Quality-of-Service Signaling in Wireless IP-based Mobile Networks

Roland Bless and Martina Zitterbart
Institute of Telematics
University of Karlsruhe, Germany
Email: {bless,zit}@tm.uka.de

Joachim Hillebrand and Christian Prehofer
DoCoMo Communications Laboratories Europe
Munich, Germany
Email: {hillebrand,prehofer}@docomolab-euro.com

Abstract— **Efficient support for multimedia communication is essential for next generation mobile wireless networks. In order to provide the necessary quality of service (QoS) on demand in mobile wireless networks, appropriate signaling has to be defined. We present a mobility-aware QoS signaling architecture that integrates resource management with mobility management. It is based on a domain resource manager concept and nicely supports various different handover types. In particular, our approach supports anticipated handover with pre-reservation of resources before the mobile node is attached to the new access point. In this paper, we discuss the signaling protocol of our architecture with focus on the different handover cases, for which QoS signaling is described in an integrated handover model.**

I. INTRODUCTION

Mobile telecommunication networks are moving towards an 'all-IP' paradigm. Users of future mobile networks expect not only access to the widely deployed and well-known traditional Internet services, but also to enhanced multimedia communications, such as real-time services. Therefore, Quality of Service (QoS) must be provided for data flows and maintained even when a mobile node (MN) changes its current point of attachment to the network. On network level there are several techniques such as Differentiated Services [1] or Integrated Services [2] to enforce QoS during packet forwarding. Additionally, a *resource management* has to take care of admission control, allotment and release of requested resources for QoS provisioning within a network. The mobile node requests resources from the resource management via a signaling protocol.

The main requirement for mobile networks is that an existing resource reservation has to be taken over to a new access point in case of a handover of the mobile node. In this case, the path of the reservation has to be adapted. The handover itself is accomplished by a *mobility management* protocol like Mobile IP [3]. Its main task is to forward packets to their destinations although the IP address of the sending or receiving node changes. This happens for example if a node moves to a domain of a different provider (macromobility) or to a different IP subnet within the same domain (micromobility), i.e., the same provider. Moreover, because Mobile IP was developed for macromobility and relatively slow-moving nodes there are also a number of micromobility enhancements—like Fast Handover [4] or Hierarchical Mobile IP (HMIP) [5]—to achieve a seamless handover. These improve the service disruption by reducing packet loss and delay while a mobile node changes its point of attachment. Our approach establishes a coupling between mobility management and resource management in order to provide the required resources along the current data path. As there are different approaches to (micro-)mobility management, we use a minimal, generic interface to mobility management.

In addition, the QoS signaling protocol must support a variety of different handover types, e.g., *hard* or *soft* handovers as well as *anticipated handovers*, which reserve resources along the new path in advance and impose new requirements on signaling. Combinations of these handover types have to be considered as well, because an MN may be forced to continue an already started anticipated handover process with a hard handover due to a sudden loss of connectivity. An enhanced form of anticipated handover uses *location management* that provides additional information about the geographic topology of the network and positions of mobile nodes. This information is especially useful in cases when mobile nodes are forced to move along certain paths, e.g., along a railroad track or a highway. In this case, movement prediction (see for instance [6]) may be used to reserve resources in advance without requiring explicit signaling by the mobile node.

This paper describes a novel, mobility-aware QoS signaling architecture that integrates resource management with mobility management and location management to provide QoS signaling for a wide variety of handover types. While most other approaches consider the handover cases individually, an integrated solution is presented that uses a single model to describe all possible handover cases. A main advantage is that this model allows one to switch between these handover cases during a handover process.

II. A MOBILITY-AWARE QOS SIGNALING ARCHITECTURE

In the following, we give an overview of *MARSP (Mobility-Aware Reservation Signaling Protocol)*. The corresponding signaling architecture provides QoS in a mobile IP-based network with the following features:

- Independence of a particular QoS technique for provisioning of QoS-based services, at IP layer as well as at link layer. The architecture does not depend on a specific QoS model, but works with various QoS solutions such as Integrated Services [2] and Differentiated Services [1].

- Independence of specific radio access technologies. It can be expected that different *radio access networks* (RANs) are used at different locations. The presented architecture does not depend on a specific radio access technology. However, it supports the utilization of RAN characteristics.
- Interworking with different micromobility concepts. The solution allows one to integrate different micromobility approaches as a base for seamless handovers.
- Support for inter-domain handovers. Even if a MN changes its point of attachment to a network that is administered by another organization, a handover is feasible.

These features provide QoS support for a wide variety of handover cases that differ with respect to different cellular environments, mobile node capabilities and connectivity, network configurations and operator needs.

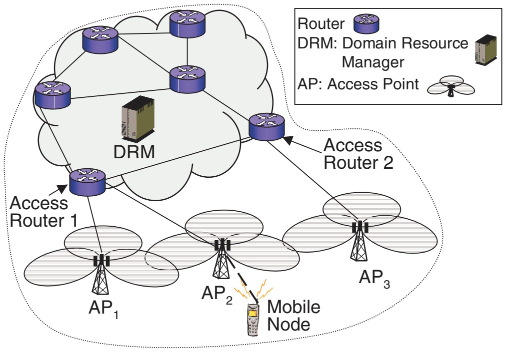

Fig. 1. Components of a domain providing wireless access

In IP-based mobile networks, an *Access Router (AR)* provides IP connectivity to a MN within a domain as illustrated in Fig. 1. An AR can be connected to multiple *Access Points (APs)*. These APs provide link layer connectivity by radio transceivers and are part of the *RAN*. Moreover, each domain has a *Domain Resource Manager (DRM)* that controls all resources at IP level within this domain. It can be viewed as a dedicated logical entity for management purposes. An actual implementation, however, may use several distributed entities.

A dedicated management entity allows one to integrate resource and location management as well as policy and accounting aspects. Moreover, it is also highly flexible, because more complex management functions can evolve independently from router implementations. Since the resource management is not necessarily directly located on the data path some basic functions have to be provided. For instance, the DRM must monitor routing protocols (e.g., OSPF and BGP) in order to perform admission control for the current data paths. Mobile nodes and DRMs must locate the next DRM along the path in order to forward signaling messages. Additionally, a DRM must configure routers by using a management interface, e.g., installation of traffic profiles for Differentiated Services by using SNMP [7], COPS [8] or command line interfaces.

The architecture separates resource management signaling from mobility management signaling, because it aims at independence of the different technologies. Resource management can be seen as an additional function to improve traditional best-effort data transport. Generally, more flexibility is achieved if signaling procedures are decoupled from each other. For instance, if a MN requests resources along the new path before it finally registers at the new AR, data packets traversing the new path will directly receive the corresponding QoS during forwarding.

Within the mobility-aware QoS signaling architecture several logical interfaces can be identified. Basically, it can be distinguished between signaling for resource management, signaling for mobility management and signaling at application level. Direct signaling between the applications may be required to let applications explicity adapt the content of their data flows to the current resource availability. The latter may change due to a handover, especially if a change of wireless access technology is involved (a so-called *vertical handover*). Thus, applications should be notified via an internal interface of QoS changes and they should signal the sender at application level, which can then adapt its sending rate.

The following section focuses on features of the QoS signaling protocol MARSP within this architecture. In addition to explicit signaling for resource reservation, location management or some of the proposed protocols for seamless mobility can be integrated into the architecture as well. Location management information may be available at a DRM because of domain-specific conditions such as a network along a highway or railroad track. In this case, movement prediction can be used by a DRM to reserve resources in advance without the need to let the MN request resources explicitly.

III. QoS SIGNALING FOR HANDOVERS

A flexible QoS signaling protocol has to deal with different handover types, e.g., hard, soft, anticipated handover, and variations of these. The approach presented in this paper is based on an analysis of different handover types in order to find a signaling solution that covers all potential cases. Most current approaches for IP-based QoS signaling protocols like RSVP [9] are not designed to support mobility [10], and other approaches provide support for a single individual handover type, such as soft handover [11] or hard handover [12]. In contrary, our proposal presents a single, integrated handover model describing all cases. A main advantage is that this model allows one to switch between these handover cases dynamically during a handover process.

This section focuses on QoS signaling for different handovers for the previously presented architecture. The particular goal is to analyze different handover types and to find a signaling solution which covers all potential cases. The result is represented by an integrated handover state model (presented in section IV) which is the basis for the specification of QoS signaling protocols. Particularly, this integrated state model allows one to describe transitions between different handover types. In other words, if the operation of one handover type fails, the handover can still be finished by using another handover type. This represents an increasingly important feature, especially when applied in heterogeneous mobile networks including short range wireless access technologies. QoS support is improved if an already initiated handover can be continued without having to start a completely new signaling sequence.

A. Handover distinctions and scenarios

Several types of handovers may be considered for IP-based mobile networks. Each of the handover variants can be applied to a particular configuration of the network. Our goal is to design signaling protocols which support multiple handover types. It is conceivable that a RAN may not support all handover variants (e.g., soft handovers). Furthermore, MNs may also show different capabilities. Some nodes are only able to maintain a single 'connection' at IP layer without being able to listen or scan for other APs simultaneously. Thus, these nodes have to drop their current connection before selecting a new AR. Other nodes are able to scan for new access points while still staying connected to the current AR. Different handover cases can be distinguished that have impact on mobility management and resource reservations. At first, a MN may perform an *intra-domain handover* (current AR and new AR in the same domain) or an *inter-domain handover* (new AR is located in a different domain). Several optimizations for intra-domain handovers are available, that result in faster signaling procedures because the signaling messages may stay local within the domain, and, there is no need to perform a full re-authentication.

A further important distinction can be made between vertical and horizontal handovers. It is possible that the currently allocated resources must be adapted if a vertical handover will be performed, because different link layer technologies offer very different capabilities (e.g., wireless LANs and GSM with respect to bandwidth). Consequently, this adaptation may require signaling at the application level as well as signaling for resource adaptation along the complete path between CN and MN, even for intra-domain handovers.

Another difference can be identified if the different roles of a MN as sender or receiver are considered. If the receiver is going to change to a new AR, it is usually not easily possible to explore the path back from the new AR to the sender in upstream direction. This results from the fact that routing information is only uni-directional, i.e., it only allows one to derive the path downstream towards a destination. Thus, it may also be necessary to signal to the sender in order to establish the correct new path.

When considering QoS, the actual handover decision depends on two main criteria:

- Signal availability—This comprises radio parameters like signal strength, signal-to-noise ratio, etc. The IP layer must be informed by lower layers about conditions and availability of radio connections.
- Resource availability—When carrying out a handover to a new AR, it has to be ensured that the available resources on and to that AR are sufficient to satisfy the QoS requirements on the MN. A handover can also be rejected or directed to another adjacent AR in order to balance the load in a group of radio cells.

Therefore, handover strategies for IP-based mobile networks should be based on both criteria. However, for some cases of anticipated handover or vertical handover no signal measurements may be available. A consideration of these handover types is especially important, as they are expected to be used more often in future, heterogeneous networks.

B. Example for anticipated handover

In this section we show an example of the operation of the protocol to emphasize the requirements and assumptions used for designing them. The example in Figure 2 shows a message sequence diagram for *anticipated handover in an inter-domain case*. The advantage of an anticipated handover is that resource management signaling can be performed before attaching to a new access point. The following example presents the interaction between DRMs of *different domains*.

At first, the MN detects new access points and selects its target AR. Subsequently, a request for changing the MNs point of attachment from AR_1 to AR_2 is issued to the current DRM_1 (RChgReq). This DRM must detect that AR_2 is located in a different domain and has to determine the responsible DRM (e.g., a DNS SRV entry may be used for this purpose).

A handover request (RExtHoReq) is subsequently sent from DRM_1 to DRM_2 in order to request resources from a downstream DRM. Depending on resource availability, DRM_2 sends a corresponding response message (RExtHoRsp) back to DRM_1, which, in turn informs the MN about the result (RChgRsp). If resource allocation has succeeded at DRM_2 it waits for a handover confirmation message of the MN (RHoCompl). If this message is not received within a certain time, the pre-reserved resources are automatically released. Thus, MN will connect to the new AR and is, then, able to confirm the completion of the anticipated handover procedure. DRM_2 informs DRM_1 that the reserved resources in the old domain are not longer used (RExtRelReq).

DRM_1 can then explicitly release unused resources. Otherwise the reservation will time out if no refresh messages are received from the MN within a certain time period.

IV. INTEGRATED HANDOVER STATE MODEL

A handover can be described with the Integrated State Model presented in Fig. 3. It forms the basis for the further design of MARSP on the interface to the MN. The state model is drawn by using a grid with connection states on the horizontal axis and reservation states on the vertical axis (see below). By using this grid, any possible handover type can be described with a path from the top left (state S) to the bottom right (state F). The intermediate handover states are denoted as $H_1$to H_{11}. The depicted transitions in the state model do not comprise reject or failure transitions in order to simplify the model. It has to be kept in mind that several additional transitions do exist but are not depicted (e.g., failure transitions due to a disconnection from layer two). In the following a

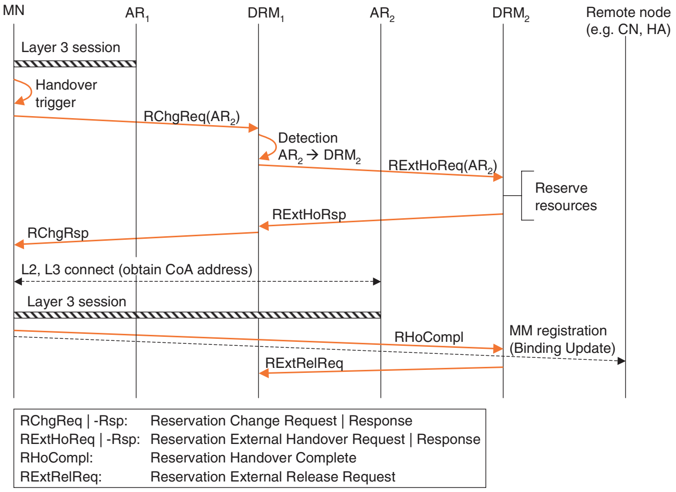

Fig. 2. Sequence diagram for inter-domain anticipated handover

description of the horizontal and vertical axis of the diagram is given.

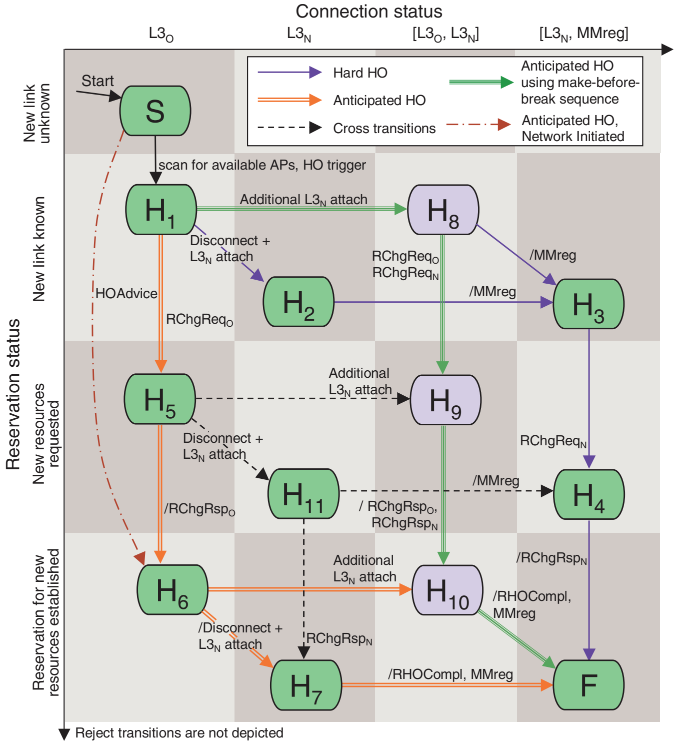

Fig. 3. Integrated Handover Model

A. Connection and Reservation States

The integrated state model describes the connection states of handover sequences on the horizontal axis. The states comprise connection with the old AR at layer three ($L3_O$), connection with the new AR at layer three ($L3_N$), and a connection with both ARs simultaneously ($L3_O$,$L3_N$). Moreover, a separate state with conducted mobility management registration (i.e., sending of binding updates) is introduced ($L3_N$, MMreg) in order to describe the trigger for the handover of data packets between the old and the new data path. In a similar manner the significant states of resource reservation are described on the vertical axis.

B. Handover types

The main benefit of this integrated state model is that in case of any network change it can be switched appropriately between handover types to preserve as much QoS as possible. When the operation of one handover type fails, the handover can still be finished by using another handover type, e.g., with a hard handover instead of an anticipated handover.

An isolated view of individual handover cases would not allow transitions between different handover types. The state model only describes transitions which are important for resource management. We only make minimal assumptions about mobility registration in order to achieve solutions for different mobility management mechanisms. Resources may be reserved for the new path before mobility management switches the data flow to it or after the MN registered at the new AR.

In the following we describe the transitions for the anticipated handover type in the state model, since this handover type represents a preferable method for efficient and fast handovers. By using the anticipated handover type the resources are already reserved when the actual handover is performed. Hence, there is no need to send packets without traffic guarantees for a certain time interval after handover. The path for anticipated handover consists of the following transitions:

- S-H_1: The MN scans for available APs or makes use of potential handover triggers.
- H_1-H_5: In the state H_1, the MN may be aware of several alternative ARs for the handover. A significant advantage of anticipated handover is that the MN has several target ARs to choose from as long as the signaling is still carried out over the old AR. Therefore, it is possible to include the whole set of potential ARs in the resource request. This gives the network the possibility to prefer a specific AR for the handover.
- H_5-H_6: Depending on the available resources the DRM answers with a reply containing the identification of the best AR fulfilling the requirements. This reply already denotes the successful resource reservation for the new path. Our approach does not include a separate query for resources without reserving them.
- H_6-H_7: Now the MN disconnects from the old AR, connects to the new AR on layer two and layer three (in case the MN supports simultaneous layer three connections: H_6-H_{10}) and triggers a mobility management registration (MMreg).
- H_7-F: The MN sends a message about completion to handover to the network (*RHoCompl*). This message is needed in case of anticipated handover to show that the MN really connects to the AR where it has reserved

resources before. Furthermore, the *RHoCompl* message may trigger the release of unused resources in the network after handover. Additionally to resource management, the MN triggers a mobility management registration in this step.

In a similar manner, the integrated state model shows transitions for

- Hard handover
- Anticipated handover using a make-before-break sequence
- Network-initiated anticipated handover
- And variations of these handover types

A hard handover shows the highest handover delay when compared to the other types, since any resource reservation has to be done after the actual handover. Thus, the use of a hard handover should be avoided if other handover types are possible. However, a hard handover has always to be considered, since the occurrence of a sudden loss of coverage would make the use of a hard handover necessary.

An advantage of the integrated handover model is, that it also shows combinations of the previously mentioned handover types. For example when layer three is unexpectedly lost to the old AR during an anticipated handover (state H_5), it is still possible to finish the handover as a hard handover without the need for a retransmission of the resource request. In a similar manner there are several other handover combinations included in the handover state model. Therefore, the state model as an extensive and compact description provides a valuable basis for the development of flexible signaling protocols on the interface to the MN.

V. CONCLUSIONS

In summary, our signaling architecture with its associated signaling protocol MARSP allows the integration of resource management with mobility management and location management. The large variety of different handover types requires an integrated signaling solution. A novel, integrated model was developed to describe the different cases in a unified model. This permits to switch between these handover cases dynamically during a handover. The designed QoS signaling protocol supports all these handover cases, including anticipated handover as well as transitions between different handover types.

There are several proposals of how to achieve QoS in mobile networks. An overview over some selected approaches is given in [13], with a nice classification of the interaction between resource management and mobility. Compared to our approach using a DRM, there are several approaches using on-path, hop-by-hop signaling such as RSVP [9] or extensions of it [14]. An overview of current research on RSVP extensions can be found in [11].

In the same direction, the IETF currently works on a new resource signaling protocol in the NSIS working group [15].

A main advantage of our approach compared to these is that network assisted advance reservation of resources, e.g., for the case of anticipated hand-over, is easier to achieve.

There are also several approaches which use central resource management entities, often called bandwidth broker, e.g., [16], [17], [18]. Compared to these, our architecture is not specific to Differentiated Services [1] and focuses on the seamless integration of different hand-over cases. Our architecture is similar to [17], which however uses implicit signaling and does not consider on specific hand-over types.

REFERENCES

[1] S. Blake *et al.*, “An Architecture for Differentiated Services,” RFC 2475, IETF, Dec. 1998.

[2] R. Braden, “Integrated Services in the Internet Architecture: an Overview,” RFC 1633, IETF, June 1994.

[3] C. Perkins, “Mobile IP,” *IEEE Communications Magazine*, May 2002.

[4] R. Koodli and C. E. Perkins, “Fast Handovers and Context Transfers in Mobile Networks,” *Computer Communications Review, ACM*, Oct. 2001.

[5] A. Campbell, J. Gomez, S. Kim, and C.-Y. Wan, “Comparison of IP micromobility protocols,” *IEEE Wireless Communications Magazine*, Feb. 2002.

[6] W.-S. Soh and H. S. Kim, “QoS Provisioning in Cellular Networks Based on Mobility Prediction Techniques,” *IEEE Communications Magazine*, Jan. 2003.

[7] J. Case, R. Mundy, D. Partain, and B. Stewart, “Introduction and Applicability Statements for Internet Standard Management Framework,” RFC 3410 (Informational), Dec. 2002.

[8] J. Boyle, R. Cohen, D. Durham, R. Rajan, S. Herzog, and A. Sastry, “The COPS (Common Open Policy Service) Protocol,” RFC 2748 (Standard), Jan. 2000.

[9] R. Braden, S. Berson, S. Herzog, S. Jamin, and L. Zhang, “Resource ReSerVation Protocol (RSVP) – Version 1,” RFC 2205 (Standard), IETF, Sept. 1997.

[10] M. Thomas, “Analysis of Mobile IP and RSVP Interactions,” Internet-Draft draft-thomas-nsis-rsvp-analysis-00.txt, IETF, Oct. 2002, (work in progress).

[11] B. Moon and H. Aghvami, “Reliable RSVP Path Reservation For Multimedia Communications under an IP Micromobility Scenario,” *IEEE Wireless Communications*, Oct. 2002.

[12] X. Fu, H. Karl, and C. Kappler, “QoS-Conditionalized Handoff for Mobile IPv6,” in *Proceedings of Networking 2002*, ser. LCNS, no. 2345. Springer, 2002, pp. 721–730.

[13] J. Manner, A. Lpez, A. Mihailovic, H. Velayos, E. Hepworth, and Y. Khouaja, “Evaluation of Mobility and QoS Interaction,” *Computer Networks, Elsevier Science Publisher*, vol. 38, Feb. 2002.

[14] A. Talukdar, B. Badrinathm, and A. Acharya, “MRSVP: A Resource Reservation Protocol for an Integrated Services Network with Mobile Hosts,” *Wireless Networks*, vol. 7, Jan. 2001.

[15] “Next Steps in Signaling (nsis) Charter,” http://www.ietf.org/html.charters/nsis-charter.html, 2003.

[16] T. Engel, H. Granzer, B. Koch, M. Winter, P. Sampatakos, I. Venieris, H. Hussmann, F. Ricciato, and S. Salsano, “AQUILA: adaptive resource control for QoS using an IP-based layered architecture,” *IEEE Communications Magazine*, Jan. 2003.

[17] V. Marques, R. L. Aguiar, C. Garcia, C. B. Jose Ignacio Moreno, E. Melin, and M. Liebsch, “An IP-Based QoS Architecture for 4G Operator Scenarios,” *IEEE Wireless Communications*, June 2003.

[18] E. Mykoniati, C. Charalampous, P. Georgatsos, T. Damilatis, D. Goderis, P. Trimintzios, G. Pavlou, and D. Griffin, “Admission control for providing QoS in DiffServ IP networks: the TEQUILA approach,” *IEEE Communications Magazine*, Jan. 2003.