The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)

A ROUTE OPTIMIZATION SCHEME FOR THE ETIMIP MICRO-MOBILITY PROTOCOL

Pedro Vale Estrela
IST-TULisbon/INESC-ID
Lisbon, Portugal

Teresa Maria Vazão
IST- TULisbon/INESC-ID
Lisbon, Portugal

Mário Serafim Nunes
IST- TULisbon/INESC-ID
Lisbon, Portugal

ABSTRACT

This paper describes a local Route Optimization (RO) scheme applicable to IP micro-mobility protocols, which increases the protocols' efficiency and scalability while maintaining their fast-handovers capabilities. This is done by reducing and decentralizing the chain of agents that participate in the data forwarding to stationary mobile nodes. This generic RO scheme is exemplified for the eTIMIP protocol, an efficient and transparent micro-mobility solution.

NS2 simulations evaluated the RO extension in comparison to basic eTIMIP and alternative solutions, showing that the extension is able to both lower the packet delay, decreasing the number of hops and improving handover efficiency, and to relieve the gateway and other internal control agents from data forwarding functions.

I. INTRODUCTION

The number of people that use wireless Local Area Networks (wLANs) is increasing at a very fast rate, being foreseen that wLANs will have a major impact on the Internet and the future All-IP 4G data networks [1]. Today, Mobile IP (MIP) [2] is the standard mobility solution for infra-structured heterogeneous networks. Available for both IP versions, MIP is a good approach for the scalable mobility in wide area networks, where handover performance is not a major issue.

Micro-mobility solutions provide local mobility inside administrative domains, shielding the frequent terminal movements to the rest of the internet. Several proposals had appeared namely the Cellular IP (CIP) [3], HAWAII [4] and Hierarchical MIP (HMIP) [5], which are able to provide efficient fast handovers using mobility agents inside the domain.

However, as the end-systems need to be modified to support any of these solutions, then the legacy terminals could be precluded from this new paradigm. Even tough this transparency problem is being addressed in the IETF netLMM group, most efficiency aspects are deferred for future work [6].

Previous authors' work "Surrogate MIP" (sMIP) coupled with "Terminal Independent Mobile IP" (TIMIP) [7] presented a global mobility architecture to solve this transparency issue in an efficient way, by supporting legacy mobile nodes (LMNs). This unique transparency support can be of major importance to future heterogeneous 4G networks, having TIMIP already been identified as part of the 4G mobile architecture [8]. Later work "enhanced TIMIP" [9] extended the protocol with additional transparency features, using an **overlay network**, which supports both LMNs and Legacy Routers (LRs), and maintains TIMIP´s fast-handover capabilities using a mobility agent tree [10].

However, eTIMIP and the other fast-handovers micro-mobility protocols may suffer from efficiency and/or scalability impairments, as their fast-handover schemes tend to force the data packets to pass through a large chain of mobility agents and/or through selected central ones, regardless if the mobile node is currently moving or stationary. Such mobility agents chain impacts efficiency, by precluding a direct, optimal routing inside the domain, which would reduce the end-to-end packet delay and provide a better network resource usage. The use of central agents impacts scalability, as they can concentrate most data traffic forwarding, becoming bottlenecks. Should the number and location of the agents involved in the communication process be reduced and/or decentralized, the above mentioned problems may be avoided.

This paper aims at solving these issues by proposing a route optimization (RO) mechanism based on the use of data path shortcuts. In the proposed scheme, as soon as the Legacy Mobile Nodes are considered stable, the border mobility agents can send the data packets directly to the current LMN location, or, in transitory situations, to its vicinity, by creating shortcuts in the overlay network's agent tree. The scheme, although exemplified for the eTIMIP protocol, extends the regular basic tree-based handover of any mobility protocol.

The remaining part of this paper is organised as follows: section II presents the related work, containing further motivation to our work; an overview of the basic eTIMIP operations is described in section III, being the new routing optimization scheme presented in section IV; the simulation studies are described and analyzed in section V; the paper ends with some conclusions and future work definition in section V.

II. RELATED WORK

Most micro-mobility protocols achieve fast handovers by hiding the frequent MN movements from the rest of the Internet. This is done using internal agents that permanently redirect the data traffic to the MN's most current location, using a full tree hierarchy (eTIMIP, CIP, HAWAII), or using a single hierarchical level (BCMP[11], hMIP, Seamless-MIP [12]).

Fast MIP (fMIP) [13] is an exception to this paradigm, by supporting fast handovers through the temporary redirection of in-flight packets at the time of the handovers only. However, as fMIP propagates all handovers to outside the domain, in practice it must be complemented with one of the mentioned micro-mobility protocols, for scalability reasons.

Previous studies [10] [14] showed that the solutions that use a full agent tree tend to provide the fastest handovers, albeit with variable performance regarding the data packets routing that is due to excessive local triangulations inside the domain. Of these, eTIMIP always forwards data through the shortest paths inside the agent tree, being particularly efficient for intra-domain traffic.

However, even with eTIMIP's tree-optimal routing capability, further efficiency and scalability gains may be achieved by supporting stationary LMNs in the optimized way illustrated in Figure 1. In the left side, the tree is always used to route data to stationary MNs, being the agents involved in the forwarding process marked as black. On the right side, the existing agent tree is extended to cover all domain borders and the new agent can send directly the data packet to the current LMN location, bypassing the agent tree.

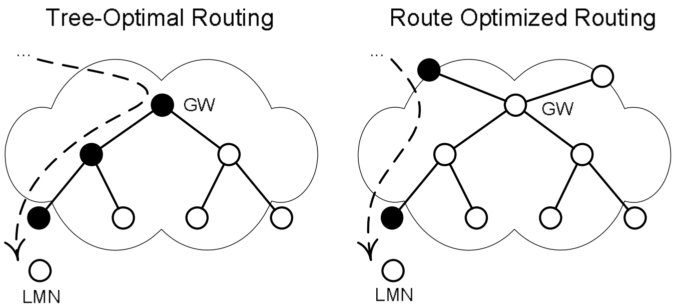

Figure 1: A) Tree based routing; B) RO routing

III. TIMIP/SMIP MOBILITY SOLUTION OVERVIEW

The eTIMIP architecture, depicted in Figure 2, is organized in two complementary networks that separate the mobile routing from the traditional intra-domain routing, named as fixed routing. The **physical network** can have any possible topology, managed by any specialised fixed routing protocol. Complementarily, the **overlay network** is solely used to perform the mobile routing, where selected routers supporting eTIMIP agents are organized in a logical tree, supporting multiple points of attachment to the outside of the domain.

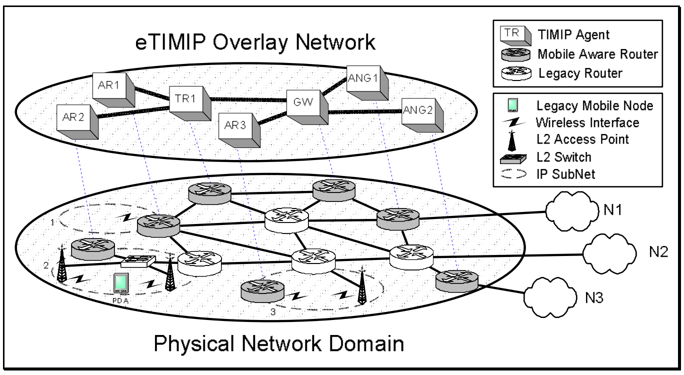

Figure 2: eTIMIP micro mobility architecture

The generic building block of the overlay network is a generic agent designated as eTIMIP Router (**TR**). Inside the overlay network, control and data packets are swapped between the TRs to support the LMN's mobility. Depending on the location on tree, their adjacency to the LMNs and the type of connections to the outside of the domain, the TRs can also play additional, possibly multiple, roles. At the top of the tree, a special TR called eTIMIP Gateway (**GW**) is used to centralize the management functions, being unique to any given eTIMIP domain. To connect to outside the domain, enhanced TRs can be used, named Access Network Gateways (**ANG)**. The ANG's dual is the eTIMIP Access Router (**AR**), which is a TR that offers IP connectivity to LMNs, being physically located in the same subnets as the LMNs. The ARs are able to detect the LMN's movements and to generate and maintain signalling on behalf of them.

The final components of this overlay network are the LMNs, which are only required to connect to the network via a suitable wireless interface. Having terminal independence, it is the network that automatically detects this event and reconfigures itself to provide the necessary IP connectivity to the terminals. On the other hand, the existing Legacy Routers are supported by establishing temporary tunnels between the agents when necessary, and by the use of a **mobile subnet**.

Using this architecture, a basic mobility scheme was defined that supports a full transparent service with good global efficiency. It is characterized by the distribution and use of next-agent routing entries to selected agents of the tree, from the AR up to the GW only. Using these, both control and data packets are transmitted using the shortest paths inside the overlay agent tree.

The first step of the mobility service is executed when a LMN arrives at an eTIMIP domain (step A, Figure 3). eTIMIP detects this initial arrival, authenticates the LMN's access to the network, and performs a ***Power-Up*** operation, that adds a new routing entry in all the agents that lie between the AR and the GW. The second step of the mobility service is executed when the LMN roams between ARs of the same domain (step B). eTIMIP detects this ***Handover*** and updates only the routing entries which become outdated, with respect to the new LMN location, in the local sub-tree that connects the involved ARs. As no information of the old AR is directly available, the network infers its location using the previous basic routing entries from this LMN.

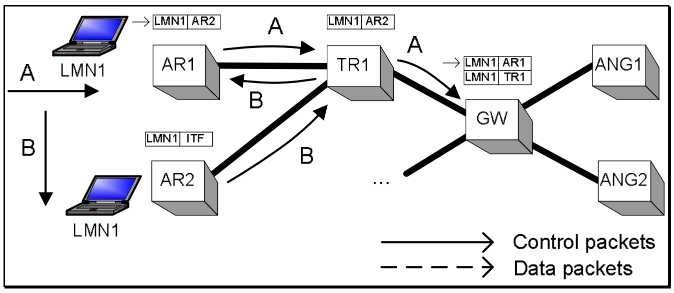

Figure 3: eTIMIP: A) Power-up; B) Handover

IV. RO EXTENSION DESCRIPTION

A. RO-extended fast handover

In order to maintain the fast handovers capabilities of the base protocol, the basic handover signalling is only extended with an extra field, containing the current AR information of the LMN, and an associated RO flag, which determines the presence of the RO extension. Using these new fields, the old AR will be able to establish a temporary local tunnel to the new AR, enabling the fast reestablishment of LMN reception of data. This is accomplished by the creation of a RO routing entry to the new AR (illustrated in Figure 3 by a small arrow in the affected routing tables).

As this fast-handover signalling is limited to the local subtree, there will be agents that are not updated with the most current LMN location (e.g., the one installed in the GW in the power up operation, which still considers AR1 as the LMN's location). However, routing consistency is maintained, as the previous locations are always updated with, at least, the most current basic information that cancels previous existing optimized entries. Thus, the combined chain of mobile routing entries always points to the correct AR starting from the GW.

B. RO-extended data forwarding

Regarding data forwarding, the addition of the RO module extends the regular data forwarding with the capability to use the optimized routing entries, which take precedence over the existing basic entries in the same agent. Such entries can then be used to skip agents inside the tree by forwarding the data packets either directly to the final location of the terminal or, in transitory situations, to locations that are close to the real LMN's location, being then re-routed to the correct destination. If no optimized entries are available, the data packet is forward as before, i.e. using the agent tree.

In the example of the Figure 4, the intra-domain traffic arriving at AR1 will be forwarded directly to the LMN's current AR2 location, instead of being transmitted up the tree as in basic routing (flow "A"). Regarding inter-domain traffic (flow "B"), it is first forwarded using basic routing from the ANG to the GW; at this point, the outdated RO entry at the GW is used to route the packet directly to AR1, which in turn uses its own RO entry to forward the packet to the LMN location (AR2). This case illustrates that eTIMIP RO uses a local triangulation scheme to retain fast handovers, which is nevertheless only temporary, using the diffusion mechanisms described in the next section that enable optimal routing.

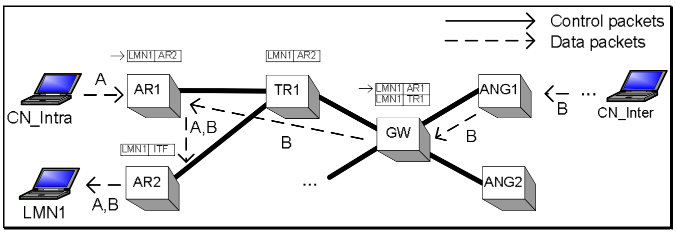

Figure 4: RO forwarding: A) Intra-domain traffic B) Inter-domain traffic

The RO entries are only refreshed if they are regularly used to forward data packets to a certain LMN. Therefore, inactive optimized entries are simply discarded, without any refreshment, reverting incoming packets to basic eTIMIP routing.

C. RO-extended agent tree shortcuts creation

Between handovers, the mobility agents can incrementally disseminate additional RO entries to the borders of the network, removing all temporary triangulations. When an agent forwards a data packet to a certain LMN using an optimized entry, it can deduce that the previous agent did not have the latest optimized routing information – if it did, the previous agent would have sent the data packet to the final TR instead. Using this information, the current agent can start an optimization update operation directed to the previous one, notifying it where it should send the LMN's packets directly to.

This process produces an RO update message, named "RO_inform", which, unlike the previous basic update messages, cannot change the pre-existing basic entries, is not guaranteed, and can be swapped between non-adjacent tree agents. The message only contains the two fields mentioned previously - RO flag and LMN location - and enables the receiver to deliver the subsequent packets to the most current LMN location. In addition, the message has a timestamp field to protect it from possible race conditions [9].

This process is illustrated in Figure 5. The initial state (step "1") of the network is the same as in Figure 4, which suffers from the local triangulation that enabled the fast handovers. When a packet is received by AR1, this agent can start the optimized update packet process towards the source of the data packet (GW). This update RO message will then modify the outdated RO entry at the GW with the latest AR information for this LMN (AR2) (step "2"). This causes future packets to be directly forwarded by the GW to AR2, and the eventual drop of the AR1's RO soft-state entry by lack of usage.

Using this mechanism in succession results in the ANG node learning the current location of the LMN; this enables direct routing inside the domain, and relieves the GW and other internal nodes from of forwarding functions (step "3").

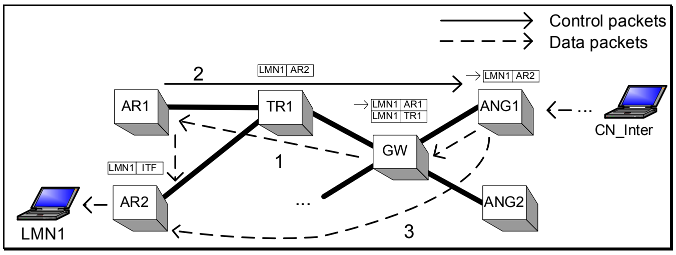

Figure 5: RO state diffusion: 1) Local triangulation 2) RO Inform message 3) Direct forwarding

This process can be subject to operator-defined rules that define the specific conditions that start and end it, in order to achieve an adequate trade-off between data forwarding benefits and extra signalling overhead (which, nevertheless, is always limited to the wired part of the domain and composed of small control packets). In particular, configured minimum and maximum values for rate, number or type of misrouted packets can be used to trigger the dissemination operation.

V. SIMULATION RESULTS

A. Simulation scenario

The simulations were carried out using Network Simulator v2.26, using implementations of HMIP, CIP and HAWAII upgraded from the CIMS v1.0 mobility suite, to which the basic eTIMIP protocol and the RO extension were added. As in our previous simulation work [10], the simulator was modified to emulate IEEE 802.11 infra-structured behaviour with multiple disjoint channels. This modification forces L2 hard handover operations, where stations only receive data packets via one Access Point at a time. For the same reasons, CIP hard-handover and HAWAII multiple stream forwarding options were used from the set of CIMS options. The simulations only focused on micro-mobility scenarios, as all protocols use MIP for macro-mobility.

Multiple sets of NS2 simulations have been carried out on the network described in Figure 6, to evaluate the RO extension in both stationary and roaming conditions, and to compare it with the base eTIMIP, CIP, HAWAII and HMIP protocols. The domain contains 8 ARs, each one managing a separate IEEE 802.11 cell. The ARs are interconnected to the GW by a node tree with additional mesh links. To connect to the outside networks, 2 ANGs are used as border routers. All internal connections are wired links of 10 Mbit/s and 5ms delay, except the outside connections that have a higher latency (indicated in the figure).

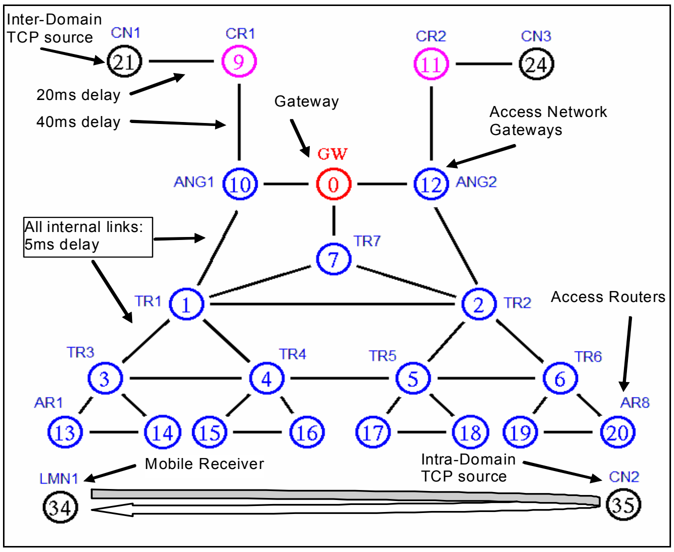

Figure 6: Simulation scenario

The network features 2 MNs connected to it; the first will move sequentially from AR to AR, starting at AR1., performing handovers at a rate of 30 handovers / min. In each test, the LMN1 will be the receiver of a CBR or FTP traffic source, generating either UDP packets of 100 bytes, at a rate of 100 packets/s, or TCP Tahoe packets of 1000 bytes each. This traffic originates from the correspondent node (CN1) outside the network, or inside the domain (CN2). All presented results are taken as the average of multiple independent runs, coupled with its 95% confidence interval.

B. *Simulation results - Stationary Tests*

The first set of simulation studies measure the paths used by the mobility protocols to forward data packets to the LMNs, to evaluate the efficiency and resource usage benefits of the RO extension for stationary MNs.

For this, the one-way downstream delay and number of hops metrics were evaluated at all possible locations in the domain, for both intra/inter-domain sources. The downstream delay is important for real-time services delivery in UDP, and accounts for the major component of TCP's Round-trip-time (RTT); the number of hops inside the domain is related with the required resources to support the mobile users.

The results of UDP traffic are summarized in Figure 7, where the X-axis refers to the MN's location in the domain (AR), and the Y-axis refers to the average one-way delay in ms, and in Table 1, presenting the average number of downstream hops inside the domain.

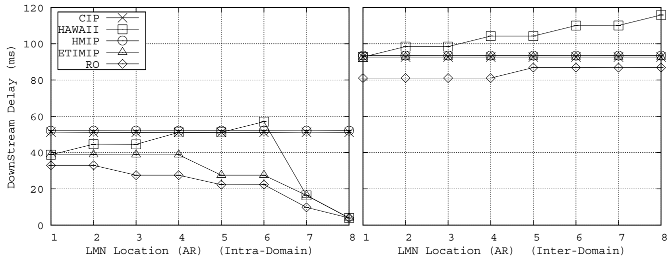

Figure 7: UDP Downstream delay, per LMN location

	eTIMIP	CIP	Hawaii	hMIP	RO
Inter Domain	5.74	5.74	8.06	5.90	4.00
Intra Domain	4.97	9.44	6.90	9.62	3.69

Table 1: Average number of hops inside the domain

Regarding the Intra-domain case, even tough basic eTIMIP already featured the most efficient tree-routing mechanism, with a corresponding lower network resource usage, the use of the RO extension improves that result by bypassing the tree in all domain locations. Thus, the most direct paths between the borders of the domain – the involved ARs - are always used, resulting in the maximum possible routing efficiency – optimal routing. The other protocols have the varying performances, previously analysed in [10]; in particular, CIP and HMIP pass all traffic through the GW, while HAWAII has increasing routing paths inside the domain.

Regarding the Inter-domain traffic, for a similar reason, the eTIMIP one-way delay is also improved using the RO extension. Here, the mobility-aware ANGs will be able to directly route the data packets to the correct AR, which also bypasses the agent's tree, in particular the single GW.

An even greater efficiency and resource usage improvement could be easily achieved by the introduction of additional direct links, which would be used to directly route the data packets to their destinations.

Similar behaviour was achieved by TCP traffic, as the routing paths are the same.

C. *Simulation results - High Speed Roaming Tests*

The second set of simulation characterizes the RO capability in terms of efficiency and scalability gains for high-speed mobile nodes, by measuring the UDP drops and TCP throughput for the former, and the number of data packets forwarded by the GW for all locations of the domain.

Figure 8 shows the first metric, for intra and inter-domain UDP traffic types. The graph shows that the RO extension has a similar amount of dropped packets per handover as the base protocol, meaning that the RO extension maintains the fast handover capabilities of the base protocol unaltered.

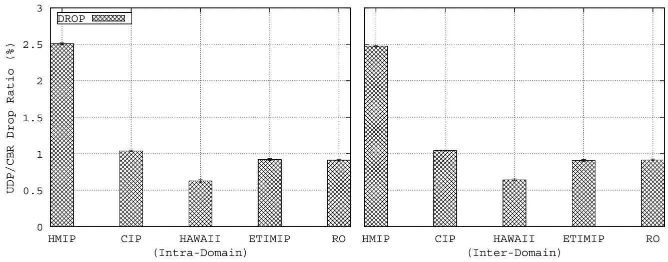

Figure 8: UDP Drop Ratio, 30 hand/min

Figure 9 shows the second metric, for intra and inter-domain TCP traffic types. Even tough eTIMIP and CIP impose similar amounts of dropped packets per handover, these are retransmitted differently, due to the different RTTs. In particular, higher RTTs will condition TCP's ability to quickly retransmit the lost packets during each handoff, being this especially evident in CIP's low result. Thus, while basic eTIMIP already had the best performance of all protocols, the RO extension results in even better performance, as the same amount of lost packets will be retransmitted quicker.

The same conclusion may be observed for inter-domain traffic. While each handover will produce the same amount of drops [10], the much higher RTTs will result in a lower throughput overall, having eTIMIP+RO the best TCP results.

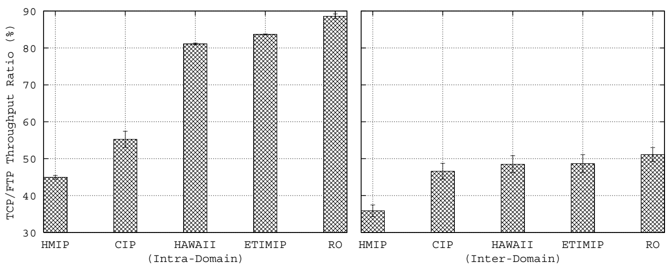

Figure 9: TCP Throughput, 30 hand/min

Figure 10 shows the average data load on the GW, by measuring the accumulated number of TCP data packets forwarded by this central point. While eTIMIP's basic routing and HAWAII do not involve the GW for intra-domain traffic forwarding [10], the inter-domain traffic stills requires the GW involvement.

However, the usage of the RO option, which enables direct optimal routing from the ANG to the LMN's AR, also has the important benefit of not forcing the inter-domain traffic data packets to be forwarded by the single GW, offloading this central node. Thus, the data paths are decoupled from the control paths, as the GW tends to perform control functions only, with important scalability gains.

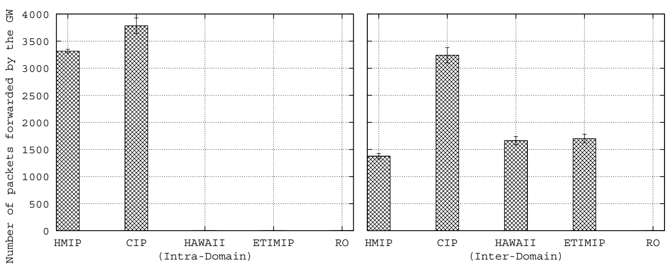

Figure 10: TCP Data load at the GW, 30 hand/min

VI. CONCLUSIONS

This paper presented a generic local Route Optimization scheme applicable to IP micro-mobility protocols that allows the forwarding of data packets directly to the MN location, or, in transitory situations, to its vicinity. As the packets can be routed directly between the domain borders, the RO scheme benefits efficiency, using fewer hops and enabling lower delays; as the packets are no longer required to pass through specific locations of the domain, the RO scheme benefits scalability, by decoupling data from control paths. By not modifying the handover procedure, the fast handover support of the base protocol is maintained, for fast roaming MNs.

NS2 simulations with UDP and TCP traffic evaluated the RO extension in comparison to basic eTIMIP and alternative solutions. The results achieved show that the RO extension provides a lower packet delay and uses fewer number of hops, while the fast handover support is maintained (UDP) or even improved (TCP) due to the lower RTTs. Scalability is also increased by relieving the gateway and other internal control agents from of data forwarding functions.

Future work comprises the application of this scheme to alternative proposals, and the application of a smooth de-triangulation option defined in [15] to this RO scheme.

REFERENCES

[1] G. Carneiro, J. Ruela, M. Ricardo, "Cross-layer design in 4G Wireless Terminals," IEEE Wireless Communications, pp. 7-13, April 2004.

[2] D. Johnson, C. Perkins, "Mobility Support in IPv6", RFC-3775, June 2004.

[3] A. Campbell, et al, "Design, Implementation and Evaluation of Cellular IP", IEEE Personal Communications, Vol. 7 Nº4, August 2000.

[4] R. Ramjee, et al, "HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks", IEEE/ACM Transactions on Networking, Vol. 10, Issue 3, June 2002.

[5] H. Soliman, et al, "Hierarchical Mobile IPv6 mobility management (HMIPv6)", RFC 4140, August 2005

[6] J. Kempf, Ed, "Goals for Network-based Localized Mobility Management (NETLMM)", draft-ietf-netlmm-nohost-req-05.txt, Oct. 2006.

[7] A. Grilo, P. Estrela, M. Nunes, Terminal Independent Mobility for IP (TIMIP)", IEEE Communications, Vol. 39 Nº12, Dec 2001.

[8] D. Saha, et al, "Mobility Support in IP: A Survey of Related Protocols", IEEE Network, vol. 18, no. 6, Nov/Dec 2004.

[9] P. Estrela, T. Vazão, M. Nunes, "Design and Evaluation of eTIMIP - Overlay Micro-Mobility Architecture based on TIMIP", ICWMC 2006, Romania, July 2006.

[10] P. Estrela, T. Vazão, M. Nunes, "Performance Evaluation of Micro-Mobility Protocols in Fail-Tolerant Mesh Networks", PIMRC 2006, Finland, September 2006.

[11] N. Georganopoulos, A. H. Aghvami, "Transport Protocol Performance over BCMP," IEEE Vehicular Technology Conference, Nov. 2003.

[12] R. Hsieh, A. Seneviratne, "A comparison of mechanisms for improving mobile IP handoff latency for end-to-end TCP", MOBICOM 2003, pages 29-41, September 2003

[13] R. Koodli, Ed., "Fast Handovers for Mobile IPv6", RFC 4068, July 2005

[14] L. Peters, et al, "Performance of micro-mobility protocols in an access network with a tree, mesh, random and ring topology", Proc. IST Summit 2003, Portugal, June 2003.

[15] P. Estrela, T. Vazão, M. Nunes, "De-Triangulation Optimal solutions for mobility scenarios with asymmetric links", International Conference on Information Networking (ICOIN 2007), Portugal, January 2007