the overlay infrastructure that implements multipath streaming. We consider both unicast and multicast scenarios. We are currently developing benchmarks for these scenarios, and implement the adaptation algorithms in the overlay nodes in the OMNET++ environment. These will be evaluated using different network topologies and scenarios. The work done in the ADIMUS project was able to produce two publications on the benchmarks, and provides a basis for the PhD thesis of Svetlana Boudko. The work on the algorithms and their comparison with the benchmarks is still work in progress.

The main achievements for the multi-access network part are as follows: The work has produced a design of the system components and functionalities required for supporting multimedia services in heterogeneous multi-access networks. The design was based on a thorough requirements analysis, and selected parts of the system were implemented into real demonstrators or simulation models. Specific solutions were developed for cross-layer signalling, mobility management as well as optimizing scalable video delivery over wireless links. Also the concept of multi-interface streaming of scalable video was addressed briefly in the project resulting in the first design for supporting the functionality in heterogeneous multi-access networks. The work done in the ADIMUS project produced several publications that provide a good basis for Tiia Sutinen for writing her PhD thesis. However, some additional work will be needed to complete the research work. At the moment, one or two articles are missing to complete this work. The plan is to utilise the existing material collected during the ADIMUS project as much as possible, and to produce some new experimental data from simulations/prototypes, if needed, to finalise the work.

While the ADIMUS architecture is not implemented as an integrated demonstrator yet, its design principles are shown for the parts it consists of.

Besides these achievements, ADIMUS contributed to a fruitful cooperation between the involved researchers. We mention that Ian Marsh served as a reviewer (tilsynsensor) of the course "INF5081 – Multimedia Coding and Applications" at the University of Oslo, Carsten Griwodz served as a committee member in the PhD defence of Ian Marsh, the advisory roles of the supervisors, fruitful project workshops, visits to the involved institutions, jointly written scientific articles, and planning of future projects. We hope that we can sustain this Nordic cooperation.

7.3 Future Work

In the current research, the ADIMUS project concentrated on the situations of using single-link streaming in the access network and performing handovers at the terminal when QoE-related events are triggered. The ADIMUS research group recognises that new opportunities are possible when applying multi-link streaming, i.e., streaming data along multiple paths simultaneously to the terminal. Multi-link streaming is a novel concept not yet deployed. While the overlay network uses multi-path streaming, the adaptation mechanisms need to be adjusted in both parts of the network for the multi-link case. This implies that the terminals receive streams from several ADIMUS proxies simultaneously. To investigate multi-path and multi-link streaming closer, we need to extend the simulation models both in the overlay network and in the multi-access network, as well as doing extensions in the implementation of these networks.

Project Publications

During the course of the project, eleven conference papers and one journal article have been published. In addition, one PhD thesis has been published, while two theses are under preparation.

The project publications include papers on the ADIMUS architecture by Leister et al. (2008), and estimating the subjective video quality from objective measurements by Leister et al. (2010), while quality of real-time voice communication is treated in the dissertation Marsh (2009). Papers related to the multi-access network of ADIMUS include the following conference papers: Backman et al. (2008); Dousson et al. (2007); Luoto and Sutinen (2008); Mäkelä et al. (2009); Piri et al. (2009, 2010); Sutinen and Frantti (2008); Sutinen and Huusko (2011), of which one was extended to a journal (Mäkelä et al., 2010), as well as one demo paper (Sutinen et al., 2010). The conference papers by Boudko et al. (2008) and Boudko et al. (2010) both show benchmarking systems that can be used to evaluate algorithms related to the overlay network.

References

Backman, W., Kordybach, K., and Sutinen, T. (2008). Implementation of WLAN part of MERCoNe system in OPNET. In *2008 International Conference on Telecommunications*, pages 1–6, St. Petersburg. 16, 19, 31

Bertsekas, D. and Gallager, R. (1987). *Data networks*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA. 24

Boudko, S., Griwodz, C., Halvorsen, P., and Leister, W. (2008). A benchmarking system for multipath overlay multimedia streaming. In *Proceedings of ICME*. 22, 31

Boudko, S., Leister, W., Griwodz, C., and Halvorsen, P. (2010). Maximizing video quality for several unicast streams in a multipath overlay network. submitted to conference. 22, 31

Castro, M., Druschel, P., Kermarrec, A., Nandi, A., Rowstron, A., and Singh, A. (2003). Splitstream: High-bandwidth multicast in cooperative environments. 25

Chiang, T. and Zhang, Y.-Q. (1997). A new rate control scheme using quadratic rate distortion model. *IEEE Trans. Circuits and Systems for Video Technology*, 7(1):246–250. Available from: http://dx.doi.org/10.1109/76.554439. 12

Chu, H.-H. and Nahrstedt, K. (1997). Multi-path communication for video traffic. In *HICSS '97: Proceedings of the 30th Hawaii International Conference on System Sciences*, page 695. IEEE Computer Society. 22

Dousson, C., Pentikousis, K., Sutinen, T., and Mäkelä, J. (2007). Chronicle recognition for mobility management triggers. In *ISCC 2007*, pages 305–310. 16, 19, 31

Golubchik, L., Lui, J., Tung, T., Chow, A., Lee, W., Franceschinis, G., and Anglano, C. (2002). Multi-path continuous media streaming: What are the benefits. Available from: citeseer.ist.psu.edu/golubchik02multipath.html. 22

Handley, M., Floyd, S., Padhye, J., and Widmer, J. (2003). TCP Friendly Rate Control (TFRC): Protocol Specification. RFC 3448 (Proposed Standard). Obsoleted by RFC 5348. Available from: http://www.ietf.org/rfc/rfc3448.txt. 21

Handley, M., Jacobson, V., and Perkins, C. (2006). SDP: Session Description Protocol. RFC 4566 (Proposed Standard). Available from: http://www.ietf.org/rfc/rfc4566.txt. 20, 28

Handley, M., Perkins, C., and Whelan, E. (2000). Session Announcement Protocol. RFC 2974 (Experimental). Available from: http://www.ietf.org/rfc/rfc2974.txt. 28

Hefeeda, M., Habib, A., Botev, B., Xu, D., and Bhargava, B. (2003). Promise: peer-to-peer media streaming using collectcast. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 45–54. 24

Huynh-Thu, Q. and Ghanbari, M. (2008). Scope of validity of PSNR in image/video quality assessment. *Electronic Letters*, 44(23):800–801. Available from: http://dx.doi.org/10.1049/el:20080522. 12

IEEE (2009). Media Independent Handover. IEEE Standard 802.21. 9, 19

Jain, R., Chiu, D., and Hawe, W. (1984). A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. Technical Report TR-301, DEC Research. 24

Kelly, F. (1997). Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8:33–37. 24

Kelly, F., Maulloo, A., and Tan, D. (1998). Rate control in communication networks: shadow prices, proportional fairness and stability. In *Journal of the Operational Research Society*, volume 49. 24

Knee, M. (2006). The picture appraisal rating (PAR) - a single-ended picture quality measure for MPEG-2. White paper, Snell & Wilcox Limited. [Online]. Available: http://www.snellgroup.com/documents/white-papers/white-paper-picture-appraisal-rating.pdf, last accessed May 14, 2010. 12

Koumaras, H., Kourtis, A., Lin, C.-H., and Shieh, C.-K. (2009). A theoretical framework for end-to-end video quality prediction of MPEG-based sequences. *International Journal On Advances in Networks and Services*, 1(1):19–29. Available from: http://www.iariajournals.org/networks_and_services/. 12, 23

Le Boudec, J.-Y. (2008). Rate adaptation, congestion control and fairness: A tutorial. Technical report, EPFL Lausanne. Available from: http://ica1www.epfl.ch/PS_files/LEB3132.pdf. 24

Leister, W., Boudko, S., and Halbach, T. (2010). Estimation of subjective video quality as feedback to content providers. In *ICSNC 2010: Proc. of the 5th Intl. Conf. on Systems and Networks Communications*. IEEE. 11, 12, 14, 31

Leister, W., Sutinen, T., Boudko, S., Marsh, I., Griwodz, C., and Halvorsen, P. (2008). An architecture for adaptive multimedia streaming to mobile nodes. In *MoMM '08: Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia*, pages 313–316, New York, NY, USA. ACM. 7, 11, 16, 18, 26, 31

Luoto, M. and Sutinen, T. (2008). Cross-layer enhanced mobility management in heterogeneous networks. In *Proceedings of ICC*. 16, 19, 31

Mäkelä, J., Luoto, M., Sutinen, T., and Pentikousis, K. (2009). Evaluation of an information service for enhanced multiaccess media delivery. In *Mobimedia '09: Proceedings of the 5th International ICST Mobile Multimedia Communications Conference*, pages 1–6. ICST, Brussels, Belgium. 16, 19, 31

Mäkelä, J., Luoto, M., Sutinen, T., and Pentikousis, K. (2010). Distributed information service architecture for overlapping multiaccess networks. *MTAP'10, to appear*. 16, 19, 31

Makela, J. and Pentikousis, K. (2007). Trigger management mechanisms. In *Proceedings of ISWPC*, pages 378–383, San Juan, Puerto Rico. 9, 19

Marsh, I. R. (2009). *Quality aspects of Internet telephony*. PhD thesis, Kungliga Tekniska högskolan. Available from: http://soda.swedish-ict.se/3801/. 15, 31

Mu, M. (2009). An interview with video quality experts. *SIGMultimedia Rec.*, 1(4):4–13. 11, 12

Ni, P., Eichhorn, A., Griwodz, C., and Halvorsen, P. (2009). Fine-grained scalable streaming from coarse-grained videos. In *International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, pages 103–108. ACM. 12

Paxson, V., Almes, G., Mahdavi, J., and Mathis, M. (1998). Framework for IP Performance Metrics. RFC 2330. [Online]. Available: http://www.ietf.org/rfc/rfc2330.txt, last accessed May 14, 2010. 12

Perkins, C., editor (1998). *Mobile IP: Design Principles and Practice*. Addison-Wesley. 9, 28

Piri, E., Sutinen, T., and Vehkaperä, J. (2009). Cross-layer architecture for adaptive real-time multimedia in heterogeneous network environment. In *European Wireless 2009: 15th European Wireless Conference - Electronic Proceedings*, pages 1–5. 16, 19, 31

Piri, E., Uitto, M., Vehkaperä, J., and Sutinen, T. (2010). Dynamic cross-layer adaptation of scalable video in wireless networking. In *Globecom, to appear*. 16, 21, 31

Pongor, G. (1993). Omnet: Objective modular network testbed. In *MASCOTS '93: Proceedings of the International Workshop on Modeling, Analysis, and Simulation On Computer and Telecommunication Systems*, pages 323–326. Society for Computer Simulation. 24

Ramesh, S., Rhee, I., and Guo, K. (2001). Multicast with cache (mcache): An adaptive zero-delay video-on-demand service. In *in Proceedings of IEEE Infocom*, pages 85–94. 25

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and Schooler, E. (2002). SIP: Session Initiation Protocol. RFC 3261 (Proposed Standard). Updated by RFCs 3265, 3853, 4320, 4916, 5393, 5621, 5626, 5630. Available from: `http://www.ietf.org/rfc/rfc3261.txt`. 28

Schulzrinne, H., Casner, S., Frederick, R., and Jacobson, V. (2003). RTP: A Transport Protocol for Real-Time Applications. RFC 3550 (Standard). Updated by RFC 5506. Available from: `http://www.ietf.org/rfc/rfc3550.txt`. 27

Schulzrinne, H., Rao, A., and Lanphier, R. (1998). Real Time Streaming Protocol (RTSP). RFC 2326 (Proposed Standard). Available from: `http://www.ietf.org/rfc/rfc2326.txt`. 27

Schwarz, H., Marpe, D., and Wiegand, T. (2006). Overview of the scalable H.264/MPEG4-AVC extension. In *Proceedings of ICIP*, pages 161–164. 9, 10, 17, 20

Sevcik, P. (2005). Defining the application performance index. *Business Communications Review*, pages 8–10. 13, 14

Small, T., Li, B., Member, S., and Liang, B. (2007). Outreach: Peer-to-peer topology construction towards minimized server bandwidth costs. In *in IEEE Journal on Selected Areas in Communications,Special Issue on Peer-to-Peer Communications and Applications, First Quarter*, pages 35–45. 25

Sutinen, T. and Frantti, T. (2008). Reference architecture for mobility enhanced multimedia services. In *Proceedings of Mobilware*. 16, 17, 18, 31

Sutinen, T. and Huusko, J. (2011). Mac-level packet prioritization and scheduling for enhanced scalable video transmission. In *CCNC'11, to appear*. 16, 21, 31

Sutinen, T., Knuutinen, J., and Luoto, M. (2010). Multi-interface streaming of scalable video. demopaper, to appear in the proc. of Mobimedia'10. 16, 20, 31

Wang, B., Wei, W., Guo, Z., and Towsley, D. (2007). Multipath live streaming via TCP: scheme, performance and benefits. In *CoNEXT '07: Proceedings of the 2007 ACM CoNEXT conference*, pages 1–12. 24

Wolf, S. and Pinson, M. (2002). Video quality measurement techniques. Ntia report 02-392, National Telecommunications and Information Administration, US Dept of Commerce. 12

Zhang, X., Liu, J., Li, B., and Yum, T.-S. P. (2005). CoolStreaming/DONet: A data-driven overlay network for peer-to-peer live media streaming. In *in Proceedings of IEEE Infocom*, pages 2102–2111 vol. 3. 26