Support of Multicast Services in 3GPP

Th. Lohmar, H. Wiemann, F. Hundscheidt, M. Meyer and R. Keller
Ericsson Research, Corporate Unit

***Abstract*— In line with the growing interest in efficient data distribution services to large mobile user groups, 3GPP is standardizing the introduction of support for multicast services into GSM and WCDMA networks. The mobile infrastructure needs to be enhanced in the wireless and in the wireline parts to support new signaling and payload transport procedures. Also mobility, charging and radio resource handling aspects have to be considered. Starting from group distribution services and their requirements, this paper motivates and describes the solutions defined and discussed within 3GPP, namely the introduction of new nodes, new bearers and new procedures to efficiently support multicast services.**

***Index Terms*—data communication, mobile communication, multi-user channels.**

I. Introduction

Since the introduction of packet data services in 3GPP (GSM and WCDMA), mobile users are always connected and reachable via their mobile devices, meaning that they are able to receive content at any time. A considerable and growing part of the content distribution services in mobile networks addresses groups of users with similar interests (e.g. sports clips, news flashes). As long as the group sizes are small and the densities are low, unicast channels can be used to distribute the content. However, the sizes of the groups are growing. In addition there is a growing tendency to require more bandwidth per content due to the shift of text to picture and video content. This implies that more efficient solutions are required to avoid that service provisioning is either not possible or too expensive for the network provider. Such solutions should use the inherent broadcast characteristic of radio signals and distribute the content only once to all users simultaneously (similar to terrestrial or satellite distribution of TV and radio broadcasts).

The IP multicast framework in IETF has defined mechanisms to efficiently distribute content to multiple users simultaneously by incorporating enhancements into the mobile infrastructure equipment. 3GPP is currently introducing support for IP multicast services into the GSM and WCDMA architecture and standardizing the necessary adaptations to cope with the particularities of mobile systems. The corresponding work item is called Multimedia Broadcast and Multicast Services (MBMS) and is covered in release 6.

This paper provides a tutorial style overview on the concepts behind MBMS. Chapter II describes the mobile group distribution services and their requirements on the content delivery. In section III, the existing IP multicast framework is introduced in order to explain the required adaptations for mobile networks in chapter IV. Chapter V elaborates on the MBMS concepts and describes the different solutions standardized and discussed within 3GPP. Finally, section VI concludes this paper and gives an outlook to future work.

II. Services and Requirements

The number and variety of content distribution services to user groups is increasing and most mobile operators today offer one or several of these in their service portfolio. In addition to download and SMS, these services are increasingly based on MMS and in the near future on wireless streaming technology. Typically, the services are accessible through the operator's WAP portal, where the mobile user can register and by that express an interest in certain information to get updated whenever new info is available. Examples are the Formula-1 racing results and football scores services.

The actual content distribution to the clients is done via unicast transmission, i.e., the network processes multiple requests for the same content sequentially. Such delivery wastes network resources and often results in (too) late content reception for some of the users. Especially for time-critical content (e.g. news, stock tickers), the value of the information will quickly diminish over time.

Besides the time-critical content services a new set of services becomes feasible and attractive. Background distribution of large files (e.g. mp3-music to update the "Top 50 Chart List" in the mobile phone) – which used to be too expensive for wireless distribution – becomes attractive for operators when a single file transfer can be used to provide the info too many users simultaneously. Furthermore, video-clip distribution services could entertain many mobile users while waiting for the train.

III. State of the Art: IP Multicast

Since the above-described services are typically IP based and IP has been introduced into 3GPP networks, it is only natural to reuse existing IP technologies IP Multicast can be subdivided into group management and efficient data delivery.

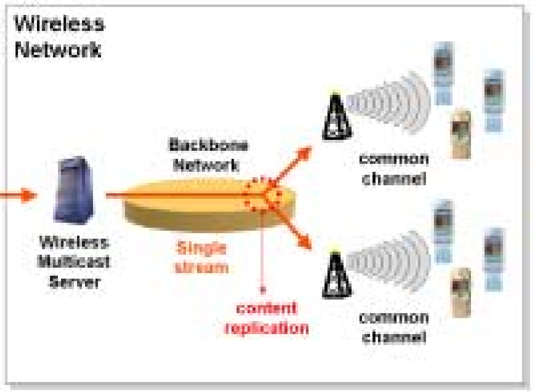

Figure 1: Multicast in Wireless Networks

A. *Group Management*

IP Multicast is a receiver-based concept, i.e. receivers join a particular multicast group by informing the multicast router on their sub-network. . A join request contains an IP multicast class D address (IPv4) identifying the multicast group. Multicast routers mark the groups for which there is at least one member within their subnetwork and periodically query the hosts in the subnetwork to update the membership. .

Membership reporting (i.e. the initial registration and the response to membership queries) is done with the Internet Group Management Protocol (IGMP) [7] in IPv4 and with Multicast Listener Discovery (MLD) [8] in IPv6. The membership reports are sent to the multicast group. Hence, not only the multicast router but also all other clients are informed that one host has already reported its membership and refrain from reporting their membership too.

B. *Efficient Data Delivery*

Neither senders nor receivers need to care about the network topology, as intermediate nodes (multicast routers and L2 switches) are responsible for data delivery. The data delivery to multiple users is optimized as a single copy of each data item is sent on a link and the replication of the data items is done as close as possible to the final destinations. This is achieved by means of multicast delivery trees, which are based on spanning trees and established and maintained by dedicated protocols like Protocol Independent Multicast (PIM) [9].

IP multicast provides best effort, unreliable data distribution mechanisms resulting in occasional packet drops. If UDP is used as transport protocol on top of IP these errors are not recovered and propagate to the application. Most audio-/video-streaming applications are able to cope with a certain packet error rate. Typically, the user perceived quality decreases with increasing number of packet losses.

However, many other applications like file download, chat or electronic whiteboard require reliable end-to-end data transmission schemes on top of IP Multicast. Those reliable multicast protocols are based on ARQ and/or forward error correction. ARQ based mechanism in particular require careful design to avoid system overload in the reverse direction (uplink).

IV. IP Multicast in Wireless Networks

The key motivation for integrating IP Multicast into mobile communication systems is to enable efficient group related data distribution services, especially on the radio interface. The basic idea is reflected in figure 1 above where the content is replicated as close as possible to the radio interface, where a common channel is used to deliver the content to all clients within the same cell.

As 3GPP networks already have a standardized and deployed architecture and mechanisms, special adaptations are required to integrate IP multicast. The main differences and corresponding adaptations are listed below:

A. *Wireless Multicast Server*

A Wireless Multicast Server is located at the edge of the wireless network and fulfills the role of the local multicast router. It is the last hop router on the user IP level although several other nodes are involved on the transport IP level. This is different from IP multicast in Ethernet based local area networks (LAN) where all end-nodes are connected via a flat and simple link layer.

The Wireless Multicast Server shall also receive unicast traffic and forward it on multicast through the mobile system. 3GPP specific protocols are used to maintain the multicast delivery tree within the mobile system.

B. *Efficient use of radio resources*

The radio resources are scarce and must be used in a most efficient way. In a LAN environment multicast traffic is broadcast and it is up to the hosts to fetch the right frames. Unnecessary broadcasting in wireless networks must be avoided and transmission resources should be allocated only when necessary. E.g. instead of regular group member query messages, the wireless multicast routers rely on group “leave” messages. Also uplink traffic must be minimized in order to avoid system overload. In particular synchronized uplink messages from several members (e.g. after session announcements) can be critical.

C. *Charging and security*

Operators of IMT-2000 networks want to charge for the multi-user service. In order to be able to charge each group member for the ser-vice, all clients must register to the multicast group. In contrast to the best effort multicast delivery in the Internet, multicast delivery in wireless network must be reliable. When charging for the service, operators must know whether a specific group member received the content. Because of the registration and charging, there is also a need for security mechanisms to avoid fraudulent usage of multi-user services.

V. System Concepts for MBMS

This section provides an introduction into the system concepts of Multimedia Broadcast and Multicast Services (MBMS). It presents and discusses how the existing IP Multicast framework is embedded into the mobile network architecture.

With 3GPP release 6, the requirement of radio and network efficient transmission of the same data to multiple users was introduced into the specifications, resulting in the work-item MBMS.

The mobile network is extended with functionality to forward multicast and broadcast traffic via the existing infrastructure and uses IP Multicast schemes for service activation and data delivery. The overall ambition is to support efficient transmission of group related traffic even on the radio interface with data rates up to 256 kbps.
3GPP differentiates between two point-to-multipoint modes:

- Broadcast Mode: data is transmitted to all users in a broadcast area
- Multicast Mode: data is transmitted to a group of users which have subscribed to become member of that group

The main difference between the two modes is, that the network shall know all participants in the Multicast Mode. In case of the Broadcast Mode, the network does not know the individual receivers.
The remainder of this section deals with the impacts the introduction of MBMS has on the system architecture. In addition, we describe in detail how MBMS is supposed to work and finally which issues need to be considered for the design of the MBMS radio access network realization.

A. *MBMS architecture and MBMS specific procedures*

The reference architecture for MBMS is depicted in Figure 4. MBMS is applicable only to the packet-switched sub-system of the mobile network. Although we concentrate here on WCDMA, both, GERAN and UTRAN are supposed to handle MBMS traffic. The picture shows besides the packet switching support nodes also the new Broadcast-Multicast-Service Center (BM-SC). This node plays a central role in the entire architecture and the associated procedures. It authenticates the content providers, which want to transmit content via MBMS bearers and it generates charging records either for the operator or for the content providers or for both.

The new Gmb interface is used to exchange control plane messages between the BM-SC and the GGSN. The BM-SC receives the join request from the clients via this interface and triggers the membership handling related procedures. It also provides the QoS attributes like bandwidth range via the Gmb interface to the GGSN.
Figure 2 shows the protocol stack for MBMS. It reveals the scope and the layering of the involved protocols. The IGMP protocol is terminated in the user equipment and the GGSN. In mobile networks, intermediate nodes between the UE and the GGSN do not process user traffic. All intermediate nodes like RNCs and SGSNs simply forward join messages without processing to the GGSN, where IGMP messages are identified. This approach leads to different membership handling procedures compared to the IP Multicast model.

The group membership information is then propagated via the SGSN into the radio network. This requires modifications of the GTP protocol on the Gn and the Iu interface. Before MBMS, GTP provided only mechanisms to handle unicast packet flows. Therefore, GTP needs to be extended to provide

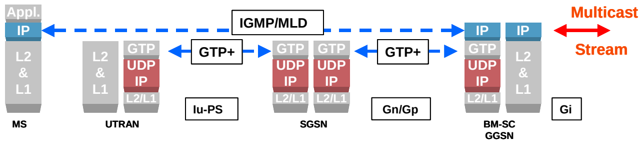

Figure 2: 3GPP Protocol Stack

group membership information to SGSNs and RNCs and to handle IP Multicast packets.

The basic principle of Multicasting is to transmit only one single copy of the packets to the entire group and replicate the packets in multicast capable routers (actually the servers) close to the receivers. In case of 3G systems, the GGSN, SGSN and the RNCs are supposed to replicate multicast packets.

B. *Split of Control and User Plane*

3G systems include a clear distinction between control and user plane. This principle is also applied to the new MBMS Multicast and Broadcast Modes. Transmission resources are allocated only when content is about to be transmitted. But since MBMS is a group communication extension, there are differences to the handling of ordinary bearers like streaming or interactive bearers. Parts of the control and user plane procedures are entire group related whereby others are individual user related.

Also, in MBMS there is a clear split between user-related procedures like mobility and membership handling and the server specific procedures like start sending content on the MBMS bearer. User and content server related procedures are independent of each other and can be executed even in parallel.

The main difference of MBMS procedures compared to the legacy IP Multicast model is, that the content provider must initiate the establishment of the user plane. Main reason for this is the necessary terminal paging procedure and the allocation of radio resources. Mobile terminals are only known on routing area level to the network. First the position of the mobile terminal is determined on cell level, before actual transmission resources are assigned. A detailed description of the activation procedure and the establishment of the user plane is given in section x.x.

In the IP Multicast framework (see chapter 3), the multicast distribution tree is kept up to date, that it is always possible for a multicast sender to simply start sending data. The forwarding and replication states are maintained and all routers

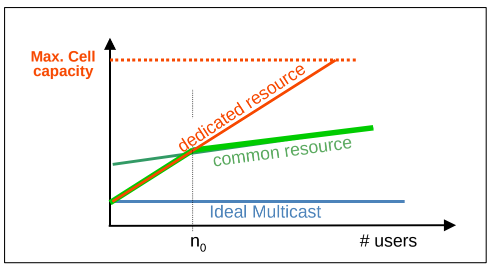

Figure 3: Resource Usage

are prepared to replicate incoming IP multicast packets according to the tree structure.

C. Group Membership and Mobility Handling

With the introduction of MBMS, the mobile network has to handle group communication issues on network layer and below. Since only one flow of packets for each group is sent through the network and is replicated at each branch, the network needs to know about group members for each branch. Traffic shall only be forwarded on branches leading to group members.

An important aspect of the membership handling is keeping track of moving members. Members change their point of attachment to the network and hence network branches and serving nodes. To ensure that all members are being reached, the membership information in each node must be up to date, when the user plane is established.

Group Membership information is maintained in each intermediate node on the transmission path. The information is stored in MBMS specific contexts, which are located in the BM-SC, the GGSN, SGSN and the RNC. Also the UE needs to store membership and lower layer specific information locally.

Two new contexts are introduced with MBMS: A MBMS Bearer context and a MBMS UE context. The latter is kept together with the mobility management contexts in the serving nodes. One MBMS UE context exists for each MBMS group that each terminal has subscribed to. It contains the IP Multicast address and Access Point Name (APN) identifying the used MBMS bearer.

Beside the IP Multicast address and the APN the MBMS Bearer context also contains a Temporary Mobile Group Identity (TMGI), the state of the bearer, QoS attributes, the list of the downstream nodes and the UEs hosted by this particular node.

The state parameter reflects the state of the user plane. In case of ‘active’, the user plane is established for this particular MBMS group and transmission resources are allocated. After establishing the control plane the state of the bearer is ‘standby’.

D. Data Transfer

In the MBMS architecture, the content provider must initiate via the BM-SC the user plane establishment process with a session start message. The content provider also indicates the end of the transmission so that the transmission resources can be released. With the session start message, the BM-SC provides quality of service attributes for the bearer and the multicast tree is created in the mobile network. A list of downstream nodes is stored in the MBMS Bearer Context. Note that the currently used cells of the group members are in general not known in the network.

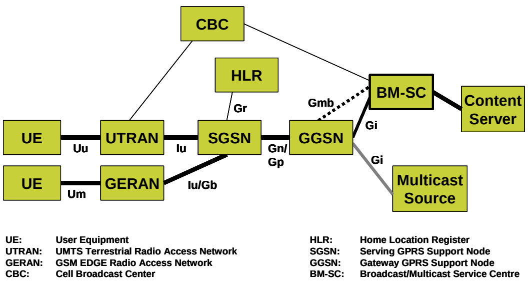

Figure 4: MBMS reference architecture

The radio link between base station and mobile terminal is the most expensive resource in a cellular network and resources are assigned very carefully in particular on that interface. Therefore, it is obvious to take advantage of the broadcast nature of radio waves to transmit to all potential receivers in a certain coverage area. Theoretically, many terminals can listen to the same radio signal and receive the same data without any additional overhead or costs. However, analysis has shown that a WCDMA common radio bearer like the forward access channel (FACH) requires more radio resources than a dedicated bearer at the same data rate and loss ratio, though being able to serve more than one user (see below for details). In other words, few dedicated bearers might outperform one common channel in terms of resource efficiency. This dependency is also depicted in Figure 3. It shows schematically the resource usage over the number of users. In case the number of users is below the threshold n_0, it is more efficient to serve all members with point-to-point bearers (dedicated resources). In case the number of users is larger than n_0, it is more efficient to use common radio resources for the data transmission. The gradient of the curves and the threshold are still a topic for further investigations.

In case of IP multicast in fixed access networks, the multicast traffic is broadcasted on the LAN segment if at least one member is present. Since the reception conditions of the connected hosts are similar, there is no need to keep track of the number of users. Consequently, the local multicast router host only monitors whether at least one member is present.

The requirement for radio resource efficiency for MBMS requires a more complex approach. At the time of activation of the MBMS user plane, the network decides to use common or dedicated radio resources. Common radio resources are used to establish a point-to-multipoint (p-t-m) bearer using e.g. the forward access channel (FACH). Dedicated resources (DCH) are applied for point-to-point (p-t-p) bearers if that is more efficient.

A new group-paging scheme is used to count the number of members for the activated MBMS bearer in each cell. The normal paging scheme cannot be used, since it requests the establishment of a Radio Resource Control (RRC) Connection from all paged terminals. This could lead to an overload situation on the Radio Access Channel (RACH). If a large number of UEs belonging to the same MBMS group camp on the same cell and if they access the channel in the same transmission time interval, then collisions occur, because the RACH is operated in a Slotted ALOHA fashion.

Consequently a more scalable concept is required. One possible approach is that the RNC does not urge all terminals to respond to the paging request. Instead, the terminals shall respond only with a certain probability to the paging message. The RNC counts the number of responses and stops, as soon as the threshold to use common radio resources is reached. The RNC only needs to know, if the necessary number of members are located in a particular cell.

E. Using common downlink radio resources

In this section we focus on using a point to multipoint channel in the downlink direction. In particular some issues are discussed, which need to be considered for the design of a WCDMA Radio Access Network.

In particular techniques to minimize the transmission power for MBMS transmissions are discussed. In the end of this section, a comparison of the strategies to achieve an error free data delivery over the air-interface is given.

1) Characteristics of the point-to-multipoint physical channel

The WCDMA Layer 1 offers two basic types of physical channels, common and dedicated. Dedicated channels are always bi-directional even if application and transport protocol generate unidirectional traffic only. The uplink direction is mainly used for fast power control schemes, which results in a nearly constant received signal to interference ratio and block error rate. Furthermore, it minimizes the transmit power consumption.

If fast power control is used for p-t-m bearers the uplink control load would increase with the number of receivers and could easily exceed the amount of downlink traffic. Therefore, it has been agreed in 3GPP to use a unidirectional common channel for p-t-m MBMS. As a consequence, the transmit power cannot be controlled as efficiently as on dedicated channels. Either a very conservative and fixed transmit power level could be defined for a p-t-m channel or it could be adapted based on (infrequent) measurement reports sent e.g. only by those terminals perceiving an unacceptable high block error rate.

Another feature optimizing transmit power requirements and quality of service is soft handover. A rather large fraction of all terminals with active p-t-p connections are typically in soft handover state, i.e., they are served by more than one cell at a time. Theoretically, soft handover could also be used on p-t-m common channels but it would require significant changes in base stations and mobiles like full site-synchronization and larger receiver buffers. However, the expected gains do not justify the necessary efforts. Therefore, soft-handover is currently not seriously considered for MBMS.

In order to compensate for the increased transmit power requirements due to the lack of soft handover and fast power control the duration of a Transmission Time interval (TTI) can be increased from typically 10 ms to 80 ms. All Layer-2 packet data units (PDUs) in one TTI are encoded in one Turbo Code Block. Coding and interleaving gain increase significantly with the length of the Turbo Code Block at the cost of a slightly increased end-to-end delay. However, the applications and services described in section II do not suffer from a higher latency. Simulations have shown that the p-t-m forward access channel can provide sufficient data rates consuming a reasonable fraction of the overall transmit power.

2) Link Layer error recovery schemes

The WCDMA Radio Link Control layer (RLC, Layer 2) can be operated in Acknowledged, Unacknowledged and Transparent mode.

Due to the uplink usage of the Acknowledged Mode, it has been decided in 3GPP to operate the radio link layer in unacknowledged mode. The RLC receiver identifies and discards erroneously received packet data units (PDU) and does not request any retransmissions from the sender. However, the QoS that can be expected from this bearer is low or at least not fully predictable. The block error rate provided by the physical layer typically varies in a wide range depending on the individual link conditions. In particular, receivers with poor radio conditions will experience significantly higher loss rates or long bursts of packet losses. In addition, link layer errors typically propagate to an even higher loss rate on IP and application layer.

Hence, some additional improvements of the MBMS bearer in terms of residual loss rate are necessary. One proposal is to use Layer-2 Outer Coding based on erasure correcting codes. After segmenting IP packets into k Link Layer PDUs, (n-k) parity packets are generated and sent in addition to the actual data. The receiving link layer entity can reconstruct the original data if at least k of n transmitted PDUs is received successfully. The additional layer-2 parity bits obviously reduce the overall throughput but also the number of residual packet losses. It is for further study, if the QoS provided by this link layer is sufficient for all applications listed in section II or if additional end-to-end mechanism e.g. for error recovery are required.

VI. Conclusion and Outlook

The current MBMS specifications and discussions in 3GPP aim at introducing new radio efficient multicast and broadcast bearers to the GSM and WCDMA mobile networks. Distribution services can be provided to large user groups even in densely populated areas. The group nature of the traffic is exploited in core- and radio network nodes.

A number of issues for further investigations have been identified: power control schemes and data protection schemes like outer coding. Besides these, it needs to be investigated to which extend the use of IP multicast on the transport IP layer can increase efficiency.

In addition to the above-described services also conversational multiparty services or multiparty gaming could use MBMS in the future. In any case, the introduction of MBMS into cellular systems will further boost existing and future data services.

References

[1] 3GPP TS 22.146 v6.1.0, *3rd Generation Partnership Project, Technical Specification Group Services and System Aspects, Multimedia Broadcast/Multicast Service*, Stage 1, (Release 6)", September 2002

[2] 3GPP TR 23.846 v6.1.0, *3rd Generation Partnership Project, Technical Specification Group Services and System Aspects, Multimedia Broadcast/Multicast Service, Architecture and Functional Description*, (Release 6), December 2002

[3] 3GPP TS 23.246 v1.1.0, *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Multimedia Broad-*

cast/Multicast Service (MBMS);Architecture and Functional Description (Release 6), July 2003

[4] 3GPP TR 29.846 0.2.0; 3rd Generation Partnership Project; Technical Specification Group Core Networks; Multimedia broadcast / multicast service; CN1 procedure description; (Release 6); June 2003

[5] 3GPP TS 25.346 V2.0.0; *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Introduction of the Multimedia Broadcast Multicast Service (MBMS) in the Radio Access Network* (Stage-2);(Release 6); June 2003

[6] 3GPP TS 23.140 V6.2.0 *3rd Generation Partnership Project; Technical Specification Group Terminals; Multimedia Messaging Service (MMS); Functional description;* (Release 6); June 2003

[7] B. Cain, S. Deering, I. Kouvelas, B. Fenner, A. Thyagarajan, *Internet Group Management Protocol, Version 3,* IETF RFC 3376, October 2002.

[8] S. Deering, W. Fenner, B. Haberman, “Multicast Listener Discovery (MLD) for IPv6”, IETF RFC 2710, October 1999

[9] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, L. Wei, *Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification.* IETF RFC 2362, June1998

[10] D. Waitzman, C. Partridge, S.E. Deering. *Distance Vector Multicast Routing Protocol* , RFC 1075 , Nov-01-1988.