An Architecture for Adaptive Multimedia Streaming to Mobile Nodes[1)]

Wolfgang Leister[1], Tiia Sutinen[2], Svetlana Boudko[1], Ian Marsh[3], Carsten Griwodz[4], Pål Halvorsen[4]

[1]Norsk Regnesentral, Oslo, Norway
[2]VTT Technical Research Centre of Finland, Oulu, Finland
[3]SICS, Kista, Sweden
[4]Simula Research Laboratory and University of Oslo, Norway

Corresponding author: Wolfgang.Leister@nr.no

Abstract:
We describe the ADIMUS architecture which addresses the problem of maintaining the subjective quality of multimedia streaming for a mobile user. In contrast to other works, the entire end-to-end path of the video stream is considered. Adaptation mechanisms for maintaining quality include time-critical handovers, overlay routing and network estimation techniques. Our architecture is built on overlays that provides the necessary functionality for a video streaming service. The paper highlights the key components that ADIMUS advocates to support quality streaming from server to mobile client.

1 Introduction

Streaming of video material using the Internet is here. Live streaming of popular events onto mobile devices will become a significant portion of the streaming video market. In particular breaking news and sports programs are very much in demand when viewed in real-time. Content providers see real-time transmissions as a valuable 'pay for now' service.

Video streaming to mobile devices is at the moment still contained in operators 3G networks. As far as we know, there are no services yet that provide content streaming from any Internet content provider to a mobile handset.

We aim at supporting end-to-end video streaming to mobile nodes and consider entertainment content with high media quality. We are therefore concerned with data delivery mechanisms that impact the quality. Since the content is streamed from live feeds/servers to the mobile devices, the interplay of backbone and access technologies needs to be considered. Therefore, this work addresses the problem of delivering video streaming with reasonable quality and startup latency. We develop methods to configure delivery systems in order to maximise the

[1)] The work described in this paper has been conducted as a part of the ADIMUS (Adaptive Internet Multimedia Streaming) project which is funded by the NORDUnet-3 programme.

Quality of Experience (QoE) and to adapt the system in response to changes in resources availability.

QoE is a purely subjective measure from the user's perspective of the overall value of the service provided. It is terminal dependent, related to a particular users expectations and the nature of the content. Modern techniques built on Mean opinion scores (MOS) have been expanded to considered problems such as jittering, blockiness and other artifacts of the video signal. Although QoE is perceived as subjective, it is the *only measure* that counts for customers of a service.

The paper continues as follows: Section 2 describes the problem statement addressed in our work. Section 3 describes the core components of the ADIMUS architecture including the backbone and access networks. Section 4 deals with the techniques we use, namely application adaption, handovers and backbone overlay techniques, and describes relevant methods to estimate network Quality of Service (QoS). We include a short discussion in Section 5 followed by the conclusions.

2 Problem statement

We consider a delivery system that consists of multiple service providers streaming content via IP networks to multiple users. We focus on mobile consumers of the video streams. Video streams transmitted from the service provider to a mobile node experience quality degradation by several factors specific to different parts of the delivery infrastructure. These are discussed next in isolation.

2.1 The backbone network

Content is streamed from multiple streaming servers through the Internet without resource guarantees. Therefore, we must use strategies that adapt to the available resources. Typical reasons for changes in resource availability are congestion and route failures. While congestion can be handled within streams, route failures lead to long sequences of lost packets.

Service providers address both challenges by deploying overlay networks where overlay nodes are operated by a service provider and deployed in the Internet. Overlay nodes rely on IP routing to communicate with their neighbouring overlay nodes; however they can forward data through one or more other overlay nodes to adapt to changing conditions.

Overlay networks have been used to allow applications to recover from Internet route failures within a few seconds by routing around network-layer problems at the application layer [1]. They have also been used to change data paths of applications in order to maintain the

QoS of applications [5] and to improve application performance by making use of distributed nodes for the transcoding and caching or content [9]. Furthermore, multipath streaming has been used to increase the total available end-to-end bandwidth and reduce the impact of congestion [10]. These works solve specific problems, but they need to be integrated and evaluated in combination.

2.2 The multi-access network

Mobile users represent a challenge as the QoS requirements of multimedia services are not well supported in state-of-the art vertical handover situations. The access network environment is characterised by technology heterogeneity: the access link used may be based on any IP-based access, e.g., fixed access, WLAN, 2/3G, WiMAX, and the simultaneous availability of several access options. Due to mobility and roaming, the characteristics of the used access link and the terminal device may change dynamically during a session. In multi-access networks, video streams may experience higher bandwidth variation and jitter than in environments supporting limited mobility such as a WLAN network [7]. Furthermore, video streams may suffer from connection outages of different lengths caused by handovers.

Due to the diversity in network technology and administration, a unified QoS architecture that provides the required support for video streaming cannot be expected in our target access network environment. Instead, the services and applications themselves need to be capable of adapting to prevailing system conditions. Moreover, to utilise the full potential of multi-access environments for supporting video stream transmission, we need to find the optimal solution for handover management with the video application in mind. In addition, an information service is needed to feed the application adaptation and handover-related decision-making with continual system status information. A reference framework for cross-layer aided handover management has been presented in [12].

3 The ADIMUS architecture

The ADIMUS architecture, shown in Fig. 1, is designed to deliver multimedia streams end-to-end. The framework is used for adaptive multimedia streams transferred from one or more servers to clients. It comprises of a delivery infrastructure based on an *overlay network*, i.e., including the service provider network and ADIMUS proxies, and a *multi-access network* supporting mobile clients. 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, an overlay network monitors connections and can make application-layer forwarding decisions to change routes.

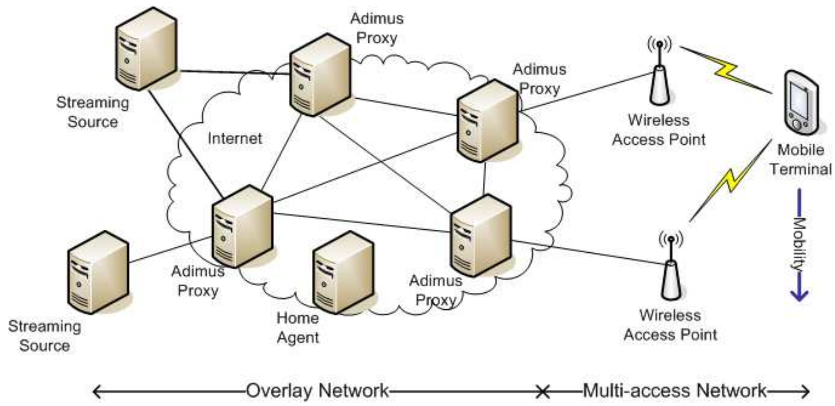


Fig. 1: Overview of the ADIMUS architecture.

- Near the client, a heterogeneous multi-access network provides application adaptation and handover mechanisms to maximize the QoE and support for different types of mobility. A cross-layer signaling system is utilized 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. It supports the exchange of media descriptions (meta-data) and current QoS values according to the selected metrics.

The system adaptation decisions are based on optimizing the QoE. Although the adaptation mechanisms are different for the overlay and access networks they are governed by the same QoS decisions. At the borders between these parts, ADIMUS proxies exchange information to perform adaptations. With respect to the QoE achieved through scalable video, we are currently performing new user studies to extend experiences from previous work [17]. The architecture will be validated at later research stages with prototype implementations of selected parts of the architecture.

3.1 The overlay network

Multimedia streaming, both live streaming and on-demand streaming, is provided using appropriate streaming protocols and source-driven mechanisms for applying quality-improving mechanisms. Note that the service provider network is a part of the overlay network, since it represents the source nodes of the overlay network.

The long-distance part of the streaming infrastructure is organized as an application-level overlay network. Streaming servers located in the service providers' home networks and ADIMUS proxies operated by the service providers spread out over the Internet to form an overlay network. Such overlays constitute fully meshed networks that allow overlay re-routing when IP-based routing from a streaming server to a terminal cannot maintain the required QoE.

The ADIMUS proxies monitor network conditions using both passive and active network measurements, and they interchange information about observed network conditions. They keep statistical information about the observed conditions of the network that is used later for estimating trends in bandwidth, latency or packet loss at a given link or path.

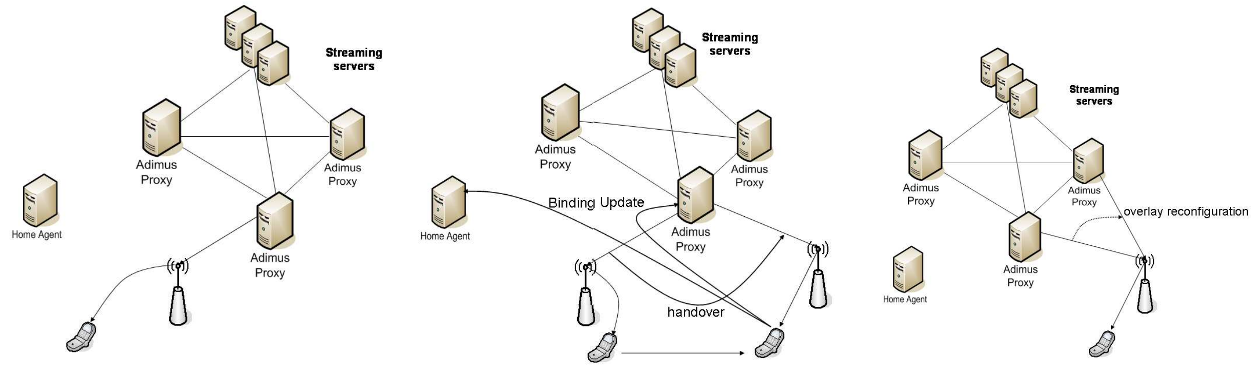
For data transport, we explore adaptive protocols that can take feedback from mobile nodes into account. The SIP protocol is used for signaling. This allows changes to a live streaming session, including the migration of streams between nodes. It requires that streaming servers register with a well-known SIP server (UAS). Terminals can then initiate a session with the streaming server. Based on the overlay's monitor function of the network, the most appropriate overlay route is chosen, and the last overlay node on the route acts as data source for the terminal. A re-invitation to the client 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 node.

Furthermore, ADIMUS will support multi-path streaming to provide failure resistance and load balancing. Thus, it is necessary that the last-hop overlay nodes use different access links for the different IP addresses of each of the mobile node's wireless devices. Aside from cross-layer information provided for faster reaction to changes at one of the wireless links, the multi-path support does thus stay entirely at the application level. This implies that multi-path streaming is only possible to mobile nodes that run application layer software that is overlay-aware and handles buffering and reordering.

3.2 The multi-access network

The multimedia streaming end point at the client-side is the mobile node (MN), which resides in the multi-access network. The multi-access network consists of different IP-based access networks of different technologies and managed by different parties. The user terminal device is equipped with multiple network interfaces, and has support for a IP mobility protocols, such as Mobile IP (MIP) [14]). The mobile node is thus capable of roaming between IP networks, provided that it can update its reachability information from the new network. If MIP route optimization is supported, the mobile node must be able to update its mobility binding from the new network at its home agent, residing in the mobile node's home network, as well as at the correspondent node, i.e., ADIMUS proxy.

The multi-access network environment allows for using handovers as means for maintaining application quality. Specifically, the mobile node is capable of selecting an alternative access where the current link does not meet the minimum QoS requirements of our video streaming service. To make an informed handover decision considering the QoE of the video stream in this environment, the mobile node collects and utilizes information related to the available



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

Fig. 2: Overlay network reconfiguration when MN roams from one access network to another.

access options' characteristics obtained through a cross-layer mechanism. Candidate cross-layer signaling frameworks for our work are the IEEE 802.21 Media Independent Handover (MIH) framework [8] and the triggering framework [13].

When there is no access option available that can support the streaming service's QoS requirements or making a handover is considered infeasible, the streaming service needs to adapt to access link conditions. For efficient adaptation, the video client application running on the mobile node is capable of reporting back to the server or the serving ADIMUS proxy about prevailing access network conditions.In these cases, the server or the ADIMUS proxy can adapt the video stream accordingly.

3.3 Handover-triggered overlay reconfiguration

The mobile node's mobility in the multi-access network means that the overlay configuration needs to be changed dynamically to use resources in an efficient manner. In Fig. 2, we show a scenario presenting a part of the signaling 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 node is overlay-aware and can inform each 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 client applications that are overlay-agnostic, the handovers occur transparently for the overlay network, and each overlay node that sends a stream to the mobile node adapts to congestion information. When service quality changes in a major way, the overlay initiates 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. 2(c), within a timescale of thenths of a second.

4 Supported adaptation and QoS-estimation methods

Our architecture employs adaptation methods that are governed by estimates of the network QoS. In this section, we describe in more detail the adaptation and optimisation mechanisms employed. Specifically, we consider application adaptation, handover management and overlay routing as the main means for maximising the QoE of multimedia delivery. The adaptation decisions are governed by estimates of network QoS at different levels.

4.1 Application adaptation

Traditional approaches for implementing dynamic adaptation support into multimedia applications involve monitoring the end-to-end QoS (throughput, delay, jitter or loss) experienced by an application flow and adjusting the operation of the application accordingly. Application adaptation can include adjusting the sending rate, buffering and QoS-informed error protection (FEC, MDC for example). We consider mechanisms for rate adaptation where bandwidth requirements are adapted to the available network bandwidth as well. In this case, spatial, temporal and SNR adaptation, as employed in H.264 SVC [16], can take place at the server or at intermediate nodes capable of processing a scalable video stream. Coarse-grained adaptation decisions that involve the addition or removal of video layers from the video stream can then be made by the overlay nodes. This is done according to the bandwidth monitoring data collected by the overlay nodes as well as the feedback information received from the mobile node. A similar approach to adaptation has been published in [15].

4.2 Handovers

For handovers triggered by terminal mobility or by explicit request from the application layer the mobile client maximizes the QoE of the multimedia service delivery. The mobile node implements an optimized handover decision-making algorithm that weighs up the information regarding the available access links, their statuses and the resulting handover implications and maps them to the application's requirements. Since handovers introduce additional signalling overhead to the network, performing unnecessary handovers is not beneficial either in the network's nor in the application's viewpoint. Tight co-operation between the application and handover management reduces unnecessary handovers and renders the application more responsive to different network situations. Finally, the mobile node is capable of utilizing simultaneous connections to multiple access links whenever possible to avoid packet loss during handovers.

4.3 Overlay routing

Overlay networks are deployed at the application layer. Software needs to be installed on the mobile node to discover the path quality to alternative overlay nodes. Inside the overlay network, re-routing decisions are made to overcome the lack of QoS support in normal Internet routing. To achieve optimal routing in the overlay, routing algorithms have to be implemented in the overlay nodes. These are taking decisions based on locally available data. In order to compare the developed algorithms running on the overlay nodes, we developed a benchmark based on globally available data [2].

4.4 Link state information

In state of the art wireless networks, the main metric used for handover decision-making is the received signal strength. In standard MIP, handover decisions are based on the reception of agent advertisements. However, neither of these metrics are capable of reflecting the actual transmission conditions in terms of available bandwidth, delay or loss characteristics of the link. Systems that use cross layer information to assist mobility-related operations in heterogeneous networks include the emerging IEEE 802.21 MIH service [8] and the triggering framework [13]. Here, MIH defines a set of link and physical layer parameters for access link selection, and also includes QoS-related parameters like throughput, Class of Service (CoS) average packet transfer delay, and CoS packet loss rate. MIH higher layer entities allows us to receive notifications about link-level events, which can be used in ADIMUS for decision-making. Furthermore, the triggering framework is more generic in the sense that it provides for cross-layer signalling that extends across the entire protocol stack. It also allows conveying additional system information for the decision points, including for example information related to terminal device capabilities. Thus, the triggering framework can be used on top of MIH to provide for the required signaling between the application and mobility management.

4.5 Bandwidth estimation

Optimal handover and routing decisions are based on estimations of available bandwidth along the paths between the overlay and mobile nodes, both on wired and wireless links. To estimate available bandwidth several active probing techniques [4, 11], and passive measurements such as collecting *Simple Network Management Protocol* (SNMP) [3] data, Syslog data and NetFlow data have been suggested. For passive measurements we use timestamps of the streams that are part of the overlay routing, while we have selected the Bandwidth Available in Real-Time (BART) scheme [6] for active probing.

5 Discussion and Conclusions

The ADIMUS architecture is a combination of the adaptation mechanisms in both overlay and multi-access networks. The challenges include the different nature of adaptation on both the client- and server side, i.e., requiring different adaptation and estimation techniques. Both sub-parts have different timing requirements when performing adaptation decisions. While the multi-access network must react very fast to handover decisions, the overlay network decisions can be slower.

In the ADIMUS architecture, the definition of the term "end-to-end" is different from the ordinary operation of the Internet, which must be reflected in the implementation. Since the overlay nodes operate on the application layer, the original source of a media stream may be hidden to the client, and only the closest overlay node will be visible. While this is explicitly done in content distribution networks, this is more hidden in ADIMUS.

In summary, we have outlined an architecture for multimedia streaming over heterogeneous IP networks. The architecture defines means for supporting adaptation of QoS sensitive multimedia applications over a best effort Internet path ranging from the video server to the mobile node. The architecture uses handover and media adaptation as means to provide the end-user with best possible QoE. To achieve this, the ADIMUS architecture uses an overlay network to enable resource efficient routing while maximizing the QoS of the multimedia delivery in the Internet backbone. The architecture also supports the user to access services from any IP-based access network that is connected to the overlay system. The user can move between different access networks, and the architecture is capable of adapting its operation with respect to the user's current location and access link characteristics.

Having defined the ADIMUS architecture, we now will look into defining algorithms for optimizing the overall system operation with respect to the quality of the multimedia delivery. This will be performed in the overlay network and the access network separately, before defining interfaces between these parts to provide an end to end optimization. To validate of the proposed systems, the algorithms employed will be simulated and compared with benchmarks of other systems, before an implementation of the architecture as a system.

References

[1] David Andersen, Hari Balakrishnan, Frans Kaashoek, and Robert Morris. Resilient overlay networks. In *Proceedings of SOSP*, pages 131–145, 2001.

[2] Svetlana Boudko, Carsten Griwodz, Pål Halvorsen, and Wolfgang Leister. A benchmarking system for multipath overlay multimedia streaming. In *Proceedings of ICME*, 2008.

[3] J.D. Case, M. Fedor, M.L. Schoffstall, and J. Davin. Simple Network Management Protocol (SNMP). RFC 1157 (Historic), May 1990.

[4] Allen B. Downey. Using pathchar to estimate internet link characteristics. In *Measurement and Modeling of Computer Systems*, pages 222–223, 1999.

[5] Zhenhai Duan, Zhi-Li Zhang, and Yiwei Thomas Hou. Service overlay networks: SLAs, QoS, and bandwidth provisioning. *IEEE/ACM Transactions on Networking*, 11(6):870–883, December 2003.

[6] Svante Ekelin, Martin Nilsson, Erik Hartikainen, Andreas Johnsson, Jan-Erik Mångs, Bob Melander, and Mats Børkman. Real-time measurement of end-to-end available bandwidth using kalman filtering. In *IEEE NOMS*, 2006.

[7] P. Ferre, D. Agrafiotis, T.K. Chiew, A.R. Nix, and D.R. Bull. Multimedia transmission over ieee 802.11g wlans: Practical issues and considerations. In *Proceedings of ICCE*, pages 1–2, 2007.

[8] IEEE. Media Independent Handover. IEEE Draft Standard 802.21, 2008, work in progress.

[9] Jingwei Jin and Klara Nahrstedt. Large-scale service overlay networking with distance-based clustering. In *Proceedings of ACM/IFIP/USENIX Middleware*, pages 396–413, 2003.

[10] D. Jurca and P. Frossard. Video packet selection and scheduling for multipath streaming. *IEEE Transactions on Multimedia*, 9(3):629–641, April 2007.

[11] S. Kang, X. Liu, M. Dai, and D. Loguinov. Packet-pair bandwidth estimation: Stochastic analysis of a single congested node, 2004.

[12] Markus Luoto and Tiia Sutinen. Cross-layer enhanced mobility management in heterogeneous networks. In *Proceedings of ICC*, 2008, to appear.

[13] J. Makela and K. Pentikousis. Trigger management mechanisms. In *Proceedings of ISWPC*, pages 378–383, San Juan, Puerto Rico, February 2007.

[14] C. Perkins, editor. *Mobile IP: Design Principles and Practice*. Addison-Wesley, 1998.

[15] Peter Schojer, Laszlo Boszormenyi, and Hermann Hellwagner. QBIX-G: a transcoding multimedia proxy. In *Proceedings of MMCN*, 2006.

[16] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the scalable H.264/MPEG4-AVC extension. In *Proceedings of ICIP*, pages 161–164, October 2006.

[17] Michael Zink, Oliver Künzel, Jens Schmitt, and Ralf Steinmetz. Subjective impression of variations in layer encoded videos. In *Proceedings of IWQoS*, June 1993.