Reference Architecture for Mobility Enhanced Multimedia Services

Tiia Sutinen and Tapio Frantti
VTT Technical Research Centre of Finland
P.O. Box 1100
FI-90571 Oulu, Finland
{tiia.sutinen, tapio.frantti}@vtt.fi

ABSTRACT

Heterogeneous multi-access networks pose new challenges for service development as supporting QoS-sensitive multimedia applications in such a dynamic environment requires additional mechanisms. The session continuity support mechanisms have to be designed so that they respect the existing communication principles. Only this way they can be sustainable solutions. This paper introduces 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.

Categories and Subject Descriptors

C.2 [**Computer-Communication Networks**]: Miscellaneous

General Terms

Design

Keywords

Session continuity, multi-access, architecture

1. INTRODUCTION

Access network heterogeneity and multi-access bring new challenges for designing multimedia applications and services. The idea of utilizing the available access options by dynamically transferring the user's communication sessions to the most appropriate network has emerged with the extensive deployment of wireless access networks and appearance of multi-mode terminals. In a heterogeneous network, the coverages of individual access networks are often partly overlapping but differing in terms of technology (2/3G cellular, WLAN, WiMAX, and fixed), network management (operator policies), and supported services, including Quality of Service (QoS); authentication, authorization, and accounting (AAA); and security. In this environment, assuming that terminal and network support exist for mobility and multi-access, access selection can be done dynamically based on factors like the user's personal preferences, applications' requirements or operator policies. Depending on the handover decision-making algorithm, any mismatch or deterioration in the conditions set for the used access may result in a handover to a new access point that can meet the requirements better than the old one.

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 terminal device may change in a handover. Handovers in heterogeneous networks introduce significant variance to the network connection and terminal 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.

Full end-to-end QoS, realized by combining IP-level mechanisms such as Integrated Services (IntServ) and Differentiated Services (DiffServ) with an access network mechanism, *e.g.*, IEEE 802.11e [9], is rare in today's Internet. Therefore, developers need to include application-level QoS support mechanisms to their QoS-sensitive applications to make them resilient to network perturbations. In a heterogeneous network, more advanced algorithms and utilization of cross-layer information are required to achieve efficient application and transport level adaptation [2][3][7]. On the other hand, multi-access enables maintaining application operation by changing the used access network whenever the currently used one is showing unacceptable performance. Handover triggering according to application requirements involves accessing and comparing application-level information and lower-level access characteristics in the handover decision-making entity [5][10], *e.g.*, a Mobile IP client [1].

The problem with services that require IP mobility and cross-layer communications is that these technologies are not inherently supported by the current Internet architecture. Thus poor or inconsistent design of mobility enhanced services can result in deteriorated communication performance (see, *e.g.*, [6] for deficiencies in cross-layer designs). During the past years, there have been a lot of short-term solutions with deteriorating effect to the communication architecture. For example, the improvements gained from in-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Mobilware'08, February 12-15, 2008, Innsbruck, Austria
Copyright 2008 ACM 978-1-59593-984-5/08/02 ...$5.00.

troducing cross-layer communication in multimedia delivery and mobility management have led to introducing new middleware functionalities to improve, *e.g.*, security and communication reliability. This has resulted in the increased complexity on cross-layer information controllers and managers as well as protocol communication and hierarchies. The new middleware functionalities may be more secure and more reliable but they might also violate existing communication architecture principles, *e.g.*, by encrypting higher layer headers thus exacerbating network management and by increasing overhead with redundant control traffic. As another example, the network address translation (NAT) solution was developed due to limited address space size but it also broke the end-to-end communication principle of the Internet architecture. Return to the end-to-end type communication is now one of the main targets on the future Internet programs, like FIND (http://find.isi.edu/) and GENI (http://www.geni.net/).

The main contribution of this paper is to present and validate a reference architecture for mobility enhanced multimedia services. The reference architecture takes into account the inherent limitations posed by each requirement domain associated with the developed service. In this paper, we define four domains that affect service development and operation: administrative, user, application, and the Internet architecture. The reference architecture is validated with use cases and detailed requirements analysis that follow a specific analysis approach described in the paper. In the next research stages, the requirements derived for service development will be utilized in validating the developed concepts with simulations and prototype implementations.

The rest of this paper is organized as follows: Section 2 introduces the concept of a reference architecture and the proposed reference architecture for mobility enhanced multimedia services. The reference architecture is validated by a session continuity example in Section 3. Discussion and conclusions are given in sections 4 and 5, respectively.

2. REFERENCE ARCHITECTURE

This section presents the reference architecture derived for mobility enhanced multimedia services. The purpose of the reference architecture is to define a set of best practices for the development of multimedia services targeted for heterogeneous multi-access networks. The use case based validation of the architecture is presented in Section 3.

A reference architecture can be defined as a consistent set of best practices for service design and development. It can be defined along different levels of abstraction, or "views", thereby providing more flexibility in how it can be used. A reference architecture provides a validated template solution and a common vocabulary and taxonomy for a particular domain. Usually, a reference architecture consists of a list of functions and some indication of their interfaces and interactions with each other and with functions located outside of the scope of the reference architecture. In this paper, the reference architecture concept is applied for multimedia service design.

When designing multimedia services that are to be used in today's Internet there are four domains that need to be taken into account in the process: the administrative, user, application, and Internet architecture. These different domains are shown as layers in the reference architecture used in this paper (see, Figure 1).

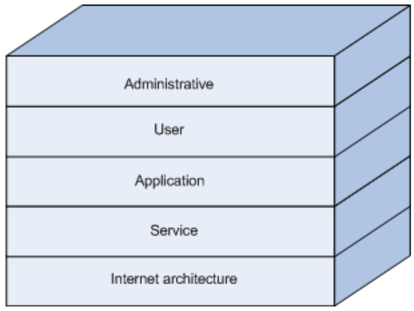

Figure 1: Overview of the reference architecture

For service design, each of the reference architecture domains pose 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. Quality of Experience (QoE) is solely seen from the user perspective, too. Finally, all the requirements or their implementations have to be designed to be in line with the existing networking (Internet) architecture.

The proposed reference architecture should enable overall improvement of the application/service implementation instead of defining proprietary solutions that may serve a valuable short-term purpose but significantly impair the long-term flexibility of the Internet. It would be odd to develop services, if the Internet itself cannot take into account their needs or if the services cannot take into account the network architecture of the Internet. As a network architecture, we mean a set of abstract principles for the technical design of protocols and mechanisms for computer communication.

3. VALIDATION OF THE ARCHITECTURE

In this section, the reference architecture is validated with a session continuity example. The validation is done by deriving the requirements for supporting multimedia session continuity in a heterogeneous multi-access network environment according to the reference architecture. The analysis includes a narrative scenario description presented in Section 3.1 and use cases presented in Section 3.2. The sequence diagram shown in Figure 2 illustrates the time dynamics of the scenario. In Section 3.3, the requirements are summarized in a table form and discussed against the existing Internet architecture. Section 3.4 summarizes the results.

3.1 User Scenario

It is close to 6 pm and Mark is sitting in his office located in Lower Manhattan, New York City. He has just started watching a talk show streamed over IP with his laptop computer connected to the office LAN. The show is particularly interesting to him since his wife Miriam is one of the evening's guests. Miriam's new movie has just come out and she is promoting it.

Soon Mark realizes that he has to get going because he had agreed to meet Miriam at a restaurant nearby the TV

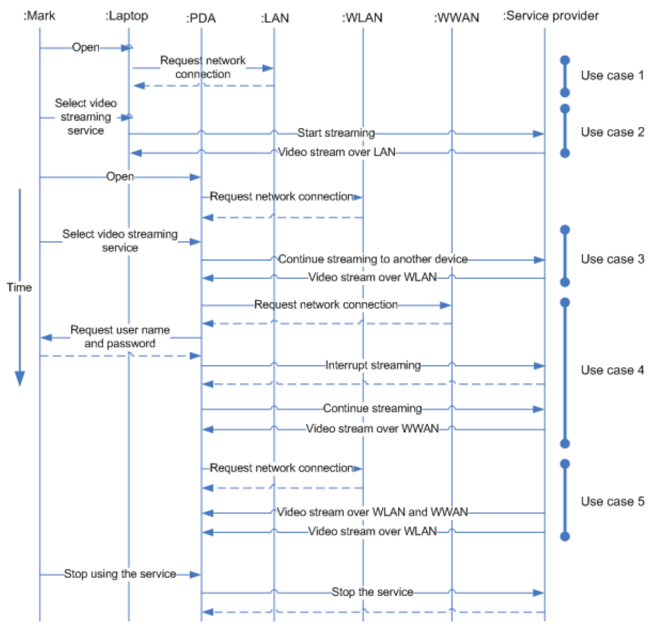

Figure 2: Sequence diagram of the scenario

studio after her show. Miriam is just about to appear in the show, so Mark decides to continue following the show using his handheld multi-access device and headphones as he starts heading out of the office. The video is now streamed over WLAN available in the office building. Mark's network selection preferences allow selecting the access network so that the QoS requirements of the application(s) he is using are always met to the highest possible degree. If multiple suitable networks are available simultaneously, the preference is for the cheapest one.

Mark grabs a taxi outside the office building. The connection to the office WLAN breaks and the multi-access device starts using a wireless wide area network (WWAN) connection, which in this case is UMTS, instead. He watches his wife's interview in the back seat while the taxi takes him from Wall Street to the Rockefeller Plaza. Already in the beginning of the journey, the taxi crawls along slowly for quite a while due to a traffic jam. At slow speeds, Mark's device can use the public WLAN available in the city centre. The taxi driver finally decides to take a detour and the taxi starts picking up speed again. Mark notices from the icon located in the right corner of the display that the used access network changes again. Despite the handovers, the IPTV transmission proceeds uninterruptedly, although the visual quality varies a bit. The interview ends just before the taxi reaches the Rockefeller Plaza, and Mark shuts down the video application.

At the restaurant, Mark manages to have one drink at the bar before his wife joins him. They head to the dinner table together and start discussing her interview.

3.2 Selected Use Cases for Multimedia Service Requirements

This section presents five use cases that give more detailed view of the technical aspects of the scenario. Due to the space limitation, only the key use cases for the analysis can be included in the paper. The outcome of the analysis is a list of multimedia service requirements, discussed in more detail in Section 3.3.

3.2.1 *Use Case 1: Open Network Connection*

Description. The goal of the use case is to describe the process of establishing a network connection. In this case, no additional user authentication is required to access the network.

Preconditions. Network is available.

Steps. Mark powers on the network interface; the laptop listens to the interface; the laptop obtains an IP address.

Issues. None.

3.2.2 *Use Case 2: Start Video Streaming Service*

Description. The use case aims to illustrate the phases involved in streaming a video from a remote server to a user terminal.

Preconditions. Network connection is established.

Steps. Mark starts a video client application on the laptop; Mark selects a video streaming service; the video client communicates the service request to the service provider; the service provider starts sending video to the client application.

Issues. Scaling the video to terminal device capabilities, network QoS.

3.2.3 *Use Case 3: Change the Viewing Device*

Description. The goal of the use case is to describe the interactions involved in transferring a video session from one device to another.

Preconditions. Video is streamed to the laptop, the laptop is connected to LAN, the PDA is connected to WLAN.

Steps. Mark starts the video client application on the PDA; Mark selects the video streaming service; the video client application communicates the request to handover the stream from the laptop to the PDA to the service provider; the service provider starts sending the video to the PDA instead of the laptop.

Issues. Scaling the video to terminal device capabilities, session mobility, network QoS.

3.2.4 *Use Case 4: Imperative Handover*

Description. The goal of the use case is to illustrate execution of an imperative handover. In this case, the old access link becomes unusable at some point, *e.g.*, due to roaming outside a network's coverage area, and the handover is thus unavoidable. Fast execution of the handover may be critical in this case, depending on how quickly the conditions in the old access link are degrading.

Preconditions. Video is streamed to the PDA over WLAN, Mark is approaching the boundary of WLAN coverage area, UMTS is available and supported but no connection is established to it.

Steps. A mobility management (MM) module running in the PDA detects that the quality of the network connection is degrading below a critical level based on network monitoring data or an explicit application-generated trigger; the MM module detects UMTS as the only suitable access alternative available in the area; the MM module opens UMTS connection; Mark enters his user name and password for the connection; if there is enough time, the MM module informs the video application that it is about to make a hard handover to UMTS; the service provider interrupts the streaming and acknowledgement is sent back to the MM module; the MM module performs a handover to UMTS and informs the video application that the handover is complete; the streaming continues over UMTS.

Non-functional requirements. The WLAN connection must be available until the service provider has received a notification about the upcoming handover. Otherwise, packet loss will occur as a result of the handover.

Issues. Proactive and accurate network connection deterioration detection, evaluating and comparing available access options against application requirements, session mobility, network QoS, signalling between the MM module and the video application, preventing unnecessary handovers.

3.2.5 Use Case 5: Alternative Handover

Description. The goal of the use case is to describe the issues related to an alternative handover. During an alternative handover the old and the new access links are both simultaneously available. Fast execution of the handover is thus not critical in this case and more profound evaluation of the benefits and drawbacks of making the handover can be done. The handover can also be done seamlessly by using the two links simultaneously.

Preconditions. Video is streamed to the PDA over UMTS, the PDA's WLAN radio is powered on, Mark enters the coverage area of a public WLAN.

Steps. The PDA connects to the WLAN; the MM module detects that a cheaper access alternative has become available; the MM module evaluates the characteristics of the WLAN (link quality, QoS support, coverage, AAA method) against the video application requirements and decides that a handover is feasible under the circumstances; the MM module makes a handover to WLAN; the streaming continues over WLAN; the MM module keeps the UMTS connection open until all the remaining packets coming through the old path have been received.

Issues. Evaluating and comparing available access and handover options against application requirements, communicating application requirements to the MM module, session mobility, simultaneous multi-access, network QoS, preventing unnecessary handovers.

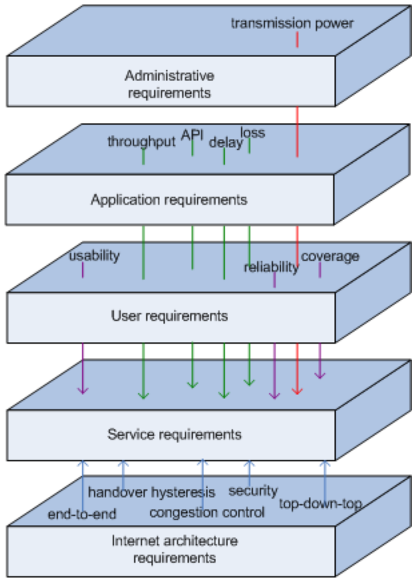

Figure 3: Requirements for a mobility enhanced multimedia service implementation

3.3 Evaluation of the Requirements

Based on the use cases presented in Section 3.2, the list of requirements for session continuity in heterogeneous networks shown in Table 1 can be derived. According to the reference architecture presented in Section 2, the individual requirements set for a service need to be in accordance with those originating from the administrative, user, application, and Internet architecture domains. These requirements are presented in more detail in Figure 3. Since the initial requirements analysis was based on a user-centered approach, this section focuses mostly in evaluating the requirements against the existing Internet architecture.

QoS-sensitive multimedia applications need the network to meet their QoS requirements (throughput, delay, and/or loss limits) end-to-end to operate properly and to provide satisfactory QoE to the user. Since the Internet architecture does not support QoS guarantees but only a best effort service, this is doable only if the networks have been (over)sized to support all possible traffic traversing them. Proposals for implementing QoS support into IP networks, *e.g.*, IntServ and DiffServ exist but their deployment is not at all unified. Moreover, IntServ that supports explicit end-to-end resource reservations has not been successful due to its complexity and poor scalability. Thus one really cannot expect to receive any end-to-end QoS support for a connection that goes beyond the boundaries of an administrative network domain.

The purpose of packet-switched communications was originally to address inefficient resource utilization of circuit switching in the case of bursty traffic. Multimedia applications such as video streaming, however, have at least a minimum throughput requirement that is constant during a usage session. One issue to consider is, therefore, that whether it is reasonable to attempt to implement QoS support into packet-based systems or should the part of communications that needs QoS guarantees be in fact transmitted over circuit-switched connections.

Table 1: Requirements extracted from the use cases and their coherence with the Internet architecture

Requirements	UC1	UC2	UC3	UC4	UC5	IP non-coh.
One or more network accesses simultaneously available	x	x	x	x	x	
Network QoS support		x	x	x	x	x
Video adaptation to terminal		x	x			
Proactive connection deterioration detection				x		
Signalling between entities belonging to nonadjacent layers					x	x
Preventing unnecessary handovers				x	x	(x)
QoS- and application-aware access selection algorithm				x	x	(x)
Estimating the consequences of a handover to applications					x	(x)
Session mobility			x	x	x	x
Simultaneous multi-access					x	x

Having wireless hops in the end-to-end path makes it even more difficult to maintain the required level of QoS for a multimedia application flow. Despite the QoS support mechanisms that can be implemented to the radio interface of UMTS or WLAN connections, the bandwidth allocated to a terminal in UMTS is hardly enough for supporting most video-based services. Setting very high QoS-requirements to applications in heterogeneous multi-access networks limits the amount of usable access networks and technologies.

The current solution for maintaining QoS-sensitive applications under varying network QoS is through adapting application QoS requirements to network conditions. In the case of video streaming, its throughput requirement can be reduced by dropping frames or enhancement layers, lowering frame rate, and increasing frame quantization. Adaptation thus has a direct effect to QoE of the user and it can be applied only up to a certain limit, *i.e.*, until the quality becomes unacceptable to the user. Application adaptation is, however, the only viable solution for implementing QoS support for applications in the current Internet. Adaptation also increases the amount of access networks and technologies that can be used for accessing multimedia services.

Another problem with the existing Internet architecture, with respect to the requirements listed in Table 1, is that it does not inherently support mobility. The IP address has two roles: it represents a node's location in the Internet and is an identifier of a connection endpoint at the application layer (*i.e.*, socket). Thus when a node moves to a new network and receives a new IP address, additional mechanisms are needed for maintaining the ongoing higher-level connections that otherwise would be broken as a result of the IP address change. Mobile IP (MIP) is the solution for extending IP to support node mobility. Over the years there have been numerous proposals for handling IP mobility above the IP layer, *e.g.*, Host Identity Protocol (HIP) [8]. HIP's approach is to allow mobility through separating the two roles of the IP address. This is done by introducing a new layer above the IP layer leaving the IP address to be only the locator of the node and using a different namespace for identifying the communication end points.

In standard MIPv4, the communications from the correspondent node (CN) to the mobile node (MN) goes always through the home agent (HA). This resorts the communications to a client-server mode and to inefficient triangle routing. The routing optimization of MIPv4 (extension) and MIPv6 allow MN and CN to communicate directly with one another in an end-to-end manner as long as CN knows of the current binding between MN's home and care-of address. That is, communications go temporarily through HA only in the case if CN initiates the communications. In current IPv4 networks, MIPv4 needs to employ even more inefficient routing caused by reverse tunnelling due to its incompatibility with certain middleboxes (NATs and firewalls). In HIP, the end-to-end principle of the Internet architecture is supported better than in MIP, that is, sessions between MN and CN are never routed through an intermediate node. Also information security is handled better in MIPv6 and HIP than in MIPv4. Thus, we can conclude that HIP or MIPv6 should be used for mobility support.

Simultaneous multi-access requires that a node supports multihoming, *i.e.*, it is connected to the Internet from two distinct locations and is reachable through two IP addresses. The Internet architecture also does not inherently support multihoming and this capability requires that HIP or yet non-standard protocol extensions of MIP are employed. Multihoming as such does not break communication principles of the Internet architecture and it can thus be enabled. Simultaneous multi-access, however, comes with the cost of using two interfaces simultaneously. In the case of wireless access, this consumes a lot of battery power, which needs to be taken into account in the design.

In the use cases 4 and 5, mobility management needs to be aware of the video application's requirements to make application-aware handover decisions. Also the use case 4 includes direct signalling between the application and the mobility management module prior and after the handover execution to minimize packet loss during a hard handover. Assuming that mobility management operates on network layer, this requires signalling between non-adjacent layers that is not supported in the currently used layered protocol architecture. As discussed in papers [2], [3], and [7] also application and transport level adaptation in wireless and heterogeneous network environments requires access to information available in non-adjacent layers, *e.g.*, link layer. This information transfer cannot be realized with the current protocol stack and API implementations. The requirement identified in the analysis for signalling that basically violates the existing layered protocol architecture challenges the capability of the layered architecture to meet the requirements of future communications in its current form. The layered architecture, however, has its indisputable advantages and making changes to it should be done very cautiously as poorly designed solutions may result in severe performance degradations. The approach proposed in this paper for enabling cross-layer signalling for multimedia services in heterogeneous networks is through the use of stan-

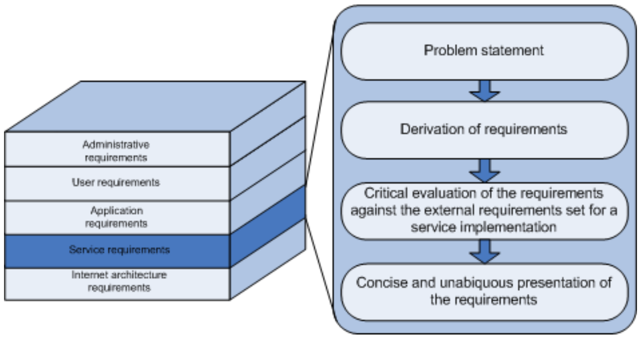

Figure 4: Reference approach

dard mechanisms, such as the upcoming IEEE 802.21 Media Independent Handover (MIH) framework [4], followed by a thorough analysis of the effects of individual design decisions to overall system performance. In long term, a new communication architecture will nevertheless be needed to meet these requirements.

3.4 Detailed Requirements

To summarize the results of the evaluation done in Section 3.3, the following set of detailed requirements can be used to replace those requirements listed in Table 1 that were identified as IP non-coherent. Those requirements whose IP non-coherence was marked in parentheses are dependent on some IP non-coherent requirement and they do not thus require any further analysis. The requirements modified based on the evaluation and their new versions that are better in line with the Internet architecture are:

- Network QoS support -> QoS-adaptive application
- Session mobility -> Mobility using HIP or MIPv6
- Simultaneous multi-access -> Multihoming and simultaneous multi-access support using HIP or MIPv6 extensions
- Signalling between entities belonging to nonadjacent layers -> Cross-layer signalling and information collection through standard mechanisms, *e.g.*, MIH.

4. DISCUSSION

In this paper, a reference architecture for multimedia services targeted for heterogeneous network environments has been presented and validated. The validation was done according to the reference approach shown in Figure 4. The reference approach starts from a problem statement which is used as a basis for deriving requirements for service operation. The requirements analysis is done according to a structured method that includes, *e.g.*, scenario, sequence diagram, and use case definitions. To ensure that the requirements identified for the service are in line with the reference architecture, the initial requirements analysis phase is followed by a critical evaluation of the requirements against the requirements posed by the individual domains of the reference architecture. The final result of the analysis is a concise and unambiguous presentation of the requirements for the service being designed.

5. CONCLUSIONS

The paper introduced a reference architecture and approach that is applicable for mobility enhanced multimedia services. Heterogeneous networks pose new challenges for service development as supporting QoS sensitive applications in such a dynamic environment calls for additional mechanisms for supporting session continuity. The reference architecture defined in this paper can be used for improving such application/service implementations since the method requires evaluating the design decisions against the existing Internet architecture as well as other associated requirement domains. The architecture is validated with a reference session continuity example. The requirements derived in this paper will be utilized in refining the concepts in a simulation study and later on with a prototype implementation.

6. ACKNOWLEDGMENTS

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 program.

7. REFERENCES

[1] C. Perkins, Ed. IP Mobility Support for IPv4. IETF Request for Comments: 3344, August 2002.

[2] L. Daniel and M. Kojo. Adapting TCP for vertical handoffs in wireless networks. In *Proceedings of the 31st IEEE Conference on Local Computer Networks*, pages 151–158, Tampa, Florida, USA, November 2006.

[3] J. Huusko, J. Vehkaperä, P. Amon, C. Lamy-Bergot, G. Panza, J. Peltola, and M. Martini. Cross-layer architecture for scalable video transmission in wireless network. *Image Communication*, 22(3):317–330, March 2007.

[4] IEEE. Media Independent Handover. IEEE Standard 802.21, 2007, work in progress.

[5] S. Kashihara, K. Tsukamoto, and Y. Oie. Service-oriented mobility management architecture for seamless handover in ubiquitous networks. *IEEE Wireless Communications*, 14(2):28–34, April 2007.

[6] V. Kawadia and P. Kumar. A cautionary perspective on cross-layer design. *IEEE Wireless Communications*, 12(1):3–11, February 2005.

[7] Y. Lin, S. Cheng, W. Wang, and Y. Jin. Measurement-based TFRC: Improving TFRC in heterogeneous mobile networks. *IEEE Transactions on Wireless Communications*, 5(8):1971–1975, August 2006.

[8] P. Nikander, J. Ylitalo, and J. Wall. Integrating Security, Mobility, and Multi-homing in a HIP Way. In *Network and Distributed Systems Security Symposium (NDSS'03)*, San Diego, CA, USA, February 2003.

[9] W. Wei and W. Seah. Evaluation of end-to-end QoS support for mobile hosts in IPv6 with IEEE 802.11e. *IEEE 58th Vehicular Technology Conference*, 3:1969–1973, October 2003.

[10] J. Ylitalo, T. Jokikyyny, T. Kauppinen, A. J. Tuominen, and J. Laine. Dynamic network interface selection in multihomed mobile hosts. In *Proceedings of the 36th Hawaii International Conference on System Sciences*, Big Island, Hawaii, USA, January 2003.