A new MIP-SIP Interworking Scheme

Karl Andersson

Division of Mobile Networking
and Programming
Luleå University of Technology
SE-931 87 Skellefteå, Sweden
+46 910 58 53 64

karl.andersson@ltu.se

Muslim Elkotob

Division of Computer
Networking
Luleå University of Technology
SE-971 87 Luleå, Sweden
+46 920 49 15 94

elkotob@ltu.se

Christer Åhlund

Division of Mobile Networking
and Programming
Luleå University of Technology
SE-931 87 Skellefteå, Sweden
+46 910 58 53 31

christer.ahlund@ltu.se

ABSTRACT

This paper proposes a new interworking scheme for Mobile IP and the Session Initiation Protocol being the most popular solutions for mobility management at the network and application layers respectively. The goal is to deliver seamless mobility for both TCP-based and UDP-based applications taking the best features from each mobility management scheme. In short, this paper proposes that TCP connections are handled through Mobile IP while UDP-based connection-less applications may use the Session Initiation Protocol for handling mobility. The two mobility management solutions are integrated into one common solution.

Categories and Subject Descriptors

C.2.2 [Computer-Communication Networks]: Network Protocols – *routing protocols*.

General Terms

Algorithms, Performance, Design, Experimentation.

Keywords

Seamless mobility, Mobile IP, Session Initiation Protocol.

1. INTRODUCTION

The Internet Protocol (IP) has been extremely successful in delivering a widespread protocol for host-to-host connectivity using the basic principle “keep the network simple”. However, merging the Internet with the ubiquitous cellular networks has proven to be a quite tough problem to solve. Mobility management is handled very well by the cellular networks at the layers below the network layer, but it has proven to be much harder to implement efficient mobility management solutions at the network and higher layers. One of the basic challenges to deal with when introducing mobility management at higher layers is that network layer addresses not only are used to identify hosts but also to find routes between hosts on the

Internet.

2. MOBILITY MANAGEMENT SOLUTIONS

2.1 Mobility at the network layer

Handling mobility management at the network layer has several advantages since applications do not need to be aware of mobility. If the network layer handles all mobility management, applications can, in theory, be used as if the user was running the application in a fixed environment since the user is reachable through a fixed IP address. The network layer is extended with a suitable mobility management module taking care of the delivery of datagrams to the user’s current point of attachment to the Internet. This mobility management solution works both for connection oriented flows (i.e. TCP connections) and connection less flows (i.e. UDP traffic).

The most well-known example of mobility management at the network layer is Mobile IP (MIP) which is defined both for IPv4 [1] and IPv6 [2].

MIP makes use of a mobility agent located in the home network, a home agent (HA), and, in MIP for IPv4, a mobility agent in the visited network, a foreign agent (FA). The HA is a specialized router responsible for forwarding datagrams aimed for the end-user at the mobile node (MN). The MN is assigned a home address (HoA) in the same subnet as the HA. The FA is responsible for assigning a care of address (CoA) for the MN and forwarding datagrams for the MN. The datagrams are transported from the originating host, correspondent node (CN), via the HA and finally tunneled through an IP tunnel using IP in IP encapsulation to the MN. When the MN changes its current point of attachment to the Internet, it sends a binding update (BU) message to the HA indicating its new CoA. Datagrams in the direction from the MN to the CN are sent directly from the FA to the CN. Route optimization techniques exist in Mobile IP enabling the CN to send datagrams directly to the FA and CN without travelling through the HA.

Unfortunately, MIP has some serious drawbacks delaying global-wide deployment, indeed some of which will be solved when IPv6 is introduced, but nevertheless making MIP not being optimal for roll-out on today’s Internet. Drawbacks include the introduction of encapsulation overhead when tunneling datagrams, the necessity of deploying mobility agents, and problems with sending datagrams directly from the visited

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MUM’08, December 3-5, 2008, Umeå, Sweden.

Copyright 2008 ACM 978-1-60558-192-7/08/12...\$5.00.

network to the CN with source addresses not being topological correct.

2.2 Mobility at the application layer

The above mentioned problems when introducing mobility management at the network layer has led researchers to seek solutions at higher layers. Descriptions of mobility management implemented at the transport layer and the introduction of a separate mobility layer above the transport layer exist in [3]. However, the idea of handling mobility at the application layer using SIP [4] is a popular idea in current research.

SIP is an end-to-end signaling protocol designed for initiating, maintaining, and terminating sessions on the Internet, mainly targeted for multimedia applications, but suitable for any type of session-oriented application. In addition to the client side, SIP user agent (UA), it makes use of three types of servers: SIP proxy servers, SIP redirect servers, and SIP registrars. SIP messages are carried both on top of TCP and UDP and are routed from endpoint to endpoint through available servers. SIP has inherited structures from both SMTP and HTTP making it easier to develop and deploy light-weight implementations when combined with email and web client software. SIP has become the state-of-the-art protocol for signaling in both IP telephony and other types of multimedia applications. SIP is also the core protocol of 3GPP IP Multimedia Subsystem (IMS), making its deployment to real applications even faster. It should also be mentioned that SIP is designed for handling both pre-session mobility and mid-session mobility.

SIP has, however, some drawbacks due to its placement in the layered protocol hierarchy. SIP can not, for example, do anything to broken TCP connections due to changes of network layer addresses at handovers. Also, if SIP is used as a general mobility management solution, already existing applications need to be rewritten completely in order to be mobility-aware.

3. SUGGESTED ARCHITECTURE

The suggested solution is based on the fact that mobility management at the network layer makes TCP connections not to break and mobility management solutions at the application layer make real-time applications like multimedia sessions handled more efficiently. Hence, this paper proposes an architecture based on a combination of MIP and SIP where MIP is responsible for handling TCP traffic and measurements of available access networks and SIP is responsible for handling real-time UDP traffic.

One important cornerstone in the architecture is to perform measurements continually on all available network interfaces at each time using MIP in a multihomed version [5] calculating a metric called the *Relative Network Load* (RNL) [6].

The formulae for calculating RNL_n and policy values U_n for access network n are found below:

$$RNL_n = \bar{z}_n + c J_n \quad (1)$$

$$\bar{z}_n = \frac{1}{h} RTT_n + \frac{h-1}{h} \bar{z}_{n-1} \quad (2)$$

$$RTT_n = R_n - S_n \quad (3)$$

$$D_n = R_n - R_{n-1} - (S_n - S_{n-1}) = (R_n - S_n) - (R_{n-1} - S_{n-1}) =$$

$$= RTT_n - RTT_{n-1} \quad (4)$$

$$J_n = \frac{1}{h} |D_n| + \frac{h-1}{h} J_{n-1} \quad (5)$$

$$U_j = w_p \ln P_j + w_c \ln C_j + w_b \ln L_j \quad (6)$$

$$L_j = RNL_j \quad (7)$$

S_i = the time of sending BU message i

R_i = the time of arrival of BU message i

c , h , w_p , w_c , and w_b are positive, real constants, and

$$w_p + w_c + w_b = 1.$$

h determines the history window for the weighted average calculations. For example, when $h = 5$, the most recent value will contribute to the calculated $\bar{z}_n$ and J_n values with 20%.

c determines the weight of the RTT in comparison to the RTT jitter value. For example, when $c = 5$, the RTT jitter value is contributing five times more to the RNL metric value than the RTT value does.

The variables $\bar{z}$, D , and J are initialized with the following values:

$$\bar{z}_0 = RTT_0; D_0 = 0; J_0 = D_1$$

Furthermore, P_j represents power consumption while C_j is the monetary cost for access network j respectively.

The actual network selection decision is made so that the network with the least value of U is chosen. However, handovers from access networks with large coverage areas (like UMTS and CDMA2000) to access networks with smaller coverage areas (like WLAN and WiMAX) are delayed until the condition: $U_{\text{new}} < U_{\text{old}} - a$ is true where a is a hysteresis avoidance constant used in order to avoid ping-pong effects.

Vertical handovers, i.e. handovers between different wireless access technologies, are performed at the network layer where MIP is sending a BU message to the HA (and possibly to connected correspondent nodes, CNs). Information about the handover is also propagated, through cross-layer signaling, to the application layer in the mobile node (MN) where a SIP re-INVITE message is sent to SIP peers in any ongoing real-time multimedia session followed by the SIP 200 OK and SIP ACK messages. Meanwhile, the HA responds to the MN with an appropriate binding acknowledgment (BAck) message and, through an extension of the HA with SIP proxy server functionality, sends a corresponding SIP re-REGISTER to the SIP registrar indicating the new CoA of the MN followed by the SIP handshake messages, see figure 2. This requires the HA to store SIP addresses and user credentials for each connected MN.

This way, handover will occur at the network layer enabling connection oriented traffic to reach the destination using MIP. In parallel any SIP peer will be noticed about the handover by sending re-INVITE messages with the new CoA address indicated in the Contact field. Hence, real-time traffic can be sent directly between peers avoiding suboptimal paths (i.e. routed via the HA). It should be noted that this solution is beneficial even when using route optimization in MIPv6 because of the network layer making the handover before the

network connection is lost completely, i.e. when performance of an access technology degrades.

Figure 2. Message sequence diagram at handover.

Cross-layer designed solutions are currently a hot topic in computer networking research. With the assumption that higher layers are allowed to take advantage of information from lower layers, the proposed architecture enables the SIP user agent in the MN to subscribe to changes of its current CoA. The SIP Proxy server functionality in the HA subscribes in a similar way to changes in changes of CoA for MNs.

4. EVALUATION FRAMEWORK

In order to evaluate and test the proposed solution a real-world prototype was developed [7]. The prototype was executed in an environment with several wireless access networks with overlapping coverage including UMTS (with HSDPA functionality), WLAN (802.11b and 802.11g), WiMAX (IEEE 802.16e), and CDMA2000 operating on the 450 MHz band. Figure 3 shows the configuration of the test bed.

The client prototype was running on top of Windows XP, while the servers were executing on top of the Linux distribution Fedora core 9. Network layer mobility management on the MN was implemented using WinpckFilter.

Figure 3. Test bed configuration.

The HA contained functionality for tunneling end points and routing and holds a binding cache with entries for each connected MN. A TUN interface acted as a common tunnel endpoint for all MNs. Outgoing traffic was decapsulated at the tunnel end-point and sent out on the home link, while incoming traffic was captured in the HA using proxy Address Resolution Protocol (ARP) functionality. This was handled by making a published static ARP entry in the Linux kernel. The HA responded to ARP requests on the home link on behalf of the MN. When an MN was added a static routing table entry was made to route traffic destined for the MN directly to the TUN interface. The captured packets were inspected to determine the

destination MN, encapsulated and sent to the MN through the tunnel. The tunneling mechanism used UDP over IP encapsulation in order to enable smooth traversal of middle boxes such as NAT and firewalls. IPv4 was used as network layer protocol. The SIP part was implemented using the SIP Express Router software. The Java Media Framework, version 2.1.1e, was used for development of a multimedia application sending audio and video over real-time protocol (RTP) [8] in both directions simultaneously (MN-SIP Peer and SIP Peer-MN).

5. RESULTS

Experiments performed in the above mentioned test bed with the described real-world prototype indicate significant performance enhancements.

Improvements in handling payload for multimedia sessions using the SIP mobility management scheme due to reduced overhead when tunneling is removed are indicated in table 1.

Table 1. Improvements of proposed architecture wrt. payload overhead.

	MIP only	MIP-SIP	Improvement
Payload overhead (RTP, UDP, IP) for multimedia sessions (bytes per packet)	68	40	28
Payload overhead (RTP, UDP, IP) for multimedia sessions using G.711 (64 kb/s) (%)	29.8%	20%	9.8%
Payload overhead (RTP, UDP, IP) for multimedia sessions using G.723.1 (5.3 kb/s) (%)	77.2%	66.7%	10.5%

6. RELATED WORK

The idea of using SIP as an application layer mobility solution is wide spread in the research community and was initially proposed by Schulzrinne et al. [9]. Later Dutta et al. [10] optimized the proposal even further.

Politis et al. [11] described the concept of hybrid Session Initiation Protocol and Mobile IP with joint mobility scheme addressed in different contexts. A MIP stack on top of which SIP resides could exploit the architecture as if using two superimposed worlds: the MIP world and the SIP world in a synchronized and aligned fashion. Two signaling mechanisms go in parallel, independent of each another. The contribution of this work is the architectural integration in conceptual form.

Wang et al. [12] presented SIP and MIP interoperating in the same architecture. The rationale behind the model was that 3GPP only adopted SIP whereas 3GPP2 adopted both MIP and SIP. A hybrid architecture that is in line with 3GPP guidelines was proposed. Mobility between a wired SIP domain and a wireless MIP domain was considered. Furthermore, a mobility broker was enriched with a translation function that converted MIP signaling messages into SIP signaling messages and vice versa. Handoff disruption time was directly linked with delay signaling in their proposal. Mean opinion score (MOS) values for SIP-MIP mixed signaling vs stateful SIP and stateless MIP signaling were compared. MOS values for speech quality perception proved to have a value of around 4.5 out of 5. Although the approach smartly used SIP and MIP jointly, the MOS values were not better than those achieved with HMIP as the paper and presentation verified. However, MIP-SIP

interworking as proposed in this paper with multi-homing removes bottlenecks, and uses a signaling scheme that mingles SIP and MIP control messages both over wired and wireless connections as supported by the architecture. Furthermore, session signaling delay is reduced by the proposed approach, and concrete numerical results on overhead reduction are provided (table 1) instead of MOS values.

Elkotob et al. [13, 14] proposed a hybrid stack that was used in different contexts. The basic idea was to combine the strengths of MIP for doing fast handovers and the strengths of SIP for powerful session adaptation capabilities. The reason for this was the fact that multimedia applications are often handled with SIP or peer protocols and WLAN environments have relatively small cells. Therefore, when traversing WLAN cells while performing VoIP calls, the session update delay caused by handovers is a significant parameter. This metric was analyzed and it was shown with real measurements what the improvement was quantitatively. Furthermore, hybrid MIP-SIP stack as a mobility solution for multimedia was utilized. MIP was used to perform fast handovers and association to a new access point (AP) and to acquire a new IP address. Then the Birdstep MIP client is a specially modified API that sends updates and event notifications upwards in the protocol stack. Then, the SIP mobility module operating adapts multimedia sessions to the new network conditions and after MIP has handled the basic mobility part.

Nasir et al. [15] provided the idea of combining TCP and MIP and UDP with SIP. Our paper provides further features such as multi-homing and improvements on the signaling. Basically it goes a step further compared to [15].

When using MIP and SIP jointly, there are many ways in which those two protocols can interact to improve system behavior and performance. The following significant trends were identified:

- 1) Network layer sending the application layer notifications upwards in the protocol stack meaning MIP is feeding SIP with information which the latter can benefit from for more adaptive application behavior
- 2) Application layer signaling to the network layer information down the protocol stack, such as when a SIP application is coordinated with location based services and triggers a MIP handover early for a seamless handover
- 3) Hard cross layered design is when on the same physical stack network and application layer protocols are integrated using APIs and are prefabricated and adjusted. Some MIP-SIP interaction schemes use that
- 4) Soft cross layered design is when inter process communication (IPC) is used with two different stacks using protocols on different layers. Message passing with triggers is used
- 5) Host multihoming is a feature when several network interfaces can be active simultaneously on the same node. Thus multiple flows can originate or terminate at the same node using different underlying access networks

Table 2. Various MIP-SIP interworking approaches.

Feature/ Approach	L3 to appl. signaling (1)	Appl. to L3 signaling (2)	Hard cross-layer (3)	Soft cross- layer (4)	Multi- homing (5)
[11]	-	-	-	+	-
[12]	+	-	-	+	-
[13, 14]	+	-	+	+	-
[15]	+	+	-	+	-
Our approach	+	-	-	+	+

To our knowledge, the integration of the described policy-based decision model using delay and jitter measurements from the

network layer with a combined MIP-SIP based mobility management solution for UMTS, CDMA2000, WLAN, and WiMAX access networks is new.

REFERENCES

- [1] C. Perkins (ed.), IP Mobility Support for IPv4, IETF, RFC 3344, August 2002.
- [2] D. Johnson, C. Perkins, and J. Arkko, IP Mobility Support in IPv6, IETF, RFC 3775, June 2004.
- [3] W. M. Eddy, At what layer does mobility belong?, In IEEE Communications Magazine, Volume 42, Issue 10, pp. 155-159, October 2004.
- [4] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, SIP: Session Initiation Protocol, IETF, RFC 3261, June 2002.
- [5] C. Åhlund, R. Brännström, and A. Zaslavsky, M-MIP: Extended Mobile IP to Maintain Multiple Connections to Overlapping Wireless Access Networks, LNCS, Volume 3420/2005, pp. 204-213, April 2005.
- [6] C. Åhlund, R. Brännström, and A. Zaslavsky, Traffic load Metrics for Multihomed Mobile IP and Global Connectivity, Telecommunication Systems, Volume 33, Number 1 – 3, pp. 155 – 185, October 2006.
- [7] K. Andersson, D. Granlund, and C. Åhlund, M^4 : MultiMedia Mobility Manager - a seamless mobility management architecture supporting multimedia applications, ACM MUM2007, December 2007.
- [8] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, RTP: A Transport Protocol for Real-Time Applications, IETF, RFC 3550, July 2003.
- [9] H. Schulzrinne and E. Wedlund, Application-layer mobility using SIP, ACM SIGMOBILE Mobile Computing and Communications Review archive, Volume 4, Issue 3, pp. 47 – 57, July 2000.
- [10] A. Dutta, S. Madhani, W. Chen, O. Altintas, and H. Schulzrinne, Optimized Fast-Handoff Schemes for Application Layer Mobility Management, ACM MobiCom 2002, September 2002.
- [11] C. Politis, K. A. Chew, and R. Tafazolli, Multilayer Mobility Management for All-IP Networks: Pure SIP vs. Hybrid SIP/Mobile IP, VTC 2003-Spring, April 2003.
- [12] Q. Wang and M. Abu Rgheff, Signalling analysis of cost-efficient mobility support by integrating mobile IP and SIP in all IP wireless networks, International Journal of Communication Systems, Volume 19, Issue 2, pp. 225-247, March 2006.
- [13] M. Elkotob and S. Albayrak, A Parameter Injection Algorithm for Real-time Traffic in 802.11 Open Access Networks, GLOBECOM 07, November 2007.
- [14] M. Elkotob and K. Andersson, Analysis and Measurement of Session Setup Delay and Jitter in VoWLAN Using Composite Metrics, ACM MUM2008, December 2008.
- [15] A. Nasir and M.-R. Mah-Rukh, Internet Mobility Using SIP and MIP, ITNG 2006, April 2006.