Wireless Netw (2007) 13:61–75
DOI 10.1007/s11276-006-1684-8

The design and evaluation of an end-to-end handoff management protocol

Antonios Argyriou · Vijay Madisetti

Published online: 8 May 2006

Abstract We introduce an efficient protocol for end-to-end handoff management in heterogeneous wireless IP-based networks. The protocol is based on the stream control transmission protocol (SCTP), and employs a soft-handoff mechanism that uses end-to-end semantics for signaling handoffs and for transmitting control messages. The design goal of this protocol is twofold—first, to reduce the home registration delay, and, second, to eliminate the tunnelling cost which exists in current proposals, such as Mobile IP and its derivatives. Furthermore, we propose successive enhancements to the initial mobility management framework for achieving better scalability. We present strong analytical and simulation-based results that show performance improvements over existing approaches.

Keywords IEEE 802.11 · Wireless LAN · Mobile IP · Mobility management · SCTP

1. Introduction

A high quality of service (QoS) is possible for mobile communications over wireless networks, if it were possible to provide high-bandwidth and tighter control over path delay, jitter, and packet loss. Mobile IP, and its various derivatives [10,19], were designed in order to provide a scalable solution that facilitates handoff between IP-based subnets using IP encapsulation and tunnelling. Its basic functionality is based on three features: Advertising a "care-of address", registering the current location of the mobile host, and tunnelling data packets. Agent advertisements, sent by specialized routers called foreign agents (FAs), are used to discover available links within a subnet. However, these mechanisms incur significant overheads like increased handoff delay, triangular routing, and IP encapsulation. All these issues constitute important performance bottlenecks that must be resolved prior to the widespread acceptability of Mobile IP. Even though a number of optimizations have been reported [10, 20], the fundamental overhead in tunnelling is unavoidable, while new problems appear mainly related to security [11].

While Mobile IP targets mobility at a macro-level, a number of supplementary mobility management schemes have been developed in order to optimize for the case of frequent handoffs due to increased localized mobility of a host (micro-mobility). Hierarchical Mobile IP [10], Hawaii [22], Cellular IP [3], are some of the approaches that fall into this category. Hierarchical Mobile IP (HMIP) [10] is an extension to the base Mobile IP that allows the creation of a localized and scalable architecture. HMIP solves the problem of excessive signalling delay when the current visited network is far away from the home network. HMIP defines a way of localizing registrations to the visited network so that they do not have to traverse through the home network. This local architecture is realized through an entity called the gateway foreign agent (GFA). The mobile host uses regional registrations to the GFA in order to reduce the number of signaling messages to the home network. Moreover, for frequent intra-domain moves, the signaling delay is reduced since the related signaling messages have only localized significance. However, while reducing some signalling costs, these localized Mobile IP-based approaches share a common overhead—tunnelling costs. Moreover, the implied dependence on specialized routing agents (FA/GFA) creates points of failure in the network. Mobile IP with route optimization [20], is used in order to route all the packets destined to a MH directly to its current

A. Argyriou (✉) · V. Madisetti
School of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
e-mail: {anargyr, vkm}@ece.gatech.edu

location and not through its HA. The authors propose protocol extensions to Mobile IP, so that a CH can cache the current binding of a MH, and then tunnels the packets for the MH directly to its care-of address, bypassing thus the MH's HA. The authors also propose a mechanism that allow packets in flight to a MH that moves, and packets sent based on an out-of-date cached binding, to be able to forwarded directly to the MH's new address binding. Another approach that shares some features with the proposed one in this paper, is homeless mobile IPv6 [17]. Homeless MIPv6 is basically intended for mobile hosts that do not need to or want to use any explicit home agent. The authors propose a number of modifications to mobile IPv6 so that it does not require, nor assume, any explicit home addresses or agents. One major drawback is that is based on IPv6, making it thus impractical for the current Internet. Moreover, it operates at the network layer and so TCP cannot handle the IP address changes. Finally, this protocol also requires the support of IPSec which incurs significant overhead and adds to the registration delay.

The first approach to handle mobility at the transport layer was TCP-Migrate [25], which is based in extensions of TCP. Location management is implemented with DNS [21] updates sent to the home network notifying it about the new IP address of the mobile host. In the meantime the TCP connection is suspended until the DNS update has taken place. If the delay is large (which is highly probable since the updates are sent to the home network), TCP might suffer from timeouts and so performance will be degraded. Another major handicap is the need for IPSec (which is penalized by a high overhead) every time the IP address is changing. Finally, additional performance degradation is attributed to the increased number of packets being lost during handoffs, because of TCP's inability to accept new socket parameters while in an existing session. Recently some proposals have been made do using the stream control transmission protocol (SCTP) [27], for mobility management [13,24,31]. These approaches simply state the observation that SCTP can handle changes in IP address, and exploit that for demonstrating a simple handoff. However, most of the aforementioned approaches, fail to provide detailed analysis of the benefits of such a solution.

A number of application layer solutions for mobility management have also been proposed in the literature. For example the session initial protocol (SIP), has been suggested as a flexible solution to the problems of terminal and user mobility [29]. Other notable approaches include protocols, that define new handoff procedures for supporting better QoS for media applications [1,2,18].

We argue that mobility handoff management would be best handled by a protocol that operates over IP while still being able to inter-operate with network layer mechanisms (i.e. MIP). Therefore, we propose a protocol that works at the transport layer, and is implemented as extensions to the IETF's standard, the stream control transmission protocol (SCTP). The proposed handoff management protocol is general enough since it operates over generalized IP-based connectionless networks (Internet). We envision users, being able to simply upgrade the software of their systems and be able to roam uninterrupted without requiring specialized IPv4 or IPv6 based mobility solutions. This feature, will be even more important in the near future, since new incompatible wireless access technologies are emerging, while old ones are changing and evolved.

The rest of this paper is organized as follows—in Section 2 we analyze the proposed protocol, and the overall system architecture. Section 3 provides implementation details in the context of the SCTP. In Section 4 we provide a detailed analytical study of the proposed protocol while Section 4.5, presents analytical results. In Section 5 a localized architecture for further optimizing the protocol operation is presented while additional optimizations with a pre-registration mechanism are presented in Section 6. Simulation results for a wide range of scenarios are presented in Section 7, while Section 8, concludes the paper.

2. Description of the protocol and related algorithms

Primary design goal of the proposed protocol is to operate under various wireless access technologies, and to provide various levels of mobility services and to support different types of QoS architectures. Additional requirements include limited modifications to the core infrastructure to a very minimum, and compatibility with Mobile IP and its various standardized derivatives. The choice of defining a transport layer based protocol appeared as solution capable of meeting these challenging requirements.

The distinguishing feature of the protocol is that it completely works around TCP's inherent inability to handle socket parameter changes for ongoing sessions. This is achieved by allowing a transport layer session to maintain more than one active addresses for both the source and the destination[1]. These addresses may be added or deleted dynamically for ongoing sessions. In this way a transport layer session is not defined by a specific IP/port number pair of the source and the destination. This feature of our protocol, removes two significant drawbacks of baseline Mobile IP: First, the overhead of IP encapsulation and tunnelling and second, the dynamic addition/deletion of addresses eliminates the need for triangular routing, since a host experiencing handoff need only to notify initially only the CH. While the protocol is new, we will use terminology borrowed from Mobile IP for convenience. This facilitates comparison with specific functionality of MIP, HMIP, and MIP with route optimization

[1] This does not imply necessarily multiple network interfaces.

(MIP-RO) in a straightforward manner. We progressively build on the simplified version of the protocol through application of successive optimizations. In this section we define in detail the operations performed by Mobile-SCTP in order to handle handoffs in an end-to-end decentralized fashion. The description of specific operations of the protocol follow.

2.1. Joining the foreign domain

When a mobile host enters a new foreign network or subnet, it must obtain a valid IP address from the visited network (e.g. via DHCP [6]) which is called care-of address (CoA). Normally the responsibility of a DHCP server is to provide an IP address to the MH. If the DHCP server provides a universally routable IP address, then the only responsibility of the foreign network is to keep track of the usage of this address. The only overhead with our protocol, is related to the processing at the DHCP server since it maintains a list of the allocated addresses. However, things are different if the MH obtains a private address, and communicates behind a NAT server [7]. In that case the NAT server effectively hides host movements inside the visited network. Even if the MH changes IP address while moving from one subnet to another, the HA and CH will still be unaware of that.

2.2. Handoff algorithm

Now in this subsection we describe what steps are followed when there is a need to handoff. The handoff process is presented in Algorithm 1. In this algorithm we assumed that the MH obtains an IP address through the DHCP protocol [6]. However, the same procedure could be followed with IPv6 by using the auto-configuration feature that IPv6 supports [5].

Algorithm 1 The handoff algorithm.

1: Mobile host enters a subnet and sends a DHCPDISCOVER message.
2: The DHCP server sends a DHCPOFFER that contains the assigned IP address.
3: The MH uses this new IP address as its primary address for any new connection.
4: When the MH receives an advertisement from a new AP, sends a new DHCPDISCOVER as before.
5: Upon reception of the new DHCPOFFER reply, it sends a DHCPACK to acknowledge the address assignment.
6: Binding update is send to the CH.
7: The CH uses the new address to reach the MH
8: The MH notifies the HA about the current location periodically.

2.3. Binding updates

As we said in the previous subsections, even with the proposed protocol, there is a need for a HA entity in order to provide the location service. However, with the proposed protocol, the CH is notified directly by the MH when handoff is about to take place through a direct binding update from the MH to the CH. Actually, the binding update procedure is separated in two parts—A binding update that corresponds to the updates sent to the CH during handoffs, and a periodic binding (that also MIP requires) sent directly to the HA. The second part corresponds to a periodic update that has limited relationship, if any, to do with updates due to host movement. This means that there will be cases where a MH has moved to a new location, and has a new IP address, but binding update need not take place, even though mobile IP would require it. We allow a level of "staleness" in the current location of the mobile host so that signaling traffic is reduced. We will later see how we specify this periodic update and analyze the relative tradeoffs. However, when a lag can be accepted in the notification of the HA, we must make sure that binding updates sent to the CH are delivered as soon as possible. These updates may happen while there is an ongoing data transmission, and more specifically in our case can be done pro-actively (see Section 3).

2.4. CH To MH communication

There are also cases when CH requires connection with the MH. Clearly, the only practical way that this can happen is with the help of the home network, which should be able to provide the current location of the MH. After the CH receives the current location of the MH it can initiate direct communication. This is implemented through a DNS function at the HA which provides the current IP for the CH. Another option would be the exchange of typical mobile IP messages between CH and HA in order to obtain initially the current address binding of the MH [19]. Homeless MIPv6 is a proposal that overcomes the need of a HA, but it is tightly couple with IPv6 routing. Since we are considering IPv4 systems in this paper, this solution cannot be applied. Note that inter-operation with Mobile IP can be smooth since we do not intervene with the inner working of the protocol.

3. The SCTP-based implementation

The recently adopted IETF stream control transmission protocol (SCTP) is a reliable transport protocol that was initially designed for signaling transport. To its advantage, it proved to be a powerful, feature-rich transport layer protocol, that can operate on top of connectionless IP-based packet networks such as the Internet. One of the most important new ideas that SCTP is using is that of explicit support for multi-homed hosts. A single SCTP association (session) is able to use alternatively anyone of the available IP-addresses without disrupting an ongoing session. However, this feature is currently used by SCTP only as a backup mechanism that provides fault tolerance in case of link failures. The state of each IP-address (path) is maintained by sending heart-beat messages and allowing thus the detect a specific path failure and switch to another IP address. Another novel feature is that the protocol decouples reliable delivery from message ordering by introducing the idea of streams. A stream is an abstraction that allows applications to preserve in order delivery within a stream but unordered delivery across streams. This feature avoids head-of-line blocking at the receiver in case multiple independent data streams exist in the same SCTP session. Despite all these differences, congestion control was defined similar to TCP, primarily for achieving TCP friendliness [27].

In the next two subsections, we describe in detail, how we implemented the proposed algorithms as part of SCTP. More specifically we present the exact message exchange procedure during handoffs and we give solutions to a number of side-effects that occurred due to the transport layer nature of the protocol.

3.1. Realizing binding updates through dynamic address reconfiguration

Recently, IETF has proposed a simple extension to the base SCTP protocol that allows dynamic reconfiguration (addition/deletion) of IP addresses of an existing association between two SCTP endpoints [26]. This SCTP extension was used in order to give the MH the ability to be able to request from its peer to reset its primary destination address to a new value. The new IP address in this case will be the IP received when registering with the new network during handoff. By doing this, the protocol is able to always maintain an active source or destination address which is topologically correct.

Figure 1 presents in detail the sequence of messages exchanged during a binding update process. The MH initially goes through a handshaking procedure with the CH that involves four SCTP session setup messages [27]. After that, a connection is established the data flow is initiated. At some point in the future, the MH may decide to move and at this moment Mobile-SCTP gets into action. Upon entering the area of a new access point the MH will start receiving advertisements. The MH immediately sends a registration request

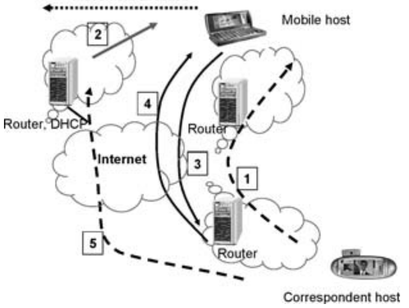


Fig. 1 Message flow of the end-to-end handoff management protocol

and obtains a registration reply. Subsequently, the MH sends a packet with an ASCONF chunk (AddresS CONFiguration [26]) to the CH requesting from it to add the new address to the current association. Upon reception of an ASCONFACK packet, the MH starts using the new IP address as its source for all its packets. Clearly, the latency of the registration procedure and the handshake with the CH, represent the total handoff delay. But as we mentioned earlier (Section 4.5), a MH must be able to move very fast or the packet arrival rate must be really high so that increased number of packet losses take place. Nevertheless, in the next subsection we delve into this matter and proposed a simple yet effective solution.

3.2. Avoiding RTO timeouts and spurious retransmissions

One important problem that arises during a handoff, is that packets that were sent to the previous access point, and are still in transit, will be lost. In order to analyze this behavior, assume a movement pattern by the MH that assures t_{OV} time of concurrent coverage by both the neighboring APs. Let also t_{ASCONF} be the delay until the CH gets the ASCONF chunk/packet from the MH and t_{RECV} be the time frame for which the MH continues the reception of data from the old AP, after it has entered the region of the new AP. If *cwnd* is the current value of the congestion window, then the percentage (%) of packets that will be lost during this transition is $p_{lost} = \frac{(t_{ASCONF}-t_{ov})}{RTT/2} \times cwnd$ if $t_{ASCONF} - t_{OV} \geq 0$.

An explanation is warranted at this point; until the ASCONF packet arrives at the CH it will approximately need RTT/2 seconds. If t_{OV} were zero, then this means that the CH would have sent a whole *cwnd* chunk of data to the old IP address before it receives the notification. However, as t_{OV} is increased, the MH will still be in the range of the old AP and its old IP will still be functional allowing it thus to receive a percentage of the packets sent there. Now if half or more of the *cwnd* is lost the sender will experience an RTO timeout. This simple analysis reveals that this problem is more complicated since it is coupled tightly with the mobility pattern that the MH follows. Our simulation results depict all these issues in an illustrative manner. Furthermore, *cwnd* in the new link will start from one segment since SCTP maintains different set of congestion control parameters. However, accumulated data at the CH for retransmission may not be able to leave fast even if they are sent first.

4. Analytical protocol modelling

In this section we define an analytical model that describes the costs related to all operational phases (registration, tunnelling, packet delivery etc.) of the protocol. We adopt notation from [4] and [30] so that comparison with Mobile

IP becomes easier. At the same time we develop analytical model for two promising mobile IP optimization mechanisms: Hierarchical MIP and MIP with route optimization. We believe that comparison of Mobile-SCTP with these two solutions represents a fair comparison with state of the art mobility mechanisms. These protocols are currently considered the best solutions for practical MIP deployment.

4.1. Model parameters and assumptions

All the parameters that are needed by the model can be seen in Table 1. As we go into more details of the protocol model into the next subsections, the meaning of each one of these parameters will become more clear. In these parameters, we can see that the model ignores the effect of queueing delays in the core network. Therefore the model will provide good estimates for the several costs related with the operational faces of the mobility protocols, when the core network is lightly loaded. This assumption is valid in today's Internet, since the bottlenecks appear usually at the access networks, while the core networks are usually under-utilized [12].

We evaluated the performance of the proposed protocols, based on the assumption that the mobile host will move randomly between the subnets [28]. Given that a network consists of *N* subnets, and that each subnet consists of *r* regional networks in case of HMIP, the expectation of a host going out of a regional network at movement *M*, requiring thus notification of the home network, is given by:

$$E[M]_{hmip} = \frac{2N - r - 1}{N - k} MR \quad (1)$$

In case $r = 1$ then each subnet consists of one regional network. As the number *r* is increased, the probability to move out of a regional network will be increased, increasing thus $E[M]_{hmip}$. In the case of Mobile-SCTP, there are no regional networks. This means that every time the MH moves out of a subnet it must inform the home network. Therefore:

$$E[M]_{msctp} = 2MR \quad (2)$$

Table 1 Parameters of the handoff protocol model

Packet service time at the FA (s_f)	10
Packet service time at the GFA (s_g)	15
Packet service time at the HA (s_h)	25
Distance cost (δ_D)	0.05
Weight factor (w_1)	0.3
Weight factor (w_2)	0.7
MH per subnet (n)	15
Wireless cost multiplier (ρ)	10
Poisson packet arrival rate (λ_a)	variable
Mobility ratio (MR) $1/T_f$	variable

4.2. Packet delivery

Packet delivery cost is crucial overhead in mobile IP's performance, as packet tunnelling is necessary even when the MH moves infrequently [28]. As mentioned earlier, the use of various optimizations like HMIP [10] are not sufficient to eliminate this problem. By moving mobility management "one layer up", the need for tunnelling is vanished since the protocol can now accommodate IP address changes.

We now calculate the overhead from packet delivery for MIP, HMIP, MIP-RO, and our end-to-end protocol. As in [4] and [30] we also define the following variables: c_h and c_g are the packet processing cost at the HA and GFA respectively. L_{ch}, L_{hg}, and L_{gf} are the transmission cost of delivering a packet between the CH to HA, HA to GFA, and GFA to FA respectively. Mobile IP with regional registration has a packet delivery overhead [30]:

$$C_{PD}^{hmip} = c_h + c_g + L_{hg} + L_{gf} + L_{ch} \quad (3)$$

Transmission related costs are calculated as follows: $L_{gf} = l_{gf}\delta_D$, where δ_D is a proportionality constant and l_{gf} is the GFA/FA distance. In the case of Mobile-SCTP $c_h = 0$ since packets do not go through the HA and so the packet delivery cost with it is simply:

$$C_{PD}^{msctp} = c_g + L_{cg} + L_{gf} \quad (4)$$

Where L_{cg} is the transmission delay Mobile-SCTP, since no tunnelling is involved or any other form of intermediate processing of data packets. If λ_a is the data packet arrival rate at the HA, then the packet processing delay is analogous to λ_a with $c_h = \lambda_a$. In addition, the lookup overhead of the IP routing table has to be calculated which is analogous to its length *e*, the number of MHs in the subnet *n*, and course the packet arrival rate λ_a. Therefore this cost is $k\lambda_a log(k)$. We also assume that the distance between CH-HA and HA-GFA is the same, making thus $L_{hg} = L_{ch}$. Eq. (3), will therefore become:

$$C_{PD}^{hmip} = \lambda_a + k\lambda_a(w_1 nk + w_2 log(k)) + (l_{gf} + l_{hg})\delta_D \quad (5)$$

MIP-RO does not suffer from triangular routing but it still has to suffer the tunnelling overhead ($c_h = 0$). So:

$$C_{PD}^{mipro} = c_g + L_{cg} + L_{gf} \quad (6)$$

Also c_g^{mipro} is the same with c_g^{hmip} because in addition to the routing overhead at the GFA, there is also the tunnelling cost present. Only c_h is avoided. Thus:

$$C_{PD}^{mipro} = k\lambda_a(w_1nk + w_2log(k)) + (l_{gf} + l_{hg})\delta_D \quad (7)$$

In addition, in MIP-RO the packets have to be routed to the mobile host, and so the logarithmic $k\lambda_a w_2 \log(k)$ IP routing overhead does exist and that is why $c_g \neq 0$.

Finally, $c_g^{msctp} \neq c_g^{mip}$, since the GFA in our case does not perform packet decapsulation leading the tunnelling cost to zero. But $k\lambda_a w_1 nk$ is zero since there is no decapsulation. So the total packet delivery cost for our protocol becomes:

$$C_{PD}^{msctp} = k\lambda_a w_2 log(k) + (l_{gf} + l_{ch,g})\delta_D \quad (8)$$

4.3. Binding updates

As mentioned earlier, binding or location updates are a requirement in Mobile IP even when the mobile host does not change its care-of address [19]. However, when there is a movement, then either the HA must be informed, or in the case of HMIP, the GFA. We first provide the cost of a binding update for these two cases. As in [4] and [30] we also define the following variables related to binding costs: C_{hg}, C_{gf}, C_{fm}, are the transmission costs of binding updates between the HA-GFA, GFA-FA, and FA-MH respectively. Finally ρ is the wireless medium cost and δ is a distance cost unit. So the HA registration cost for HMIP is given by:

$$C_{uh} = 2s_f + 2s_g + s_h + 2C_{hg} + 2C_{gf} + 2C_{fm} \quad (9)$$

Under the HMIP case, the regional registration cost is:

$$C_{ur}^{hmip} = 2s_f + s_g + 2(l_{gf} + \rho)\delta \quad (10)$$

and so the total location update cost per unit time is [30]:

$$C_{BU}^{hmip} = \frac{E[M]_{hmip}C_{ur} + C_{uh}}{E[M]_{himp} * T_f} \quad (11)$$

For the MIP-RO case C_{ur} is the same since we assume that MIP-RO operates on top of HMIP. So C_{ur} is again:

$$C_{ur}^{mipro} = 2s_f + s_g + 2(l_{gf} + \rho)\delta \quad (12)$$

$$C_{BU}^{mipro} = \frac{E[M]_{mipro}C_{ur} + C_{uh}}{E[M]_{mipro} * T_f} \quad (13)$$

In Mobile-SCTP, the regional registration cost (C_{ur}) is non-existent since the protocol does not send any kind of updates to any local FA or GFA. But it sends directly updates to the CH which can be quite far away and so it has to suffer the cost $C_{u,ch}$ which is higher than C_{ur}. That is why we introduce the aforementioned U_f variable which captures the update frequency to the HA so that we balance out the increased signaling cost due to the direct notification of the CH. The rationale behind U_f being less than 1 is that since we notify directly the CH, the HA does not have to be notified right away for the current connection as we stated earlier (thus the term "lazy" update). Future connections will need the new location of the MH and that is why the higher lag is acceptable. So, when $U_f = 1$ the usual binding frequency is established. The general equation for binding updates (per unit time) in our protocol is:

$$C_{BU}^{msctp} = \frac{E[M]_{msctp}C_{u,ch} + C_{uh}U_f}{E[M]_{msctp} * T_f} \quad (14)$$

with the transmission cost of binding update between the FA and MH being:

$$C_{u,ch} = s_{ch} + 2C_{ch,g} + 2C_{gf} + 2C_{fm} \quad (15)$$

In this case the packet service time at the correspondent host s_{ch}, is zero since the transport layer endpoint costs are not included, and all the intermediate packet processing costs $2s_h$, $2s_f$, $2s_g$ are equal to zero. The HA registration cost C_{uh}, is the same as in vanilla MIP.

4.4. Packet drop cost

As mentioned in the introduction, the proposed protocol represents a soft-handoff mechanism. This implies that we make the assumption that there are cells with overlapping coverage that remove the possibility of complete detachment from the network. This overlapping means that there are no blackouts due to restricted coverage, these are only because of the nature of the wireless channel. So, in our case, packets may be lost in two cases—due to wireless channel errors, and during handoff when the MH is receiving data to the one address but while starts entering the region of a new AP and it notifies the sender that it must send new data to the new address. However, if connectivity is lost with the previous access point (and FA), data packets that were in flight from the CH for this old destination will be lost. Note that the problem is not as severe as in hierarchical mobile IP where the fast or proactive handoff mechanisms are used in conjunction with bicasting [23]. In our case, the old FA continues to be active and still receives packets while the MH is performing the necessary steps for registering with the new FA. We treat this problem (later) carefully, since our protocol operates at the transport layer and increased losses could lead to unnecessary pipe stalls and large RTO timeouts. However, we also provide some good theoretical estimations concerning the packet loss rate.

We showed that the average binding update cost per unit time is C_{BU}^{msctp}. If we consider the packet arrival rate as a Poisson process with average value λ_a, then the average number of packet drops per unit time due to binding updates (registration delay) is:

$$Drop_{avg}^{msctp} = \lambda_a C_{BU}^{msctp} \quad (16)$$

and so the packet drop probability will be:

$$Drop^{e2e} = C_{BU}^{msctp} \quad (17)$$

This is very important in our case since we are interested in maintaining an average number of packets lost as being less than one half of the current window, so that the SACK algorithm used by SCTP can recover from the losses without RTO expiration [8].

4.5. Analytical results

In this section we present analytical results that compare the hierarchical mobile IP and the proposed protocol. We present results for various packet arrival rates, and user mobility. All the parameter values defined above were set as in [30] for ease of comparison with HMIP and MIP-RO.

All the model parameters were defined in Table 1. An explanation of their exact value follows. The packet processing costs wre set to the values 15, 15, and 10 for the HA, GFA, and FA, so that they represent the fact that HA could be more loaded since it can also act as a local GFA, and of course the processing loads for the GFA, and FA are gradually smaller. The weight factors for the routing overhead, w_1 and w_2, were set to 0.3 and 0.7 to represent the fact that the packet IP-decapsulation is more costly over the parsing of of the routing table with size *k*. Finally, the wireless cost multiplier represents the increased cost that is incurred due to the physics of the wireless medium.

So essentially the signalling cost expresses the latency due to processing and networks delays, that a specific mobility management scheme has to suffer. Essentially this means that when the signalling cost is lower, the handoff procedure is faster, and the load on the mobility infrastructure is lighter. Therefore, we will use this metric in order to present analytical results based on the previously developed analytical formulas.

For small packet arrival rates ($\lambda_a = 0.1$) the proposed protocol obviously exhibits nearly the same total signaling cost with mobile IP regional registration (Fig. 2(a)). This is of course something to be expected, since IP tunneling is not utilized at all. However, with increasing packet arrival rate ($\lambda_a = 10$) the benefits of our protocol accrue as Fig. 2(b) indicates, and the signaling overhead presents a significant decrease in the order of 300%. This behavior is exactly what we would expect from our protocol, since it gets rid of tunneling (which dominates the overall signaling cost). Practically, this means that the protocol is highly scalable with respect to the packet arrival rate λ_a—No matter how high is the data rate, it would not degrade overall performance. One other important element of these figures is the effect of the user cell presence time, which as it is increased results in high degradation of the signalling cost.

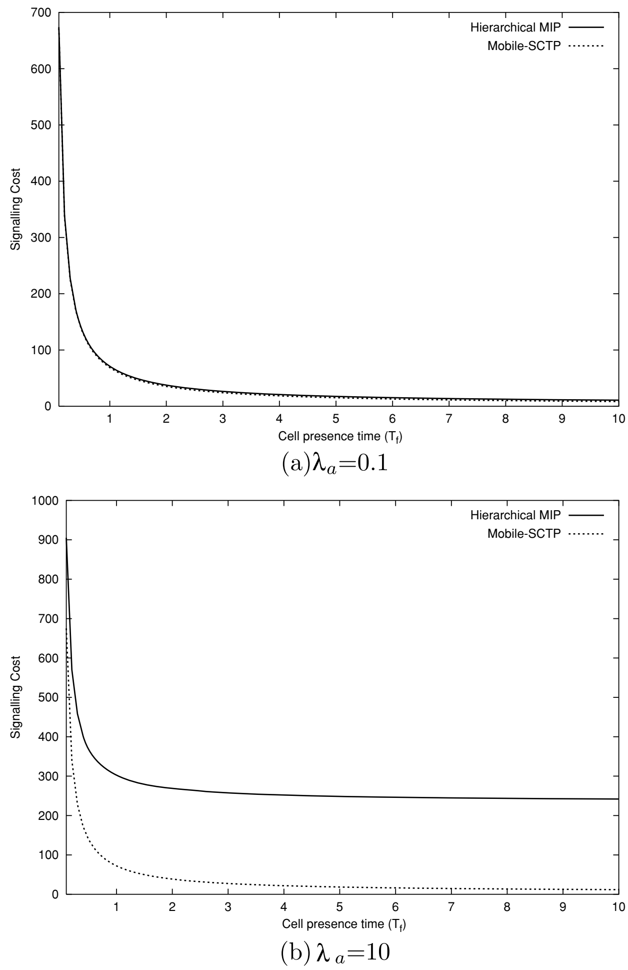
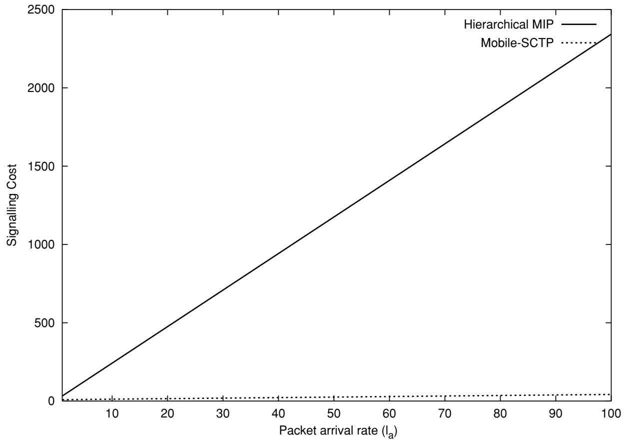


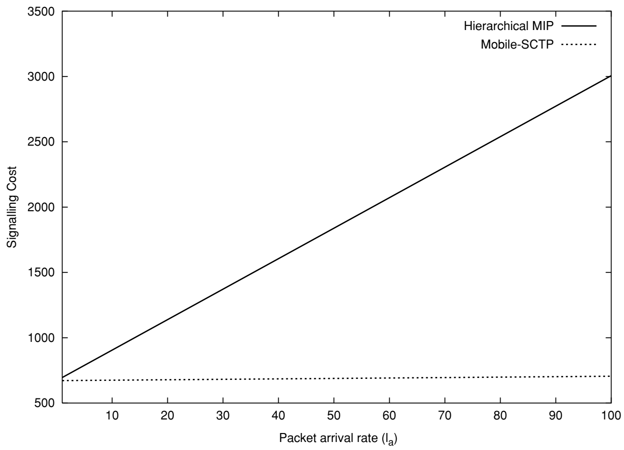
Fig. 2 Signalling cost for fixed packet arrival λ_a

Figure 3 compares the total signaling cost for cell presence time (T_f) of 10 and 0.1 units respectively. We can clearly see the other important advantage of our protocol—As the packet arrival rate λ_a is increased, we observe a linear increase in total signaling cost in the standard HMIP and MIP-RO. This is to be expected since an increase in the number of arriving packets results in higher encapsulation/decapsulation costs. However, our protocol only incurs a minimal overhead that is related to HA binding updates, routing overhead at the GFA, and visitor list maintenance. These results underline the benefits offered by our proposed protocol.

Results concerning packet drop probability are depicted in Fig. 4. Packet drop probability for our protocol is much

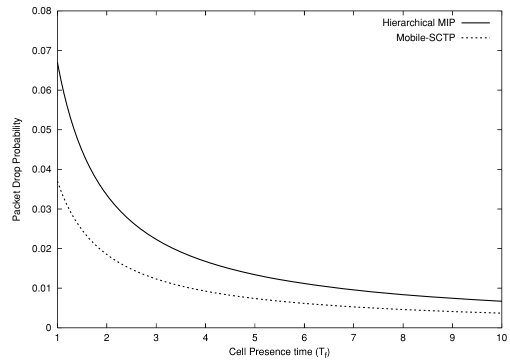


(a) T_f=10

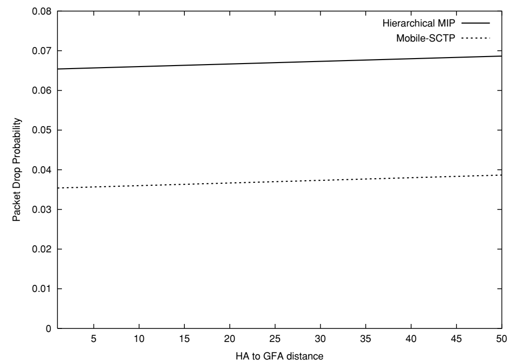


(b) T_f=0.1

Fig. 3 Signaling cost for fixed two values of MH residence time T_f



(a) Packet drop probability with respect to host cell presence time



(b) Packet drop probability with respect to HA-GFA distance (Ch-GFA=25, T_f = 10)

Fig. 4 Packet drop probability for HMIP and Mobile-SCTP

lower than HMIP and MIP-RO which both have the same probability (Fig. 4(a)). For a very high packet arrival rate of 100, the average number of packets being lost would be 3.7 for Mobile-SCTP and 6.7 for hierarchical mobile IP. This means that if our protocol (during a handoff) loses 4 packets, it must have a current window of at least 8 so that it will be able to recover without experiencing RTO expiration [8]. This requirement is not severe, as this event would occur only for a very high packet arrival rate ($\lambda_a = 100$) and if the value of the current congestion window is very low.

5. Gateway/Proxy assisted handoff

Our claims have emphasized our protocol's end-to-end optimization, justified because FA/GFAs act as regular routers. However, in order to provide an additional degree of scalability and create a really localized architecture like HMIP, we added one more component to our system. It can be thought as an transport layer gateway (TLG) and proxy, that implements basically protocol translation between the Mobile-SCTP capable MH and the CH. The purpose of this TLG/Proxy is twofold: (1) Act as local anchor point and localize the MH binding update signalling. (2) In a later stage, translate between Mobile-SCTP and TCP/UDP for achieving wider inter-operability in the Internet. This enhancement, is completely different than MSOCKS [15] since end-to-end semantics are preserved and we do not split the connection in a pure proxy fashion.

The TLG/Proxy is co-located with the GFA, and plays the exact same role as the GFA in HMIP but with functionality tailored to our end-to-end protocol. It receives packets from the CH and sends local Heartbeat messages to the MH so that these messages do not flow through the whole MH/CH route. In addition it maintains a binding between the possibly multiple addresses (during handoff) of the MH, and the connection with the CH. In this way the CH sees one address for the MH during the handoff. The gateway does that by intercepting ASCONF packets locally without informing the CH. Similar to Hierarchical MIP, a binding update is sent to the HA when the MH moves out of the administrative domain of the TLG.

The detailed handoff procedure is depicted in Fig. 5. The MH sends as usual an ASCONF packet to the CH when it is about to handoff. This packet is intercepted by the TLG/Proxy, which instantly replies with a *Local_ASCONF_ACK*. In the meantime until a "real"

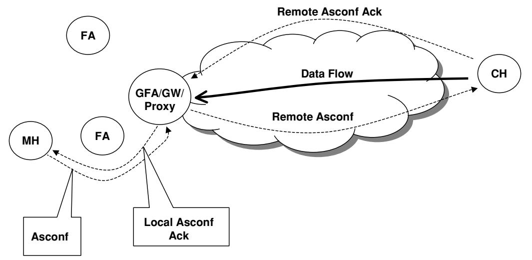


Fig. 5 Message exchange for Gateway/Proxy assisted handoff

Remote_ASCONF arrives from the CH the TLG/Proxy buffers the packets, waiting to forward them to the MH. Usually this buffering time is very short since the time it lasts is the time required for handoff $t_{handoff}$. However, the time of handoff is already minimized for our system, and depends of the overlapping degree of the neighbor cells. So if the time that the TLG/Proxy does not have connection with the MH is $t_{handoff}$, and the data rate is *cwnd/RTT*, then the amount in bytes that have to be stored will be:

$$data\_size = t_{handoff} \times \frac{cwnd}{RTT\_mh\_ch} \tag{18}$$

We use *RTT_mh_ch* since the amount of data that the CH is committing to the network depend on the delay of the end to end connection. If we want to elaborate more the calculation of $t_{handoff}$, we will also must add the $\frac{RTT\_gw\_mh}{2}$ time delay component because the MH will enter the handoff procedure after it sends a *Local_ASCONF*. In addition the *data_size* that we earlier calculated represents the worst case since when the MH starts the handoff stage, it does not mean that it has completely lost contact with the old AP. So finally the buffer size that would be needed is:

$$data\_size = \frac{t_{handoff} - t_{OV}}{RTT\_mh\_ch} \times cwnd \tag{19}$$

Note that we could not use this buffering and let some packets get lost as in the fully end-to-end version of the protocol. However, since we already maintain a TLG/Proxy it is very efficient to improve throughput by using buffering for very short time.

5.1. Analytical results

In the same spirit as before, we also present here some analytical results. The packet delivery cost when the gateway is added is the same. However binding updates are only sent to the HA as in the case of Hierarchical MIP. So $C_{u,ch}^{e2e\_gw} = 0$. But $C_{ur}^{e2e\_gw} \neq 0$ since there are additional components in the total cost:

$$C_{ur}^{msctp\_gw} = 2s_f + s_g + 2(l_{gf} + \rho)\delta \tag{20}$$

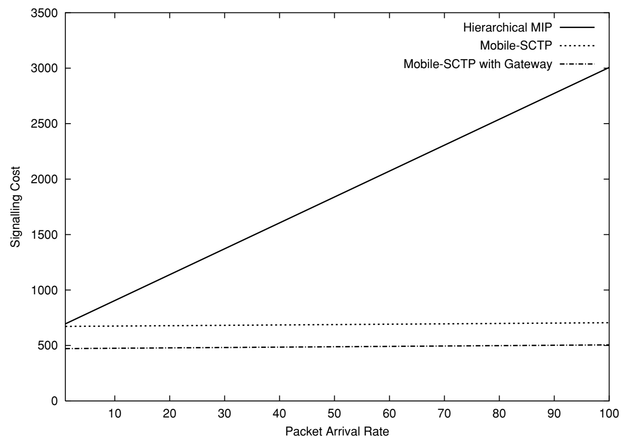
In this case packets are being processed only at the GFA/TLG and so $s_f = 0$. Moreover, C_{uh} is the same as with Hierarchical MIP (Eq. (9)) and so the total binding update cost becomes:

$$C_{BU}^{msctp\_gw} = \frac{E[M]_{msctp} C_{ur}^{msctp\_gw} + C_{uh}}{E[M]_{msctp} * T_f} \tag{21}$$

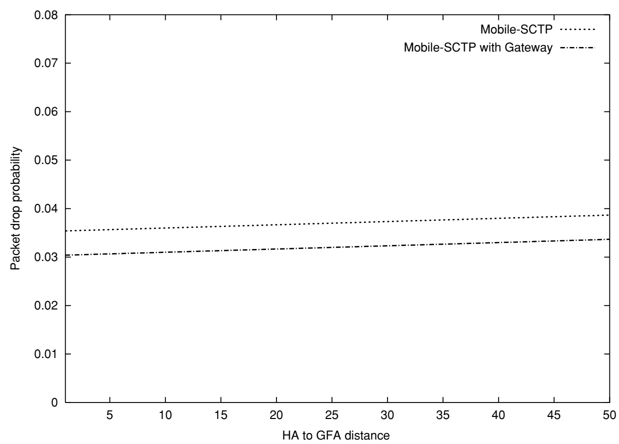
And as in Section 4.4 the packet drop probability will be:

$$Drop^{msctp\_gw} = C_{BU}^{msctp\_gw} \tag{22}$$

Figure 6 presents results when the TLG/Proxy was used. We can see in Fig. 6(a), that we have further reduction in the signalling cost when the TLG/Proxy is used. It is evident that this addition, provides as a local anchor point that helped quite a lot (signalling reduction is nearly 33%). Moreover, the packet drop probability is also reduced as Fig. 6(b) indicates. Nevertheless the reduction does not have the magnitude expected. However, this happens, because Mobile-SCTP has



(a) Signalling cost with respect to packet arrival rate ($T_f = 0.1$)



(b) Packet drop probability with respect to HA-GFA distance (CH-GFA=25,$T_f = 10$)

Fig. 6 Mobile-SCTP with Gateway/Proxy already substantially lowered handoff delay due to its proactive operation.

6. Enhancing gateway performance with intelligent address distribution (IAD)

In this section we propose another, final enhancement to the protocol that is based in the idea of exploiting the mobile host's neighborhood of Access Points (AP). With the term neighborhood, we mean all the Access Points that could possibly be the next point of attachment of the MH. This of course means that we assume the existence of overlapping cells, which is not a limitation since the same assumption is used by "hard" handoff protocols [14,28].

The following simple example qualitatively explains the main operation features of the IAD algorithm. A specialized server (e.g. DHCP), colocated with the TLG, in this neighborhood should be able to provide the MH with a *list_of_addresses* that correspond to different AP/FA in the wired network. For example, when a MH is attached to a AP/FA, it obtains addresses from a specific DHCP server, and when it moves to a new administrative domain requests a new address from a new DHCP server. Even in the case where the host is moving inside the domain it should be able to change its address so that it corresponds to a new AP/FA. This neighborhood of IP addresses corresponds to different physical paths that reach the AP in each case. However, these paths are not all active at the same time, but they can be activated in the near future if the mobile host performs handoff. So if the distribution of addresses from the server is intelligent in the way we explained before, the mobile host can have available a number of addresses already registered before it handoffs to a new AP/FA. The rationale behind this, is that the mobile host can save registration time if it already has the address pre-registered. The only thing that the mobile host has to do, is to identify which one of the addresses to use after it suffers handoff. Our algorithm requires that all contiguous cells to the currently active cell, will be automatically included in the *neighborhood_list*, of addresses. This kind of lookahead additions to the *neighborhood_list* proved to be important for registration delay improvement, at the cost of slightly increased network signaling traffic. A summary of the above procedure can be seen in Algorithm 2.

Now, when the MH moves to a new point of attachment (experiences handoff), it will use one of the pre-allocated addresses. It is fairly obvious now that this form of allocation saves the MH valuable time of requesting address every time it has to suffer handoff. Additionally, the MH is able to pre-initiate data transfer from the new AP. After the MH has decided to handoff and to use a *new_primary_address*, it must notify the CH about the new addresses from which the other endpoint (MH) is reachable now. Our protocol requires from the MH to inform the CH about the *new_primary_address* and the *list_of_addresses* (though ASCONF packet) that it has as soon as the MH receives them. This means that the CH will be able to use any of the remote addresses, that correspond to different paths, in order to reach the MH. In this way direct communication takes place directly between the CH and the MH at all times. The detailed algorithm is shown in Fig. 2.

Algorithm 2 The IAD-based handoff algorithm.

1: MH enters a subnet and sends a DHCPDISCOVER message.
2: The DHCP server provides a *list_of_addresses* based on the *neighborhood_list*.
3: The MH uses as its *primary* address the one of the *list_of_addresses* that corresponds to its current AP.
4: Upon MH movement to another AP two cases arise:

   1. When the MH movement is inside the subnet, it can switch to another IP address from the ones in its *list_of_addresses*.
   2. When MH is moving to a new network, it sends a new DHCPDISCOVER message.

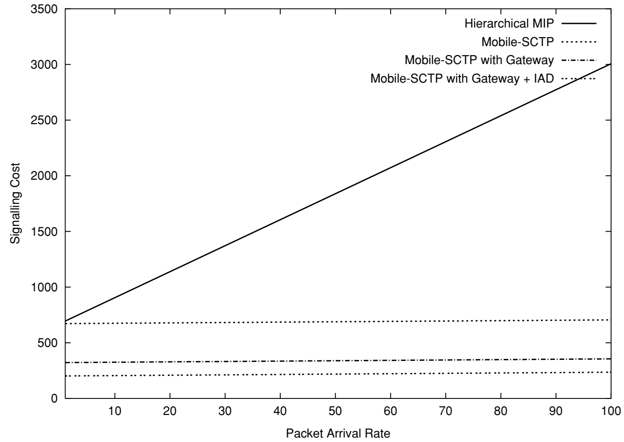
5: The MH decides when it has switched AP by using movement detection info from either L2 or L3 information.
6: Upon its last decision, the MH discards the old *list_of_addresses*, and starts using a new *primary_address* from the newly allocated *list_of_addresses*.
7: Finally the MH notifies the HA about its current location periodically.

6.1. Analytical results

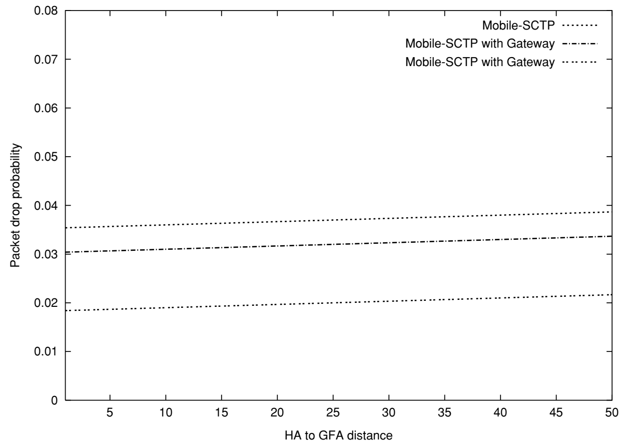
Assuming a hexagon cell topology [28], every host must use one current, plus six addresses for its immediate neighborhood. While the packet delivery cost is the same, the binding updates will change. More specifically a_f will still be zero, but the value of a_g will have to be different (it was 15). This is due to the fact the the further optimized system will not suffer a full registration from the MH which also includes the agent advertisements. Because the number of provided IP addresses will be six, the small overhead of 0.5 that we set [30] will be multiplied by six. So the total binding update cost will become:

$$C_{ur}^{msctp..gw..iad} = 2s_f + s_g + 2(l_{gf} + \rho)\delta \quad (23)$$

It is obvious from Fig. 7(a), that we are achieving even lower signalling cost when compared with all the previous approaches. As we explained, the reason for that is that we completely avoid registration delay since the mobile host has already pre-registered the address that it is using. Furthermore, the reduction in packet drop probability drops even lower when compared with the reduction obtained when we added the TLG to the initial version of Mobile-SCTP. The reason for this significant drop is that the handoff delay is reduced to the minimum possible since a single notification is only sent by the MH to the GFA. Clearly this scheme is tradeoff of reduced registration delay with the number of



(a) Signalling cost with respect to packet arrival rate ($T_f = 0.1$)



(b) Packet drop probability with respect to HA-GFA distance (CH-GFA=25,$T_f = 10$)

Fig. 7 Results for Mobile-SCTP with Gateway and IAD algorithm

used IP addresses. Nevertheless, it could be used when the WLAN is not loaded with many users allowing thus better service to the ones that currently use it.

7. Performance evaluation and comparison through simulations

Figure 8, depicts the topology used throughout our experiments in this section. The values for delay and bandwidth of the links shown in this figure, were set to various values according to the specific experiment. The simulation scenario that we used is the following: Initially at time X the MH initiates an FTP data flow from the correspondent host (CH). According to the scenario, at time X the MH starts moving away from the first AP, at a speed of 10 m/sec, and is heading towards the other AP (see Fig. 8). MH follows the procedure mentioned earlier in order to notify the CH. At time Y, contact with the old AP is lost and the MH can only the new IP. Note that there is a degree of overlapping of the two neighbor

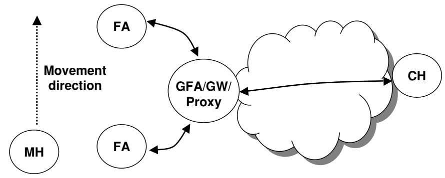


Fig. 8 Topology used for simulations

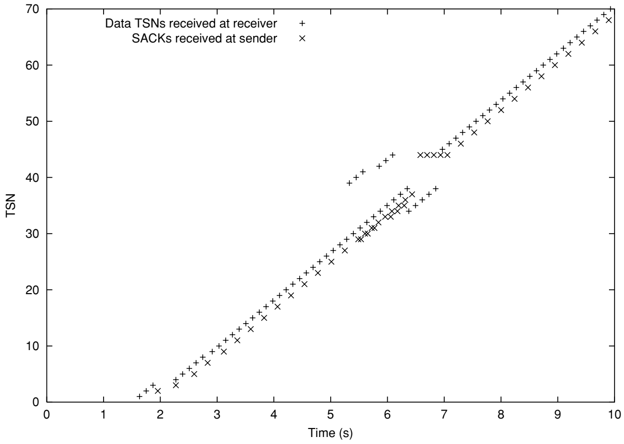
cells which corresponds to a few seconds when a host moves with speed of 10 m/sec.

We modified the ns-2 SCTP implementation [16], by adding our own algorithms and modifications to the protocol. We used 802.11b as the MAC protocol, while we disabled Mobile IP support when we tested Mobile-SCTP. In addition, we made another modification to the Mobile IP implementation in ns-2 so that it supports smooth handoffs. Packets are buffered in the old AP during handoff, and they are forwarded to the new AP upon completion of the MH registration. The smooth handoff principle was proposed in Mobile IP route optimization draft (MIP-RO) [20]. In this way our comparisons are against current state of the art MIP approaches. Generally, our simulations were favoring HMIP. This is because this process has to be done securely [11]: a registration key is established between a foreign agent and a mobile host during the registration process. As a result the whole handoff process is slowing down because this additional delay. However, we did not included this delay in the simulations because we did not have any real life measurement for setting it. On the other hand SCTP, does not need this kind of mechanism since it inherently supports a cookie-based mechanism for securing association between two parties [27].

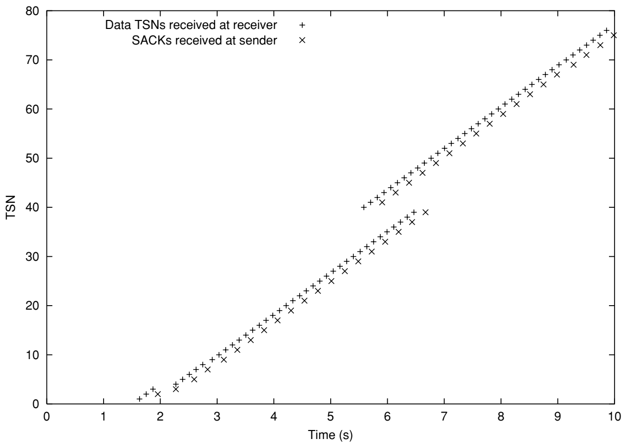
7.1. Results for uplink data flows

Few results are generally presented for data transfer on the uplink direction. Even if the problem is not so severe, as in the case of downlink data flows, we present here results that indicate what happens in this case.

Figure 9 shows results when a data flow was initiated in the uplink direction, and both old and new links from main router to each FA, are symmetric (100 Kbps/200 ms). We can easily see that both the new and old link are symmetric since the data reception at the receiver continues with the same rate throughout the 10 seconds in Fig. 9(a). The MH is initially attached to one FA and starts the data transfer there. At time 5 sec, the MH is associated with the new FA/AP, finishes the registration and starts sending TSNs starting with numbers around 40 through the new AP. But at the receiver, packets being in transit through the old link and packets sent through the new link, are received out-of-order for a little while resulting in SACK generation for every gap observed in TSNs [27]. However, since the links are symmetric packets are inter-changeably received from the old and new link.



(a) vanilla SCTP over MIP-RO



(b) Mobile-SCTP

Fig. 9 Sequence number progression uplink data flows in low bandwidth symmetric links. Data flow is on the uplink direction (MH → CH)

Nevertheless, when the first Data Chunk transmitted from the new link (TSN 40), arrives at the receiver, all the lower TSNs (sent through the old AP) have not yet been received, resulting into generation of SACKs with gap reports. For the TSNs that have number close to 40, and were transmitted through the old link, the number of gap reports at the sender soon reaches the limit of four. This results into fast retransmissions of five Data Chunks, starting at time 6.4 sec. Even worst, at time 6.6 sec the receiver observes the duplicate reception of Data TSNs and sends successive SACKs to the sender, based on the assumption that the other SACKs have been lost [27]. It is obvious, that even if the links are symmetric, and handoff delay was set to zero, we still had side-effects during the handoff process, resulting into spurious retransmissions and wasted bandwidth.

With Mobile-SCTP (Fig. 9(b)), we see that despite the fact that out of order packets are received, the is neither spurious retransmissions nor reduction of the congestion window. In this case, SACK packets are only acknowledging (for the transitional handoff period) packets sent from a specific source address, and not for the whole association. We described this procedure in details in Section 3.2. This interesting form of Fig. 9 should be expected from our explanation before: Packets sent form the old address are still either in the interface queue or in the network in transit. On the other hand the first packet with TSN 40 is being sent from the new address, and arrives at the receiver sooner than the final packets sent from the old address. Since SACK packets are acknowledging only packets sent from a specific address, no spurious retransmissions, duplicate SACK generation, RTO timeouts occur as before. Moreover, we see that it is possible to deliver a few more chunks, during this time period. If the bandwidth was higher, this mechanism would have resulted into even more bandwidth savings and better throughput.

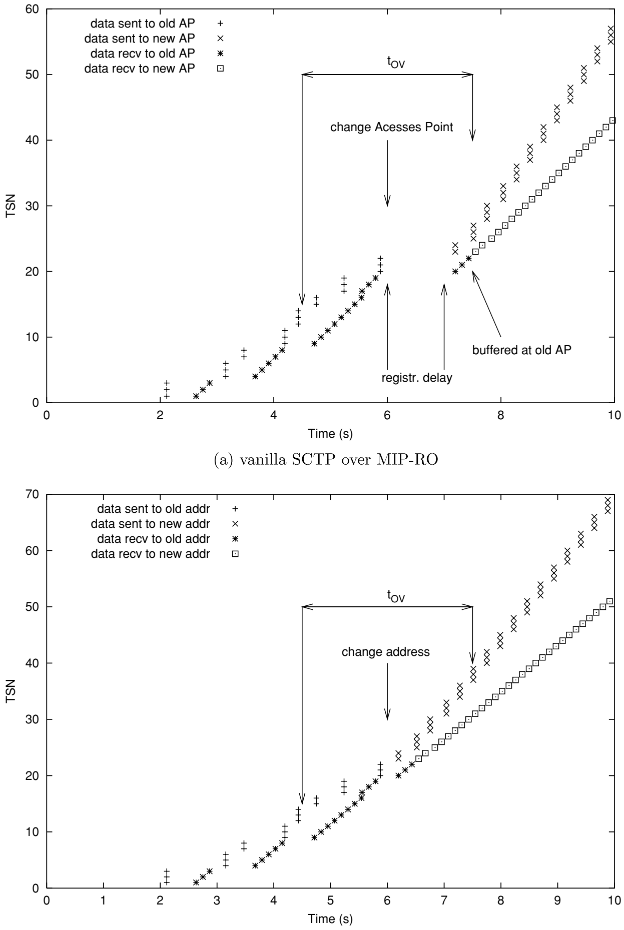
7.2. Results for downlink data flows

Data flows on the downlink, pose the most difficult problems since they require some sort of buffering in case a MH moves out of the range of its old AP. Figure 10 shows how the sequence number progressed when vanilla SCTP was employed under Mobile IP with a downlink flow present. We can see indeed from Fig. 10(b) that our system has no losses, while in Fig. 10(a) even if we used buffering at the old AP and then data were forwarded to the new AP after the registration phase was finished. Even in this case, with buffering, we see that registration delay is a problem. Even a small delay can cause spurious re-transmissions and even lead to *RTO* timeouts. This situation could also happen if the FA is located far from the MH, adding thus more to the handoff delay.

7.3. Results for downlink data flows and variable movement detection delay

We obtained additional results for nearly the same setup, but for higher link bandwidth. In addition we added a number of 100 more hosts trying to register with frequency of around 1 host per second. Moreover, we added this time a delay component modelling the movement detection delay [9]. In [9] they claimed correctly, that in case of HMIP and MIP-RO the handoff delay is dominated by the movement detection delay and not the registration delay. In the previous experiments presented in this section, we did not included the movement detection delay. However, this is not the case, especially when there is no L2 support, for notification of imminent handoffs.

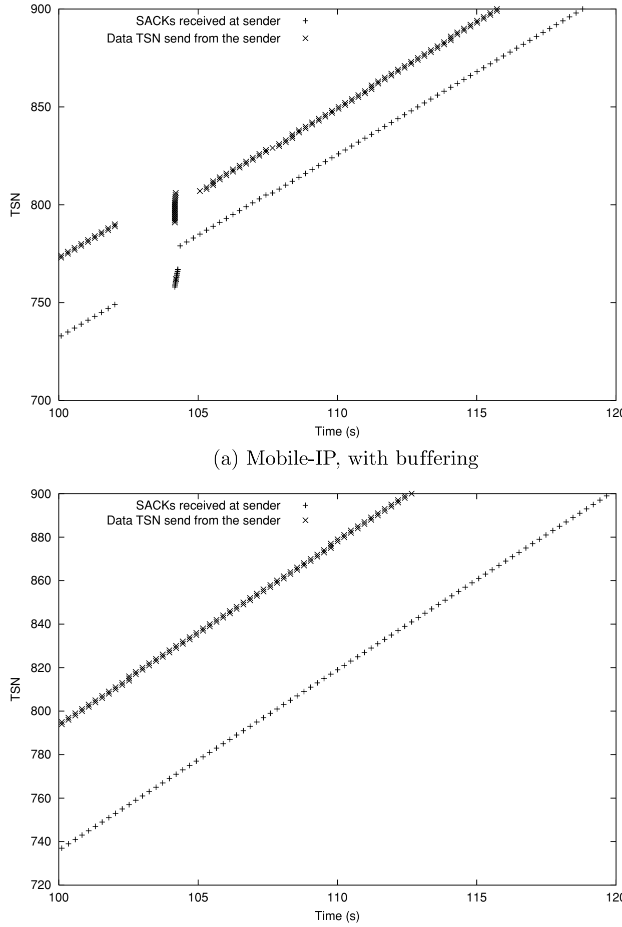
We experimented with the lazy cell switching (LCS) [9] algorithm for movement detection. This algorithm simply dictates that the MH should look for a new FA and register with it after it has missed three advertisements from the old FA. After the MH has obtained a new CoA it should register



(a) vanilla SCTP over MIP-RO

(b) Mobile-SCTP

Fig. 10 Sequence number progression for downlink data flows in low bandwidth symmetric links. Host speed 20 m/sec



(a) Mobile-IP, with buffering

(b) Mobile-SCTP

Fig. 11 Sequence number progression for downlink flows with LCS movement detection algorithm (host speed 10 m/sec)

with its HA. Basically, this algorithm chooses not to handoff even in the presence of other agents (i.e. overlapping cells).

So, we tested this algorithm both with HMIP, MIP-RO and Mobile-SCTP. We set the agent advertisement period to 1 sec. Results are shown in Fig. 11. In the first experiment (Fig. 11(b)), we used buffering at the APs so that we implement the smooth handoff optimization. The effect of the LCS algorithm still results into relatively high delay. Nevertheless, the packet buffering avoids retransmissions and significant reduction of the congestion window. Finally in Fig. 11(c), we show Mobile-SCTP, which despite the use of LCS, maintains a smooth rate for its outgoing traffic at the CH. Until the MH starts accepting data at the new address, provided that LCS will switch to it, it continues receiving data at the old address.

7.4. Evaluating protocol scalability

For evaluating the ability of the protocol to scale as the number of mobile hosts per subnet is increased, we performed a series of simulation experiments. More specifically, the topology in Fig. 8 was used again, only this time we changed the capacity of the wired links to 10 Mbps and the delay to 10 ms. Concerning the wireless 802.1b link we set its effective capacity to 60%. In addition, we use the random waypoint movement model for setting the movement pattern of the MHs, that exists in the ns-2 simulator. The experiments were repeated 100 times for each scenario, that involved the use of specific number of mobile hosts per subnet. Figure 12 presents throughput results for this set of simulations.

As the number of mobile hosts is increased in the subnets, contention for the wireless medium is increased, leading thus to an increase of the packet loss rate. This phenomenon will lead not only to the reduction of the actual goodput, or offered load, for a MH but also to loss of binding updates and even bigger throughput reduction. That is the reason we observe an increased in the number of binding update messages, as the number of mobile hosts is increase

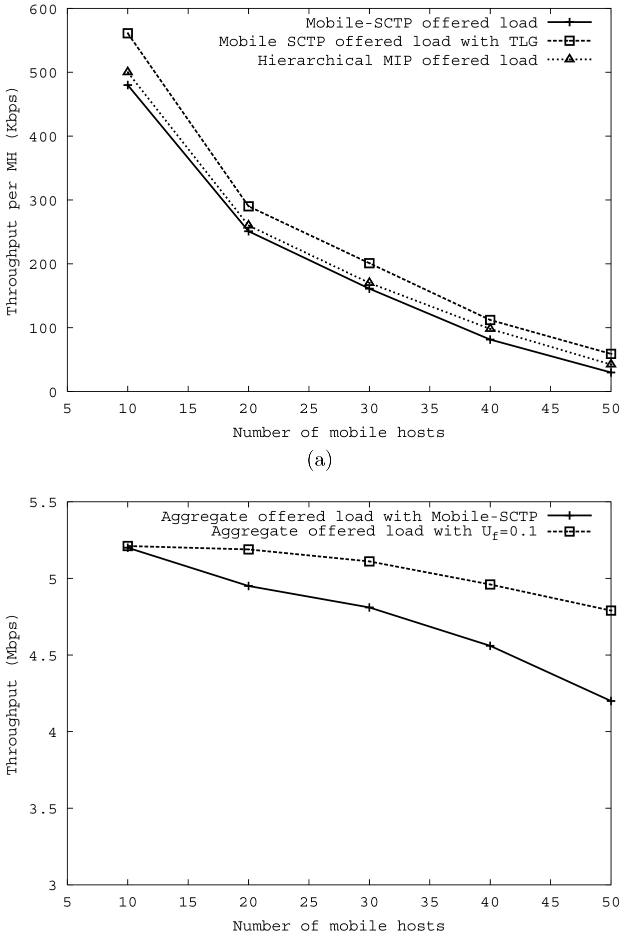


Fig. 12 Throughput for different handoff management protocols

in Fig. 12(a). As a general observation though, we can say that the load due to the binding updates to the correspondent host, is less than 1% of the offered load or actual data rate.

In addition we evaluated how the effect of lazy binding updates can affect scalability, and the overall performance of the system. Figure 12(b) presents results for this part of our simulation. It is interesting to note that while the number of mobile hosts is increased, the baseline Mobile-SCTP protocol suffers from degradation in the overall throughput. Additional packets were lost in this case, since the MH with Mobile-SCTP, was waiting for the binding update to reach the CH. However, when we set $U_f = 0.1$, which means that the MH notifies the HA after 10 new registrations, we observe a throughput improvement. The reason is now that the MH basically only communicates with the CH and occasionally notifies the HA. This result stresses the significance of the bindinding updates to the HA, which should be kept to a minimum, unless there is a need to keep the precise location of the MH at all times.

8. Conclusions

In this paper we presented a lightweight end-to-end mobility management protocol. Overall, the four key characteristics of our protocol are—(1) direct pro-active binding updates to the correspondent host, (2) end-to-end signalling, (3) soft-handoff between successive APs, and (4) "lazy" binding update of the HA. We also proposed a lightweight enhancement by introducing the idea of a transport layer gateway/proxy. This feature allows the realization of a localized micro-mobility solution that contributed to lowering even more the overall signaling load, and increasing the offered load to a level similar to hierarchical mobile IP. Therefore, we can conclude, that since the proposed protocol can achieve performance similar to HMIP, it can help for the creation of a more scalable mobility infrastructure, since it can push functionality and intelligence at the endpoints.

References

1. A. Boukerche et al., A two-phase handoff management scheme for synchronizing multimedia units over wireless networks, in: *IEEE International Symposium on Computers and Communication (ISCC)* (2003).
2. A. Boukerche and I. Harold Owens, Media synchronization and QoS packet scheduling algorithms for wireless systems, Mob. Netw. Appl. 10(1–2) (2005) 233–249.
3. A. Campbell, J. Gomez, S. Kim, A. Valko, C. Wan and Z. Turanyi, Design, implementation, and evaluation of Cellular IP (2000).
4. W. Chen and J. Ho, Performance analysis of adaptive location management for mobile IP, Technical Report 97-CSE-13, Southern Methodist University (1997).
5. C. P. D. Jonhson and J. Arkko, Mobility support in IPv6. RFC 1975, (2004).
6. R. Droms, Dynamic host configuration protocol, RFC 2131 (1997).
7. K. Egevang and P. Francis, The IP network address translator (NAT), RFC 1631 (1994).
8. K. Fall and S. Floyd, Simulation-based comparisons of tahoe, reno and SACK TCP, ACM Computer Communications Review 26(3) (1996) 5–21.
9. N. Fikouras, K. Malki and S. Cvetkovitc, Performance evaluation of TCP over mobile IP, in: *International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*. Osaka, Japan (1999).
10. E. Gustafsson, A. Jonsson and C. E. Perkins, Mobile IPv4 regional registration, *draft-ietf-mobileip-reg-tunnel-07.txt* (2002).
11. W. Haitao and Z. Shaoren, The security issues and countermeasures in Mobile IP, in: ICII Beijing Vol. 5 (2001) pp. 122–127.
12. International Engineering Consortium, Digital Subscriber Line (DSL) and the Evolution of Asynchronous Transfer Mode (ATM) Networks. (2004) http://www.iec.org/online/ tutorials/dsl_atm/topic01.html
13. S.J. Koh, M.J. Chang and M. Lee, mSCTP for soft handover in transport layer, in IEEE Communication Letters 8(3) (2004).
14. Y.-B. Lin and A.-C. Pang, Comparing soft and hard handoffs, in: IEEE Transactions on Vehicular Technology 49(3) (2000) 792–798.
15. D.A. Maltz and P. Bhagwat, An architecture for transport layer mobility, in: *INFOCOM* (1998) pp. 1037–1045.
16. S. McCanne and S. Floyd, Network simulator, http://www.isi.edu/nsnam/ns/

17. P. Nikander, J. Lundberg, C. Candolin and T. Aura, Homeless mobile IPv6, http://www.tml.hut.fi/ pnr/publications/draft-nikander-mobileip-homelessv6-01.txt (2001)
18. D. Park, W. Yoon and D. Lee, An efficient handoff management for mobility support to H.323, in: *International Workshop on Mobile Multimedia Communications (MoMuC)* (2000).
19. C. Perkins, IP mobility support for IPv4. RFC 3220 (2002).
20. C. Perkins and D.B. Johnson, Route optimization in mobile IP, draftietf- mobileip-optim-11.txt (2001).
21. J. Postel, Domain name system structure and delegation, RFC 1591 (1994).
22. R. Ramjee, T.F.L. Porta, S. Thuel, K. Varadhan and S.Y. Wang, HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks, in: *ICNP* (1999) pp. 283–292.
23. P. Reinbold and O. Bonaventure, A survey of IP micromobility protocols, citeseer.nj.nec.com/reinbold02survey.html (2002).
24. M. Riegel and M. Tuexen, Mobile SCTP, draft-riegel-tuexen-mobile-sctp-02.txt (2002).
25. A. C. Snoeren and H. Balakrishnan, An end-to-end approach to host mobility, in: *6th Annual International Conference on Mobile Computing and Networking (MobiCom)*. Boston, MA (2000) pp. 155–164.
26. R. Stewart, Stream control transmission protocol (SCTP) dynamic address reconfiguration, draft-ietf-tsvwg-addip-sctp-07.txt (2003).
27. R. Stewart, Q. Xie, K. Morneault, C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang and V. Paxson, Stream control transmission protocol, *RFC 2960* (2000).
28. I. Stojmenovic, *Handbook of Wireless Networks and Mobile Computing* (John Wiley and Sons, 2002).
29. E. Wedlund and H. Schulzrinne, Mobility support using SIP, in: *ACM/IEEE Internationl Conference on Wireless and Multimedia (WOW-MOM)* (1999).
30. J. Xie and I. F. Akyildiz, A novel distributed location management scheme for minimizing signaling costs in mobile IP, IEEE Transactions on Mobile Computing 1(3) (2002) 163–175.
31. W. Xing, H. Karl and A. Wolisz, M-SCTP: Design and prototypical implementation of an end-to-end mobility concept (2002).

Antonios Argyriou is a Ph.D. candidate in the school of electrical and computer engineering, Georgia Institute of Technology. He received his M.S. degree from the Georgia Institute of Technology in 2003, and the diploma from Democritus University of Thrace in 2001, both in electrical and computer engineering. His research interests spawn in all aspects of computer networking while specific interests include wireless networks and multimedia communications. He is a student member of IEEE and ACM.

Vijay Madisetti is a professor of electrical and computer engineering at the Georgia Institute of Technology. He splits his time among teaching, research and entrepreneurship. His interests are design, prototyping, and packaging of electronic systems, virtual prototyping, embedded software systems, and computer networks. He obtained his Ph.D. in electrical engineering and computer science from the University of California at Berkeley. He is a member of the IEEE and the Computer Society.