QoS-aware Fast Handover Optimization Supported by Multicast Networks

Nuno João Sénica, Rui L. Aguiar, Susana Sargento

Abstract— This paper presents three different solutions for handover optimization covering three different scenarios. The first scenario suits the needs of an operator driven network with no degree of liberty on the choice of the new Access Router (AR) by the terminal. An higher degree of liberty is achieved in the second proposed solution, where a supporting multicast network grants the non predictability of the target AR. The multicast network allows the reduction of the bandwidth usage inside the operator network assuring resource optimization, and the delivery of the packets to the surrounding ARs, and thus to the roaming terminal. Nevertheless, these two methods (which are operator driven) depend on an entity in the network for handover permission and control. To avoid this in a high mobility network, we propose a third solution where there is no admission control and always assures available resources in the surrounding *ARs*.

***Index Terms*— Fast mobility, QoS, multicast**

I INTRODUCTION

Nowadays, operators feel the pressure to provide the best service they can to their customers. The deployment of heterogeneous networks is thus a pressing reality to them. Mobility is a 'must' in those heterogeneous networks and as a consequence of its heterogeneity, mobile nodes can potentially roam between different types of access network (e.g. WLAN, 3G).

The Internet Protocol version 6 (IPv6) has a vital role in these heterogeneous environments for data traffic as well as for multimedia applications, providing a convergence layer for seamless mobility, Quality of Service and multicast. IPv6 already includes basic mobility support. However, in order to achieve fast, efficient and seamless mobility it is required that no packet loss is felt, no interruption or degradation should be noticed by the user or its corresponding nodes. With the growing number of wireless users, scalability is also an issue when designing new architectures since a large number of handovers may potentially occur at the same time.

With all these requirements in mind, we present three fast handover architectures. All three architectures aim to provide seamless handovers although they have different applications and characteristics. The first two proposals are adequate for operator-driven networks, combining the integration of mobility with Quality of Service, granting seamless mobility with QoS support, with multicast networks in the case of the second architecture. The third proposed architecture envisions a high mobility network supported by a multicast network, also to achieve seamless mobility.

The paper is organized in the following way. In section II we summarize the mobility, Quality of Service and multicast solutions considered and its integration. Section III details the considered architectures and its qualitative evaluation. Finally in Section IV we present our conclusions.

II BACKGROUND

This work bring together Mobility solutions associated to QoS, using Multicast Technologies as a supporting tool.

Mobility Solutions

The Mobile Internet Protocol version 6 (MIPv6) [1] is the current IETF standard to provide global mobility management and to enable mobile nodes (MN) to roam across different networks, maintaining its reachability to and from other nodes in the Internet. MIPv6 creates a new care-of-address (CoA) that represents the mobile node's new location and advertises this to its correspondent nodes and to a mobility manager (Home Agent, HA) in the home network. To support mobile Internet users, a MN has thus two IP addresses assigned, one fixed (the "identification" home address), and the other changing (the "topologically correct" CoA). Even if MIPv6 potentially enables mobile Internet users to be always reachable regardless of the specific access network technology, increasing multimedia demands from mobile users highlighted MIPv6 shortcomings. Real time audio/video applications underline the need to have in place mechanisms minimizing the large handover latency and service degradation (eg. packet loss) usually associated with MIPv6. In [2] different micro mobility management schemes, such as Cellular IP [3] and Hierarchical Mobile IP [4], offering fast and seamless local mobility are discussed and compared. This comparison is purely based on the evaluation of local mobility management schemes taking into account handover latency, packet loss and scalability issues without integration concerns. Other mobility mechanisms enhancing Mobile IPv6 to account for performance issues are further defined in the IETF, such as Fast Handover [5], recently announced as experimental RFC. The Fast Handover (FHO) proposal, which represents the initial influence for this work, is based on the "make before break" approach, where the terminal signals its handover with the new network using its current connection through the old network. Moreover, during handover, the packets are sent to the mobile node both via the old and the new network to prevent the packet loss during the handover period. One other issues, is that in next generation networks, fast mobility has to be considered along with QoS profiles. In [6], an enhanced fast handover stack was designed and implemented as an extension to MIPv6, also exploiting the FHO basic ideas.

Integrating Fast-Handovers with a QoS Subsystem

The basic fast handover signalling [5] has been extended to support the integration of QoS, with a QoS Broker, a resource manager. The adopted "make before break" philosophy allows to prepare an handover by informing the new point of attachment in the network previous to the handover. In that process, inter access router communication enables the transfer of user related information such as security information and user profile. Achieved results show that handovers do not last for more than 30 msec, assuming idealized QoS components (e.g. only local computation), and that there is no performance degradation during handover in real-time UDP traffic and TCP data transfer [7]. The complete handover process, from the moment the terminal decides to handover until it performs the binding in the new network, including real QoS delays, does not last more than 130 msec [8], a major improvement over standard MIPv6. Other similar proposals have been discussed in the literature. As an example, [9] presents an end-to-end QoS architecture that enables roaming services over heterogeneous wireless access networks. The proposed scheme is also based on a resource manager approach where each autonomous system implements a Domain Resource Manager. The authors present an integrated state model aiming at run time switching between different kinds of handovers in case of failure while preserving reservations. Several types of handovers are here supported: inter- and intra-domain, vertical and horizontal handover, but no mechanism is provided to achieve no packet loss.

Multicast IPv6

IPv6 was developed from the start taking in consideration IP multicast. IPv6 multicast protocols evolved from their IPv4 counterparts, creating a solid base for the tight deployment of IPv6 and IP multicast.

IPv6 multicast is supported by several fields and protocols: a 128 bits group address space, a scope identifier for domain control of the multicast group, a protocol-independent routing protocol, designated by Protocol Independent Multicast (PIM) [10], and group membership mechanisms, designated by Multicast Listener Discovery (MLD) [11]. Group membership in IPv6 multicast is handled by MLD. Its purpose is to enable terminals to communicate the multicast group they which to subscribe to the multicast enabled router. Periodically, the router queries the terminal on the groups it wishes to maintain subscription.

Multicast routing uses PIM and its variants: Sparse Mode (SM) and Source Specific Multicast (SSM). PIM is used to construct the multicast tree used to forward the multicast packets from the source to the terminals. These trees can be based in two different approaches. PIM-SM employs a special configured router, denoted by Rendezvous Point (RP), that serves as a meeting (common) point for multicast senders and listeners. Leaf routers that detect multicast listeners via MLD generated join messages and send them in unicast to the RP's. The PIM-SM also supports the Any-Source Multicast (ASM) model. The ASM model is appropriate for multicast applications such as multiparty videoconferencing, in which multiple sources transmit to the same group

PIM-SSM only supports source-routed deliver trees, and therefore does not use or require an RP. The leaf router learns, via MLDv2 [12], the IPv6 multicast group address and the sender's IPv6 unicast address. This combination of source unicast and group multicast addresses *(S,G)* identifies a channel in the SSM model. Broadcast media applications are therefore natively supported by the SSM model.

III CONSIDERED ARCHITECTURES

In this section we present three proposed architectures, designed to achieve fast mobility and minimization of the real-time session degradation. The first proposal considers a fast mobility scheme enhanced from FMIPv6 extension used in the Daidalos IST Project [13], with mobile terminal and network initiated handover. The second proposal considers the previous enhanced FMIPv6 proposal with multicast supporting networks to enhance the handover efficiency. Finally, a novel mechanism is proposed that uses a supporting multicast network to guarantee an "always on" paradigm on fast moving mobile nodes.

Fast Mobility in the Daidalos Project

Envisioning environments with high level of mobility requires the minimization of the overhead required for signalling and focus on access resource control (typically considering the Diffserv model). Besides being integrated with QoS, the fast handover approach presented in this paper extends the previous ones in its ability: (1) to be independent on the mobility protocol in use, even if it is implemented with basis on the FHO; (2) to address both mobile initiated and network initiated handover; and (3) to potentiate the existence of interface selection entities through the information provided by existing network discovery mechanisms such as [14] and an intelligent decision module in the mobile terminal.

The QoS reference architecture is *DiffServ* based and relies on a central resource management entity, the QoS Broker (QoSB). It performs admission control and manages network resources, controlling the mobile node's services, the network routers and its reservations. The QoSB is also responsible for handover authorization, verifying if the node can use the requested resources on the new link. A Performance Manager module located in the QoSB gathers reports (link availability, signal measurements, etc.) from Performance Attendants located in every access router (AR). By means of this information it computes an algorithm in order to optimize radio link resources, and determines if and which mobile node(s) need(s) to change their current point of attachment. Several mobile nodes then receive a notification from the QoSB to change their point of attachment: this process is denoted as **Network Initiated Handover**. Thus, the network can impose an handover because of network performance and geographical mobility. The former aspect mainly deals with resource optimization and network load balancing. The latter addresses the connectivity problems caused by signal level degradation. Similar mechanisms apply to nowadays 2G/3G networks.

The mechanism so far described mainly considers access network specific operations. However, the proposed architecture takes into account user preferences and requirements, by providing an interface selection scheme able to guarantee communication capabilities according to the "always best connection" paradigm. Thus, the MN can independently decide to request an handover taking into account terminal related conditions (e.g. wireless signal level fast degradation) and user preferences (e.g. a better and cheaper available connection). This operation is identified as **Mobile Terminal Initiated Handover**. Notice however that the network, more specifically, the QoSBs in charge of the old and new access networks, needs to finally authorize this mobile terminal request, and that the handover is only performed after this authorization. In the next sub-sections we detail both the mobile terminal and network initiated handovers.

IV MOBILE TERMINAL INITIATED HANDOVER

In the first phase of the handover process, the MN needs to bootstrap a handover preparation mechanism to discover available candidate access routers. For this purpose, the CARD protocol is used - CARD components are located on the access routers and MNs. The fast handover preparation and execution processes are depicted in Figure 1. Upon receiving information on the available ARs (eventually offering access in different technologies), a MN can decide to roam to another AR, e.g. because of user preferences, by sending a *HandoverRequest* message (message 1) to its current AR containing an ordered list (up to three), of selected candidate ARs. This message has a flag indicating if the Handover (HO) is imminent (e.g. lost of communication is imminent) or not. Upon the reception of this message, the AR sends a request for handover approval to the QoSB (*HandoverRequest* – message 2). Thus, the QoSB verifies whether the resources are available on the indicated ARs and answers (authorizes) with the first occurrence of the list matching the user's current services requirements. For supporting this authorization, the QoSB sends a *HandoverDecision* message (3, 4) to both the old and new ARs (oAR and nAR). The nAR books the reservations and starts to buffer the packets sent to the new nCoA. The oAR processes the *HandoverDecision*, starts the duplication of the streams directed to the old mobile node's location and triggers the Context Transfer. Finally, it informs the MN that it can now move to the selected AR. As soon as the MN receives the *HandoverResponse* (5) which contains the decision, it performs the necessary internal checks and sends a *Fast Binding Update* (FBU) message to the oAR (6), which then is reported to the QoSB (7). Then, the MN configures the layer 2 connection on the new link (e.g. layer 2 handover) and sends a *Fast Neighbor Advertisement* (FNA) (8) message in order to populate nAR neighbor cache where buffered packets may be already waiting to be delivered. Finally, the nAR sends to the QoSB the information about the successful handover (9). This information is then transferred to the oAR (10) to inform that the handover already occurred. At this time, the packet duplication process is stopped, and the oAR reports to the QoSB (11) that every information about this terminal was deleted, handing over the MN control to the nAR.

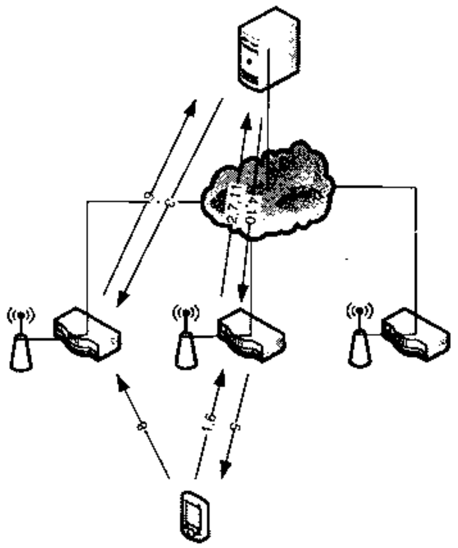

Figure 1 - Fast Mobility Scheme in Daidalos Project

Notice that in the case of the old and new ARs belonging into different access networks, when the QoSB in the old network receives the HandoverRequest message (2), it needs to contact the QoSB in the new network to ask for available resources and to transfer the context related to users, sessions and QoS. The new QoSB answers with the resources availability and the old QoSB can then send the HandoverDecision message to the old AR.

V NETWORK INITIATED HANDOVER

In this case, no preparation phase is required. The network is both the decision and selection point. The QoSB communicates to the AR by means of a *HandoverDecision* message which MN(s) have to roam. Thus, the AR sends a *HandoverResponse* with a special flag set. This flag indicates that the message is unsolicited. Upon the reception of this datagram, the MN must perform the necessary steps to attach to the new target AR, since it knows that its current point of attachment will not be available in the near future. The MN, in case the QoSB indicated a candidate no longer available (i.e. out of the coverage area), can still request an abort to the handover procedure by sending the FBU with a negative acknowledge. The roaming steps are the same mentioned above.

Fast mobility supported by a Multicast Network with QoS integration

Multicast networks are the best choice to transport the same traffic inside a network without using duplication mechanisms. With the assumption of a multicast network and the previous fast mobility process in mind, an integrated architecture was designed. The integration of these two techniques includes an extra step in the target selection. The MN does not need to rely on additional protocols to discover surrounding networks, which is a time consuming operation and may cause an interruption on the current connectivity (since it has to disconnect, survey the wireless channels and connect again). This operation is done by the network: since we are considering handovers inside the same domain, the network administrator has the complete knowledge of the network topology. Using this knowledge, the administrator can configure the QoSB with the network topology. The QoSB can then select the proper surrounding AR when a *HandoverRequest* is made, in mobile node and network initiated handovers.

In order to guarantee that no packet is lost in this process, a *any source multicast* network is established using the known network topology; this *any source multicast* network is formed by each AR (as source) and its surrounding neighbours at the network's boot up.

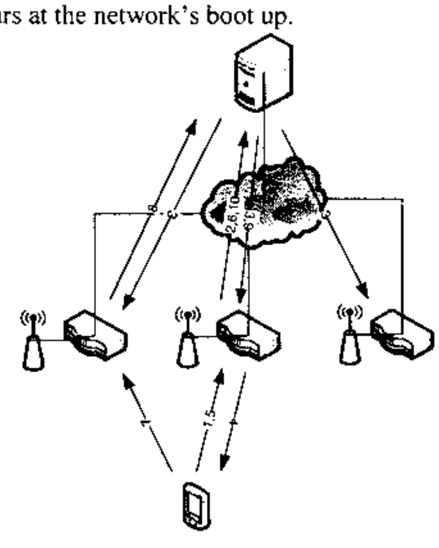

Figure 2 - Fast Mobility Scheme supported by a Multicast network integrated with a QoS System

This fast mobility process is depicted in Figure 2. The MN sensing lower signal in the current AR, performs a *HandoverRequest* (message 1) directed to the AR which is then forwarded to the QoSB (2). The QoSB looks up the MN surrounding networks, checks which networks can handle the current MN's connections and answers back with the possible targets (3). These targets are then informed of the possible handover, and that they are candidate targets and need to be prepared to handle the MN (3).

The current AR, upon the reception of this response from the QoSB, forwards the information to the MN (4). If the handover is allowed, it starts to intercept the traffic directed to the MN inserting it in the previously established *any source multicast* network. At this point, all the surrounding ARs receive the traffic directed to the MN, buffering it for delivery when the MN attaches or until they are instructed by the QoSB that the handover procedure is complete. At this point, the MN can now freely move to any of the candidates listed by the QoSB. Before leaving its current network the MN sends a *FastBindingUpdate* to the AR (5) which is then reported back to the QoSB (6). As soon as it is attached to a network it sends a *FastNeighbourAdvertisement* (7) and the AR starts delivering the packets to the MN which also needs to send a *Binding Update* to its corresponding nodes (CN) (not depicted for readability issues). The CNs then send the *Binding Update Acknowledgment* to the MN. At this moment the AR triggers the information of reception of the MN to the QoSB (8). The QoSB forward this message to the previous AR (9) in order to inform if the handover was successful and, if so, to stop inserting any remaining traffic in the *any source multicast* network. The previous AR reports the successful handover back to the QoSB (10). Each of the ARs informed of the handover have a handover time frame for its success; if the FNA message does not arrive in that time frame, the buffered packets are discarded and they start discarding all the incoming traffic directed to the MN.

After all these steps, the MN is directly communicating with its CNs with no interruption of the current communication.

In a Network Initiated Handover scenario, the MN receives the order to move to one of the candidate targets following the previously described procedure.

Fast mobility supported by a Multicast Network

The fast mobility mechanism presented in this section is addressed to a fast mobility network, where MNs are always moving with a very high probability to be in a low signal coverage or in overlapping areas. With these requirements, a fast mobility scenario without any intervention of bandwidth management mechanisms was designed. This mechanism is presented in Figure 3.

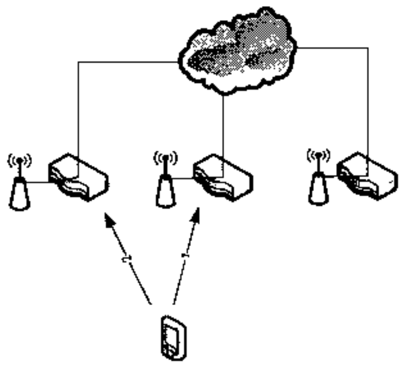

Figure 3 - Fast Mobility Scheme supported by a Multicast network

As in the previous presented solution, there is an *any source multicast* network previously established between each AR and its neighbours. All the traffic directed to a MN is intercepted and inserted in the *multicast* network corresponding to that AR. At this point, all its neighbours receive the traffic, which has a small lifetime in the surrounding ARs buffers. For a small period of time, the ARs keep the traffic in order to guarantee the delivery to the roaming MN as soon as it attaches and signals the attachment. Periodically, the MN sends *Keep-Alive* messages to its current AR (message 1). These messages signal the AR that the MN is still connected to that AR; as long as the AR is receiving this signal, it inserts the traffic into the multicast network. When the MN moves, it senses new ARs. To attach to a new AR, the MN signals the attachment with a *Keep-Alive* message (2). It also sends the *BindingUpdate* message to the CNs. When a AR receives the *Keep-Alive* message, it starts to insert the unicast traffic directed to the MN into its *any source multicast* group, preparing a future handover of the MN. After a time out of 3 *Keep-Alive* messages, the previous AR stops introducing the traffic into the multicast network. At this stage, the handover procedure is concluded.

Architectures Evaluation

These three fast handover mechanisms have advantages and disadvantages, which make them best suited to specific situations and/or scenarios.

The first two methods are very similar. The differences between both rely on the way how the packets get to the new ARs and the selection of the new target AR. Comparing the first two methods, it is possible to find some similarities, as both depend on a central entity to control the QoS, which is also responsible for the admission control and therefore for the handover authorization. One of the large advantage of having an *any source multicast* network to support duplicated packets is the previous knowledge of the MNs' surrounding ARs. Also, the multicast groups already assigned allows the MN to move to a finite set of those ARs in the neighbouring, deciding according both to signal and user preferences. Although this is an advantage for the classic fixed network where the ARs are fixed for a long period of time, it is not a good solution for moving ARs, since it is required to continuously update the topology in the QoSB (create the *any source multicast* networks requires some time to set up and balance). This problem is not present in the first presented solution, since the MN communicates which ARs it can attach to, and the QoSB decides which of the ARs can handle the MN. However, this process limits the MN choice of ARs, and it is subject to problems in highly dynamic networks where the signal level can change very fast.

Due to the existence of a previous multicast group including the neighbouring routers, the second approach has a large advantage in the handover time. These source/group multicast networks may also be controlled by the QoSB, since it can inform each AR of its multicast network and its surrounding ARs. With this information, the AR can start the join process and establish the multicast networks at boot up (or when it is informed by the QoSB).

In terms of overhead, the second approach has a significant signaling overhead in the wired network. This is due to the existence of a control entity. However, the overhead in the wireless link is low, since the traffic is only inserted in the multicast group upon handover request.

The third approach does not contain a control entity, and therefore, there is no access control and no QoS guarantees in the new network, both for the new flow and for the ones already present in the network. Also, it requires the complete knowledge of the network topology in order to establish the multicast networks. However, this procedure requires very small signalling overhead, and provides a handover really fast without packets loss and additional signalling. In terms of data overhead, all the ARs belonged to the multicast group receive the same data stream, which increases the resource usage in the core networks. However, the core network is usually not the bottleneck compared to wireless link. Moreover, this is the only way to grant a continuous stream to wandering MNs. This scenario is the best suited one for very large mobility scenarios.

VI CONCLUSION AND FUTURE WORK

This paper presented three different solutions for handover optimization covering three different scenarios. The first scenario suits the needs of an operator driven network with no degree of liberty on the choice of the new AR by the MN. The MN always depends on the QoSB decision on the next network to move. This degree of liberty is achieved in the second proposed solution, where a supporting multicast network grants the non predictability of the target AR (in this case a set of neighbouring ARs are prepared to receive the MN). The multicast network allows the reduction of the bandwidth usage inside the operator network assuring the resource optimization, and the delivery of the packets to the surrounding ARs, and thus to the roaming MN. Nevertheless, these two methods (which are operator driven) depend on an entity in the network for handover permission and control. To avoid this in a high mobility network, we proposed a third solution where there is no admission control and where there are always available resources in all surrounding ARs.

As future work, it is planned to evaluate in details these three solutions within different scenarios, in order to understand which suits better in a particular scenario. Extensions and updates to the presented architectures will be performed as results of the simulation results.

REFERENCES

[1] D. JOHNSON, C. PERKINS. *MOBILITY SUPPORT IN IPV6*.JUNE 2004, RFC 3775

[2] A. Campbell, J. Gomez, S. Kim, and C. Wan. *Comparison of IP micro-mobility protocols*. In IEEE Wireless Communications, vol 9, pages 72–82, February 2002.

[3] A. Campbell et al., *Cellular IP*. Internet draft, draftietf-mobileip-cellularip-00, work in progress, Dec. 1999.

[4] Hesham Soliman et al. *Hierarchical mobile IPv6 mobility management* (hmipv6). June 2003. http://www.ietf.org/internet-drafts/draft-ietf-mobileiphmipv6-04.txt.

[5] Rajeev Koodli (ed). *Fast handovers for mobile IPv6*, July 2004. http://www.ietf.org/internet-drafts/draft-ietf-mipshop-fast-mipv6-03.txt.

[6] V. Marques, R. L. Aguiar et al. *An IP-based QoS architecture for 4G operator scenarios*. IEEE Wireless Communications, June 2003.

[7] T. Melia, R. Schmitz, and T. Bohert. *TCP and UDP performance measurements in presence of fast handovers in an ipv6-based mobility environment*. In 19th World Telecommunications Congress (12-15 September 2004, Seoul, Korea), September 2004.

[8] V. Marques. *Multimedia services for heterogeneous networks*. In Phd Thesis, October 2004 Universidade de Aveiro.

[8] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, L. Wei, *Protocol Independent Multicast – Sparse Mode (PIM-SM): Protocol Specification*, RFC 2362

[9] R. Bless, J. Hillebrand, C. Prehofer, and M. Zitterbart. *Quality-of-service signaling for next generation ip-based mobile networks*. volume 42 of Communications Magazine, pages 72–79. IEEE, June 2004.

[10] S. Deering, W. Fenner, B. Haberman, Multicast Listener Discovery (MLD) for IPv6, RFC 2710

[11] R. Vida (Ed), L. Costa (Ed.), Multicast Listener Discovery 2 (MLDv2) for IPv6, RFC 3810

[12] The Daidalos Consortium. November 2003. http://www.ist-daidalos.org

[13] Marco Liebsch (ed). *Candidate access router discovery*, Sep 2004. //www.ietf.org/internet-drafts/draft-ietf-seamoby-card-protocol-08.txt.