SASHA—A Quality-Oriented Handover Algorithm for Multimedia Content Delivery to Mobile Users

Bogdan Ciubotaru, Student Member, IEEE, and Gabriel-Miro Muntean, Member, IEEE

Abstract—The convergence of the existing network access technologies to a common IP-based architecture and the increase in popularity of accessing video content over the Internet makes IPTV a promising solution for media and entertainment industries. Additionally, video content delivery to the increasingly popular mobile devices over heterogeneous wireless networks makes IPTV even more appealing. However the distribution of multimedia content over heterogeneous wireless networks to mobile devices involves significant technical challenges related to mobility management and quality of service provisioning. The existing solutions do not consider quality of service as a decision making parameter for mobility management in general and handover management in particular.

This paper proposes the Smooth Adaptive Soft Handover Algorithm (SASHA), a novel quality-aware approach to handover based on load balancing among different networks using a comprehensive, Quality of Multimedia Streaming (QMS), function for decision making. SASHA represents the handover management solution at the core of the more comprehensive Multimedia Mobility Management System (M3S), a quality oriented mobility management framework for multimedia applications which maximizes user perceived quality by efficiently exploiting all available communication resources.

Simulation-based testing results are presented, outlining the performance of SASHA in different mobility scenarios. The evaluation is performed for different number of nodes performing handover simultaneously and for various situations in terms of networks’ overlapping area. The results show that SASHA outperforms other three mobility management solutions in terms of quality, scalability and resilience to the dynamics of the networks’ overlapping area.

Index Terms—Handover, heterogeneous networks, IPTV, mobility, multimedia streaming.

I. INTRODUCTION

While the trend in current network access technologies is to converge to a common IP-based architecture (all-IP), the increasing popularity of accessing video content over the Internet makes IPTV a promising solution for media and entertainment industries. Moreover by using DVB-IP gateways, IP networks are a viable alternative for traditional multimedia content distribution networks [1].

The large number of mobile handheld devices, with increasing capabilities to communicate in a heterogeneous wired and wireless environment opens the door to a more appealing IPTV service: the Mobile IPTV. Several solutions are available to broadcast multimedia content to mobile devices: digital video broadcasting—handheld (DVB-H), third generation cellular telephone networks (3G with MBMS), broadband metropolitan area network (WiMax) or broadband local area networks (WiFi) [2].

However delivering multimedia content in heterogeneous network environments in a quality-oriented manner involves certain technical challenges. Heterogeneous access technologies, mobile devices with various characteristics relating to computing power, display capabilities, and communication resources [3] and the very different user preferences require special techniques for media encoding, transport streams encoding, rate adaptation and quality provisioning.

Fig. 1 presents a typical situation involving multimedia content distribution to diverse mobile devices roaming through an area covered by different wireless access technologies.

As the Internet was originally designed without considering mobility as a core feature (the traditional TCP/IP model was designed for communication between fixed nodes [4]), in order to accommodate mobility, which is required in heterogeneous network environments, additional development is necessary. In this context a mobility management system has to provide services like handover management, location management, multihoming and security.

Handover management is required for the mobile node (MN) to be able to switch its point of attachment to the network while preserving connection with the corresponding nodes (CN). Location management is employed when a fixed or mobile node initiates a connection (call) to a MN.

Multihoming has to be supported by the protocol stack in the context of heterogeneous networks which provide the MN with network access using multiple communication technologies. Security is a major concern for mobile networks protocol designers as MNs which change their points of attachment while roaming...
through different networks present additional security risks in comparison with the traditional fixed nodes.

Several mobility management solutions like Mobile IP [5], [6], Mobile SCTP [7], Mobile DCCP [8], MOBIKE [9] and Mobile SIP [10] were proposed at different layers of the protocol stack [4]. The proposed solutions provide completely or partially the services required to support mobility in Internet, but their main drawback is the lack of a quality oriented approach. To provide maximum quality of service (QoS) while performing handover, two important decisions have to be made: optimum network selection and the choice for the most appropriate moment to switch networks. This decision making algorithm is usually not specified by the mobility management solutions currently proposed. Several network selection algorithms are proposed in the literature [11] and using the right network selection algorithm has a major impact on the application QoS and on the ultimate user perceived quality of experience (QoE).

This paper proposes the Smooth Adaptive Soft Handover Algorithm (SASHA), a novel quality-aware approach to handover based on load balancing among different networks which considers QoS and QoE in the handover decision making process. SASHA makes use of both the old and the new connections to transfer multimedia data when the user is crossing two networks’ overlapping area. In this context SASHA transfers gracefully multimedia streaming process from the old fading connection to the new improving one. This operation is performed efficiently without data duplication.

SASHA represents the handover management solution at the core of the more comprehensive Multimedia Mobility Management System (M3S) a quality oriented mobility management framework for multimedia applications which uses multiple simultaneous connections to efficiently exploit all the available communication resources, aiming at maximizing QoS and user perceived quality. M3S make use of a comprehensive Quality of Multimedia Streaming (QMS) function for decision making, which combines QoS, QoE, cost, energy and user preference components. M3S is designed as an application layer module used by the multimedia applications to efficiently deliver high quality multimedia content to mobile users.

This paper focuses on comparing SASHA performance with three other schemes proposed for different network layers: Mobile SIP, Mobile DCCP and Mobile IP. Simulation results clearly show how SASHA outperforms these other solutions.

The structure of the paper is as follows: Section II presents three existing mobility management schemes along with some technologies which enable this solution. In Section III M3S is presented and in Section IV SASHA is introduced. Section V details the simulation environment and scenarios and Section VI presents testing results and performs result analysis. At the end conclusions and possibilities for further work are described.

II. RELATED WORK

A. Mobility Management at Different Network Layers

1) Network Layer Mobility Solutions: Mobile IPv4 (MIPv4) [5] and Mobile IPv6 (MIPv6) [6] are the main mobility management solutions at network layer. Mobile IP [5], [6] enables transparent routing of packets to mobile nodes. Each MN is assigned a permanent IP address known as home address which corresponds to the home subnet. While roaming through different foreign subnets, MN acquires new IP addresses (care-of-address) corresponding to each visited subnet. The MN sends a binding update to its home agent (HA) which tracks the current node’s location (care-of-address) and tunnels the incoming traffic from the corresponding node (CN) to MN.

Mobile IP latency can range between 2 s and 10 s depending on the bit-rate and the number of the MNs in the network [12], latency which can be unacceptable for real-time multimedia applications.

Several handover enhancements were proposed for Mobile IP [12] mainly using routing optimization, hierarchical and anticipation techniques.

To avoid the inefficient triangular routing that is involved by the HA tunneling the traffic to the MN’s foreign AP a binding update is also sent to any CN to inform it about the new care-of-address in order for the CN to route the packets directly to MN [6].

Hierarchical Mobile IPv6 (HMIPv6) [12], [13] uses a network organization based on domains which contain several access routers (AR) and a Mobility Anchor Point (MAP) which connects the domain to the Internet. The MAP receives the packets from the CNs and tunnels them to the domain level care-of-address of the MN, Mobility within the domain is managed by the MAP. This solution reduces the handover delay and loss by performing a micro-level address registration which takes less time for binding updates. There is still the macro-level handover (when MNs pass from one domain to another) which involves high latency.

Fast Handover Protocol (FMIPv6) [14] uses Link Layer events (triggers) to improve the handover performance in terms of packet loss by anticipating the handover and tunneling the packets to the new AR until the binding update is received by the HA and CN. In the same time the MN will advertise its presence and availability to the new AR and will start receiving data to the new care-of-address. This solution provides a substantial improvement of handover latency and packet loss. The main drawback of this solution is the precise coordination required between the MN, old AR and new AR and high unpredictability of packets arriving at the APs.

As different approaches, FMIPv6 outperforms HMIPv6 in terms of handover latency and packet loss, but a solution combining both approaches will give better performance than each of them separately [15].

2) Transport Layer Mobility Solutions: Several enhancements for mobility support using TCP and UDP, which are still the main transport layer protocols in the Internet, were proposed in [16]–[18].

The newly developed Stream Control Transmission Protocol (SCTP) [19] provides multihoming support by allowing each endpoint of an association to use several IP addresses. Mobile Stream Control Transmission Protocol (mSCTP) [7] uses SCTP’s ADDIP [20] extension to allow each of the endpoints of an association to change the primary IP address without interrupting the current data transfer. Although mSCTP can provide seamless handover the exact conditions when the primary address should be changed remains an open issue.
A mobility extension for the Datagram Congestion Control Protocol (DCCP)—another recently introduced transport layer protocol—is presented in [8]. Mobile DCCP uses a generalized connection that includes several normal DCCP connections. During the handover a new connection is added using the new IP address while the old connection is deleted. This solution can provide seamless handover although an efficient algorithm for managing the traffic over this group of normal connection is not specified.

3) Application Layer Mobility Solutions: Mobility support at the application layer has also been developed. Two proposed solutions are discussed: one which uses the Session Initiation Protocol (SIP) [21] and one which employs the Internet Key Exchange version 2 (IKEv2) [22].

The basic idea of handover using SIP involves the MN sending a RE-INVITE message to the CN when it acquires a new IP address. The RE-INVITE message informs the CN about the new address of the MN. The new packets will be sent directly to the new IP address or tunneled to the MN until MN sends a REGISTER message to the home SIP server to update the new location. Handover operation using SIP can involve latency for signaling and overhead for IP encapsulation [4]. Some enhancements to SIP mobility were proposed. A solution for reducing handover latency by proactively processing the new address allocation and session update is presented in [23].

MOBIKE [9] was developed as an extension to Internet Key Exchange version 2 (IKEv2). MOBIKE allows both MN and CN to have several IP addresses. When the MN changes its IP address it sends a notification to the CN from the new address. After new location notification the CN starts using this address as destination. MOBIKE permits the MN to move but does not specify how the decision is made to change the IP addresses used for data communication.

B. Multimedia Mobility Management

Two integrated mobility management solutions for multimedia applications are presented in [24], [25].

M4: MultiMedia Mobility Manager [24] uses Multihomed Mobile IP for handover support and a simplified version of Relative Network Load (RNL) for network selection. RNL grades are computed based on the round-trip time (RTT) and RTT jitter values of binding updates. The main drawback of using RTT only for decision making is the lack of information related to network bandwidth and packet loss which have a great impact on multimedia quality.

The multimedia mobility management solution proposed in [25] uses proactive buffering to perform seamless handover and select networks based on received signal strength indicator (RSSI). The main drawback of using RSSI for decision making is represented by the impossibility of detecting the AP’s level of congestion and bandwidth capacity and also values of delay and packet loss.

C. Multi Channel Communication and Software Radio

To reduce handover latency many solutions rely on proactively setting up de data flow on the new AP and eventually duplicating data flow to both APs. These practices raise the necessity of communicating in parallel with two different APs which is basically impossible using only one NIC.

In case of vertical handover, parallel communication is possible because of the necessity of equipping the MN with a different NIC for each technology (e.g. WiFi, WiMax, GPRS etc.). Having a MN with multiples NICs for the same technology is prohibitive due to device cost and power consumption.

Despite the above mentioned drawbacks, parallel communication enables throughput enhancements [26] and also handover enhancement techniques [27]. The emerging technology of software radios [28] can be seen as a solution to allow for highly flexible radio communications for both handover and throughput enhancement techniques.

III. MULTIMEDIA MOBILITY MANAGEMENT FRAMEWORK

In a heterogeneous wireless network environment the mobile device (MN) has access to several networks using different wireless technologies. Therefore the applications running on the mobile device can access a certain content (multimedia content in particular) or service via different communication channels (paths).

Device movements, variable network conditions and variable application’s demands in terms of data traffic, determine a certain level of dynamicity regarding path availability, QoS, cost, and stability. To maximize user’s perceived quality in such a dynamic environment a quality oriented mobility management solution with efficient resource allocation is required.

A. Multimedia Mobility Management System Architecture

Multimedia Mobility Management System (M3S) is an application level framework for delivering high quality multimedia content to mobile clients in the context of a dynamic heterogeneous network environment.

To maximize the application QoS and consequently user QoE, M3S efficiently exploits all the communication resources available to the mobile device.

As presented in Fig. 2, M3S uses several communication channels (links), each channel using a different communication interface. To be eligible for a new active communication channel, the network interface has to be within the range of an
AP (BS) and to be able to register and exchange data packets with it.

Having several active communication channels available, M3S is able to split and efficiently distribute the application’s data traffic over them. Each channel follows a separate (or partially separate) communication path and, as mentioned before, each channel has a certain level of dynamicity that influences its QoS and consequently the QoE.

For efficient traffic distribution and maximization of the overall system QoS and user perceived QoE, each communication channel has to be constantly monitored and the traffic has to be efficiently balanced. Therefore a novel Quality of Multimedia Streaming metric (QMS) is introduced to describe and quantify the impact of different network parameters to the multimedia delivery quality. The QMS score is computed by the M3S server-side module for each channel separately and the traffic balance is performed consequently.

The information required to compute QMS scores is harvested by the M3S client-side module and is reported as feedback to the server. The client measures the QoS related parameters for each separate communication channels and also harvest user related parameters like user preferences and QoE.

Lower level signaling procedures are used to harvest information related to network availability. IEEE 802.21 standard can be used to monitor link parameters and also to search and setup new communication channels (links).

Fig. 3 presents the architecture of the M3S framework. Two main building blocks can be identified: the M3S server side module and client side module.

The server side module is composed of several sub-modules. The Traffic Splitter and Allocator (TSA) sub-module is responsible for splitting the main data traffic into sub-streams according to the available connections and corresponding rates. SASHA Controller is responsible with QMS scores computation and rates allocation over the available connections. SASHA Connection Manager maintains the connection pool and generates the sampling traffic.

The client side module contains four sub-modules. SASHA QoS Monitor (SQM) is responsible with monitoring the QoS parameters on the available connections. SASHA Data Harvester (SDH) communicates directly with the application and is responsible for QoE evaluation as well as gathering information about user preferences, power consumption and network costs. SASHA Feedback Controller (SFC) centralizes the information received from the SQM and SDH modules and sends it to the server module. The Traffic Merger is responsible with re-synchronizing and merging the sub-streams.

M3S provides handover management and efficient quality oriented resource allocation using the innovative Smooth Adaptive Soft-Handover Algorithm (SASHA). SASHA performs handover between different networks by smoothly transferring the load from one network (communication channel) to the other. QMS components will be detailed in the next section.

B. Quality of Multimedia Streaming Metric

QMS is described by the function from (1) and is dependent on the characteristics of the communication channel $i$.

$$QMS^i = w_1 \cdot QoS^{i}_{grade} + w_2 \cdot QoE^{i}_{grade} + w_3 \cdot Cost^{i}_{grade} + w_4 \cdot PEf^{i}_{grade} + w_5 \cdot UPr e^{i}_{grade}$$

(1)

For maximum efficiency and flexibility weights are associated with each component. These weights are set based on user preferences and application requirements. Weights normalization is required, so the condition from (2) has to be respected.

$$\sum_{1}^{5} w_i = 1$$

(2)

The following subsections detail each of the QMS’s components. The components are expressed on a 100 point scale from 0 to 100, with maximum grade 100.

1) Network Quality of Service Grade: $QoS^{i}_{grade}$ represents the grade which assesses the network QoS for the communica-
The throughput component is computed using the formula presented in (5). $\text{Throughput}^i_t$ represents the throughput received by the communication channel $i$ in the time interval $t$. $SRate$ is the application’s average streaming rate in the same time interval.

$$\text{Throughput}^i_{\text{grade}} = \frac{\text{MaxGrade} \times \text{Throughput}^i_t}{SRate} \quad (5)$$

The loss component is described by the functions from (6) and (7). $\text{Loss}^i_t$ represents the average loss recorded by the client on communication channel $i$ on the time interval $t$ and is expressed in Mbps. $SRate$ is the application's current streaming rate. $\text{RateShare}^i_t$ represents the percent of the $SRate$ that is currently allocated to communication channel $i$. $\text{MaxLossRate}^i_t$ represents the maximum allowed loss expressed as a fraction of the streaming rate. $Q_L$ is a quality factor and is set by the application to specify the required sensitivity to loss.

$$\text{Loss}^i_{\text{grade}} = \frac{\text{MaxGrade}}{1 + \left( \frac{\text{Loss}^i_t}{\text{MaxLossRate}^i_t} \right)^{Q_L}} \quad (6)$$

$$\text{LossRate}^i_t = \frac{\text{Loss}^i_t}{SRate \times \text{RateShare}^i_t} \quad (7)$$

The delay and jitter components are described by the functions in (8) and (9). $\text{Delay}^i_t$ and $\text{Jitter}^i_t$ represent the average delay and jitter measured by the client on communication channel $i$ during the time interval $t$. $\text{DTThreshold}$ and $\text{JThreshold}$ are thresholds specified by the application and represent the maximum delay and jitter accepted while still preserving a minimum multimedia quality. $Q_D$ and $Q_J$ represent quality factors which denote application’s sensitivity to delay and jitter respectively.

$$\text{Delay}^i_{\text{grade}} = \frac{\text{MaxGrade}}{1 + \left( \frac{\text{Delay}^i_t}{\text{DTThreshold}} \right)^{Q_D}} \quad (8)$$

$$\text{Jitter}^i_{\text{grade}} = \frac{\text{MaxGrade}}{1 + \left( \frac{\text{Jitter}^i_t}{\text{JThreshold}} \right)^{Q_J}} \quad (9)$$

2) Power Efficiency Grade: $\text{Peff}^i_{\text{grade}}$ represents the energy efficiency score of communication channel $i$ with respect to the MN power usage. $P Mb$ is the power consumed by the transceiver to receive 1 Mb of data. $P$ is a power efficiency factor and denotes the application’s sensitivity to power consumption.

$$\text{Peff}^i_{\text{grade}} = \frac{\text{MaxGrade}}{e^{P \cdot Mb \cdot P}} \quad (10)$$

An exponential relationship between the power efficiency grade and the power consumption was chosen based on the fact that battery lifetime exponential decreases with the increase in load [29].

3) Cost Grade: $\text{Cost}^i_{\text{grade}}$ is a cost related component and is computed based on the user cost-utility rating of the provided service. This component is described in (11) and (12) with $\text{MaxC}$ representing the maximum cost that the user is willing to pay for viewing the specified multimedia content, $C$ being the total cost of streaming the multimedia content over channel

$$\text{Cost}^i_{\text{grade}} = \frac{\text{MaxGrade}}{max(C, C - C_t)} \quad (11)$$

$$\text{Cost}^i_{\text{grade}} = \frac{\text{MaxGrade}}{e^{C_t}} \quad (12)$$
Fig. 6. QMS values for horizontal handover using SASHA with 150 m, 160 m and 170 m between APs respectively.

Fig. 7. Simulation scenario involving two WLANs with access points AP1 and AP2 and a mobile node MN.

\[ \text{Cost}_{\text{grade}}^i = \begin{cases} \text{MaxG} - \frac{\text{MaxG}}{\text{MaxC}} \times C, & C \leq \text{MaxC} \\ 0, & C > \text{MaxC} \end{cases} \quad (11) \]

\[ C = \text{SRate} \times \text{CMb} \times \text{EstimatedPlayTime} \quad (12) \]

4) User Preference Grade: \( U_{\text{Preference}_{\text{grade}}}^i \) represents the score given by the user and express the users preference for the network interface used by the communication channel \( i \). \( U_{\text{Preference}_{\text{grade}}}^i \) is expressed on the same scale as the other QMS components.

5) User’s Perceived Quality Grade: The \( \text{QoE}_{\text{grade}}^i \) represents the user’s perceived quality and is computed based on the received content’s video quality, assessed using no-reference video quality metrics [30], [31]. QoE grade for each communication channel is determined by distributing the overall QoE according to the channel’s corresponding rate share. The function described in (13) outlines the procedure of assessing the contribution of each of the currently used communication channels to the overall QoE grade \( \text{QoE}_{\text{overall}}^i \):

\[ \text{QoE}_{\text{grade}}^i = \text{QoE}_{\text{overall}}^i \times \text{Rate.Share}^i \quad (13) \]

IV. SMOOTH ADAPTIVE SOFT HANDOVER ALGORITHM FOR MULTIMEDIA STREAMING TO MOBILE USERS

OVER WIRELESS NETWORKS

M3S maximizes user’s QoE by efficiently exploiting all the communication resources available to the mobile device. Various technologies, variable network conditions, diverse QoS requirements and device mobility patterns impact on user perceived QoE. To preserve and maximize the level of quality as perceived by the user, M3S employs a novel handover management scheme: the Smooth Adaptive Soft-Handover Algorithm (SASHA).

A. SASHA—A Novel Handover Management Scheme

SASHA performs a quality oriented handover by gracefully transferring the load from one network to another. The handover decision-making algorithm uses the QMS grades computed for each communication channel (network) separately. By taking several QoS and QoE related parameters into account QMS represents a more comprehensive metric aiming to a handover management solution oriented on user’s perceived quality.

Although in the following example the handover is triggered by a QMS drop due to link fading the same algorithm can be employed in case of network congestion, non efficient energy consumption, change in user preferences etc. Moreover the same algorithm is used when several networks are available (APs or BS’s) using different technologies (vertical handover).

Fig. 4 presents schematically a horizontal handover performed using SASHA. The scenario involves two networks using infrastructure modes and having AP1 and AP2 as access points. The mobile node is traveling from AP1’s coverage perimeter to AP2’s, crossing the two networks overlapping area. The handover process is divided in three stages.

When the mobile node resides exclusively within AP1’s coverage area, all the multimedia content is routed over the only available communication channel (AP1).

In stage 1 MN enters the overlapping area. When the link via AP2 becomes available, MN sets a new communication channel to the server and the server sends a low bitrate sampling stream over the new channel to gather QoS information and to compute QMS. The QMS metric is now evaluated for the two communication channels and due to the high distance to AP2, QMS2 is very much lower than QMS1. Consequently all multimedia traffic is transferred via AP1.

In stage 2, MN moves towards AP2 determining the AP1 link to start fading, while AP2 link starts to increase, consequently QMS decreases for AP1’s path and increases for AP2’s link. Based on QMS values, SASHA server starts transferring gradually the multimedia content from AP1 path to AP2 communication channel. This load transfer is an adaptive process which is performed based on the dynamics of QMS values which are
computed for each communication channel separately and updated periodically.

In stage 3 MN is approaching the boundary of the overlapping area and is about to enter exclusively in the AP2 coverage area. In this situation, QMS value for the AP1 link decreases significantly, whereas the QMS value for the AP2 path becomes very high, the SASHA server is forced to route all multimedia traffic over AP2 channel. While the AP1 link is still available, channel sampling is performed and QMS values are computed allowing the handover process to be reversed, if the MN moves back towards AP1.

In case of a MN roaming within the overlapping area of two or more networks the multimedia content will be continuously shared between the available communication channels depending on the values of the QMS grades. Consequently, SASHA’s dynamic behavior in this stage accommodates any travel of the MN within the network overlap area, including u-turns.

B. Quality Monitoring and Decision Making

Fig. 5 presents the pseudo-code of a simplified version of SASHA rate adaptation algorithm. Rate update (Update\_Rate) is performed each time QMS related feedback is received from the client, or new information is harvested from the lower network layers. If the variation in QMS is significant according to the required algorithm sensitivity (a threshold value was introduced), the rate adaptation procedure is triggered.

The first step consists of communication channel selection. Based on QMS values the first best channels are selected which gather enough efficient traffic capacity to deliver high quality multimedia content at the target bitrate.

In the next step the rate share is computed for each communication channel according to the QMS scores and application requirements. The QMS scores are expressed on a 100 point scale and represent the estimated share (expressed in percentage) of the total streaming rate that a certain connection can transport at high quality. The rate share (RS) associated with a connection represents the fraction of the total streaming rate which can be transported at high quality over that connection and is calculated according to the connection’s QMS score. The actual sending rate (R) expressed in Mbps is computed from the target bitrate and the previously computed RS parameter.

The last step distributes the traffic load according to the rate shares computed in the preview step.

For increased performance the sensitivity and reaction speed of the algorithm has to be correlated with environmental factors like network dynamics, size of the networks overlapping areas and also MN speed and trajectory. These aspects of algorithm tuning are beyond the scope of this paper.

The QMS scores are computed by the server side module while the QMS parameters are harvested by the client side module. Consequently the proposed solution involves a certain network overhead determined by QMS feedback sent by the client to the server. The QoS and QoE parameters are sent more frequent while the other QMS parameters (i.e. user preferences, cost, etc.)—seldom. Solutions like MIP and Mobile DCCP present less network overhead as the decision is made by the mobile device (client) and only a location update is required. However if these mobility management solutions are used in conjunction with a feedback-based adaptive multimedia streaming scheme, when using M3S there is a significant advantage of sending the feedback information only once and therefore reducing the overall overhead.

Fig. 6 presents the evolution of the QMS grades for each communication channel when the mobile node is crossing the overlapping area as presented in Fig. 4. The evaluation was performed for four different overlapping areas, with the APs being positioned 150 m, 160 m and 170 m apart from each other.

V. SIMULATION-BASED TESTING

A. Simulation Environment

The behavior and the performance of the proposed mobility solution was evaluated based on simulations conducted using the NS-2 Network Simulator (v2.29) [32]. To evaluate the solution in a scenario as close as possible to a real life situation, the realistic radio patch developed by Marco Fiore [33] was used to enhance the simulation platform.

The simulated environment is presented in Fig. 7. Two wireless APs are connected to an intermediate router which is further connected to the multimedia server.

The two APs were positioned close enough to each other to provide a coverage overlapping area. At the beginning, the mobile devices are positioned outside the coverage areas of the two APs. To evaluate the scalability of the proposed solution with the number of MN’s a number of maximum three nodes were considered to cross the two AP’s coverage areas simultaneously.

M3S and its core handover algorithm, SASHA, are also evaluated regarding the resilience to overlapping area variations. Three situations were considered, with a distance between the APs of 150 m, 160 m, and 170 m. The throughput received by the MN while crossing the APs coverage areas in each of the three situations is presented in Fig. 8. As it can be seen in Fig. 8, for a distance of 150 m between the APs the throughput is almost continuous. A small throughput gap appears when the distance is increased to 160 m, leading to a significant throughput gap when the distance is further increased to 170 m.

Fig. 8. Throughput received by the MN via AP1 and AP2 with a distance between APs of 150 m, 160 m and 170 m.
Fig. 9. Handover loss using SASHA, Mobile SIP Mobile DCCP and MIP with 150 m between APs.

B. Simulation Models

For simulation-based performance comparison several simulation purpose handover models were considered. The next sections will describe each of the simulation model used.

1) Mobile IP: To evaluate Mobile IP, the standard implementation distributed with NS-2 was used. The simulation scenario consists of a Home-Agent (HA) and a Foreign-Agent (FA) which are base-station nodes capable of communicating over wired and wireless links. A mobile node, with only wireless communication capability, represents the Mobile-Host (MH). The MH has its home address set to the HA’s address and it achieves new care-of-addresses when roaming between it’s HA and FA. The MH sends binding updates to it’s HA to inform it about its new location. The handover is performed according to the Mobile IP tunneling-based algorithm.

2) Mobile DCCP: The Mobile DCCP multihoming-based simulation model was also developed under NS-2. Each MN was enhanced in order to be capable of alternative communication with two APs. The decision on the appropriate moment to switch the traffic to the new AP was made off-line. The throughput received by the MN from the APs was evaluated for the specific node mobility scenario and the optimum moment for traffic switching was determined.

The delays involved by switching the data flow from one AP to the other were not considered. Based on the previous considerations and the off-line handover decision making, an optimistic simulation model for Mobile DCCP was created.

3) Mobile SIP: The Mobile SIP simulation model was developed under NS-2 based on an older version of SIP patch developed by NIST and ported to NS 2.27 [34]. The mobility support was added by allowing the mobile client to send a RE-INVITE message to inform the server about the new address of the mobile client (mobile node). The precise timing of the handover was determined by an off-line evaluation of the throughput for the specific mobility scenario and the optimum decision was made.

4) SASHA: To implement SASHA on NS-2 simulator, a mobile node capable of communicating in parallel over two different wireless channels was necessary. As NS-2 simulator v2.29 does not support multiple wireless channels, the implementations of the MN and the ad-hoc routing agent had to be changed. The resulting enhanced solution involves each node having several wireless interfaces and the active channel being able to be set for each node separately in the mobile roaming agent.

The M3S’s core mobility component, SASHA, was deployed in an application which emulates a multimedia streaming server. The application is capable of sending a constant bitrate multimedia content using SASHA for mobility management. SASHA determines the corresponding communication channel, which will be used to send each of the data packets, depending on channel’s QMS score.

In this paper, the following assumptions were made regarding the QMS’s components: the cost of all alternative communication links is considered to be the same, the power efficiency is similar for both network interfaces and as the same technology is used, there is no difference in user preference. QoS component uses average loss, delay and jitter, computed by the client application based on traffic statistics. The QoE component uses peak signal-to-noise ratio (PSNR) to assess user perceived quality.

C. Simulation Scenario

The performance evaluation is accomplished by comparing three mobility solutions: SASHA, Mobile SIP [10], Mobile IP (MIPv4) [5] and Mobile DCCP [8].
The mobility solutions are evaluated regarding the scalability with the number of mobile nodes performing handover simultaneously as well as the resilience to dynamic overlapping areas. For scalability, three different node mobility scenarios were considered, when a single, two and three nodes respectively cross the overlapping area simultaneously. The resilience to dynamic overlapping areas is evaluated by considering three different AP positioning scenarios.

The MNs are crossing the APs coverage areas at a constant speed of about 5 km/h. When passing through the networks’ overlapping area the MNs perform handover according to one of the three solutions.

The goal of the simulations is to determine the capability of each handover technique to maintain high level of user perceived multimedia quality in different node mobility scenarios with variable network overlapping area. Consequently the multimedia server is streaming multimedia content at a constant rate of 1.5 Mbps. No multimedia adaptation techniques are employed.

VI. SIMULATION RESULTS AND ANALYSIS

A. SASHA Performance Assessment

Figs. 9–11 compare the performance of SASHA, Mobile IP, Mobile DCCP and Mobile SIP related to packet loss. Figs. 12–14 present the performance evaluation in terms of user perceived quality estimated based on PSNR.

Fig. 9 presents the packet loss recorded by MNs in each of the three mobility scenarios with a network overlapping area determined by a 150 m distance between the APs.

Mobile SIP and Mobile DCCP performs very well for one and two nodes, presenting insignificant loss rates, but encounters peak loss rates of around 0.4–0.5 Mbps (26%–33%) for almost 5 seconds in case of Mobile DCCP and almost 10 seconds in case of Mobile SIP when three nodes are performing handover simultaneously.

Mobile IP experiences frequent loss rates as high as 1.5 Mbps (100%) for short periods of time (1–2 seconds). The time intervals with high loss rates increase when the three nodes mobility scenario is employed.

Although Mobile SIP and Mobile DCCP outperforms both Mobile IP and SASHA for mobility scenarios involving only one or two nodes, SASHA scales better outperforming both Mobile SIP and Mobile DCCP for the scenario involving three mobile nodes. As it can be seen in Fig. 9 SASHA encounters loss rates around 0.3 Mbps (20%) for periods of time no longer then 1 second.

Figs. 10 and 11 present the mobile nodes’ packet loss for the same mobility scenarios, but with increased distance between the APs (160 m and 170 m respectively), leading to a decrease in network overlapping area size.

From the point of view of mobility scenarios the same scalability trend can be observed when the networks overlapping area is decreased. Mobile SIP and Mobile DCCP outperform SASHA and Mobile IP for one and two mobile nodes crossing the overlapping area simultaneously. For a three node mobility scenario SASHA scales better, outperforming Mobile DCCP and Mobile SIP as well as Mobile IP.

Figs. 12–14 compares the performance of SASHA, Mobile IP, Mobile DCCP and Mobile SIP in terms of user perceived quality estimated by PSNR.
Fig. 11. Handover LOSS using SASHA, Mobile SIP, Mobile DCCP and MIP with 170 m between APs.

The average PSNR score usually achieved by all mobility solutions outside the overlapping area is around 55 db. Next the four solutions’ scalability with the increase in the number of nodes is studied.

For example, as seen in Fig. 14, in the two node handover scenario with a 170 m between APs, Mobile DCCP and Mobile SIP present a PSNR score as low as 20 db for almost 10 seconds while SASHA has its lowest score of 35 db for no more then 3–4 seconds. In the case of the three node scenario, Mobile DCCP and Mobile SIP experiences a drop in PSNR to 10 db for almost 5 seconds and a period of 14 seconds with PSNR of around 30 db. In the same scenario SASHA presents very high PSNR scores, which drop to 30 db for shorter periods of time only.

Mobile SIP performs similar to Mobile DCCP with a slightly better scaling with the number of nodes as it can be seen in Figs. 13 and 14.

B. The Impact of Network Overlapping Area Size on Handover Performance

SASHA’s resilience to different network overlapping areas can be observed by analyzing the evolution of loss rates in Figs. 9–11 and PSNR in Figs. 12–14. Because the impact of the overlapping area size on the handover performance is similar for all three mobility scenarios, the results will be further discussed only for the three node scenario.

Mobile IP presents the same frequent, short term (1–2 seconds), very high loss rates (up to 100%), with a longer period (5 seconds) of high loss (55%) when the distance between APs is increased to 170 m. The trend of the dependency between overlapping area size and Mobile IP performance in terms of loss and PSNR cannot be clearly stated for the one node scenario. As it can be seen in Figs. 9–11, the loss rate when there are 150 m between the APs is higher then the loss rate encountered when there are 170 m between the APs. This can be due to the simulated random fluctuations which are meant to appear in wireless communications and are reproduced by the simulation tool.

When Mobile DCCP is employed, the loss rate is about 0.5 Mbps (33%) for around 5 seconds when the distance between APs is 150 m. When the overlapping area is decreased by increasing the distance between APs to 160 m the loss rate is still around 0.5 Mbps (33%), but the period of time this is encountered for increases to 10 seconds. By further decreasing the overlapping area by increasing the distance between APs to 170 m, a loss rate of about 33% is recorded for 12 seconds with a peak of 73% encountered for almost 5 seconds, significantly affecting user perceived quality.

The performance of Mobile SIP is similar to Mobile DCCP in terms of resilience to variable network overlapping area size. The loss rate is approximately constant when the distance between APs is increased from 150 m to 160 m but the duration of the period when loss occurs is increasing. When the distance is further increased to 170 m both the duration and the loss rate increase.

SASHA is more resilient to the decrease in the overlapping area determined by the increase in the distance between the two APs. For the highest overlapping area size (150 m between the APs) the loss rate reaches 20% for 1–2 seconds only. When the overlapping area is decreased (160 m between the APs) some very short term (1–2 seconds) 20% loss rates appear with a peak of 33% for about 1 second. For the smallest overlapping area
(170 m between the APs) a 13% loss rate is encountered for around 10 seconds with a peak of 26% for 4 seconds only.

In conclusion Mobile DCCP and Mobile SIP perform very well for large network overlapping areas and reduced number of mobile nodes which perform simultaneous handover. Mobile IP encounters short term PSNR drops with longer periods of low scores when the number of nodes is increased and the overlapping area decreased.

Although a certain trend cannot be determined for Mobile IP all three schemes affect significantly their users’ perceived quality when the overlapping area decreases or the number of users performing simultaneous handover increases.
TABLE I

<table>
<thead>
<tr>
<th>Nodes No.</th>
<th>Average PSNR (dB)</th>
<th>Throughput (Mbps)</th>
<th>Loss (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SASHA</td>
<td>64.52</td>
<td>1.50</td>
<td>0.0080</td>
</tr>
<tr>
<td>Mobile SIP</td>
<td>56.60</td>
<td>1.49</td>
<td>0.0194</td>
</tr>
<tr>
<td></td>
<td>62.45</td>
<td>1.49</td>
<td>0.0195</td>
</tr>
<tr>
<td>Mobile DCCP</td>
<td>56.11</td>
<td>1.50</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>56.23</td>
<td>1.50</td>
<td>0.0051</td>
</tr>
<tr>
<td></td>
<td>53.65</td>
<td>1.46</td>
<td>0.0451</td>
</tr>
<tr>
<td>Mobile IP</td>
<td>55.19</td>
<td>1.50</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>55.47</td>
<td>1.50</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>52.34</td>
<td>1.45</td>
<td>0.0235</td>
</tr>
</tbody>
</table>

In these conditions, SASHA outperforms both Mobile IP Mobile DCCP and Mobile SIP recording lower loss and higher user perceived quality.

C. Results Analysis

Tables I–III present average PSNR, throughput and loss for the four mobility solutions, SASHA, Mobile SIP, Mobile DCCP and Mobile IP when used to perform handover between networks whose APs are located at distances of 150 m, 160 m and 170 m respectively.

The impact of the number of nodes and overlapping area size on multimedia streaming performance can be clearly observed. For example, Mobile DCCP presents 0% average loss rate for the one and two MN scenario with a 1.5% average loss rate in case of three MN’s for the biggest overlapping area (Table I). The average loss rate increases to 3.94% for three nodes with the decrease of overlapping area (Table II). For the smallest overlapping area considered, the average loss rate increases to 19.5% for three node mobility scenario (Table III).

A similar trend can be observed for Mobile SIP, the average loss rate being 3% for the largest overlapping area and 18.9% for the smallest one, when the three node mobility scenario is employed.

The performance of SASHA in terms of scalability and resilience to overlapping area variations is clearly depicted by the loss rates. SASHA presents a 1.3% average loss rate only for...
three mobile nodes with the largest overlapping area (Table I), increasing no further than 11.9% when the overlapping area is minimal (Table III).

As the same trend can be observed by analyzing the throughput and the PSNR scores presented, it can be concluded that SASHA maintains a high user perceived QoE during handover even when load increases and network overlapping area is minimal.

VII. Conclusion

Mobility is becoming a crucial component for the future Internet. As IP-based networks were originally designed for fixed IP nodes, mobility solutions have a most important part to play in the future envisaged heterogeneous network environment.

As delivering multimedia content to mobile devices over IP networks becomes increasingly popular this paper presents Multimedia Mobility Management System (M3S) a quality-oriented mobility solution for multimedia applications.

This solution aims at maximizing the end-users perceive quality when streaming multimedia content by efficiently using all the communication resources available. M3S uses the novel Smooth Adaptive Soft Handover Algorithm (SASHA) to gracefully and dynamically distribute the load over the available communication channels based on their estimated contribution in order to deliver high quality multimedia content.

Simulation-based tests show how SASHA offers good scalability with the number of mobile nodes, presenting a 32% improvement in terms of loss compared with Mobile IP and 7% compared to Mobile DCCP and Mobile SIP. In terms of throughput SASHA presents a 21% improvement compared to Mobile IP, 13% compared to Mobile DCCP and 9% compared with Mobile SIP. In terms of PSNR SASHA presents an improvement of 15% compared with Mobile SIP and 26% compared with Mobile DCCP.

M3S and its core mobility solution, SASHA, also present higher performance regarding resilience to variations in network overlapping areas.

Future work will consider assessing the performance of SASHA with even higher number of nodes and with mobile nodes traveling at different speeds. Using adaptive multimedia streaming techniques will also be considered as well as the effect of background traffic on the overall performance of SASHA. SASHA performance in terms of user perceived quality will be evaluated using subjective user quality assessment techniques.

REFERENCES


TABLE III

<table>
<thead>
<tr>
<th>Nodes No.</th>
<th>PSNR</th>
<th>Throughput</th>
<th>Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average (db)</td>
<td>Average (Mbit)</td>
<td>Average (Mbit)</td>
</tr>
<tr>
<td>SASHA</td>
<td>1</td>
<td>60.90</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>55.20</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>44.90</td>
<td>1.36</td>
</tr>
<tr>
<td>Mobile SIP</td>
<td>1</td>
<td>56.50</td>
<td>1.49</td>
</tr>
<tr>
<td>Mobile DCCP</td>
<td>2</td>
<td>48.64</td>
<td>1.39</td>
</tr>
<tr>
<td>Mobile IP</td>
<td>3</td>
<td>38.18</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>54.00</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>47.20</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>32.90</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>45.60</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>45.59</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>38.48</td>
<td>1.18</td>
</tr>
</tbody>
</table>


Bogdan Ciubotaru (S’08) is a Ph.D. researcher with the Performance Engineering Laboratory, School of Electronic Engineering, Dublin City University, Ireland. He was awarded the B.Eng. and M.Sc. degrees in System Engineering from the Computer Science Department, “Politehnica” University of Timisoara, Romania in 2004 and 2005 respectively. His research interests include wireless mobile networks, multimedia streaming over wireless access networks as well as wireless sensor networks and embedded systems. He is student member of IEEE and Research Institute for Networks and Communications Engineering (RINCE) Ireland.

Gabriel-Miro Muntean (M’02) is a Lecturer with the School of Electronic Engineering and co-Director of the Performance Engineering Laboratory at Dublin City University, Ireland. He obtained his Ph.D. degree from Dublin City University, Ireland for research in quality-oriented adaptive multimedia streaming over wired networks in 2003. He was awarded the B.Eng. and M.Sc. degrees in Software Engineering from the Computer Science Department, “Politehnica” University of Timisoara, Romania in 1996 and 1997 respectively. Dr. Muntean’s research interests include quality and performance-related issues of adaptive multimedia streaming, and personalized eLearning over wired and wireless networks and with various devices. Dr. Muntean has published over 80 papers in top-level international conferences and journals and has authored a book and five book chapters and edited two books. Dr. Muntean is Associate Editor with the IEEE TRANSACTIONS ON BROADCASTING and reviewer for important international journals, conferences and funding agencies. He is a member of IEEE and Research Institute for Networks and Communications Engineering (RINCE) Ireland.