Investigation of Multicast-Based Mobility Support in All-IP Cellular Networks

Andreas Festag[1], Holger Karl[2], and Adam Wolisz[3]

[1] NEC Europe, Network Laboratories, festag@netlab.nec.de
[2] University Paderborn, Fachgebebiet Rechnernetze, holger.karl@upd.de
[3] Technical University Berlin, Telecommunication Networks Group (TKN), awo@ieee.org

***Abstract*— To solve the IP mobility problem, the use of multicast has been proposed in a number of different approaches, applying multicast in different characteristic ways. We provide a systematic discussion of fundamental options for multicast-based mobility support and the definition and experimental performance evaluation of selected schemes. The discussion is based on an analysis of the architectural, performance-related, and functional requirements. By using these requirements and selecting options regarding network architecture and multicast protocols, we identify promising combinations and derive four case studies for multicast-based mobility in IP-based cellular networks. These case studies include both the standard any-source IP multicast model as well as non-standard multicast models, which optimally utilize the underlying multicast. We describe network architecture and protocols as well as a flexible software environment that allows to easily implement these and other classes of mobility-supporting multicast protocols.**

Multicast schemes enable a high degree of flexibility for mobility mechanisms in order to meet the service quality required by the applications with minimal protocol overhead. We evaluate this overhead using our software environment by implementing prototypes and quantifying handoff-specific metrics, namely, handoff and paging latency, packet loss and duplication rates, as well as TCP goodput. The measurement results show that these multicast-based schemes improve handoff performance for high mobility in comparison to the reference cases: basic and hierarchical Mobile IP. Comparing the multicast-schemes among each other the performance for the evaluated metrics is very similar.

As a result of the conceptual framework classification and our performance evaluations, we justify specific protocol mechanisms that utilize specific features of the multicast. Based on this justification we advocate the usage of a source-specific multicast service model for multicast-based mobility support that adverts the weaknesses of the classical Internet any-source multicast service model.

***Index Terms*— Mobile computing; multicast; handoff; performance measurements**

I. INTRODUCTION

THE GROWING importance of mobile communication requires support for host mobility in IP-based cellular networks. A common architectural approach is to assign a mobile host, at every wireless access point with which it associates, a new IP address that is topologically correct. Such an address assignment allows use of standard IP routing, the disadvantage is that existing sessions are interrupted as a session is associated with the old, now invalid IP address. This dichotomy – an IP address representing both host identification and location – is the fundamental problem of mobile IP-based networks and needs to be overcome by mobility concepts.

The classical solution for support of host mobility is Mobile IP [1]. It uses additional agents to separate host identification from current location and ensures that arbitrary hosts can communicate with a mobile host in an uninterrupted way even while the host moves around. Despite this achievement, Mobile IP has been criticized for its performance problems, including triangular routing and its effect on end-to-end delay, high protocol overhead due to encapsulation of data, and, more importantly, a high handoff latency in scenarios with distant foreign and home agents. Hierarchical Mobile IP [2] has been proposed to reduce the handoff latency by differentiating between local and global handoff.

Besides Mobile IP and its hierarchical variant, other mobility concepts have been proposed to separate a host's identity and location. One approach that inherently uses such a separation is multicasting: Here the identity of a host is the identifier of a multicast group to which a message is destined and the location is represented by a set of conventional addresses. This set of conventional addresses could comprise the address(es) under which the mobile host is currently reachable, and as this set can be modified, multicast provides a natural mapping from a fixed identity (equating a mobile host with a multicast group) to a changing set of locations. One advantage of multicast-based mobility in comparison to Mobile IP is the support of local handoff. With a local handoff the rerouting host is a multicast router close to the mobile host resulting in a shorter service interruption caused by handoff. While this feature is similar to the concept of hierarchical Mobile IP, the multicast-based approach benefits from a high degree of flexibility of mobility mechanisms that includes soft and predictive handoff types to reduce the handoff latency to an absolute minimum. Also, in scenarios where simultaneous distribution of data to multiple locations is advantageous, a multicast-based solution already provides efficient mechanisms to set up and maintain distribution trees with less data overhead in comparison to mobility solutions that integrate similar handoff mechanisms using replicated unicast.

This work has been partially funded by the German Ministry of Education and Research (BMBF) in the project *Flexinet* under grant number 01AK019C and by Siemens AG in the project *SeQoMo*.

This work has been done while Andreas Festag and Holger Karl were working at Technical University Berlin, Telecommunication Networks Group.

Since we expect a diversity of mobility solutions for future IP cellular networks, we consider the multicast-based approach solely for mobility support in an access network (i.e. *micro mobility*), similar to other mobility approaches, such as Cellular IP [3]. The advantage of the multicast-based approach in comparison with these micro mobility solutions is the fact that a multicast infrastructure is reused that potentially exists anyway in future IP networks to provide multicast service to mobile users.

The utilization of multicast for mobility support poses a number of challenges. In particular, today's IP multicast faces some problems which have prevented its wide-spread deployment. These problems include the necessity for global address allocation, lack of receiver and transmission authorization, and complexity. In addition, new challenges arise from the use of multicast for mobility support, such as very dynamic join and leave operations or scalability in number of multicast groups. Finally, multicast does not offer all the functions which are required or useful from a performance point of view.

As these challenges can be overcome in several different ways, considerable freedom remains in choosing multicast options to solve the mobility problem. Examples of design options, which will be described in this article, are the location of the multicast endpoint and the scope of the mobility protocols: The mobile host itself could be the addressee of multicast packets, or multicast traffic could already be terminated at the access point. Multicast could be used for macro mobility in the Internet at large, or only within access networks for micro mobility. Additionally, multicast has to be used to support mobility as such; examples would include address translation or paging of inactive hosts. To structure this large option space, we introduce an abstract framework for multicast-based mobility support. The framework might not be all-encompassing, but does reflect the important practical options for multicast-based mobility support.

On the basis of the requirements and protocol options, we derive a set of preferable instantiations that will be investigated in detail. While the selected case studies have similarities in their choice of basic protocol options, such as the location of the multicast endpoint, they differ in other, such their choice of employed multicast service model employed. Consequently, the case studies provide different mobility functions and optimally utilize the specific features of the multicast they are based on.

We have developed a software environment that turns the abstract framework into a practical platform. While fixed in some architectural assumptions, the platform can utilize a number of different multicast concepts to provide mobility functions in a flexible and extendable manner. Using this software environment we evaluate the performance of three of the selected case studies and justify the multicast mechanism offered by different service models for host mobility.

In summary, this article aims at a systematic discussion of the options for multicast-based mobility support in all-IP cellular networks. We define a set of case studies and design protocols that optimally utilize the features of the multicast for mobility support. The performance of the case studies is evaluated. Analyzing strength and weaknesses of the case studies and the results of the performance evaluation, we give some recommendations for features of future multicast schemes that are beneficial for supporting mobility.

The remainder of the paper is organized as follows: After this introduction we give a high-level protocol overview that describes how multicast-based mobility support works and discuss possible caveats. Section III summarizes the requirements for multicast-based mobility support, including architectural, performance-related, and functional demands. In Section IV we identify the protocol options and give our rationale for the selection of candidates. Section V describes details of the protocol design and Section VI presents the testbed and results from our performance evaluation. Finally, Section VII describes related work and in Section VIII we draw our conclusions.

II. Protocol Overview of Multicast-Based Mobility Support

A. Basic Principles of Multicast-Based Mobility Support

MULTICAST-BASED host mobility attempts to utilize mechanisms for data distribution to certain locations, where the locations are identified network attachment points, e.g. access points. In principle, one multicast group is created per mobile host.[1] A mobile host can subscribe to a multicast group at its current location. In the case of IEEE 802.11 the access point with which the mobile host is currently associated – or at multiple locations by an appropriate mechanism (e.g. pre-registration). Data is distributed via a multicast tree with branches reaching the current potential locations of the mobile host (Figure 1). The branches of the multicast tree can grow and shrink to follow the mobile host's trajectory. Hence, in a wireless network with multiple access points arranged as cell clusters of neighboring cells, data can be efficiently forwarded to multiple access points simultaneously. When a mobile host moves to a new wireless cell within the cell cluster, the data destined for the mobile host are already available in the access point. The new access point can immediately start forwarding data over the wireless link to the mobile host. Consequently, the service interruption is shortened. For support of host mobility predictive handoff is an obvious benefit of utilizing multicast.

B. Discussion of Possible Caveats

APPLYING multicast for mobility raises some common objections that have to be considered in protocol design.

1) "It does not scale.": Insufficient multicast addresses with IPv4 and, closely related, the need for a global address allocation scheme to manage this scarceness, is the first scalability concern. However, using multicast only within an access network where these IP addresses have local significance, allows a static mapping from unicast to locally managed multicast addresses suffices to make this a non-issue.

[1] More generally, this is per *user*, thus also supporting multiple devices and users change the device(s) they use.

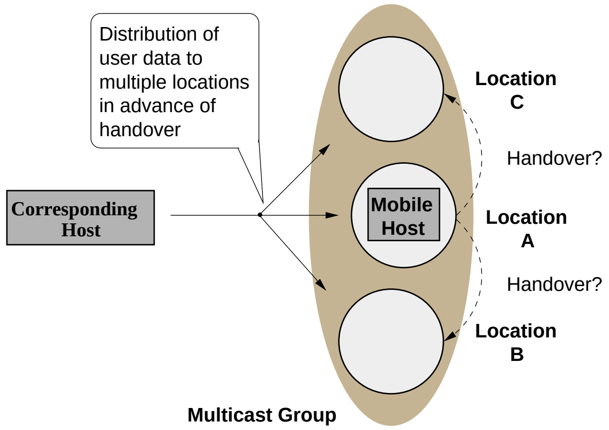


Fig. 1. Principle of multicast-based mobility

A second scalability aspect is the amount of per-group/-source state maintained by a multicast router. With multicast, routing state is generated in each router along the distribution tree. Unlike unicast where routing state is aggregated by means of the network prefix (domain/subnetwork prefix), the aggregation of multicast routing is more difficult. In principle, multicast routing protocols using *explicit-join* only requires per-group state (unlike broadcast-and-prune protocols with per source/group state). For this reason we limit our attention to the family of "explicit-join" protocols. While for aggregation of multicast routing state some potential solutions have been proposed (e.g. bit-wise aggregation in [4]), large multicast routing state can also be reduced if we distinguish between active and inactive mobile hosts and multicast trees are only setup for active ones. This reduces the state size considerably, in cases when many of the mobile hosts are inactive.

A third scalability issue pertains to the use of multiple root nodes of multicast trees to avoid a single point of traffic concentration and failure. For multicast trees that originate at different root nodes the assignment of root nodes to mobile hosts can be controlled. This facilitates the distribution of load among the roots of the tree and constitutes a redundancy for potential failures. Alternatively, the multicast address space can be split up between multiple root nodes using a fixed assignment scheme. In fact, some multicast schemes (e.g. PIM-SM [5]) support multiple cores or rendezvous points (RPs) for a given group where only one group-to-core mapping is active at a given time. Moreover, existing commercial proprietary extensions provide mechanisms to select RPs dynamically.

2) "It is insecure.": Mobile networks are by nature more vulnerable as far as security is concerned and (unfortunately) multicast creates better opportunities for malicious users. These opportunities include the misuse of the soft handoff capability for eavesdropping by simply joining the mobile host's multicast group, DoS attacks by generating unwanted traffic that is distributed to the multicast group and exhausts resources, as well as betraying the multicast paging mechanism in order to force the unnecessary establishment of multicast trees to inactive mobile hosts, causing high signaling traffic and routing state. Restricting today's any source multicast service model would result in improved security. Source-specific multicast simplifies the design of sender authentication and authorization; the small number of group members enables the use of key management protocols for securing multicast even in mobile environments and hampers eavesdropping and DoS attacks. To limit the scope of the paper, we do not discuss multicast security in detail. For a detailed discussion of evaluation criteria and recent proposals for secure multicast see [6].

3) "It is unreliable.": In general, reliability and congestion control in IP multicast suffer from feedback implosion problems caused by multiple receivers and multicast-specific solutions have been developed (e.g. RMTP [7]). However, when multicast is used for host mobility, the mobile host is the only TCP receiver per group. Hence, using standard reliability protocols, even TCP or SCTP, is possible without modifications. Additionally, performance-enhancing last hop protocols can easily be employed this way [8]. However, the proxy should typically be before or at the end of the multicast tree.

In summary, while the caveat of unreliability does not apply, others, such as scalability and security at least partly remain. After definition and evaluation of the case studies we will discuss in the conclusions of this paper whether the caveats are met.

III. Requirements for Multicast-Based Mobility Support

We identify the requirements for mobility support that are essential for choosing among multicast protocol options, to judge existing approaches, and to derive new ones. Many of the requirements seem to be evident for mobility support [9], such as small handoff latency and low packet loss. However, multicast-specific aspects can be identified and have consequences for the design of a multicast-based mobility scheme. We structure the requirements into two main categories: A) requirements that pertain to network architecture and performance and B) functional requirements.

A. Architectural and Performance-Related Requirements

Support of heterogeneous access networks. Typically, a mobile host will be equipped with multiple interfaces. With the progress of software radio technology wireless interfaces are able to adapt to the current environment and switch between different modes if demanded/beneficial. Hosts will switch between wireless networks with very different characteristics regarding bandwidth, delay, error rate, cost, etc. (vertical handoff). Hence, a multicast scheme cannot assume that access network parameters or even organization remains constant before and after a handoff. A multicast protocol must be able to select the optimal interface with respect to available bandwidth, coverage, costs, and other criteria.

Reliable transport of data. A full-scale communication network requires reliable services. With the assumption that the mobile host is the only receiver per multicast group (as discussed in Section II-B.3 by terminating the multicast in the access point and the TCP connection in the mobile), TCP or

SCTP can provide a reliable service. This also enables the usage of wireless-enhanced TCP.

Privacy. Besides general security concerns for multicast (especially, denial of service and eavesdropping), location privacy is important in mobile networks yet constantly violated by today's mobility schemes as temporary IP addresses are tracked and communicated to corresponding hosts. Multicast might be a promising candidate to remove this weakness as it separates identity and routing addresses.

Adaptivity. Constructing a single solution that supports all kinds of application requirements might be possible. However, such a system would provide strong and expensive guarantees even to data flows/applications that do not require them, e.g., seamless handoff for WWW browsing would incur overhead and hence cost that is only necessary for real-time data flows. Therefore, some adaptivity within a single mechanism or an adaptable choice from among a number of mechanisms will be a more economical while providing a satisfactory solution.

Short handoff latency and low packet loss. Handoffs with short latency and low packet loss require a fast execution of multicast join operations. Unmodified multicast protocols would – in the case of handoff – indeed ensure that a mobile host re-joins the multicast group from its new access point, but this would only happen using the slow membership query/report process. Hence, explicit, unsolicited re-join operations are required. Also, multicast management protocols are optimized to reduce the signaling overhead at the expense of increased join latency in order to support large groups of receivers.[2] The impact of such mechanisms should be small if they are employed at all.

Scalability Using multicast for host mobility implies a unique multicast group per mobile. Hence, mobile systems potentially having a very high number of mobile hosts require a multicast scheme which scales with the number of multicast groups, each with typically only a few ($\sim$ 1) members. Today's IP multicast, however, is designed for scalability with the number of hosts per multicast group and additional aspects like the availability of multicast addresses, address assignment, multicast router state, signaling overhead, and route aggregation for reduction of the router state must be taken into account.

Small signaling overhead. Multicast signaling was originally designed for wire-line networks, but the frequent location update and handoff characteristic for mobile networks can result in a considerable load of multicast signaling. Wireless networks require that the signaling on the wireless hop consumes only a small portion of the overall bandwidth. In particular, most IP multicast routing protocols provide soft state maintenance where the routing state needs to be refreshed and expires otherwise. Hard state maintenance reduces the signaling overhead, but is less robust for stale states which are likely to occur in error-prone wireless and mobile environments.

[2]For example, IGMP and MLD defer sending the host's membership report by a random delay. This mechanism reduces the signaling load since usually only a single host sends the report and membership reports of other hosts belonging to the same group are suppressed. However, in the case of a single receiver there is no reason to delay.

Small data overhead. Redundant packet transmissions caused by delays in maintaining the multicast tree (branches not being removed immediately) should be kept low. Broadcast-and-prune algorithms exhibit lots of redundant traffic and are unlikely to be feasible for mobile environments. With multicast routing protocols based on explicit-join operations, data packets are only sent on links where they have been requested. This is the second reason to focus on this type of multicast protocols. However, also with explicit-join multicast protocols slow leave operations may contribute to data overhead as well.

B. Functional Requirements

The functional requirements associated with mobility support can be classified into several categories, each providing a basis for constructing a variant for a multicast-based mobility protocol. The most relevant functional requirements are described here, additional functions are shown in Figure 2.

Detection of link availability. Access points may advertise their availability on their local links. A multicast management protocol may provide this functionality (e.g. IGMP membership query/report scheme). Optionally, a mobile host may also solicit advertisements from access points.

Registration. On top of existing link-layer connectivity a mobile host registers with the network to update its current location information, enabling tracking. Registration can be based on a request/reply scheme initiated by the mobile host or on an invitation by the access point. Alternatively, a mobile host can also be indirectly registered by another access point. Again, the multicast group management protocol can be used for implicit registration.

Address translation. Due to the difference between unicast and multicast addresses, an address translation might be necessary when a multicast-based micro-mobility approach complements a unicast-based scheme for macro-mobility. This functionality is usually performed in a gateway interconnecting the access network with the Internet and also in the mobile host/access points to reverse the translation. Address translation works either by Network Address Translation (NAT) [10] or by encapsulation. Single source multicast necessitates translating the source address of the unicast packet in the gateway. This ensures that the data packets sent over the multicast distribution tree carry the source address of the tree root. In order to keep the original source of the packet as well, tunneling is the preferred option (though may cause fragmentation).

Packet delivery. Packets can be forwarded in advance to access points that the mobile host is likely to reach to be buffered there. Dynamic trees and movement prediction reduces the number of access points involved. Alternatively, packets arriving at an access point where the mobile host has already lost connection can be forwarded to the currently active access point.

Rerouting. A rerouting operation changes the network path of packets for a mobile host in a certain access point. A rerouting operation is based on adding and pruning branches of an existing multicast tree. The appropriate multicast operations

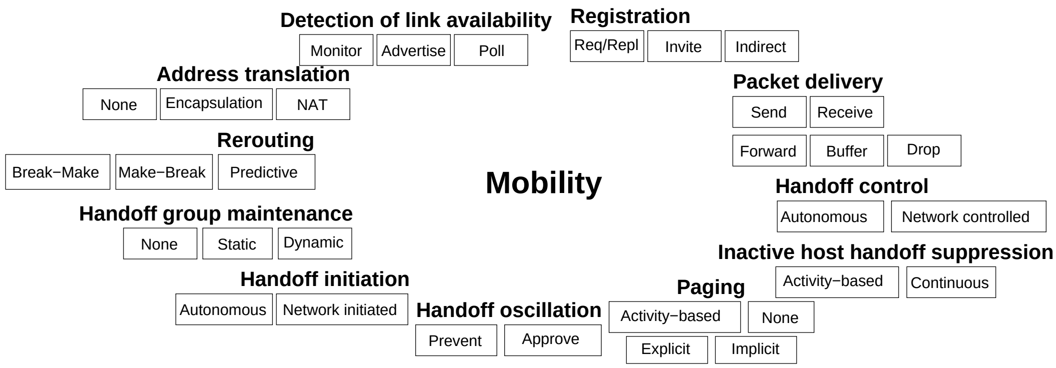


Fig. 2. Functional categories for mobility support.

can be executed in a *break-make* and *make-break* order: old branches are deleted before new ones are added or vice-versa. Additionally, access points can be added to a multicast tree in advance, implementing predictive handoff.

Handoff initiation and control. Both the network or the mobile host could initiate and control a handoff from the old to the new access point. Controlling a handoff means orchestrating the sequence of multicast group manipulation functions. A related issue is the prevention of handoff oscillation (being handed back and forth between two access points).

Paging. Inactive mobile hosts reduce their frequency of handoff registration and location updates, saving wireless resources. Paging locates such mobile hosts and multicast can be used to efficiently distribute paging requests to a paging area identified by a multicast group. Note that the paging group can be distinct from the multicast groups that represent individual mobile hosts. Paging can be done explicitly by sending paging requests to the access points in the paging area or implicitly when data packets are distributed to access points of the paging area.

IV. Selection of Protocol Options

THE requirements summarized in Section III can be met in a variety of network environments. We list different architectural assumptions and alternative protocol options. In fact, a combination of every possible option would result in a huge number of options for protocol and system design. Therefore, we rather derive a set of possible instantiations for multicast-based mobility support that will be investigated in detail in the remaining sections.

A. Description of Protocol Options for Multicast-Based Mobility

The main alternative protocol options are the following:

Multicast type and service model. Main options are network layer (including IP multicast and unicast-based solutions) and link-layer multicast (especially, ATM multicast). The multicast type determines the service model, such as the receiver-oriented model of IP, the sender-oriented model of ATM or the *MCall* model in [11].

Location of multicast endpoint. Selecting the mobile host as multicast endpoint enables multicast protocols to work across the wireless link. This requires multicast protocols which are optimized for economical usage of the scarce wireless resources. Alternatively, the access point might be selected as the multicast endpoint and act as a multicast proxy. The latter option facilitates the usage of optimized signaling protocols on the wireless link that are better adapted to the bandwidth-limited wireless links.

Micro- vs. macro-mobility. Using multicast for macro-mobility allows a uniform solution but requires global scalability of the multicast scheme. If, on the other hand, multicast is only used for micro-mobility, an additional solution (such as Mobile IP) for migration between access networks is necessary but the scalability requirements are reduced. However, coupling two different mobility solutions will (in general) necessitate some form of address translating (between unicast and multicast), e.g. performed in a gateway between an access network and Internet.

Multicast tree directionality. A multicast scheme can provide either unidirectional or bidirectional trees. A unidirectional tree is set up to transport downlink packets from a corresponding host/gateway as the root of the tree towards the mobile host while uplink packets use unicast. With a bidirectional multicast tree, traffic is carried on the tree for both up- and downlink. Bidirectional trees suffer from the fact that uplink packets destined to the gateway/corresponding host are also sent to the access points that belong to the same multicast group and cause unnecessary traffic in the backbone.

X+Multicast. Multicast based on location-independent addressing and routing can be applied as a sole mechanism for mobility support, but not all multicast schemes enjoy this property (e.g., Small Group Multicast (SGM)/Explicit Multicast (XCast) [12], [13]). Nevertheless, these schemes might be utilized for augmenting specific functionalities in other mobility approaches (expressed by *X*). As an example, multicast can distribute packets to multiple Mobile IP foreign agents.

Dynamic tree. The multicast tree can be static or dynamic. In the first case, the access points belonging to a pre-established multicast tree covering a geographical area and the tree does not need to be modified as long as the mobile host remains within this coverage. In the second case, the tree follows the trajectory of the mobile.

Multicast adaptation. Existing multicast protocols can be used “as is”, without modifications. Alternatively, the protocols might be adapted to better meet the requirements of mobility support.

B. Candidate Selections

The combination of various protocol options allows us to define a mobility concept that potentially meets the architectural, performance-related, and functional requirements.

Multicast in today’s IP networks is based on the any-source multicast (ASM) service model. Its usage for mobility support has first been proposed by *Mysore and Bharghavan* [14]. The ASM service model is defined as the “transmission of an IP datagram to a host group, a set of zero or more hosts identified by a single IP destination address” [15] where any host can send to a group and the sender does not need to be a member of the group. The proposal in [14] is based on broadcast-and-prune mechanisms (DVMRP). This routing protocol is optimized to support large groups with sparsely distributed members, an assumption that does not match mobility support. A modification of this proposal is to use explicit join/leave protocols, as has been described by Wu and Maguire [16]. We follow this latter approach in its basic protocol option choices (e.g., placing the multicast endpoint in the access point, dynamic tree, no adaptivity) and extend it by some additional mobility functions (esp. support for inactive hosts, paging). This option is described in Table I, which contains the options common to all our case studies, and as case study MB-ASM in Table II.

This approach does not exploit all capabilities of its underlying multicast service model. In particular, all the necessary multicast functionalities can also be provided by a simplified multicast service model. One example would be source-specific multicast. In fact, for fixed networks a trend to a source-specific service model (SSM) can be identified, particularly driven by the *EXPRESS* [17] proposal and by availability in commercial routers by, e.g., Cisco IOS-based routers. Using this reduced service model instead of the full model corresponds to using a source-based tree (SBS) as the multicast type, resulting in our next candidate, described as case study MB-SSM in Table II. The benefits of SSM in comparison to ASM are: a) lower protocol complexity and easy deployment, b) inhibits denial of service attacks from unwanted sources and c) averts the problem of address allocation. In this way, SSM alleviates some of the main problems associated with ASM-style multicast as a prerequisite for its utilization for host mobility. Moreover, there are other reasons which make attractive the use of SSM: It is ideally suited for tree-like topologies of access networks with a gateway as the root. Since SSM sets up source-based forwarding trees, there is no need for a shared infrastructure with core routers. Finally, the problem of security is aggravated in mobile networks and SSM fairly solves the source access control problem by itself. Additionally, it provides the same actual protocol actions as would result from the use of case study MB-ASM. Hence, the performance of both these case studies is expected to be practically identical.

The usage of the *MCall* service model motivates the third case study. In contrast to the ASM and the SSM service models, a multicast group in this model is represented by a *multi-point, multi-connection communication channel* termed *call*, where only end points being member of the call are allowed to send and receive to/from the call (closed group). End points (i.e. hosts) can be dynamically added and dropped to/from the call, as well as connections between the members of the call. The multicast group is non-anonymous and operations can be executed by the host itself or by a surrogate host. Particularly important for mobility support is the fact that the MCall service model facilitates, unlike the ASM and SSM service models, a much better management of the members of the multicast group, including the senders, and the communication between them. The addition and dropping of end points to/from the call allows precise control of rerouting operations for mobility, such as break-make and make-break rerouting. Additional protocol mechanisms such as third-party registration, resource reservation in advance, sub-casting and others are useful and may potentially improve the handoff performance. Such mechanisms provide a larger design space and increased possibilities for the design of multicast-based mobility concepts, but are not available in current IP-based multicast protocols. In order to investigate such mechanisms, the third instantiation of our framework hence uses a connection-oriented, link-layer type multicast protocol providing such features and is summarized in Table II as case study MB-CMAP, where CMAP represents the protocol realizing the *MCall* service model [11].

One of the main objections to multicast is the scalability problem. Most IP multicast protocols are optimized to scale with the number of participants per multicast group. However, such an optimization does not reflect the needs of a multicast protocol that is to support mobility: Here, only very few participants belong to a single group, but the number of multicast groups is going to be very large. Therefore, a multicast protocol for mobility support should much rather scale with the number of groups; scalability with the number of group participants is only a secondary concern. One example for such a multicast protocol is *small-group multicast (SGM)* multicast [18]. This concept introduces a new protocol layer between IP and transport layer. The sender encodes a list of IP unicast addresses in the SGM header and SGM-capable routers scan the complete list of addresses, determine the outgoing interface for each of the addresses and copy the packet to the outgoing interfaces accordingly. However, this protocol does not separate location and identity and must hence be supplemented by a mobility mechanism. Choosing basic and hierarchical Mobile IP results in case study MIP-SGM of Table II. The multicast is used to efficiently distribute data to multiple foreign agents and enable soft handoff.

Considering the common protocol options in Table I, we have assumed that the multicast terminates in the access point. Thereby, the mobile host does not need to have any knowledge about multicast.[3] This approach also better integrates with existing IP-based protocols such as TCP or ARP and facilitates the deployment of performance-enhancing proxies [8] in the

[3]Additionally, for approaches using IGMP the problem of high handoff delay is solved which occurs when the mobile host waits for the next IGMP membership query instead of sending an unsolicited join.

TABLE I

BASIC PROTOCOL OPTIONS (I.E. OPTIONS USED IN ALL CASE STUDIES)

Protocol options	
Multicast endpoint	Access point
Dynamic tree	Yes
Mobility-specific adaptation	No
Mobility support functions	
Detection of link availability	Advertise/solicit (including link layer trigger)
Handoff initiation and control	Autonomous
Handoff oscillation	Prevent

TABLE II

CLASSIFICATION OF CASE STUDIES

Case studies	MB-ASM	MB-SSM	MB-CMAP	MIP-SGM
Protocol options[a]				
Micro vs. macro	Micro	Micro	Micro	Micro/Macro
Tree directionality	Uni	Uni	Bi	Uni
X+Multicast	Only MC	Only MC	Only MC	MIP/HMIP
MC Type	NWL	NWL	LL	NWL
	Shared	SBT	Shared	NA
MC service model	ASM	SSM	MCall	XCast/SGM
Mobility support functions				
Registration	Req/Reply & Indirect	Req/Reply & Indirect	Req/Reply & by surrogate reg.	Req/Reply
Address translation	Yes	Yes	Yes	Yes/No
Packet delivery	Send, receive, buffer, forward, drop			Send, receive
Rerouting	Make-break Predictive		Break-make Make-Break Predictive	Break-make Make-Break
Paging	Activity-based, explicit			None

[a] Abbreviations: MC = Multicast, MIP = Mobile IP, HMIP = Hierarchical Mobile IP, NWL = Network Layer, LL = Link layer, SBT = Source-based tree, NA = Not applicable.

access point improving the protocol performance over error-prone wireless links. We also make use of dynamic multicast trees that follow the mobile host's movement and cover a small number of access points instead of a geographical region. Finally, we attempt to utilize existing multicast schemes without mobility-specific modifications, but do not limit ourself to the standard multicast service model of IP.

V. DETAILED PROTOCOL DESCRIPTION

IN the following sections we investigate the utilization of different multicast types and present the protocol design common to these schemes. We concentrate here, as well as in the performance evaluation, on the case studies MB-ASM, MB-SSM, and MB-CMAP. For a full description of the protocol design and specification see [19], [20].

A. Network Architecture

The core components of the network architecture (Figure V-A) are mobile hosts, access points, multicast-capable network nodes (routers, switches), and gateways: respectively running a mobile agent (MA), mobility-enabling proxy (MEP), multicast routing daemon (MRD), and a gateway proxy (GWP). These components (with exception of mobile hosts) comprise an access network, which is a network under the control of a single authority. The gateway interconnects the access network with the global Internet. The access network provides wireless communication service in a geographical region. The access points may use different wireless technologies, each offering different service in terms of bandwidth and spatial coverage. A mobile may roam freely within the coverage of an access network having IP connectivity to more than one access point simultaneously. The mobile host associates with one or more access point(s) at its current location and (implicitly) subscribes to the multicast group. Within this access network a multicast tree is setup with the gateway as the root of the tree and at least a single branch with the current access point as the leaf. Data packets are distributed via this dynamic multicast tree with branches reaching the current locations of the mobile host on its movement. The branches of the multicast tree grow and shrink, and hence follow the mobile's location.

The coverage of the access network is further subdivided. First, we group multiple access points – with neighboring and overlapping wireless cells – into an *access point group* identified by a multicast address with the current access point(s) referred to as the *active* access point(s) and the others as *passive* access points. When a mobile host registers with an access point, a multicast tree is established to all access points in that group as leaves. Second, multiple cells are arranged into paging areas, which are also identified by a multicast address. When the exact location of a mobile host is not known, the mobile host is searched within a paging area. Paging areas have typically a larger spatial extent than the spatial coverage of access point groups.

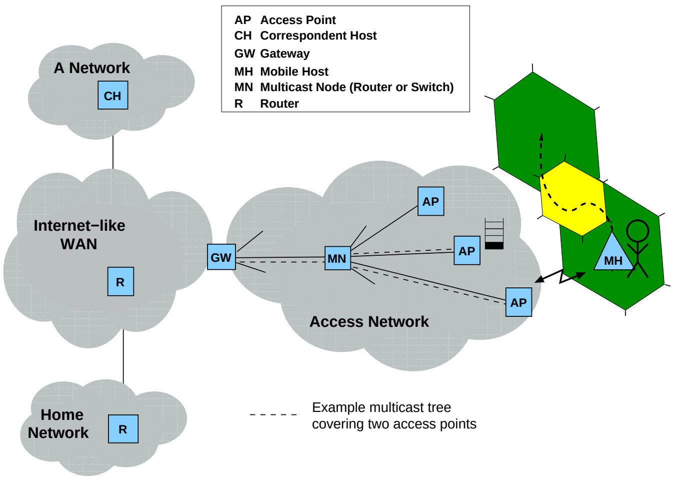


Fig. 3. Network architecture

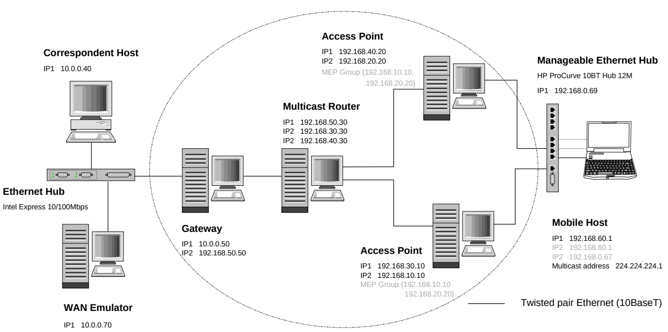


Fig. 4. Exemplary testbed setup (Case study MB-ASM)

B. Addressing and Routing

When a mobile host enters an access network the first time or is switched on, it is assigned a globally valid unicast IP address which remains unchanged as long as the mobile host stays in the access network. This unicast IP address can be used by the macro-mobility protocol to address the mobile host (e.g. as a Mobile IP care-of-address). Within the access network, the unicast address is mapped to a multicast address. In traditional IP multicast (ASM), the multicast IP address identifies a multicast tree; with source-specific IP multicast, both the gateway's unicast address and the mobile's multicast address are used for identification; within MCall multicast the *call-id* serves as an identifier for the multicast group. Address translation functionality from unicast to multicast addresses/identifiers and vice versa is necessary and is provided by the gateway and the access points. Three types of address translation are supported: Network Address Translation (NAT), packet encapsulation, and Segmentation and Reassembling (SAR). Using NAT for address translation reduces the overhead in terms of header/payload ratio. However, source-specific multicast requires that the source address of packets is the gateway's IP address; therefore, for this multicast type IP encapsulation and tunneling is provided rather than NAT. With the MCall service model, packets are mapped to the call-id and sent on virtual circuits that make up the call, whereas the necessary segmentation and reassembling (SAR) from IP packets to cells (and vice versa) is regarded as the third type of address translation.

ASM and SSM provide routing of packets in the downlink direction. After the address translation, packets are sent along a multicast distribution tree and, after reversing the address translation, these packets are delivered to the mobile host by the access point. In the uplink direction packets are sent by unicast (without translating the addresses) and there is no need of using the tree. In contrast, the MCall model provides a bidirectional tree where both up- and downlink packets are transported via the multicast tree; thus leading to unnecessary data traffic to number of access points - 1.

C. Handoff

In principle, a handoff consists of multicast subscribe and un-subscribe operations. We distinguish between hard, soft, and predictive handoff and define these handoff types (in contrast to the definition e.g. in [21]) with respect to the simultaneous connectivity to access points, handoff detection scheme, sequence of rerouting operations, and the buffering capability in the access point. Table III defines these handoff types as options. Considering hard handoff as the standard case, soft handoff attempts to minimize the service interruption and predictive handoff aims at reducing packet loss. We provide policies to select the optimal functionality among multiple possible ones and therefore the capability to control certain system behavior. This capability includes handoff types as described above, but also paging schemes and policies to determine the optimal network service in terms of bandwidth and coverage. The four basic mechanisms in Table III could also be combined differently into other handoff types, but we believe that our combinations are the practically important ones.

With hard and soft handoff (see the message sequence chart in Figure 5(a)), the mobile host detects the availability of the new access point by means of the advertisements sent by the access point and registers with the new access point – either after the lifetime expiration of the old access point's advertisement (hard and predictive) or directly after receiving the advertisement form the new access point (soft). For soft handoff the new access point will now subscribe to the mobile host's multicast group. As long as the mobile host's registration in the old access point has not expired, both the old and the new access point will receive and forward packets to the mobile host. Optionally, the mobile host may de-register with the old access point after registering with the new one. For hard handoff, the mobile host first un-registers from the old access point and then registers with the new access point.

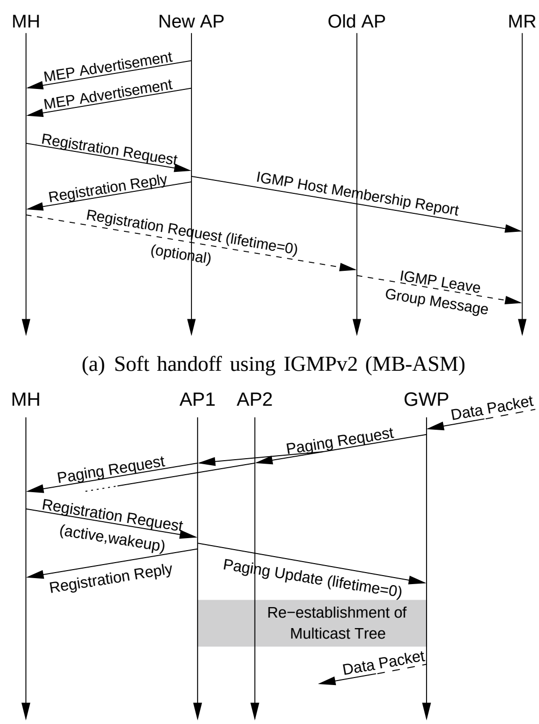
With predictive handoff, the mobile host is directly registered with the *active* access point(s) which forwards packets to the mobile host and registers the mobile host indirectly

TABLE III
HANDOFF TYPES

	Hard	Soft	Predictive
Simultaneous connectivity	to a single access points	to multiple access points	to a single access point
Handoff initiation	Lazy[a]	Eager[b]	Lazy
Rerouting for handoff	Break-Make	Make-Break	Make-Break
Buffering	No	No	Yes

[a] The mobile switches to a new access point when the lifetime of the advertisement from the old access point has expired.
[b] The mobile immediately switches to a new access point when it receives a new advertisement.

– either by indirect registration (MB-ASM, MB-SSM) or by surrogate signaling (MB-CMAP) – with the other access points belonging to the access point group. These *passive* access points subscribe to the mobile host's multicast group as well and buffer incoming packets. When a handoff occurs, the mobile host registers with the new access point which is likely to be one of the formerly passive access points from this access point group. This access point switches from buffering to forwarding mode and forwards the buffered as well as newly received packets to the mobile.



(a) Soft handoff using IGMPv2 (MB-ASM)

(b) Paging

Fig. 5. Selected signaling procedures

The benefit of the predictive handoff scheme is that the packets are already available at the mobile host's new location and the access point is already subscribed to the mobile host's multicast group. The new access point can then immediately forward the buffered packets which the mobile host has probably lost during the handoff process as well as newly arriving packets. As a consequence, the service interruption is minimized and the number of lost packets is reduced. The predictive scheme causes costs in terms of buffer space and data processing in the access points, signaling overhead (over wired links only) for indirect registration among the access points and – more importantly – transport of duplicated packets over the wireless link.

D. Paging

When a mobile host moves freely through the coverage area of the access network, the current location of the mobile host is tracked and the multicast distribution tree is updated. In order to save resources in terms of signaling on the wireless link as well as in the wired segments of the access network, we differentiate between active and inactive mobile hosts. A mobile host switches from active to inactive mode when it has not sent or received data packets except signaling for a predefined duration. This allows the mobile host to update the location information in the network less frequently, reducing power consumption. Since the multicast distribution tree is released, the multicast signaling overhead in the network is reduced and the number of states in the multicast routers decreased. On the other hand the network has only with uncertain location information and once packets arrive for an inactive node, paging is used to find the current location of the mobile host within a limited coverage identified by a multicast group.

When a gateway receives a packet for an inactive mobile host (Figure 5(b)), the packet is buffered and a *PAGING_REQUEST* is sent out to the multicast group identifying the last known paging area of the mobile host. On receiving a *PAGING_REQUEST* the mobile host becomes active and registers with its current access point. The access point subscribes to the multicast channel and sends a *PAGING_UPDATE* with a lifetime of 0 to the gateway. When the multicast distribution tree is established, the buffered packet(s) are forwarded to the mobile host. The case where the mobile host is in inactive state and wishes to send data packets is simple: the mobile host registers and starts sending packets.

VI. TESTBED AND EVALUATION

IN order to evaluate the performance of our case studies we have implemented a software platform, set up testbeds, designed a set of experiments, and conducted measurements to analyze the protocols.[4] We have selected measurements as the main technique for performance evaluation because a) there was a lack of experimental results in this area (except [16])

[4] In this section we focus on the case studies MB-ASM, MB-SSM, and MB-CMAP. The fourth case study MIP-SGM is considered separately.

and b) in comparison to network simulations, measurements can give more accurate results, though it requires considerable efforts to get statistically confident results and allows only limited investigation of scalability effects.

A. Prototype Implementation

We have designed and implemented a generic platform for multicast-based mobility support[5] [22] which can be regarded as a toolkit that implements the mobility functions complementing multicast protocols. The main software components are MA, MEP, and GWP (see Sec. V-A). They were implemented for Linux systems based on IPv4 as daemons running in user space. The software environment provides a generic interface to the multicast protocol and is therefore targeted for examination of different multicast types; the access points are able to execute multicast operations by means of IGMP [23], [24] or other multicast management protocols, such as CMAP [25]. Moreover, the software platform does not depend on a specific multicast scheme. It makes use of the multicast support in the standard Linux kernel[6] and therefore existing and probably also future implementations of multicast routing daemons can be used. The prototype provides hooks for policy handlers to control certain system behavior, such as time and destination of handoff, buffering and flushing of packets, retrieving signal quality indicators for handoff triggers and paging strategies. From this software platform we have derived software prototypes for the case studies MB-ASM, MB-SSM, and MB-CMAP.

B. Testbed Setup

We have investigated a network topology as depicted in Figure V-A: the corresponding testbed setup is illustrated in Figure V-A. In addition to the prototype components MA, MEP, and GWP common to all case studies, for the case study MB-ASM the gateway and multicast node execute a multicast router daemon PIM-SMv2 and the access points use IGMPv2. For the case study MB-CMAP, the multicast node is represented by a multicast-capable switch with a switch controller[7] and the MEP and GW use CMAP as a multicast management protocol.

In the testbed we replace the wireless link by standard Ethernet. The benefit is that the mobility-specific performance characteristics of various protocols can be studied in a controlled manner in isolation of the potentially error-prone wireless link. The mobile host and the access points are interconnected by means of a managed Ethernet hub. The handoff is triggered by switching off/on the hub's port to the respective access point. We have verified that the delay introduced by the operations to manage the Ethernet hub are comparable with the handoff delay caused by the re-association procedure in IEEE 802.11.

In order to investigate the SSM-style multicast case study, we in fact used an ASM-style multicast implementation. This is indeed feasible and the measurement results are valid as the additional ASM functionality is not used and hence incurs no additional overhead; the differences between the usage of IGMPv2/PIM-SM and IGMPv3/PIM-SSM only marginally affect the performance. Moreover, we are able to basically confirm the initial measurements for the handoff latency published in [16] of the approach using also IGMPv2/PIM-SM with the same protocol options, but with enhanced functionalities. More importantly, we have conducted our experiments in a common evaluation environment that facilitates comparing performances.

Putting our measurement results into the right perspective requires measuring a standard (non-multicast) mobility solution as well. Thus, we have repeated the same set of experiments for basic and hierarchical Mobile IP (MIP and HMIP) using the *Dynamics* Mobile IP implementation[8] in a comparable testbed setup with MIP agents replacing the corresponding multicast-based components of our prototype. For fair comparison between Mobile IP (basic and hierarchical) and the multicast scheme, we have set the parameters in the WAN emulator such that effectively the delay between mobile and corresponding host is almost the same. This ensures that the triangular or bi-directional routing between mobile and corresponding hosts for basic and hierarchical Mobile IP does not effect the measured (i.e. TCP) performance.

C. Handoff Performance

The objective of this experiment was to analyze the performance of the different handoff policies provided by the case studies and to investigate the impact of handoff on UDP and TCP performance.

1) Handoff Latency: The handoff latency is defined as the service interruption caused by handoff and is measured in the mobile host at the IP layer. A continuous packet stream is sent downlink from the corresponding host to the mobile host. During the receive process, the mobile host executes periodic handoffs between both access points. The duration between two subsequent handoffs (cell dwell time of a mobile in a particular cell) is exponentially distributed with a mean of 10 s plus offset[9] by 5 s. Each run took 1 hour resulting in about 240 handoffs. A handoff is controlled by switching off the Ethernet port to the old access point and switching on the port to the new one. The order of these operations is important since it results in a small gap of connectivity of about 100 ms which corresponds with non-overlapping wireless cells in a real system. The handoff latency is assessed as the duration between the last received packet before a handoff and the first received packet after the handoff. A handoff is detected using the 16-bit identification field in the IP header – either by a gap in the successive packet identifiers or by duplications. Clearly, the granularity of handoff latency with this method is limited by the inter-arrival time of packets. In the experiment, the

[5]The software environment is open software under the GNU public license and is available from `http://www.tkn.tu-berlin.de/research/mombasa/mse.html`.

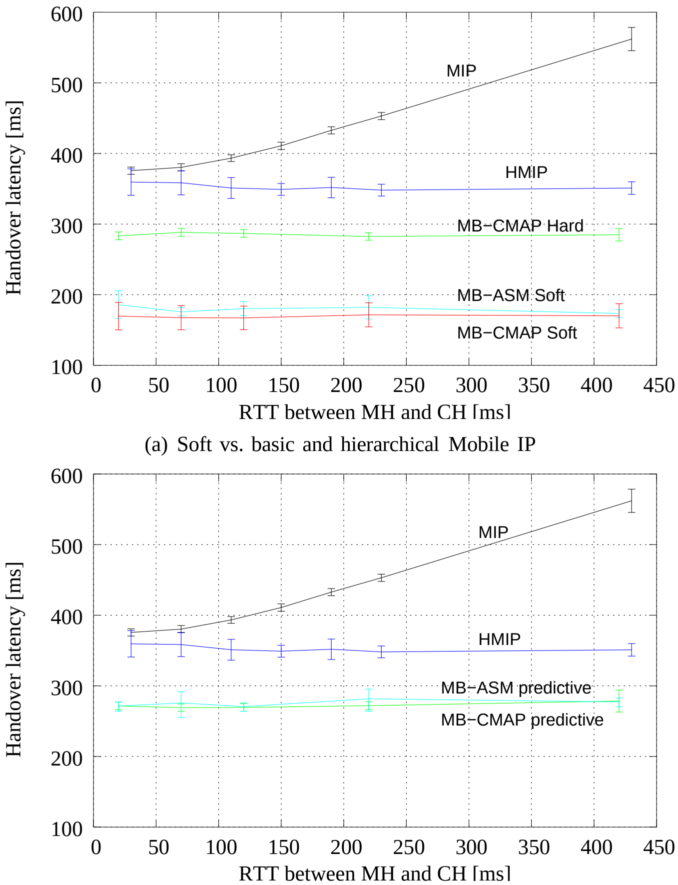
[6]However, modifications of the Linux kernel were necessary for address translation between unicast and multicast addresses and for paging.

[7]These components belong to the Washington University Gigabit Switch Kit, an open, non-proprietary networking package. `http://www.arl.wustl.edu/gigabitkits`.

[8]`http://www.cs.hut.fi/Research/Dynamics`

[9]The offset prevents a mobile host from executing two subsequent handoffs without registering with the first access point.

inter-packet time in generating the packets is set to 10 ms. We have validated that the inter-arrival time deviation (jitter) due to other reasons is very small. Therefore, we can indeed assume a inter-arrival time of packets also amounts to 10 ms.



(a) Soft vs. basic and hierarchical Mobile IP

(b) Predictive vs. basic and hierarchical Mobile IP

Fig. 6. Handoff latency versus round trip time between MH and CH for horizontal handoff (≈ 240 handoffs, 99 % confidence level)

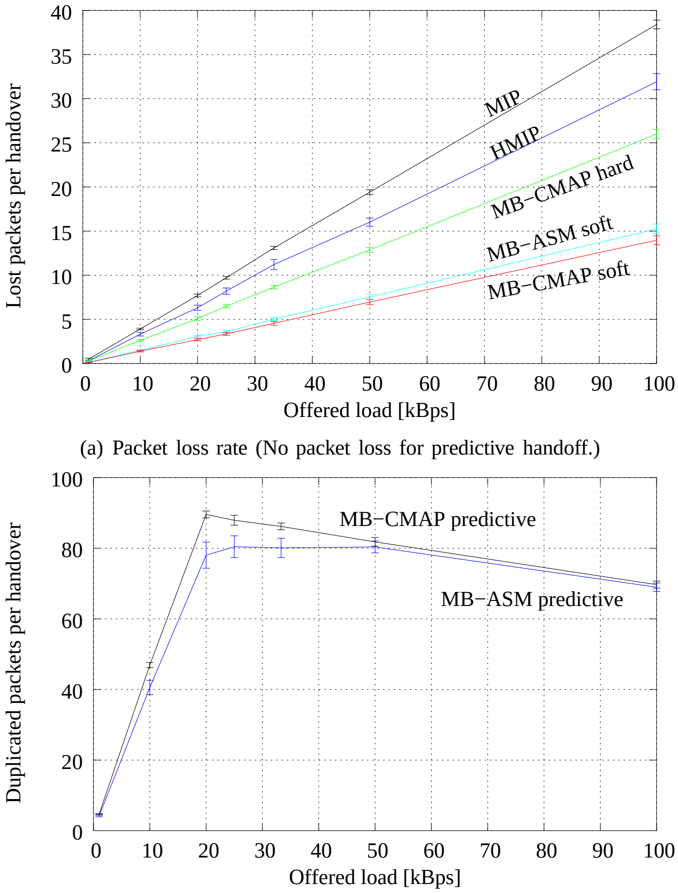
In Figure 6 the mean handoff latency for the case study MB-ASM and MB-CMAP and the reference case basic and hierarchical Mobile IP is plotted. Each point in the graph is obtained by averaging the handoff latency over about 240 handoffs. Principally, the handoff latency for the multicast-based handoff (MB-ASM, MB-CMAP) and hierarchical Mobile IP is independent of the round-trip time (RTT) between the corresponding host and the mobile host. This seems to be trivial for multicast-based mobility (the rerouting node for handoff is close to the mobile), however, it is an important measure in comparison with basic Mobile IP and emphasizes the ability of our multicast-based schemes to select the optimal network node for rerouting. In all cases the handoff latency is composed of the duration for handoff detection and for handoff execution. The duration for handoff detection is determined by the advertisement interval of 100 ms, whereas the advertisement lifetime of 300 ms impacts the handoff detection only when lazy cell switching is used. As a result from the measurements the mean handoff latency for soft handoff is less than 200 ms (MB-ASM and MB-CMAP), for predictive and hard handoff less than 300 ms. In fact, all multicast-based schemes are superior to basic and hierarchical Mobile IP, where the soft schemes cause the shortest handoff latency – mainly due to the reduced latency for handoff detection (eager cell switching) combined with a fast make-break rerouting. Interesting is the difference between Hierarchical Mobile IP and MB-ASM though both use comparable mechanisms:hard handover: The multicast-based schemes maintain one timer per event. In contrast, the Mobile IP schemes implement the timing such that a function is called at equidistant time intervals and processes expired events. Hence, it is more efficient but causes additional latency for handoff. Finally, it can also be stated that there are no significant differences between the different handoff mechanisms of the case studies MB-ASM and MB-CMAP.

2) UDP Packet Losses and Duplications: In this experiment we examine the number of lost and duplicated packets due to handoff. The corresponding host continuously generates 1024 byte UDP packets. We vary the inter-packet times to change the offered load; other parameters are set to the same values as in the handoff latency experiment but with a fixed RTT between the mobile and corresponding host of about 100 ms. To calculate the number of lost and duplicated packets, packet traces are recorded in the corresponding host and in the mobile host. We obtain the number of lost and duplicated packets by comparing both traces and marking lost and duplicated packets. The mean number of lost packets (and similarly, the mean number of duplicated packets) per handoff is calculated by dividing the overall number of lost packets by the number of handoffs. Since this method may also incorporate *any* packet loss, we have conducted the measurements without handoff and verified that packet loss due to other reasons than handoff is negligible. The results are plotted in Figure 7(a) and 7(b). For soft handoff (MB-ASM, MB-CMAP) the packet loss rate increases constantly and reaches about 15 packets per handoff, while for hard handoff about 25 packets per handoff are lost. The reason that for any packet loss with any handoff is the (emulated) gap of connectivity of about 100 ms between the neighboring wireless cells in the investigated scenario. This 'physical' gap cannot be bridged by the soft handoff. We expect, however, better results in case of overlapping wireless cells.

For predictive handoff, we have set the buffer size to 100 kB per mobile in each access point and use a buffering policy, which forwards only new packets (not older than 5 s) and drops outdated packets. It can be seen in Figure 7(a) that the packet loss rate is very small and can be neglected (packet loss rate < 0.1 %) though the gap in wireless connectivity. However, the reliability of predictive handoff causes some costs, mainly due of the duplication of packets, as explained below.

In comparison, the packet loss rate for basic and hierarchical Mobile IP amounts to more than twice the loss rate for soft handoff.

While the number of duplications can be neglected for soft handoff, for predictive handoff the duplications grow with the offered load up to a certain point at 20 kBps with about 80 duplicated packets per handoff. Beyond a load of 20 kBps the number of duplicated packets decreases slowly. The reason for



(a) Packet loss rate (No packet loss for predictive handoff.)

(b) Packet duplication rate (No packet duplication for hard and soft handoff.)

Fig. 7. UDP Packet loss and duplication versus offered load for horizontal handoff (≈ 240 handoffs, 99 % confidence level)

the shape of the curve is the limitation of the buffer size (set to 100 kB per mobile in each access point) and the buffering policy, which forwards only new packets (not older than 5 s) and drops outdated packets. These (relatively large) values were chosen with the assumption that the handoff latency can not be predicted and packet loss should be avoided. Clearly, reducing these values (e.g. by considering aspects specific to the wireless technology in order to predict the handoff latency or state transfer between the old and new access point) can considerably reduce the duplication rate.

The results for packet loss can be verified by calculating the number of lost packets by $T * r_{\text{Data}}$, where T is the handoff latency from Figure 6 and r_{Data} is the data rate. For example, the measured handoff latency of about 140 ms for MB-ASM with soft handoff theoretically results in a packet loss of 7 packets per handoff.

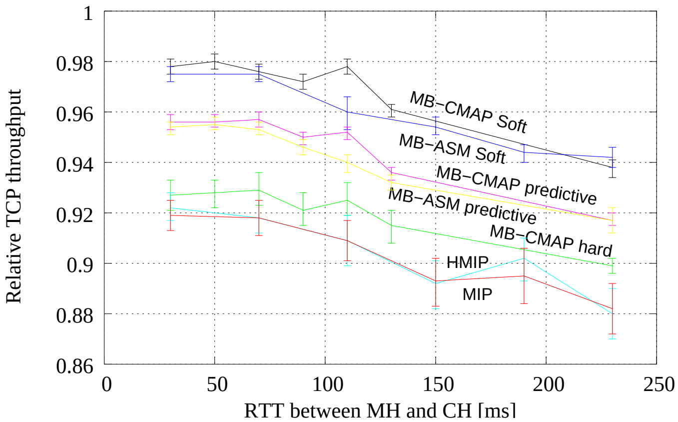
3) TCP Goodput: In this experiment we study the impact of handoff on the throughput of standard TCP (Reno implementation in Linux). We define the measure *Relative TCP goodput* as the TCP goodput with handoff normalized to the TCP goodput without handoff for the particular RTT. First, we measure the relative TCP goodput for a short-lived TCP connection (60 s) with a single handoff event. Here, we have set the advertisement interval/lifetime to 1 s and 3 s, respectively. We observe (Figure 8(a)) that the relative TCP goodput declines only slightly with the RTT. The differences between the case studies and the reference case basic and hierarchical Mobile IP are small (less than 10 %).

In order to evaluate the impact of multiple handoffs on TCP where the subsequent handoff events impact each other, we also examine a single long-lived TCP connection with subsequently executed handoff events for a fixed RTT between the mobile host and corresponding host of 100 ms and set the handoff frequency by varying the mean cell dwell time. With handoff events successively executed and a high handoff rate TCP has less time to recover from packet loss. Nevertheless, it can be seen in Figure 8(b) that the goodput for soft handoff (MB-ASM and MB-CMAP) decreases moderately even for frequent handoff. The relative TCP goodput for predictive handoff is reduced to 0.6 for the high handoff frequency of 6 handoffs/min, whereas the TCP throughput of basic and hierarchical Mobile IP considerably degrades to less than 0.35. Regarding predictive handoff, this is not an evident benefit, since the duplication of TCP segments triggers the mobile host to send duplicated acknowledgments, which in turn forces the corresponding host to retransmit TCP segments. Overall, the combination of TCP's slow start mechanism and subsequently received duplicated data segments finally results in a TCP goodput that is still better than the TCP goodput of basic and hierarchical Mobile IP. However, it is well-known that standard TCP shows a bad performance over wireless links. A modified TCP can provide a better performance, as shown with performance-enhancing proxies. The question of whether a predictive handoff policy with buffering and forwarding of packets is a promising option for a modified TCP will be further investigated.

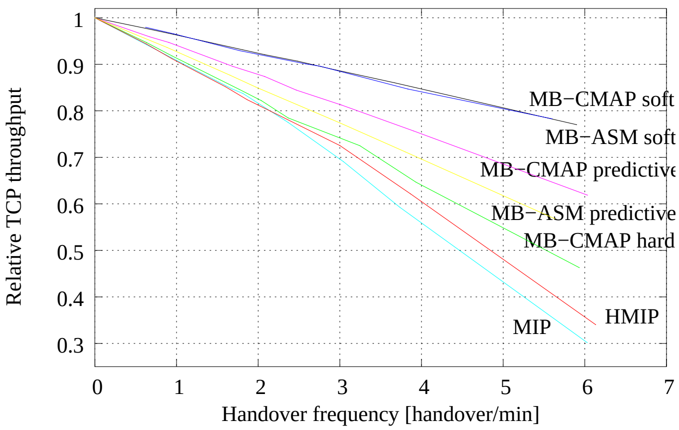
D. Paging Performance

The support of idle connectivity causes some costs, such as buffer space in the gateway proxy, signaling overhead, and delay in packet delivery. The latter metric considers the deferred delivery of packets of newly established data streams generated by a corresponding host. While for the case study MB-CMAP the delay in packet delivery is mainly caused by sending the signaling message, we observe a worse behavior in MB-ASM: In an experiment we measured this delay by sending single *ICMP echo requests* to a mobile host in the inactive state. The duration between sending the *ICMP echo request* and receiving the *ICMP echo reply* measured in the corresponding host gives the round trip time including the paging delay and is referred to as RTT_{Paging}.

The mean paging delay of 100 observations amounts to about 1.2 seconds with a standard variation of about 0.5 seconds. The histogram for the paging latency is depicted in Figure 9 and shows peaks at multiples of 0.5 seconds. The reason for this shape is the polling algorithm of standard Linux to notify the multicast forwarding cache of a successfully resolved multicast entry. In detail, after the paging request had been sent and the paging daemon has received a paging update, the entry in the multicast forwarding cache is marked as



(a) Short-lived TCP connection with single handoff (≈ 240 handoffs, 99 % confidence level)



(b) Long-lived TCP connection with multiple handoff (≈ 240 handoffs)

Fig. 8. Relative TCP throughput for single and multiple handoff events

unresolved.[10] This triggers a cache miss report to the multicast routing daemon. The multicast routing daemon is polled until the routing entry for the corresponding mobile host could be resolved. The polling interval is set to 0.5 seconds. This is a reasonable value taking into consideration that the polling is done per multicast group and the number of mobile hosts can be high. It is well known that polling is an inefficient mechanism and alternative mechanisms would likely give better results. Nevertheless, the polling mechanism has been used since it requires only minimal changes in the multicast routing demon and kernel support itself. Furthermore, it is worth noting that the paging delay is relatively high in relation to a round trip time of an *ICMP echo request* and *ICMP echo reply* (which would be about 20 ms without invoking the paging procedure). We have identified the IGMP and PIM join operations as contributing with the major part to the latency. This high delay opens some potential for improvement. However, the paging latency applies to the first packet(s) of a newly established data stream only. For most applications this paging delay is acceptable.

[10]This is the usual Linux mechanism to resolve a multicast routing entry which is not in the routing cache.

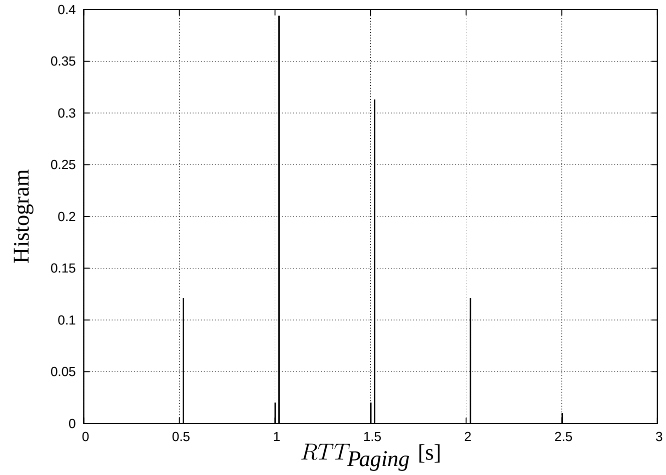


Fig. 9. Histogram of delay in packet delivery due to paging (RTT of IGMP echo request/reply)

VII. Related Work

THE combination of multicast and mobility has been proposed shortly after the introduction of IP multicast [26] and mobility as separate problems. It is widely accepted that multicast has some attractive features for mobility support, particularly in highly mobile environments with very small cells. Nevertheless, a systematic effort for investigating multicast-based mobility support was missing and has been provided by this paper. Here we give a brief overview of existing solutions and add some examples to the classification in the previous chapters. In general, the existing approaches have different motivations and requirements and are based on different assumptions about the networking architecture.

One of the first steps was taken by Keeton et al. [27] who proposed a *multicast-based re-establishment scheme* for re-routing of connections for mobile hosts where the access point performs multicast operations for mobility support on behalf of the mobile host. In [14] the mobile host is assigned an IP multicast address and joins an IP multicast group. A handoff is executed by means of the Internet Group Management Protocol (IGMP) [23] join and leave operations by the mobile host itself. The proposed protocol includes that a mobile host can pre-register and packets can be delivered to the next cell in advance of handoff.

Approaches following this initial work can be categorized into multicast-based mobility support in connection-oriented or connection-less environments. Ghai and Singh [28] propose an architecture for pico-cellular networks and a connection-oriented protocol for seamless mobility support. Acampora and Naghshineh [29] introduce a *virtual tree concept* where a multicast connection tree is pre-established covering a number of access points in a geographical region. Data packets are delivered across the branch of the virtual tree to the access point actually serving the mobile host. The signaling load is reduced as a handoff mainly consists of the activation/deactivation of a pre-established branch. Based on the work in [27], Seshan [30] applies the multicast approach to Mobile IP in order to perform fast handoff. The Mobile IP home agent encapsulates packets destined for the mobile host into multicast packets and sends these packets to multiple Mobile IP foreign agents. This approach has been extended in [31] for improvement of handoff in *wireless overlay networks* with a hierarchy of wireless cells of different bandwidth and coverage where the mobile host usually changes the network interface.

Recently, a few new approaches have emerged. Wu and Maguire [16] have proposed a network architecture with mobility-supporting agents (MSA) running in access points and using standard IP multicast protocols (IGMPv2 and PIM-SM [5]) and protocols for agent discovery and pre-registration. Mihailovic et al. [32] have focused on multicast to be used as a micro-mobility protocol in a mobile access network and on using IGMPv2 and CBT [33]. Finally, Helmy, Jaseemuddin, and Bhaskara have investigated different handoff options for multicast-based handoff – termed CAR[11]-set protocols. Helmy [34], [35], [36] evaluates the performance of multicast-based mobility support in large-scale simulations focusing on system evaluation.

Closely related to the problem of using multicast for mobility is the problem of delivering multicast messages to mobile hosts. Despite some solutions to parts of this problem (e.g. Acharya and Badrinath [37] address the problem of reliable delivery of multicast messages to mobile hosts), this is still a challenging problem, but not the main focus of the present paper. Rather, we focus on utilizing multicast for mobility support.

The systematic investigation of options for multicast-based mobility support as described in this paper is essential to optimize existing solutions and to discover the design space for mobility support with potential future multicast solutions. In particular, this paper considers also multicast approaches beyond the standard Internet ASM service model. This is in line with the current trend to develop a new multicast service model for the Internet that complements the ASM model. In addition, up to now a comparative quantitative evaluation did not exist. This includes a comparison among multicast-based as well as between multicast and Mobile IP (except [34] for large-scale system simulation).

VIII. Conclusions and Outlook

The large number of possible approaches to use multicast to solve the mobility problem in IP networks made it necessary to structure all these options. Based on a list of requirements for a mobility concept, we developed a matrix of possible protocol options of how to deploy multicast for mobility and a number of possible mobility functions that can be implemented using multicast. Requirements, options, and functionalities constitute a framework in which existing approaches can be classified and new approaches can be derived by sensible combinations of decisions for each individual aspect.

To substantiate this classification with examples, we have derived four case studies for mobility support. These case studies are based on different multicast service models that are not limited to today's standard model of IP multicast. The multicast protocols themselves used in the case studies were unchanged but augmented with support functions for mobility. In order to quantify relevant performance metrics for multicast-based mobility protocols, we developed a software environment that allows to easily implement a variety of multicast protocols and to experiment with them. Experiments showed that our protocols provide efficient micro mobility and location management along with smooth, fast handoffs. In particular, multicast-based mobility support offers a great flexibility of mobility mechanisms that meet the service quality required by applications at a minimal protocol overhead. For multicast-based mobility support an alternative multicast service model is advocated that averts the weaknesses of the classical Internet any-source multicast (ASM) service model.

The success of a future mobile communication system leveraging multicast-based mobility support is strongly associated with the success of multicast in general in future IP networks. The multicast service model and multicast protocols that will be used depends on how these approaches overcome the problems of today's classical ASM service model. Potential candidates were evaluated and it could be shown that no single perfect multicast exists. Without being committed to multicast protocols as they exist today, a multicast-based mobility support would highly benefit from several multicast features: A multicast with source-specific trees and closed groups is ideally suited for the topology of an access network and improves security. Non-anonymous groups simplify the management of a multicast tree for soft and predictive handoff and facilitates the resource reservation in advance and the sub-casting of data to a subset of access points for predictive handoff.

[11]CAR stands for Candidate Access Router and refers to a group of access points.

References

[1] C. Perkins, "IP Mobility Support for IPv4," RFC 3220, January 2002.

[2] D. Forsberg, J. T. Malinen, T. Weckstroem, and M. Tiusanen, "Distributing Mobility Agents Hierarchically Under Frequent Location Update," in *Proc. of Sixth IEEE International Workshop on Mobile Multimedia Communications (MOMUC 99)*, San Diego, CA, USA, November 1999, pp. 159–168.

[3] A. Campell, J. Gomez, S. Kim, A. Valko, C.-Y. Wan, and Z. Turanyi, "Design, Implementation, and Evaluation of Cellular IP," *IEEE Personal Communication*, vol. 7, no. 2, pp. 42–49, August 2000.

[4] A. Helmy, "State Analysis and Aggregation Study for Multicast-Based Micro-Mobility," in *Proc. of IEEE International Conference on Communications (ICC 2002)*, New York, USA, May 2002, pp. 3301–3306.

[5] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, and L. Wei, "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification," RFC 2362, June 1998.

[6] M. Moyer, J. R. Rao, and P. Rohathi, "A Survey of Security Issues in Multicast Communications," *IEEE Network*, vol. 13, no. 6, pp. 12–23, November/December 1999.

[7] S. Paul, K. Sabnani, J. Lin, and S. Bhattacharyya, "Reliable Multicast Transport Protocol (RMTP)," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 3, pp. 407–421, April 1997.

[8] H. Balakrishnan, V. Padmanabhan, Seshan.S., and R. Katz, "A Comparison of Mechanisms for Improving TCP Performance Over Wireless Links," *IEEE ACM Transactions on Networking*, vol. 5, no. 6, pp. 756–769, 1997.

[9] A. Festag, H. Karl, and G. Schäfer, "Current Developments and Trends in Handover Design for All-IP Wireless Networks," TKN, TU Berlin, Tech. Rep. TKN-00-007, August 2000.

[10] P. Srisuresh and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)," RFC 3022, January 2001.

[11] E. Gaddis, R. Bubenik, and J. DeHart, "A Call Model for Multipoint Communication in Switched Networks," in *Proc. of IEEE International Conference on Communications (ICC 92*, Chicago, IL, USA, June 1992, pp. 609–615.
[12] R. Boivie, "A New Multicast Scheme for Small Groups," IBM T. J. Watson Research Center, Research Report RC21512(97046)29Juni1999, June 1999.
[13] R. Boivie, N. Feldman, Y. Imai, W. Livens, D. Ooms, and O. Paridaens, "Explicit Multicast (XCast) Basic Specification," Internet Draft Work In Progess, August 2000.
[14] J. Mysore and V. Bharghavan, "A New Multicast-Based Architecture for Internet Mobility," in *Proceedings of ACM MOBICOM 97*, Budapest, Hungary, October 1997, pp. 161–172.
[15] S. Deering, "Host Extensions for IP Multicasting," Internet RFC 1112, August 1989.
[16] J. Wu and G. Maguire, "Agent-Based Seamless IP Multicast Receiver Handover," in *Proc. of Conference on Personal Wireless Communications (PWC 2000)*, Gdansk, Poland, September 2000, pp. 213–225.
[17] H. Holbrook and D. Cheriton, "IP Multicast Channels: EXPRESS Support for Large-Scale Single-Source Applications," in *Proceedings of ACM SIGCOMM 1999*, Cambridge, MA, USA, 1999, pp. 65–78.
[18] R. Boivie, N. Feldman, and C. Metz, "Small Group Multicast: A New Solution for Multicasting on the Internet," *IEEE Internet Computing*, vol. 4, no. 3, pp. 75–79, May/June 2000.
[19] A. Festag, "Mobility Support in IP Cellular Networks – A Multicast-Based Approach," Ph.D. dissertation, Technical University Berlin, Germany, June 2003, http://edocs.tu-berlin.de/diss/2003/festag_andreas.htm.
[20] A. Festag and L. Westerhoff, "Protocol Specification of the MOMBASA Software Environment," TKN, TU Berlin, Berlin, Germany, Tech. Rep. TKN-01-014, May 2001, http://www.tkn.tu-berlin.de/research/mombasa/.
[21] T. Ojanperä and R. Prasad, *WCDMA: Towards IP Mobility and Mobile Internet*, ser. Artech House Universal Personal Communications Series. London, UK: Artech House, 2000, ISBN 1-58053-180-6.
[22] A. Festag, L. Westerhoff, and A. Wolisz, "The MOMBASA Software Environment – A Toolkit for Performance Evaluation of Multicast-Based Mobility Support," in *Proc. of Performance Tools 2002*, London, GB, April 2002, pp. 212–219.
[23] W. Fenner, "Internet Group Management Protocol, Version 2," RFC 2236, November 1997.
[24] B. Cain, S. Deering, I. Kouvelas, B. Fenner, and A. Thyagarajan, "Internet Group Management Protocol, Version 3," RFC 3376, October 2002.
[25] K. Cox and J. deHart, "Connection Management Access Protocol (CMAP) Specification," Washington University, St. Louis, MO, USA, Applied Research Laboratory Working Note ARL 94-21, July 1994.
[26] S. Deering, "Multicast Routing in a Datagram Internetwork," Ph.D. dissertation, Stanford University, CA, USA, December 1991.
[27] K. Keeton, B. Mah, S. Seshan, R. Katz, and D. Ferrari, "Providing Connection-Oriented Network Services to Mobile Hosts," in *Proceedings of the USENIX Symposium on Mobile and Location-Idependent Computing*, Cambridge, MA, USA, August 1993, pp. 83–102.
[28] R. Ghai and S. Singh, "An Architecture and Communication Protocol for Picocellular Networks," *IEEE Personal Communications*, vol. 1, no. 3, pp. 36–46, 1994.
[29] A. Acampora and M. Naghshineh, "An Architecture and Methodology for Mobile-Executed Handoff in Cellular ATM Networks," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 8, pp. 1365–1375, October 1994.
[30] S. Seshan, "Low-Latency Handoffs for Cellular Data Networks," Ph.D. dissertation, University of Berkeley at California, March 1996.
[31] M. Stemm and R. Katz, "Vertical Handoffs in Wireless Overlay Networks," *ACM MONET, Special Issue on Mobile Networking in the Internet*, vol. 3, no. 4, pp. 335–350, Winter 1998.
[32] A. Mihailovic, M. Shabeer, and A. H. Aghvami, "Multicast for Mobility Protocol (MMP) for Emerging Internet Networks," in *Proc. of International Symposium on Personal Indorr and Mobile Radio Communications (PIMRC 2000)*, London, UK, September 2000.
[33] A. Ballardie, "Core Based Trees (CBT) Multicast Routing Architecture," RFC 2201, September 1997.
[34] A. Helmy, "A Multicast-Based Protocol for IP Mobility Support," in *Proceedings of ACM Second International Workshop on Networked Group Communication (NGC 2000)*, Palo Alto, CA, USA, November 2000, pp. 49–58.
[35] A. Helmy, M. Jaseemuddin, and G. Bhaskara, "Multicast-Based Mobility: A Novel Architecture for Efficient Micromobility," *IEEE Journal on Selected Areas in Communications (JSAC), Special Issue on All-IP Wireless Networks*, vol. 22, no. 4, pp. 677–690, May 2004.
[36] ——, "Efficient Micro-Mobility Using Intra-Domain Multicast-Based Mechanisms (M&M)," *ACM SIGCOMM Computer Communications Review (CCR)*, vol. 32, no. 5, pp. 61–72, October 2002.
[37] A. Acharya and B. Badrinath, "A Framework For Delivering Multicast Messages in Networks With Mobile Hosts," *ACM MONET*, vol. 1, no. 2, pp. 199–219, 1996.

Andreas Festag received a M.S. degree in electrical engineering in 1996 and a PhD in 2003, both from Electrical Engineering. Between 1996 and 2003, he was research assistant at the Telecommunication Networks research group (TKN) of Technical University Berlin. Since 2003, he works as senior researcher for the Network Laboratories of NEC Europe Ltd and. His scientific interests are in communication systems, in particular protocol design, specification, and performance evaluation of cellular and ad hoc networks.

Holger Karl studied computer science at the Technical University Karlsruhe from 1990 to 1996; he obtained his PhD in 1999 from Humboldt University of Berlin. Between 2000 and 2004, he was assistant professor (wissenschaftlicher Assistent) at the Telecommunication Networks research group of Technical University Berlin. Since October 2004, he is a professor of computer science at University of Paderborn, heading the computer networks research group. His scientific interests include mobile and wireless communication, especially wireless sensor networks, architecture and protocols of future mobile systems, cross-layer optimization of protocols and performance evaluation of such systems.

Adam Wolisz obtained his degrees Dipl.- Ing. in Control Engineering, Dr.-Ing. and Habilitation (both in Computer Engineering) respectively in 1972, 1976 and 1983, all at the Silesian Technical University in Gliwice, Poland. Since 1972-1989 he was with the Institute of Complex Control Systems of the Polish Academy of Sciences working initially on real-time industrial computer systems and since 1980 on computer networks and distributed systems. From 1990-1993 he was with the Research Institute GMD-Fokus in Berlin heading activities on quantitative aspects of high-speed networks and multimedia systems. Since 1993 he has joined the Technical University Berlin (TUB) where he is chaired Professor of Electrical Engineering and Computer Science. He has founded and is leading the Telecommunication Networks Group (TKN) (see http://www.tkn.tu-berlin.de). Currently he is executive director of the Institute for Telecommunication Systems, grouping the activities in Communications, Networking and Distributed Systems. Effective July 1st, 2005 he has been additionally appointed Adjunct Professor at the EECS Department, University of California, Berkeley. His research interests are in architectures and protocols of communication networks as well as protocol engineering with impact on performance and Quality of Service aspects. Recently he is focusing mainly on mobile networking and sensor networks. He is author of over 150 papers in technical journals and Conference Proceedings.