

Mobility Management for Multicast Sessions in a UMTS-IP Converged Environment

Georgios Leoleis, Lila Dimopoulou, Vaggelis Nikas, Iakovos S. Venieris
National Technical University of Athens
School of Electrical and Computer Engineering
9 Heroon Polytechniou str, Athens, 157 73, Greece
E-mail: gleol, lila, enikas@telecom.ece.ntua.gr, ivenieri@cc.ece.ntua.gr

Abstract

The support of multicasting as a network layer service poses stringent requirements to next generation mobile networks. But for the fact that IP Multicasting has been considered in wireless networks, not until recently has there been a first attempt for the Multimedia Broadcast/Multicast Service (MBMS) support in UMTS. The investigation of MBMS architectural and functional design in respect to session and mobility management is the major focus of this paper. We approach mobility management support for idle- and connected- mode users with activated MBMS sessions as an extension to the existing mechanisms applicable to point-to-point communication. Moreover, we move one step further to the integration of the UMTS with the IP world, by replacing the UMTS core network with a fully IP-enabled network and propose mobility management mechanisms, in this integrated environment, for users receiving MBMS data.

1. Introduction

One of the basic promises of next generation wireless and mobile networks is the support of multimedia applications with stringent QoS requirements being uninterruptedly provided to a significant number of users on the move. Multicasting appears as a straightforward solution to facilitate one- or many-to-many delivery of packet switched services in a resource efficient manner. However, the support of multicasting in wireless environments is more complicated considering that the frequent change of the users' access point requires the dynamic reconstruction of the distribution tree. In IP-based wireless networks, the exploitation of Mobile IP mechanisms to handle mobility at network layer [1], [2], introduces two alternatives for the support of multicasting to mobile recipients. The first one assumes the tunneling of multicast data from the user's home network, while the

second proposes the from-scratch subscription of the user from its new access point. Both alternatives have many pros and cons related to the inherent problems coming along with Mobile IP [3] but also to the fact that no mechanism to enable uninterrupted service provision is foreseen.

In UMTS, access to IP multicast services is anticipated to be provided over the framework introduced by Packet-Switched (PS) Domain as already done for unicast ones [4]. However, no consideration has been made with respect to supporting network layer multicasting till Release 6 of the specifications. The definition of the MBMS architectural and functional design has been introduced to fill the gap of providing multi-destination PS services, addressing session and mobility management functionality. MBMS covers both broadcasting and multicasting by presenting two distinct modes of operation (only the latter is considered hereafter).

Going a step further, a beyond 3G architecture bringing the IP world closer to the UMTS is introduced in this paper. In this new environment, traditional mechanisms employed in the UMTS core network are replaced by pure IP mobility and multicast functionality. A more flexible and easily deployable IPv6 network comes to complement the UMTS access network by means of a Radio Access Support Node (RASN), serving as the first IP router to mobile terminals and also hiding the access network individualities from the IP network. The provision of multicast services to users moving across IP access points, i.e. RASNs, requires the orchestration of IP and UMTS mobility mechanisms along with multicast support at IP and UMTS bearer level.

The rest of the paper is structured as follows; section 2 provides the fundamental principles of the MBMS architecture, as presented in 3GPP specifications [5]. In section 3, core network level mobility procedures, applicable to users with activated MBMS services, are proposed and analyzed while the next section investigates

these procedures for a beyond UMTS network. Last, section 5 concludes the paper.

2. MBMS service Description

MBMS is defined as a network layer UMTS service that adopts a unidirectional (downstream) point-to-multipoint data delivery mechanism, featuring bearer-level multicasting for data transmission towards groups of mobile users. The service should not be viewed as the "IP multicast" counterpart for the UMTS network, since it is designed to tackle more complex issues, such as QoS (Quality of Service) and user mobility within the UMTS. However, interoperability between both services is deeply considered with MBMS advancing on efficient data transport and performing optimized management of radio and network resources.

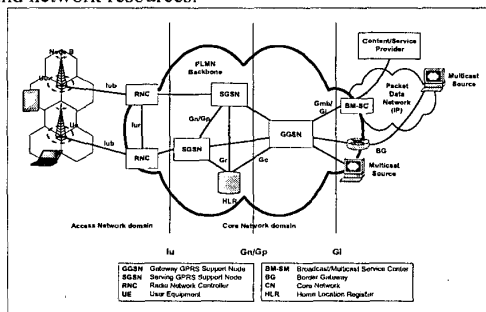


Figure 1. MBMS Reference Network Architecture

Point-to-point communication serves as a starting point for the definition of the MBMS service architecture. Thus, the minimization of modifications to already defined procedures and protocols, as illustrated in the reference architecture of Figure 1, is aimed. Focusing on the functionality assigned to each network node, the UE is responsible to request the activation or deactivation of particular MBMS bearers carrying multicast data and to listen for session start notifications indicating forthcoming transmissions. UTRAN (UMTS Terrestrial Radio Access Network) enables the delivery of MBMS data to terminals in a radio resource efficient manner relying on the utilization of common and dedicated radio channels for both user and control planes [6]. Access to network services is orchestrated by the Core Network (CN) nodes (SGSN and GGSN) managing user and service specific information. SGSN is in charge to perform MBMS bearer management, preserve mobility management contexts for its serving users and control authorization and accounting information. Furthermore, GGSN, being the network's MBMS data entry point, polices the incoming traffic based on QoS agreements and caters for the establishment of per service data tunnels carrying MBMS data to SGSNs.

MBMS data are provided by particular multicast sources or specific content providers located within or outside the network's administrative boundaries. A new network node, namely the Broadcast-Multicast Service Centre (BM-SC), is designated to control information flows exchanged with content/service providers of MBMS data and manage Authentication, Authorization and Accounting (AAA) information associated with third parties, through standardized application level interfaces (OSA; Open Service Access). Additionally, it provides service and user specific control information to GGSN over the introduced Gmb reference point, manages the QoS of incoming MBMS data flows, performs error resilience tasks and administers service advertisement and discovery mechanisms.

In principle, access to PS domain services is supported by establishing network bearers with pre-defined QoS attributes on a per session and user basis. With the introduction of MBMS, data transfer over the so-called MBMS bearers is performed for multiple users on a per service basis only. The following list, presents the phases constituting a complete MBMS service provision procedure (more information is available in [5]):

- Subscription
- Service Announcement
- Joining – Bearer Activation
- Session Start
- Notification
- Data Transfer
- Session Stop
- Leaving – Bearer Deactivation

2.1. MBMS Contexts

Multicasting features a many-to-many association among users and utilized bearers/services (there is a one-to-one association between provided services and activated end-to-end bearers). Namely a single bearer may be activated by many users camping on the same cell or more likely served by the same RNCs/GSNs while a single user may potentially join multiple multicast groups by activating multiple MBMS bearers. This complexity necessitates the disassociation of service and user information within the network. The definition of MBMS contexts as containers of MBMS service- and user-related data provides the appropriate information segregation required for the MBMS session and mobility management. MBMS presents two distinct contexts for this purpose; the UE Context and the Bearer/Service Context.

MBMS contexts are exchanged among network nodes that require service and user control information for managing the MBMS bearers. A non-exhaustive list of the parameters contained in each context type is presented

in Table 1. Special emphasis should be put on the 'State' attribute. Once the MBMS Bearer is activated, and given that there is no data transfer in progress, the Bearer Context state is set to *Standby*, meaning also that no radio or network resources are allocated. When data transfer is ongoing, the bearer state changes to *Active* initiating the respective reservation of resources. The execution of the Session Start and Session Stop procedures sets up the triggers for setting the bearer's state from *Active* to *Standby* and vice-versa.

Table 1. UE and Bearer Information Contexts

Context Parameter	Type
<i>IP multicast address</i>	Both
<i>Access Point Name</i> (on which the IP multicast address is defined)	Both
<i>Temporary Mobile Group Identity</i> (TMGI, assigned for the service/bearer)	Both
<i>UE NSAPI</i> (for the UE PDP context used to carry group membership signaling)	UE
Activation State of the MBMS bearer ('Active' or 'Standby')	Bearer
QoS of the MBMS bearer	Bearer
Area over which MBMS is provided	Bearer
List of nodes where the bearer is activated or notifications/data are forwarded	Bearer
Number of UEs hosted by a node and have joined an MBMS	Bearer

The MBMS UE context for each user is created during the execution of the Joining procedure and deleted after Leaving. On the other hand, the MBMS Bearer Context is created when a user requests the first activation of an MBMS service and is deleted when the last user being a member of the group leaves it. In case where an MBMS Bearer Context is not existent in a node, a particular upstream Registration procedure is executed resulting in the extension of the distribution tree to include the node originated the request (by communicating the Bearer Context). If the MBMS Bearer Context is available, only the association (linking) of it with the corresponding UE Context is needed. Respectively, a node triggered by a user-initiated service deactivation or a mobility event, may request its parent serving node to remove it from the "list of downstream nodes" (Deregistration) maintained in its Bearer Context.

3. UMTS Extensions for MBMS

Mobility Management (MM) support for MBMS relies on existing UMTS mechanisms specified for the mobility handling of a user participating in unicast services. Furthermore, it takes into account the QoS nature of a multicast bearer along with its one-to-many relation to users registered for the multicast service. Special focus is given on how mobility is handled within the core

network, i.e. when the movement of a user affects the CN data path of the established distribution tree (inter-SGSN movement). Mobility management procedures are executed while the mobile station is in either (connected or idle) GPRS MM state [7]. In the former case, the user is participating in sessions and keeps the network aware of its exact location (serving RNC) while in the latter, it only provides coarse-grained information (in the granularity of a routing area) to the network concerning its whereabouts. However, the possibility of an idle user receiving multicast data is currently being specified within 3GPP [5]. Even in this case, the described MM procedures are not affected since the handling/update of the MBMS bearer in the CN is always performed and is independent of the respective functionality in the UTRAN.

3.1. Idle Mode

The UMTS Access Network is organized in Routing Areas (RA -set of RNCs under a single SGSN-), identified by a Routing Area Identifier (RAI), in order to differentiate the MM procedure for idle users (power consumption reasons) and allowing them to provide location information in a larger time scale. The RAI information is communicated to the user's serving SGSN via the Routing Area Update (RAU) procedure, and is used for paging the mobile station when there are data addressed to the latter but no PS connection is available. A mobile station, when in idle mode, listens to broadcast beacons transmitted by base stations, communicating the respective RAI. Upon detection of a change in its RA (either under the same or different SGSN), the user informs its serving SGSN via the RAU procedure. This procedure is performed periodically, each time a corresponding timer expires at the mobile station. In this way, RAU ensures that the user is always reachable by the SGSN, when there appears a need for it.

The RAU procedure (see [6]) is extended with the required functionality for handling the mobility of an idle user having activated an MBMS service, meaning that MBMS UE Context is already established at the SGSN and GGSN along with the MBMS Bearer Context for that service. As aforementioned, data transfer is not in progress and therefore we are only concerned with establishing/updating multicast-related contexts in CN nodes. Thus, context transfer of the MBMS UE information - along with unicast SM information if any - towards the new SGSN shall be performed (SGSN Context Response). This will result in the new SGSN updating the respective MBMS UE context in the GGSN. The establishment/removal, however, of UE context from CN nodes is mapped to MBMS bearer-level procedures, which handle the potential changes in the multicast

distribution tree. Therefore, if the new SGSN detects that it is not part of the corresponding multicast bearer (i.e. it maintains no bearer context for the service), it requests it from the GGSN (MBMS Registration Request). The GGSN is resolved from the APN carried in the MBMS UE context. This results in the GGSN updating its list of downstream nodes (Table 1) with the new SGSN. Likewise, when the old SGSN detects that it no longer serves users for an MBMS service, it sends an MBMS Deregistration Request for being removed from the multicast bearer. However, if these two conditions are not met, then no extra signaling is required for updating the MBMS bearer(s) and consequently the multicast distribution tree within the CN.

3.2. Connected Mode

The basic concept adopted within 3GPP for point-to-point communication is that a temporary GTP tunnel is established between the source and target RNCs, over the Iu interface, for the downlink forwarding of data, till the full path is reconstructed. The relocation procedure takes care for a lossless handoff, by allowing the buffering of diverted packets to the target RNC and moreover the synchronization of unacknowledged PDCP PDUs (for both uplink and downlink) between the UE and the target RNC. However, this solution should be differentiated for the case of MBMS since more parameters come into play. Firstly, it is assumed that MBMS services are more tolerant to packet loss although sensitive to delay (service disruption). Secondly, it is possible that data are already available in the new path (or part of it), if other users registered for the same MBMS service are located there.

The relocation procedure needs to respond to the new requirements imposed by the multicast sessions. As illustrated in Figure 2, the procedure starts with the decision to perform SRNS relocation for a particular user. The request for relocation is forwarded from the old SGSN to the new one carrying the user's MBMS and PDP contexts. The target RNC is next instructed to start the relocation of the unicast Iu bearers (GTP tunnels between RNC-SGSN) via the Relocation Request. The latter may as well serve the trigger for the update of the multicast distribution tree and also the reservation of resources, if needed. Actually, it is possible that none, only the new SGSN, or both the new SGSN and target RNC are part of the multicast bearer. In the first case, procedures 3 and 4 are carried out while in the rest cases these procedures are partially or entirely skipped.

At this point, the MBMS Bearer Context is in Active state and data distribution already follows the path towards the target RNC. At the completion of step 5 (unicast case), data are already directed over the temporary GTP tunnel towards the target RNC, where

they are buffered (for lossless RABs) and finally delivered after the relocation of the radio bearer is performed (step 6). However, we argue that this tunnel is not needed for the MBMS services case, since MBMS data are more tolerant to packet loss and in addition to that, they already follow the new path (after procedure 4). Certainly, the mobile station needs to reconfigure itself (i.e. reset the relevant protocol entities) for being able to receive data from the target RNC, which has assumed the serving RNC role after step 5. The relocation complete signal triggers the update of the MBMS UE and PDP contexts, resulting in the switch of the user plane (for unicast sessions) from the old to the new SGSN. At this point, the release of the unicast Iu bearers as well as the conditional pruning of the multicast distribution tree (via procedures 10 and 11) may start. Last, the new SGSN and HLR are informed of the user's new RA and currently Serving RNC respectively.

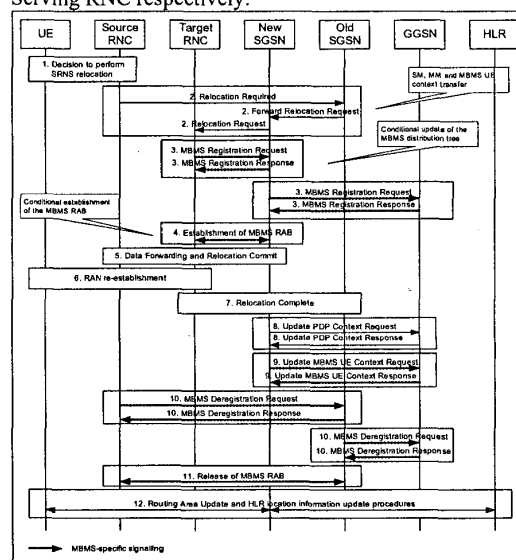


Figure 2. MBMS-enabled Serving RNC Relocation procedure

4. MBMS Support in a Converged UMTS-IP Environment

As afore-described, mobility management at UMTS CN level is accomplished by means of the traditional GTP tunneling across the GPRS support nodes. Going a step further to the evolution of the UMTS Core Network towards a fully IP-based mobile-aware network, a hybrid mobility management model based on both UMTS and Mobile IPv6 (MIPv6) mechanisms is proposed. More succinctly, the replacement of SGSNs by special nodes

situated at the border between the RAN and the CN is performed in order to support the inter-connection of the UMTS protocols with MIPv6 [1]. These nodes are referred to as Radio Access Support Nodes (RASNs) and are specified in the framework of the IST SAILOR project [8]. It is noted that the concept of incorporating MIP functionality in the UMTS CN but also the integration of GSNs into a single node has also been addressed in [9]. In this regard, mobility management in the Access Network is based on legacy UMTS protocols, while Core Network mobility is handled with MIPv6, resulting in MIP as the only mobility management signaling protocol in the Core Network at bearer level. The benefits are multiple; the access network mobility mechanisms are reliable and adapted for fast handoffs in order to support the frequent mobility events at this level. The interworking with IP-based external networks is eased due to the fully IP-compatible Core Network. Further, since Mobile IP substitutes GTP, the mobility management complexity in the Core Network is significantly lessened, the network deployment is facilitated due to the use of simpler IPv6 routers instead of GSNs, and the overhead of packets is also reduced (IP packets are no longer carried over a GTP/UDP/IP header).

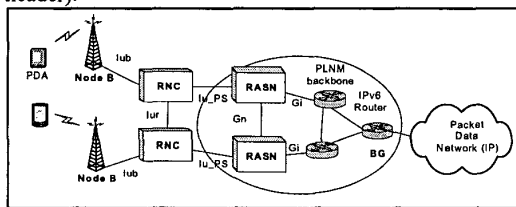


Figure 3. Target network architecture

The RASN's role is twofold; on the one hand, it provides radio access unawareness to the fully IP-fashioned Core Network by completely hiding the specific aspects of the radio access technology, and on the other, it terminates the access stratum and the non-access stratum UMTS protocols. The target is to develop a system able to support IP multicast connectivity to mobile users over a UMTS-capable access network, by incorporating already developed UMTS mechanisms along with the MBMS extensions introduced above, and yet MIPv6 and IP multicasting functionality introducing in this way a complete solution, concerning global mobility and unicast or multicast session management aspects.

4.1. Mobility Management Procedures

Mobility events are discriminated in two main categories: those that take place under the same RASN and those that demand RASN relocation. In the first case,

mobility is handled with the existing UMTS mechanisms provided by the access network, without any implication of IP mobility mechanisms. Since the mobile terminal remains under the control of the same RASN, it is served by the same multicast router and, therefore, no reconstruction of the multicast spanning tree is needed. The RASN assumes the role of the SGSN from the access network's point of view, and handles the mobility of the user by means of the standard Iu_PS protocols. As a consequence, the intra-RASN Routing Area Update and Serving-RNC relocation procedures are completely transparent to CN elements other than the serving-RASN.

Concerning the mobility events involving two RASNs, both UMTS and Mobile IPv6 MM mechanisms are used, in order to enable seamless inter-RASN handoff. The basic concept adopted is that the whole procedure is comprised of two sequentially executed main stages, i.e. a layer 2 and a layer 3 handover. During the layer 2 handover, the Iu bearers are relocated, resulting in reconstructing a part of the path (from the new RASN to the target RNC). In the framework of this stage context transfer between the old and the new RASN and establishment of a short time lived GTP tunnel among them, eliminating packet loss until the layer 3 handoff is completed, is needed to be carried out. It is underlined that the layer 2 handover is reusing the same UMTS mechanisms presented in the previous section, in regards to the Iu bearer relocation and the conditional establishment/release of the MBMS bearer.

During the layer 3 handover, the reconstruction of the distribution tree allows the mobile station to directly join the multicast group through the new RASN, by issuing an MLD -Multicast Listener Discovery- "Unsolicited Join Report". Since RASN assumes the role of the multicast router in its serving domain, it can become part of the multicast distribution tree. This method, denoted as 'remote subscription mechanism' [1], is founded on the fact that the Care-of Address (CoA) is used as the source address for packets emanated from the Mobile Node.

In Figure 4, the inter-RASN relocation procedure is illustrated, giving special focus on the enhancements made to the native UMTS approach. When stage 1 is completed, the path from the new RASN towards the mobile station is restored and all relative procedures, such as the conditional establishment and release of the MBMS bearers, the relocation of the unicast Iu bearers and the context transfer between the two RASNs have been performed. However, it is possible for the new RASN not to be part of the IP multicast spanning tree. Until that takes place, a short time lived tunnel between the old and the new RASN is established, so that the former will forward incoming MBMS traffic to the latter. The exploitation of already developed UMTS mechanisms for the establishment of a GTP tunnel

between the involved RASNs may be performed. The tunnel is set up, after context transfer is performed, by means of an MBMS Registration Request message issued by the new RASN (playing the new SGSN's role) towards the old RASN (assuming the GGSN's role). Another option is the establishment of an IP-in-IP unidirectional tunnel. Moreover, context transfer between the two RASNs can alternatively be performed at the generic IP layer, where an abstraction of the AN-specific mechanisms is feasible, enabling in this way mobility across heterogeneous access networks. However, these procedures have not yet been standardized within IETF and, therefore, the respective UMTS mechanism is selected.

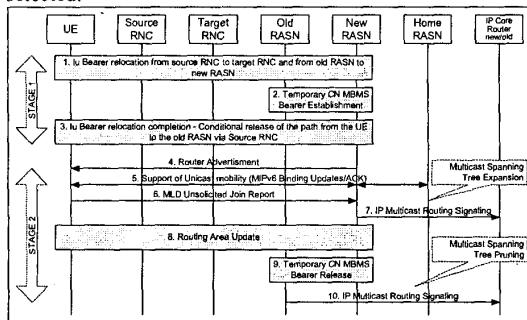


Figure 4. RASN relocation in the target architecture

Stage 2 is executed so that the distribution tree is reconstructed. The new RASN is triggered after the RNC relocation finalization for sending a MIPv6 Router Advertisement message to the mobile node. The latter realizes that it has entered a new IP subnet and initiates the standard IPv6 procedures to obtain a new CoA (acquired CoA either by stateful auto-configuration, using the DHCPv6 protocol, or by stateless auto-configuration by means of the Neighbor Discovery mechanism). The obtained CoA, is used in an unsolicited group membership report, issued by the mobile station to the IP Multicast Listener Protocol module (the MLD in our case) of the new RASN. The arrival of the Join report triggers the new RASN to signal its designated IP core multicast router in order to become member of the multicast spanning tree. The procedure is completed by performing a Routing Area Update. Once the new RASN has become member of the multicast tree, the release of the temporary tunnel between the two RASNs via an MBMS Deregistration procedure may be performed. Incoming data for the respective multicast address or the protocol functionality may serve the trigger for the tunnel release. In case the mobile station was the last terminal receiving MBMS traffic at the old RASN, the latter initiates the appropriate actions, in order to prune itself from the tree. Alternatively, the RASN may take no

specific action and just wait for a timeout in the IP multicast listener protocol module for the particular group membership.

5. Conclusions and Future Steps

In this paper we have described the principles of the multicast mode MBMS architectural design as presented in specifications and sketched the basic concepts associated with the establishment of network level group communication in mobile networks. Basing on UMTS procedures for point-to-point communication, we presented a number of appropriate extensions to enable mobility management of MBMS sessions by incorporating the peculiarities associated with multicast data and the concepts followed in the MBMS framework. Furthermore, we have presented a UMTS-IP converged architecture, obviating the need for the RASN node, situated in the border of a fully IP-enabled CN and a UMTS AN. Our driving force has been the replacement of traditional and heavy UMTS protocols between GSNs with a lighter IP network, handling mobility and multicast support at the IP level. Inter-RASN mobility to multicast sessions is provided through the specified RNS relocation procedure enhanced with the establishment of a short time tunnel for preventing packet loss till the full reconstruction of the multicast distribution tree. The benefits introduced by the proposed architecture are straightforward: use of native IP multicast inside the CN instead of carrying multicast data over GTP tunnels; easy deployment of the core network with simple IPv6 routers; CN independence of the radio access technology thanks to the RASN node; IP mobility in the CN instead of having the GGSN as an anchor point.

6. References

- [1] David B. Johnson et al., "Mobility Support in IPv6," Internet draft, draft-ietf-mobileip-ipv6-21.txt, Feb. 2003.
- [2] H. Gossain et al., "Multicast: Wired to Wireless," *IEEE Comm. Magazine*, vol. 40, no. 6, June 2002, pp. 116-123.
- [3] Christian Bettstetter et al., "Interoperation of Mobile IPv6 and PIM-DM", in proc. of IPDPS2000, Aug. 2000.
- [4] 3GPP TS 29.061 v5.7.0, "Interworking between the PLMN supporting Packet Services and PDN," Sept. 2003.
- [5] 3GPP TS 23.246 v6.0.0, "Multimedia Broadcast/ Multicast Service (MBMS); Architecture and Functional Description," Sept. 2003.
- [6] 3GPP TS 25.346 v2.4.0, "Introduction of the MBMS in the Radio Access Network; Stage 2," Nov. 2003.
- [7] 3GPP TS 23.060 v6.2.0, "General Packet Radio Service (GPRS); Service description; Stage 2," Sept. 2003.
- [8] SAILOR web site, URL: <http://www.ebanet.it/Sailor.htm>
- [9] 3GPP TR 23.923 v3.0.0, "Combined GSM and Mobile IP Mobility Handling in UMTS IP CN," May 2000.