An Evaluation of SIP Based Mobility in the IP Multimedia Subsystem

Richard Good, Tapfuma Mvere, Phillippa Wilson, Keoikantse Marungwana, Neco Ventura

University of Cape Town, Rondebosch, South Africa

Email {rgood, tmvere, pippa, marungkoa, neco}@crg.ee.uct.ac.za

***Abstract*— Mobility will be the cornerstone of any Next Generation Architecture. Users will expect seamless mobility across available access networks without visible degradation of ongoing multimedia services. The IP Multimedia Subsystem (IMS) is an IP communications framework that provides an environment for the rapid development of innovative and reusable services; SIP has been adopted as the signaling protocol for this architecture. SIP is independent of the underlying transport protocols and type of session being established making it an ideal candidate for supporting terminal mobility in the IMS. In this paper we present two solutions for terminal mobility in the IMS; the first uses standard SIP session re-negotiation to transfer IMS sessions between different interfaces, this approach involves full IMS Registration on the new interface. The second approach maintains two active connections – the signaling remains on the original interface while the media traffic is transferred to the new interface; this approach requires that the terminal be under the coverage of overlapping access networks for the duration of the session. The performance and tradeoffs involved with these solutions are analysed. Both solutions are implemented in the form of a practical IMS test-bed where they are subjected to realistic use case scenarios.**

***Index Terms* – Mobility, SIP, IMS, Test-bed**

I. INTRODUCTION

TEchnological advances and the miniturisation of computing equipment has resulted in increasingly powerful portable devices. At the same time we have seen the rise of IP multimedia applications driven by the revolution of Web 2.0 Internet services.

This has created the need for a high speed, ubiquitous network capable of catering for diverse application domains; with mobility being a central requirement of this so called Next Generation Network (NGN). Supported behaviour should include personal, terminal, service and session mobility between the same and across different access technologies. Terminal mobility is of particular importance in the NGN environment where access technology will be irrelevant and users will expect seamless mobility across available access networks without visible degradation of the ongoing services. To cater for the explosion in multimedia services, network operators have recognised the benefits of an IP based communications framework. To this end the 3GPP have defined the IP Multimedia Subsystem (IMS) as a platform for rapid service development and a vehicle for advancing fixed and mobile convergence [1]. The IMS provides an environment for efficient development of innovative, reusable applications that are integrated horizontally as opposed to the traditional stovepipe architecture. Other standards bodies have adopted the IMS as a service control element within their broader architectures, most notably ETSI TISPAN, 3GPP2 and the ITU-T.

The rapid creation of services requires dynamic and scalable network management that controls the complex signaling, QoS, security and mobility issues previously handled by circuit switched protocols. The IMS is session based and uses the Session Initiation Protocol (SIP) for session control; comprehensive security and authentication mechanisms are defined but mobility remains an open area. 3GPP2 have focused on Mobile IP as their mobility solution. On the other hand, 3GPP are currently working on specifications for SIP-based IMS-level mobility mechanisms that cater for vertical handovers [2], this will handle the complex mobility requirements introduced by multimedia rich IMS applications.

SIP is independent of the underlying transport protocols and the type of session that is being established. It provides support for all four mobility aspects (Personal, Terminal, Service and Session) and is ideal for the envisaged heterogeneous access environment. The development and integration of SIP handover into the IMS for multimedia rich applications is currently a very active and interesting area of research, and is still in its infancy. Several different approaches have been suggested for implementing SIP based mobility in the IMS architecture. These can be grouped into categories based on where the mobility solution is implemented; that is: Host-based or “device-centric” approaches [3], Middleware solutions which use intermediary devices such as session border controllers and proxies to handle mobility [4, 5] and server based approaches which implement mobility at the Services and Applications layer of the IMS [6].

In this paper we examine the ability of SIP to handle terminal mobility in the heterogeneous access IMS environment – that is the ability to handover between different access networks while maintaining ongoing IMS multimedia sessions. A SIP based approach is first examined and then integrated into the authentication framework of the IMS to provide a partial mobility solution. Both approaches are implemented in a practical test-bed framework where they are subjected to realistic use case scenarios. Specifically, the delay experienced when moving between access technologies and the effect that this has on the user's ongoing session is examined.

II. RELATED WORK

A. *Network Mobility*

IP mobility as defined in [7] refers to the ability of a mobile device to overcome the location-dependent nature of IP addresses. Various protocols have been put forward to handle seamless IP mobility and service continuity across both homogenous and heterogeneous access networks. These protocols can be broadly classified by their scope of operation, micromobility protocols operate within a single administrative domain while macromobility operate across domains.

The IETF has standardised Mobile IP (MIP) [8] to provide macromobility at the network layer of the OSI model. While the network layer is architecturally the best suited layer to perform IP level mobility [7], MIP has well known shortcomings [9]. Several works [10, 11] have been proposed to address the limitations of MIPv4, including the move towards MIPv6. Proposals such as the hybrid scheme of Fast MIPv6 and SIP and cross-layer architectures interwork the different layers of the protocol stack to achieve mobility.

B. *Application Layer Mobility*

Application layer mobility allows for transparent mobility that offers true end-to-end semantics. This means that applications can run oblivious of the underlying network technology allowing for portability across different networks. Numerous works have investigated SIP as the protocol of choice in providing application layer mobility.

While SIP incurs a large overheard due to application layer processing [7], it overcomes a number of drawbacks encountered in lower layer mobility offerings. This has led to the 3GPP's adoption of SIP as the signaling protocol for the IMS [23], making it a key focus area in research.

C. *Proposed SIP Mobility Solutions*

A middleware-based implementation is evaluated in [4] which employs a Session Border Controller (SBC) that has built-in network address translation (NAT) functionality and also acts as a back to back user agent (B2BUA). The SBC is further responsible for managing all the mobility related functions through an entity termed a Mobility Management Server (MMS). All SIP signaling and media flows from the Mobile Host (MH) pass through the MMS/SBC, at which the MH's address is replaced by the MMS/SBC's address. The MH has dual interfaces which can connect to different networks simultaneously. Handover is performed by sending a SIP request to the MMS/SBC containing the address of the new interface, which is connected to a different access network. This approach localises mobility procedures to the MH, thereby shielding the Corresponding Host (CH) from any architectural enhancements. This approach, however, breaks the intended end-to-end semantics of SIP.

A similar approach is presented in [5] where the MH, through the second interface, sends an **INVITE** message with the JOIN header to perform handover to the new network. During the transition, a middleware packet replicator sends duplicate packets to the MH's two interfaces until handover to the new interface is completed; the MH in turn sends duplicate packets through both interfaces until handover is completed. For practical purposes, the main drawback of this approach is the transmission of duplicate packets between each MH and the middleware, an inefficient use of resources.

In general, middleware solutions are network-centric, requiring the addition of extra network equipment, an approach less flexible than host-based solutions. It should be noted that none of these solutions were implemented in an IMS environment.

D. *3GPP Standardisation*

One of the key requirements of 3GPP's Evolved Packet System (EPS) is Service Continuity [13]. This entails the continuity of multimedia sessions across different access networks, different devices and different administrative domains. SIP-based mobility has emerged as the solution for multimedia session continuity at the IMS layer, and it is aimed at complementing existing lower layer intra-network handover mechanisms. 3GPP Release 8 feasibility studies are currently underway for SIP-based mobility solutions in the IMS, while standardised interworkings of non-3GPP access systems such as Wireless Local Area Networks (WLAN) and WiMAX with 3GPP systems are also key areas [14, 15]. 3GPP2 have adopted MIP as their mobility solution.

III. SIP BASED HANDOVER SOLUTION

In this section we discuss the proposed architecture for the SIP based approach to terminal mobility in the IMS.

A. *Architectural Requirements and Design*

The SIP based handover solution achieves terminal mobility through the MH issuing a **re-INVITE** through a new interface to the CH. The **re-INVITE** contains an updated SDP that specifies the new IP address through which the media and signaling should traverse. However, when the MH wishes to use a new IP address the IMS requires that it Register the new IP address and de-Register the old one.

To mitigate delays incurred when the MH is acquiring a new IP address through the dynamic host configuration protocol (DHCP) or the packet data protocol (PDP) context activation we assume an overlapping wireless network architecture as illustrated in Fig. 1. This allows the MH to bring up and configure the new interface prior to the handoff event. It should be noted that the MH requires coverage from both networks when acquiring the new IP address and during the handoff event only. Once the handoff has occurred the MH may move out of the coverage area of the old network.

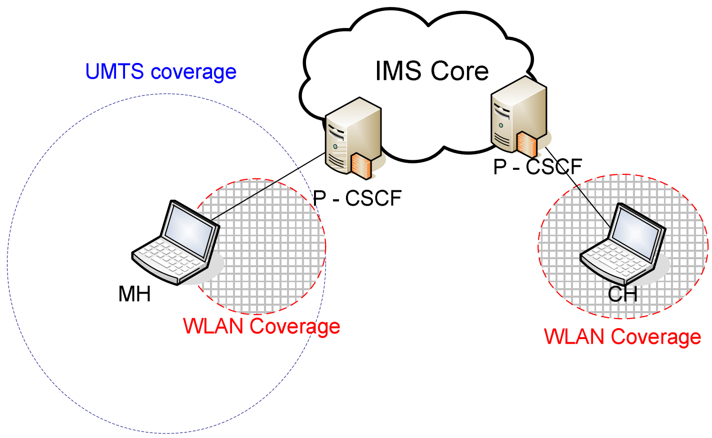

Fig. 1 Standard SIP Mobility Handover Scenario.

B. *Handover Scenarios*

The SIP based approach caters for scenarios where the MH is moving out of the coverage area of one network while having access to another network.

Consider a user moving from his office to his car while on a call. It would be desirable to continue the call session when the user moves away from the WLAN coverage but has access to the wider coverage of UMTS.

The MH will first de-Register the old interface with the IMS core and issue a Register message through the new interface. Upon completion of the registration the MH will send a **re-INVITE** with an updated session description protocol (SDP) message informing the CH of the new IP address to use from this point onwards. The media will then be directed from the CH to the new IP address of the MH. The corresponding signaling flow diagram is shown in Fig 2.

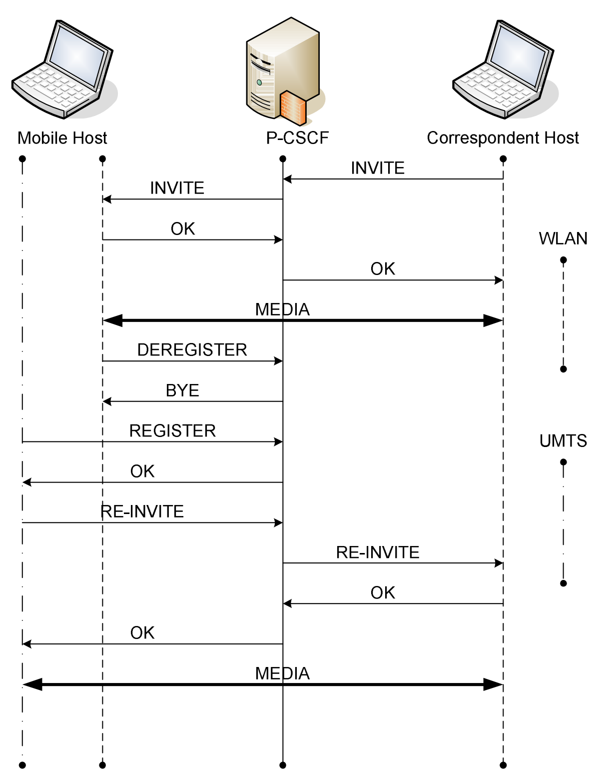

Fig. 2 Signaling Flow diagram for Standard SIP Mobility.

C. *Advantages/Disadvantages of the Architecture*

The scenario described above allows the user to completely move a session to a new access network and move independently of the old access network. In this respect full terminal mobility is achieved.

However a large delay is incurred during the re-authentication process with the IMS core. This delay results in a user perceivable break in communication. So while offering terminal mobility the solution would perform poorly when used with real-time services such as voice and video calling where minimum delays are vital.

The solution also requires user interaction to initiate the handoff event. It would be desirable to automate the handoff so the MH knows when it is leaving the coverage area of one network and entering another. Various methods have been put forward to perform this function but investigating them was deemed beyond the scope of this paper [16].

IV. PARTIAL IMS HANDOVER SOLUTION

The SIP based handover solution presents a problem since it requires re-Registration with the IMS core and hence results in a noticeable cut in the call as well as a large latency during terminal mobility. This section introduces the partial handover solution which addresses these shortcomings.

A. *Architectural Requirements and Design*

The partial mobility solution is a handover scenario that allows a handover to take place from one access network to another without interrupting a session during terminal mobility. This is possible since the partial handover solution makes use of both the old and the new interface in a soft handover setup. The signaling continues on the old interface after the handover, however, the real-time transport protocol (RTP) traffic, which constitutes the majority, uses the new interface of the new access network. Since the signaling remains on the old interface, there is no need to re-Register a new IP address with the IMS core. All that is required is to send a **re-INVITE** containing the new IP address specified in

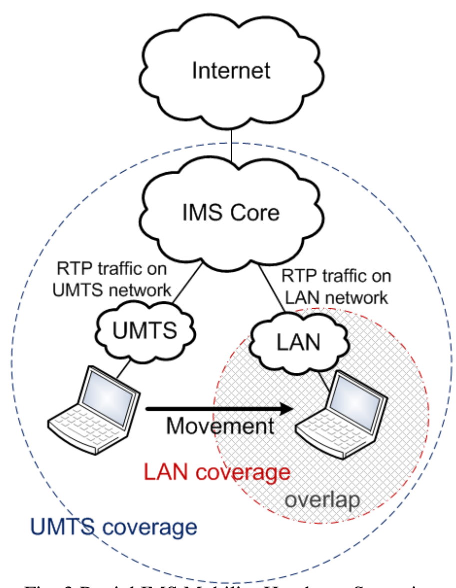

Fig. 3 Partial IMS Mobility Handover Scenario.

the Connection Information field in the SDP. This specifies the address on which the RTP traffic is to be received.
Since both the old and new interfaces are used simultaneously in this solution, the access networks' coverage areas must overlap for the entire duration of the session. This is shown in the architecture diagram in figure 3.

B. *Handover Scenarios*

The partial handover solution supports a number of useful scenarios. One such scenario is the business case when a user is under the wide coverage area of a UMTS network and moves inside an office and is now under the coverage of a WLAN in addition to the UMTS network. The user may wish to transfer the RTP traffic, which constitutes the majority, to the WLAN to save on cost. The partial handover solution supports this with minimal delay and without causing a break in the session. This is vital for real-time applications.
The media, which constitutes the RTP traffic, will flow through the cheaper WLAN network while the signaling will remain on the more reliable UMTS network. The corresponding signaling flow diagram is shown in figure 4.

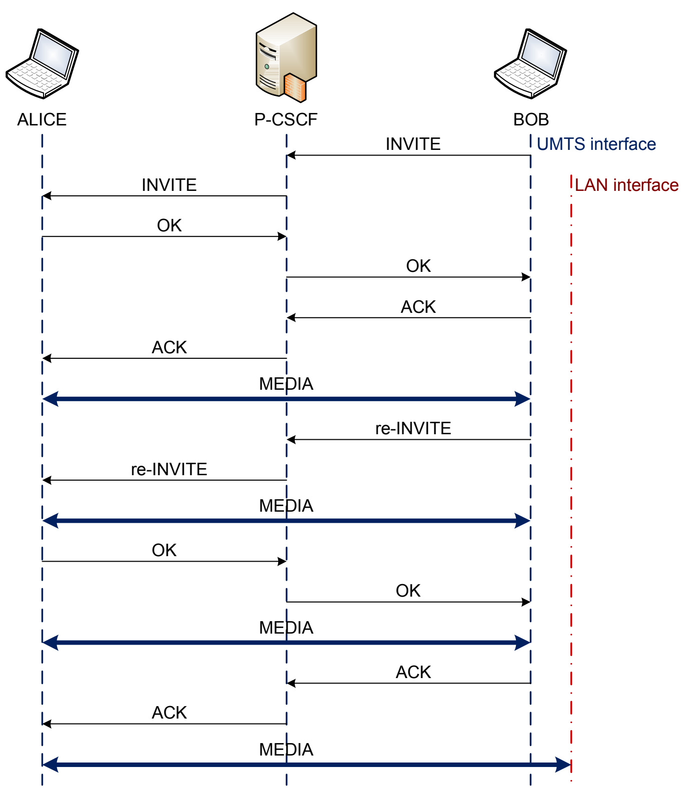

Fig. 4 Signaling Flow diagram for Partial IMS Mobility.

C. *Advantages/Disadvantages of the Architecture*

The partial handover solution has some novel advantages. Firstly, it allows a seamless handover to take place during terminal mobility, without requiring re-Registration with the IMS core. This results in a significantly lower handover delay and no cut in the session which is vital for real-time services. It also allows a user to save on cost by transferring the bulk of the traffic flow to a cheaper network while keeping the minimal but vital signaling traffic on a more expensive but reliable network.
However, the partial handover has a number of disadvantages associated with it, possibly making a full SIP handover a more desirable solution. The first disadvantage is that it requires the user to be under the coverage of two overlapping access networks for the full duration of the session and may only be used in soft handover scenarios. This may not always be possible in a situation where a user wishes to handover to a cheaper network and remain mobile. The second limitation is that the user's device requires two interfaces which have to be used simultaneously. This is undesirable in the case where either the battery life of the device is limited or the user's device only has one interface, for example a cellular phone.

V. Evaluation Framework

The described mobility architectures were implemented in the form of a practical IMS test-bed where they were subjected to realistic use case scenarios and contrasted regarding performance, handover delay and trade-offs involved. Free and Open Source Software was used for all elements of the test-bed; this ensures reproducibility and provides a convenient point of departure for future work in the field.
The FOKUS Open Source IMS Core was used to create the base IMS environment [17]. This package provides software implementations of the three Call Session Control Functions (CSCF) and the Home Subscriber Server (HSS). This package is a standards based and widely used IMS implementation and provides a reliable and comprehensive testing environment.
IMS terminals were represented using the UCT IMS Client – a software based emulation tool that implements full IMS signaling and some services [18]. The source code of this client was modified to implement the SIP based and Partial IMS handover solutions already described. With SIP based handover the client must de-Register the current IP address, Register the new IP address and then re-establish the session with a **re-INVITE** transaction. With the partial IMS handover solution both interfaces are used – the primary interface continues to carry the IMS signaling while the RTP traffic is transferred to the secondary interface; re-Registration is not required.
The IMS Core elements: the Proxy, Serving and Interrogating CSCF, and the HSS ran on a single high performance Intel 3.0GHz Dual Core Server machine, while the IMS clients ran on 1.6GHz Intel Centrino Laptops. The test-bed supported 4 different access technologies: High-speed Downlink Packet Access (HSDPA), Enhanced Data Rates for GSM Evolution (EDGE), Ethernet and 801.11g WiFi. The 802.11g and Ethernet access technologies terminals were connected directly to the IMS Core, however when using the HSDPA and EDGE access technologies the terminals connected via the wireless medium and traversed the public Internet to reach the laboratory hosted IMS core. This was not ideal as longer round trip times were incurred due to public Internet traffic delays. However these delays were incurred by both EDGE and HSDPA and comparative analysis of the two architectures was still possible.
A corresponding client was hosted on the home network connected to the IMS core via Ethernet; the mobile client

established a session with the corresponding node and performed handovers across all combinations of access technology, e.g. handing off from 3G to LAN, 3G to 802.11g, LAN to 802.11g etc. As mentioned previously the handover delay does not include the process of acquiring a new IP address through DHCP or PDP context activation as this can be performed prior to the handoff event. In all test scenarios both interfaces had registered and publicly routable IP addresses. Note that the mobile client could not handover from EDGE to 3G or vice versa because the hardware was limited to a single cellular device per client.

A. *SIP Based Handover*

This section analyses the handover when using standard SIP to support terminal mobility.

The delay involved full IMS de-Registration and Registration, session re-establishment and media stream initiation. Measurements were performed using the built in UCT IMS Client timer – the delay was measured from the first Domain Name System (DNS) lookup for the de-Registration up until the initiation of the media stream after the receipt of the 200 OK in response to the **re-INVITE** transaction. Each session represented a typical session request with an audio and video component and 20 measurements were taken for each test. Figure 5 shows the mean handover delay between each of the different access networks. Table 1 shows the detailed handover delays recorded between HSDPA and 802.11g; this is noted because it is this scenario that provides the most compelling business case for the partial IMS handover solution.

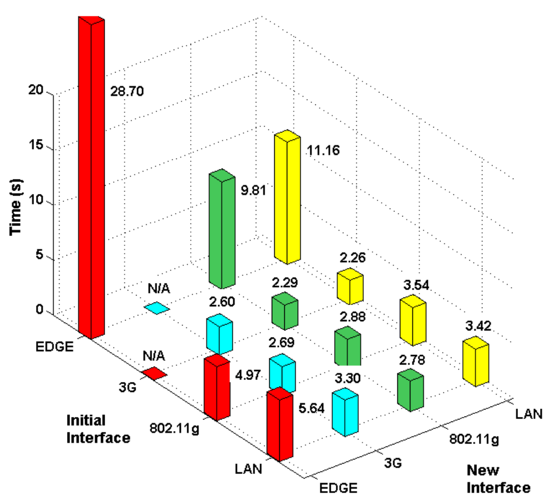

Figure 5 Handover Delay results for Standard SIP Mobility.

TABLE 1
STANDARD SIP MOBILITY HANDOVER DELAYS BETWEEN HSDPA AND 802.11G

	HSDPA to 801.11g (s)	802.11g to HSDPA(s)
Minimum	1.788	1.805
Mean	2.288	2.694
95th Percentile	3.542	5.008
Std Dev	0.525	0.889

B. *Partial IMS Handover Architecture*

The handover delay for the partial IMS terminal mobility solution is analysed in this section. This delay involved only the session re-establishment and media stream initiation as the original interface still carried all the IMS signaling. Measurements were performed using the built in UCT IMS Client timer – the delay was measured from the first DNS lookup up until the initiation of the media stream after the receipt of the 200 OK in response to the **re-INVITE** transaction. Each session represented a typical session request with an audio and video component; measurements were taken when moving between each of the access networks and 20 measurements were taken for each test. Figure 6 shows the handover delay between each of the different access networks. Table 2 shows the detailed handover delays recorded between HSDPA and 802.11g access technologies.

TABLE 2
PARTIAL IMS MOBILITY HANDOVER DELAYS BETWEEN HSDPA AND 802.11G

	HSDPA to 801.11g(s)	802.11g to HSDPA()
Minimum	0.920	0.658
Mean	1.062	0.873
95th Percentile	1.175	0.971
Std Dev	0.077	0.074

C. *Discussion*

The standard SIP handover showed consistent handover delays of over 2.5 seconds across all access technologies; EDGE in particular showed unacceptable behavior which can largely be attributed to SIP retransmission timeouts that were triggered too soon. The handover from HSDPA to 802.11g, which is currently the most compelling business case for mobility, showed a mean delay of 2.228 seconds. For user initiated and not automatic and seamless handover this mechanism could be feasible.

The partial IMS solution showed some interesting results, while the handover delays incurred were greater than the ITU-T specified real time media limits the LAN, 802.11g and HSDPA technologies showed promising results consistently around or below 1 second. While the break in media transmission would be noticeable to the user this solution could be used for automatic but not seamless handover. However this solution needs constant coverage of both access networks. A hybrid architecture combining both systems is a compelling alternative that needs to be further investigated.

VI. CONCLUSIONS

This paper has presented two SIP based solutions for terminal mobility in the IMS. The first uses standard SIP session negotiation while the second maintains two active connections and the RTP streams are transferred between interfaces while the signaling remains on the reliable primary interface. This solution provides a compelling business case for mobility between HSDPA and 802.11g access technologies.

Both standard SIP mobility and the proposed partial IMS architectures were implemented in a practical IMS test-bed where they were subjected to realistic use case scenarios. The results showed a significant handover delay with standard SIP

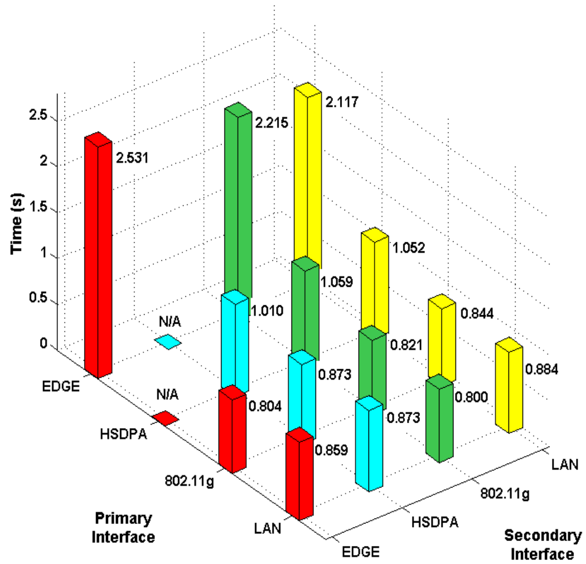

Fig. 6 Handover Delay results for Partial IMS Mobility.

mobility that would be noticeable to the user. The partial IMS solution exhibited better performance but with this architecture the end user needs constant coverage from both access networks. Future work includes the design of a hybrid architecture that combines the standard SIP and partial IMS mobility solutions. The use of a Session Border Controller to ensure minimum disruption to the RTP traffic during handover also needs to be further investigated.

References

[1] 3GPP TS 23.228: "IP Multimedia Subsystem (IMS)"

[2] 3GPP 23.893: "Feasibility Study on Multimedia Session Continuity"

[3] Sujeet Mate, Umesh Chandra, Igor D.D. Curcio, Movable-Multimedia: Session Mobility in Ubiquitous Computing Ecosystem, MUM'06 December 4-6, 2006, Stanford, CA, USA.

[4] Stefano Salsano, Luca Veltri, Andrea Polidoro, Alessandro Ordine, "Architecture and testbed Implementation of vertical handovers based on SIP Session Border Controllers", Wireless Communications: An International Journal, Vol 43, Issue 3, pg 1019-1034, Nov 2007.

[5] Nilanjan Banerjee, Arup Acharya, Sajal K. Das, "Seamless SIP-based Mobility for Multimedia Applications", IEEE Network, March/April 2006

[6] Mehdi Mani, Noel Crespi, "Session Mobility between Heterogeneous Accesses with the Existence of IMS as the Service Control Overlay", ICCS, October 2006

[7] N. Banerjee, W. Wu, S.K. Das, S. Dawkins, J. Pathak "Mobility Support in Wireless Internet" IEEE Wireless Communications, pg 54-61, October 2003

[8] C. Perkins, "IP Mobility Support for IPv4," IETF RFC 3344, Aug 2002.

[9] D. Johnson and C. Perkins, "Mobility support in IPv6," Internet Draft, Internet Engineering Task Force, Apr. 2000. Work in progress.

[10] C. Perkins and D. Johnson, "Route optimization in mobile IP," Internet Draft, Internet Engineering Task Force, Feb. 2000. Work in progress.

[11] E. Gustafsson, A. Jonsson, and C. Perkins, "Mobile IPv4 Regional Registration," draft-ietf-mobileip-reg-tunnel-06.txt, Mar. 2002, work in progress.

[12] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A.,Peterson, J., Sparks, R., Handley, M. and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

[13] 3GPP TS 22.278: "Service Requirements for the Evolved Packet System"

[14] 3GPP TS 23.234: 3GPP to Wireless Local Area Network (WLAN) interworking; System Description"

[15] 3GPP TS 23.402: "Architecture enhancements for non-3GPP accesses"

[16] Chen, et. al., A Smart Decision Model for Vertical Handoff, 2004

[17] D. Vingarzan, P. Weik, T. Magedanz: "Development of an Open Source IMS Core for emerging IMS Testbeds", Special Issue on IMS, Journal on Mobile Multimedia (JMM), Vol.2 No.3, Rinton Press, Princeton, USA, 2006

[18] D. Waiting, R. Good, R. Spiers, and N. Ventura, "Open Source Development Tools for IMS Research," 2008 4th International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, March 2008.

Biographies

Richard Good received his B.Sc. Eng. (Hons) degree from the University of Cape Town in 2005. He is currently working towards his Ph.D. at the same institution. He is an active open source software contributor and has developed several open source IMS tools. His research interests include next generation resource management, service provisioning and mobility in heterogeneous networks.

Tapfuma J Mvere completed his B.Sc. Eng. (Hons) degree at the University of Cape Town in 2007. He is currently reading for an MSc degree in Electrical Engineering in the Communications Research Group at the same institution. His research interests are in mobility in next generation networks and its effects on QoS provisioning.

Phillippa R. Wilson obtained her BSc Eng. degree from the University of Cape Town in 2007. She joined the Communications Research Group at the University of Cape Town as a Telkom Centre of Excellence student where she is currently working towards her MSc (Elec Eng) degree. Her research interests are security and the IMS.

Keoikantse O.A Marungwana obtained his BSc (Eng) Electrical degree from the University of Cape Town and is currently doing his MSc degree in Electrical Engineering. His research interests are IP Multimedia Subsystems and Mobility Management