Research Article
MBMS—IP Multicast/Broadcast in 3G Networks

Frank Hartung, Uwe Horn, Jörg Huschke, Markus Kampmann, and Thorsten Lohmar

Ericsson GmbH, Eurolab R&D, Ericsson Allee 1, 52134 Herzogenrath, Germany

Correspondence should be addressed to Markus Kampmann, markus.kampmann@ericsson.com

Received 1 March 2009; Accepted 1 December 2009

Recommended by Sooyoung Kim

In this article, the Multimedia Broadcast and Multicast Service (MBMS) as standardized in 3GPP is presented. With MBMS, multicast and broadcast capabilities are introduced into cellular networks. After an introduction into MBMS technology, MBMS radio bearer realizations are presented. Different MBMS bearer services like broadcast mode, enhanced broadcast mode and multicast mode are discussed. Streaming and download services over MBMS are presented and supported media codecs are listed. Service layer components as defined in Open Mobile Alliance (OMA) are introduced. For a Mobile TV use case capacity improvements achieved by MBMS are shown. Finally, evolution of MBMS as part of 3GPP standardization is presented.

Copyright © 2009 Frank Hartung et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Mobile networks have emerged from voice telephony networks to multimedia delivery networks. Mobile TV services have become quite popular data services during the past two years. Apart from live TV channels, the offerings often include special mobile editions with highlights from the weekly TV program, such as series and comedies, delivered as looped channels.

The majority of today’s mobile TV services are delivered over existing 3G networks [1, 2] since this is the fastest and easiest way to deploy mobile TV services. Existing 3G operators have enough capacity in 3G networks to scale up for a mass market of mobile TV services. Latest 3G technology such as HSDPA (High-speed data packet access) [3] gives room for several steps of capacity increases. This allows for more users while benefiting both the diversity and the quality of mobile TV services. The underlying technology is called packet-switched streaming (PSS) [4, 5]. PSS is nowadays supported by all UMTS terminal vendors and offers high-quality streaming services for live or on-demand services. Further quality improvements have been achieved by the introduction of the advanced H.264/AVC video codec [6, 7] and by the introduction of streaming bearers with specific Quality-of-Service (QoS) support.

Nevertheless, the increasing popularity of mobile TV and similar services may lead to situations in which many users want to watch the same content at the same time. Examples are live events of high interest like soccer matches, game shows, and so forth. For those cases, multicasting, known from the internet, or broadcasting is clearly more appropriate technologies.

The work on adding broadcast/multicast support to 3G networks started back in 2002 when both 3GPP and 3GPP2 created work items for broadcast/multicast services in GSM/UMTS and CDMA2000, respectively. In 3GPP the work item was called Multimedia Broadcast and Multicast Service (MBMS). In 3GPP2 it was named Broadcast and MultiCast Service (BCMCS). The specifications of 3G broadcast services were functionally frozen in 2005. Both MBMS and BCMCS introduce only small changes to the existing radio and core network protocols. Therefore, MBMS and BCMCS can be introduced by a pure software upgrade in general as long as the underlying hardware provides sufficient processing power which is the case for all state-of-the-art network nodes. This reduces the implementation costs both in terminals and in the network. It makes cellular broadcast a relatively cheap technology if compared to noncellular broadcast technologies [8–10] which require new receiver hardware in the terminal and significant investments into a new network infrastructure.

MBMS and BCMCS focus on the transport aspects of broadcast and multicast services. They provide transport bearers over which IP multicast [11] packets can be delivered.
They also provide protocols and codecs for the delivery of multimedia files and streams on top of the IP multicast bearers provided by MBMS [12].

Service layer functionality related to multicast/broadcast transport services has been defined by the Open Mobile Alliance (OMA), in the Mobile Broadcast Services 1.0 (BCAST 1.0) specification, which was finalized in 2007 [13, 14]. BCAST 1.0 addresses features like content protection, service and program guides, transmission scheduling, notifications, and service and terminal provisioning. It enables, besides linear TV services, also Podcast services. OMA BCAST 1.0 is agnostic with respect to the underlying broadcast/multicast distribution scheme and applies to MBMS, BCMCS and other noncellular broadcast systems like DVB-H.

Focus of this paper will be an introduction into MBMS as specified by 3GPP. We will start with highlighting the main features of MBMS. After that we give an architectural overview, followed by an explanation in which capabilities MBMS adds to the various protocol layers, starting with the physical layer and then moving up in the protocol stack. In addition to this, capacity and media quality issues will be addressed.

2. MBMS Overview

This section starts with a history and a high-level summary of the key features of MBMS. After that we explain how MBMS fits into the 3G architecture. We will describe in a bit more detail the so-called BM-SC (Broadcast/Multicast Service Centre) which acts as a container for the new functions MBMS provides. At the end we will describe the phases in a typical MBMS service delivery scenario.

2.1. History and Key Features. Standardization of MBMS started in 3GPP and 3GPP2 back in 2002. The first version of the standard was functionally frozen in 2005 and finalized during 2006. Already beginning of 2007 the first MBMS prototype was demonstrated at the 3GSM World Congress in Barcelona (Figure 1(a)). A next version of an MBMS prototype terminal was presented at MWC 2008 (Figure 1(b)). Market introduction of MBMS in UMTS networks is expected during 2010/11.

MBMS has been specified as an add-on feature to existing cellular network technologies, namely, GSM/EDGE [1] and UMTS [2]. Existing transport and physical radio channels of those systems were reused to a large extent in order to keep the implementation effort low. In the ongoing standardization of LTE [15–18] very efficient support of MBMS is taken into account from the start as a main requirement.

MBMS supports two basic transmission modes for delivering IP packets: broadcast and multicast.

The MBMS Broadcast mode can be used to deliver IP packets to all terminals in a certain area or the whole network. If the MBMS broadcast mode is used, a transmission bearer is setup for all cells in which the service should be available and is continuously transmitting as long as the service is up and running. In broadcast mode MBMS does not require an uplink connection and can thus be used like any other “downlink-only” broadcast technology such as DVB-H and DMB.

The MBMS Multicast mode works very similar to IP multicasting. A terminal which wants to receive information related to a particular multicast channels “joins” one or several content channels (e.g., expresses interest to receive content associated with this channel). This information is processed in the routing layer of the core network and is used for optimizing the data delivery path. “Optimizing” means that over connections shared by receivers of the same multicast channels, data is transmitted just once. The only drawback of multicasting is the additional delay when switching from one channel to another one. Therefore MBMS multicasting is less suitable for mobile TV services, which usually require a low TV channel switching delay. The main application of MBMS multicasting is for download or Podcast services.

MBMS was specified such that broadcast/multicast services can be used together with voice and data services within the same radio carrier. This gives the greatest flexibility to cellular operators.

In MBMS it is possible to define so-called MBMS service areas (MSAs). An MSA is defined as a collection of cells. In each of the defined MBMS service areas different services can be delivered. In this way MBMS services can be restricted to geographically limited areas.
The various network configurations MBMS supports range from multicast/broadcast transmission in a single cell, over locally restricted areas and up to a nationwide, single frequency network, broadcasting the same content (e.g., TV channels) across the whole country.

The specification of MBMS in the UMTS radio access network (UTRAN) contains an interesting mechanism called “Counting”. “Counting” means that an uplink signaling channel is used to inform the radio access network when a terminal requests reception of a content channel. Based on this information, the radio access network can keep track of the number of terminals currently receiving a particular content channel. As we will explain later in more detail, this allows the radio access network to make an intelligent decision on whether a particular content channel needs to be transmitted in a cell or not and if yes which type of radio bearer should be used. “Counting” has originally been developed for MBMS multicast services in UMTS but it can also be combined with the MBMS broadcast services. We will come back to the counting mechanisms later in this Section.

The new bearer type which was introduced by MBMS is the so-called point-to-multipoint (P-t-M) radio bearer. While a point-to-point (P-t-P) bearer can only be received by one terminal, the new P-t-M bearer can be received by several terminals in a cell simultaneously.

In UMTS not only the new P-t-M bearer but also existing P-t-P bearers can be used for MBMS. A P-t-P bearer has the advantage that the network can adapt the radio resource usage to the terminal’s reception conditions. In good reception conditions, less radio resources are allocated, whereas in bad conditions more might be needed. In contrast to this, a P-t-M bearer always requires a relatively high amount of radio resources to provide the wanted coverage also for terminals in bad reception conditions. Therefore under certain conditions one or even several P-t-P bearers (one for each receiving terminal) might consume less radio resources than a single P-t-M bearer. The decision which bearer type—P-t-P or P-t-M—uses for a particular UMTS MBMS transmission is made at the radio network controller (RNC). The information for making an optimal decision is obtained from the already mentioned “Counting” process which will be explained in more detail in Section 4.2.

MBMS provides support for both streaming and download services. Hence different types of services will benefit from the introduction of MBMS. One example is mobile TV services in which several operators have launched over existing UMTS networks and which will greatly benefit from the capacity enhancements MBMS brings. Besides mobile TV, other services will benefit, too. One example is audio or video podcasting where mobile users subscribe to a certain podcast provider. MBMS can then be used for the download delivery of the provider’s audio and video clips to a large number of subscribers at the same time. Another example is public safety. Here, MBMS could be used for emergency alerts in certain areas, for example, in case of a tsunami or earth quake.

MBMS streaming and download services can be confidentiality and integrity protected, by using the Secure Real-Time Transport Protocol (SRTP [19]) for streaming and a special container format for downloaded data. The keys used for encrypting and decrypting the SRTP, called MBMS traffic keys (MTKs), change frequently and are transmitted in a key hierarchy where the MTKs are protected using MBMS service keys, MSKs. The MSKs are in turn protected using MBMS User Keys (MUKs) which are derived in a bootstrapping procedure between the smartcard (UMTS integrated circuit card, UIICC, often also denoted as Subscriber Identity Module (SIM) card) and the device. Thus, the SIM card, more exactly its UMTS version, the UMTS SIM (USIM) card, is the basis for service protection in MBMS.

MBMS employs three packet error recovery schemes to cope with packet losses which could occur during bad radio conditions. The most important one is the use of Forward Error Correction (FEC) coding, which allows recovery of lost packets without any server interaction. FEC can be used both for MBMS download and streaming delivery. For the MBMS download service, two additional file repair mechanisms were defined. The first one uses a point-to-point connection to explicitly request the missing parts of a file. The second uses a point-to-multipoint bearer to deliver missing parts to several terminals at the same time.

The quality of a multimedia service is mainly determined by the compression efficiency of the used audio and video codec and the data rate at which audio and video data are encoded. MBMS uses standardized, state-of-the-art audio and video codecs. The video codec is H.264/AVC [6, 7]. The latest MBMS release supports image resolutions up to QVGA (320 × 240 pixels) and frame rates up to 30 frames per second. For audio, the High-Efficiency AAC version 2 codec [20] and the various AMR codecs (narrow band [21], wideband [22], and enhanced wideband [23]) are supported. In the current MBMS release the transmission rate for streaming delivery is limited to 256 kbps. Higher bitrate bearers are under development.

MBMS defines basic components for the encryption of the delivered audiovisual content. However, MBMS does not specify additional service enablers like a program guide or support for interactive services such as voting delivered together with the audiovisual content. Instead those additional functionalities were specified by the Open Mobile Alliance (OMA) under the term “Mobile Broadcast Enabler”, or abbreviated BCAST 1.0 [13, 14]. In Section 7 we will explain BCAST 1.0 and its application to MBMS in more detail.

If the same multimedia service shall be provided over geographical areas comprising several cells or even an entire nation, additional advantages can be taken from the inherent single frequency operation of the UMTS network. In single frequency network (SFN) mode all cells use the same carrier frequency and the terminal is therefore inherently able to receive signals from multiple adjacent cells simultaneously. This leads to a significant increase in spectral efficiency and thereby throughput on the used radio resources.

This efficient exploitation of the single frequency network nature for MBMS is called MBSFN in 3GPP. MBSFN operation is currently being specified by 3GPP as an enhancement for the WCDMA technology. MBSFN is also
a main driver in the 3GPP work item on the long-term evolution (LTE) [15–18]) of UMTS. Apart from this, performance improvements of the P-t-P and P-t-M bearers used in MBMS are addressed as well.

2.2. Architecture. MBMS does not require architectural changes to existing 3G networks. The new functionalities which MBMS provides to operators and service providers are grouped in a new functional node called the Broadcast/Multicast-Service Centre (BM-SC). The BM-SC can be regarded as a functional interface between content delivery services and the MBMS service offered by a cellular network. Towards the core network the BM-SC controls the set-up and release of the MBMS transport bearers and the scheduling of MBMS transmissions.

Figure 2 shows how the Broadcast/Multicast-Service Centre (BM-SC) fits into the existing 3G architecture. The new functions introduced by the BM-SC require a few extensions to the existing Core Network and RAN nodes and their associated protocols. The GGSN (Gateway GPRS Support Node) is extended such that it can establish, manage, and release a point-to-multipoint distribution tree within the mobile network upon notification from the BM-SC. In case of the MBMS Multicast mode, the GGSN includes only those SGSNs (Serving GPRS Support Node) into the distribution tree, which serve subscribers for that bearer. In case of the MBMS Broadcast Mode, the BM-SC provides a list of SGSNs, which shall be included in the broadcast distribution tree. Furthermore, the GGSN is extended by the capability to receive IP multicast traffic from the BM-SC function and to forward it to the SGSNs participating in the multicast or broadcast session using the existing GPRS Tunneling Protocol (GTP). Currently, the multicast traffic is tunneled over unicast which minimizes the impacts onto the existing 3G architecture. The use of IP multicast on transport level between an evolved GGSN and the radio network nodes is currently discussed in LTE and will be discussed in more detail in Section 9.3.

To both radio access networks supported by 3G—UTRAN (UMTS Radio Access Network) and GERAN (GSM/EDGE Radio Access Network)—capabilities were added to efficiently deliver MBMS data to the designated MBMS service area; that is, the geographical area in which the related MBMS service is active.

In GERAN, MBMS may use up to 5 timeslots in downlink for a single transport channel. Depending on the modulation scheme and the network dimensioning, a transport channel capacity between 32 kbps and 128 kbps can be achieved. The total cell capacity depends on the number of supported frequencies in that cell.

MBMS in UTRAN may use up to 256 kbps per transport channel. Several MBMS bearers can be active at the same time in a single UTRAN carrier. Details of the radio bearer realizations in UTRAN will be discussed in Section 3.

2.3. The Broadcast/Multicast-Service Centre (BM-SC). The sub-functions of the BM-SC and their relations are depicted in Figure 3. It is not necessary to provide all BM-SC sub-functions via a single physical node. It is possible to host the Service Announcement, the Session and Transmission, and the Key-Management functions on different physical nodes.

2.3.1. Discovery and Announcement Functions. The User Service Discovery/Announcement function provides service announcements to end-devices. These Service Announcements contain all necessary information, such as a multicast service identifier, IP multicast address, time of transmission, and media descriptions which a terminal needs in order to join an MBMS service. The Service Announcement information is encoded in XML [24, 25] and SDP [26] format and is structured into information units, which are usually called fragments. MBMS supports several ways for delivering Service Announcements:

(i) Interactive retrieval of Service Announcements: HTTP [27] may be used to fetch Service Announcement fragments (XML encoded information element, see above) from the User Service Discovery/Announcement function.

(ii) Service Announcement using MBMS bearers: the Service Announcement fragments are distributed using the MBMS Download delivery method.

(iii) Push announcement using OMA PUSH [28]: the Service Announcement fragment is pushed using OTA-WSP or OTA-HTTP to the terminal. In the simplest way, that Service Announcement fragment may be encapsulated in one or more SMS messages.

2.3.2. Session and Transmission Functions. The Session and Transmission Functions include all data transmission related functions including data encryption. The function is further subdivided into the MBMS Delivery Function and the Associated Delivery Function. The MBMS Delivery Function basically includes the Multicast Sender, which either delivers files via the MBMS Download services (see Section 5.1) or streams via the MBMS Streaming service (see Section 5.2).

The Associated Delivery Function adds auxiliary procedures such as file repair or reception reporting. The purpose of the File Repair Procedure is to guarantee error-free reception of files delivered over MBMS. File Repair Procedure will be explained in more detail in Section 5. The Reception Reporting Procedure allows the BM-SC to collect reception statistics.

2.3.3. Membership Function. The Membership function is only necessary for MBMS Multicast Bearers. It is used to authenticate the “join” request from a mobile receiver. The authentication request is send by the GGSN using the Gmb interface (see Figure 3) to the BM-SC. The membership function may optionally be connected to the key management function.

2.3.4. MBMS Key Management. The MBMS Key Management function is used to provide authorized terminals (e.g., terminals subscribed to a particular MBMS service)
with the necessary keys to decrypt the received files or streams.

Figure 4 shows an overview of the MBMS security functions and data flows. The BM-SC is responsible for the generation and distribution of the MBMS keys to terminals. A terminal requests a key when it needs to decrypt the data. This request may also be initiated by a message from the BM-SC to indicate a key update. The key management system is based on the use of SIM or USIM cards. SIM cards are commonly understood as the smartcard in 2G systems, and USIM as the smartcard in UMTS/3G systems. First, a bootstrapping procedure as defined by GBA (Generic Bootstrapping Architecture) [29] is executed. The bootstrapping procedure consists of several steps, during which the terminal and the BM-SC authenticate each other and agree on a shared secret, called MBMS User Key (MUK). Note that the MUK is never transmitted over the air. Two different variants of GBA exist, depending on the capabilities of the UMTS Integrated Circuit Card (UICC). The UICC is understood as the hardware smartcard containing either a SIM or a USIM application. In the UICC based version GBA_U, the MUKs are stored and further used for decryption of other keys within the UICC. In the terminal based version GBA_ME, the MUKs are stored and used in the device, but outside of the UICC. In both cases, the MUK is subsequently used to protect MBMS Service Keys (MSKs), which are transmitted
in MIKEY (Multimedia Internet KEYing) messages [44]. MSKs can be either pushed to, or requested by the device. The MSKs are used to protect MBMS Traffic Keys (MTKs), which are also transmitted in MIKEY messages, interleaved with the actual content transmission on the MBMS bearer. The MTKs are finally the keys which protect the content streams and/or files. The keys that are used to protect the transmitted data in an MBMS User Service should be regularly changed to give potential hackers less time to extract and distribute them. This ensures that only legitimate users can decrypt the delivered data. In particular, frequent re-keying acts as a deterrent for an attacker to pass the MBMS keys to other users to allow them to illegitimately access the data in an MBMS User Service.

Subscription information is maintained in the BM-SC. The specification of how this agreement is obtained is out of the scope of the 3GPP specifications (it could be configured for instance on a web portal). The subscription should not confused with the actual registration of the end-users to a service. A service registration (e.g., security registration) is possible for both, the multicast and the broadcast mode.

Service Announcement. In this phase the terminal receives all information needed for the reception of a specific service, for example IP multicast addresses and associated port numbers, used codecs and codec configurations. Service announcements can be delivered over interactive (HTTP [27], MBMS and OMA-Push [28] bearers).

User Service Initiation. In this phase the receiver initiates the reception of a certain MBMS service. The MBMS service may use one or several MBMS Bearer Services. In case of an MBMS Multicast bearer, the initiation on the receiver triggers the “Joining” phase in the network. During this “Joining” phase the receiver becomes a member of the multicast group.

Session Start. During this phase, the sender requests the network to activate an MBMS data bearer. This happens independently of initiation or termination of the service by the receiver—for example, a receiver may activate the service before or after bearers for the MBMS service were activated. Furthermore, during this phase transmission resources for the upcoming MBMS data transfer are established.

MBMS Notification. In this phase the receiver is informed about upcoming and ongoing MBMS transmissions. In the radio access network a battery conserving signaling method is used for delivering those notifications to the receivers.
Service Termination. In this phase the receiver stops the reception of a certain MBMS service and its associated bearers. If reception of a multicast bearer is stopped the leaving phase in the network is triggered. During this phase the network removes the receiver from the multicast group.

This concludes the high level overview of MBMS. In the next section we will look at the various protocol layers in more detail. We will start with radio bearer realization and then move up in the protocol stack until we have reached the MBMS User Service layer. In addition we will also describe the complementing service enablers which were defined by OMA BCAST.

3. Radio Bearers for GSM/EDGE and UMTS/WCDMA

3GPP has defined MBMS radio bearer realizations for GSM/EDGE and UMTS/WCDMA. The MBMS radio bearer realization obviously has to differ for these radio access technologies. The present section provides an overview and also presents MBMS radio bearer performance results from simulations.

3.1. GSM/EDGE. In GSM systems, MBMS uses GPRS and EDGE modulation and coding schemes (that is, CS1-4 and MCS1-9 [1]). MBMS also uses the GPRS and EDGE packet data channel (PDCH) for MBMS P-t-M bearer, and the radio link control/medium access control (RLC/MAC) protocols on layer 2. MBMS bearers also support multi-slot operation. In this case, the radio network may use up to four timeslots per MBMS session.

Two enhancements have been introduced to further increase the performance of MBMS in GSM:

(i) RLC/MAC with automatic repeat request (ARQ)—also called Packet Downlink ACK/NACK (PDAN) mode. In this mode, ACK/NACK feedback is provided from up to 16 terminals in a given cell. This way, the RLC data blocks that a terminal did not receive correctly are re-broadcasted over the MBMS radio bearer so that the terminal can use incremental redundancy techniques to reconstruct an error-free data block from the rebroadcasts.

(ii) RLC/MAC without ARQ—also called blind repetition mode. In this mode, RLC blocks are repeated a pre-defined number of times, using an incremental redundancy technique, before the next RLC block is sent.

Simulations results [31] indicate that IP broadcasting at 40 kbps requires two to four timeslots with EDGE channel coding in PDAN mode depending on the number of users whereas four timeslots are needed in blind repetition mode. Note that a regular point-to-point EDGE channel can provide streaming at the same bit rate using two timeslots, but only to one user. For three users, six timeslots are required; for four users, eight, and so on. Therefore, if there are two users in the cell, the MBMS point-to-multipoint bearer is as efficient as a regular point-to-point connection and becomes much more efficient with an increasing number of users.
MBMS terminals will probably be based on existing EDGE hardware with a software update to support MBMS signaling procedures.

3.2. WCDMA. The radio bearers of UMTS are implemented using WCDMA technology. In WCDMA, MBMS reuses existing logical and physical channels to the greatest possible extent. In fact, the implementation in WCDMA requires only three new logical channels, called MBMS control channel (MCCH), MBMS traffic channel (MTCH), and MBMS scheduling channel (MSCH) and one new physical channel called MBMS notification indicator channel (MICH).

The MCCH carries details concerning ongoing and upcoming MBMS sessions. The MTCH carries the actual MBMS application data. The MSCH provides information on the data scheduled on MTCH. The MICH can be configured with 40ms or 80ms interleaving depth (TTI). The selection of a longer TTI provides greater diversity in the time domain by spreading user data over the fading variations. This, in turn, yields improved MBMS capacity.

The MBMS notification indicator channel (MICH) is the new physical channel by which the network informs terminals of available MBMS information on MCCH. This approach was chosen in order to save battery power. A single bit contained in the MICH indicates the terminal whether there is new or changed information on the MCCH. As long as the information on the MCCH is not changed, the terminal does not need to receive the MCCH and can therefore shutdown the receiver to save battery power.

MCCH, MSCH and MTCH reuse the forward access channel (FACH) transport and secondary common control physical channel (S-CCPCH) in WCDMA.

P-t-P transmission within MBMS reuses the existing WCDMA shared or dedicated channel types and will be discussed in the following subsection.

3.2.1. P-t-P Realization Based on HSDPA. HSDPA [3] is designed to serve users in the most efficient way based on fast channel feedback from the terminal. Based on the feedback HSDPA avoids to schedule terminals when they are experiencing fading dips. Furthermore, the modulation, QPSK or 16QAM, and the channel code rate is dynamically selected to exploit to the greatest extend possible the theoretical instantaneous capacity of the radio channel perceived by the respective terminal. Finally, hybrid ARQ (HARQ) using incremental redundancy allows to limit the amount of transmitted channel coding redundancy to just the level that is instantly required.

For video streaming there is typically a small initial delay of about 1 second acceptable after the request for a particular stream has been send by the UE. This delay budget allows efficient channel dependent scheduling of the terminals and HARQ retransmissions; therefore a higher capacity can be achieved with HSDPA.

Performance simulations for a video streaming service have been reported in [32], for a typical network deployment with inter site distance of 1500 m in an urban environment with user speeds of 3 km/h. Table 1 summarizes those results and shows the Erlang capacity per cell sector for 95% satisfied users. A user is satisfied if in 99% of the cases a needed data packet is available from the streaming playout buffer. In other words, it means, a user is satisfied if a playout interruption occurs less or equal 1% of the time. Shown are results with and without receiver antenna diversity in the terminal (“RxDiv”). Terminal receive antenna diversity is not a requirement in the first MBMS release but is likely to be introduced during 2007/2008 in order to increase the availability of high data rates under locally varying coverage. It can be seen that receiver antenna diversity almost doubles the capacity. Apparently the capacity scales almost linearly with the data rate.

3.2.2. P-t-M Bearer Realization. In contrast to P-t-P radio bearers, a Point-to-Multipoint (P-t-M) bearer does not employ feedback. Since the enhancements provided by HSDPA rely on feedback from the terminal, HSDPA cannot be used for P-t-M bearers. The P-t-M transmission parameters need to be statically configured to provide the desired coverage in a given cell. The transmitted signal is lowest at the cell border and therefore the P-t-M bearer can greatly benefit from exploiting also the signals from adjacent cells transmitting the same service, that is, from soft-combining. For soft combining, it is required that transmissions are synchronized with an accuracy of 1 TTI plus 1 slot (= 0.66 msec).

Capacity results have been compiled by 3GPP [33] for radio bearers of 64 kbps for the case without and with soft combining assuming the terminal has capabilities to combine 2 or 3 radio links, one per nearby base station. The results are reproduced in Figure 6, which shows the percentage of the total transmit power (20 W) of a base station that is required to achieve a certain coverage percentage.

The results show that soft combining with 3 radio links significantly reduces the transmit power requirement by 6.5 dB. The capacity can be further increased by receiver antenna diversity in the terminals. Simulations for a different channel model (so called “3GPP case 3”) have shown that terminals with antenna diversity require 3–5 dB lower signal to interference plus noise ratio.

From the fractional power requirements, the capacity in terms of the total number of simultaneous bearers can be calculated directly for a desired coverage percentage, taking into account that about 20% of the power needs to be reserved for control channels. For 95% coverage a total MBMS cell capacity of 1.7 Mbps can therefore be calculated assuming receivers with antenna diversity and soft combining with three radio links. Not shown in Figure 6 is that the required power scales approximately linearly with the bitrate up to the maximal rate of 256 kbps supported.

<table>
<thead>
<tr>
<th>Data rate [kbps]</th>
<th>64</th>
<th>128</th>
<th>256</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity without RxDiv [Erlang]</td>
<td>24</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Capacity with RxDiv [Erlang]</td>
<td>40</td>
<td>19</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 1: P-t-P (HSDPA) capacity (for 95% satisfied users).
4. MBMS Bearer Services in UMTS

MBMS in general provides two basic transmission modes for IP packets: broadcast mode and multicast mode (3GPP calls them “MBMS bearer services” [34]). In this section we will explain how those two modes are realized in UMTS. During this we will explain the “Counting” mechanism in UMTS and the possibility to adaptively select between the P-t-P or P-t-M WCDMA bearers mentioned in the previous section.

4.1. Broadcast Mode. A transmission in MBMS Broadcast mode can be received by all terminals which have activated the reception of the specific broadcast service locally and who are in the MBMS Services Area defined for that particular service. No information is sent in uplink direction. A terminal which likes to receive an MBMS broadcast service, just listens. The UMTS network has no information about active receivers in the MBMS Service Area. Reception of data sent in MBMS broadcast mode is not guaranteed. However, the receiver may be able to recognize data loss. The MBMS Broadcast Mode is very similar to existing Broadcast systems like DVB-T/DVB-H and can even be implemented using only downlink radio transmissions, that is, the uplink functionality of UMTS is not required.

4.2. User Counting for Enhanced Broadcast and Multicast Bearer Services. Within the UMTS Radio Access Network (UTRAN) the so-called “counting” or “re-counting” procedures can be executed to determine the number of terminals in each cell. “Counting” is initiated by the RNC as soon as the RNC needs to know the amount of active UEs that want to receive a specific MBMS service. This is used to determine the optimum transmission bearer, Point-to-Multipoint (P-t-M), Point-to-Point (P-t-P) or no transmission at all for a given MBMS service in the considered cell.

The need for counting is indicated on the MCCH channel per MBMS service. For the counting, the UEs need to reveal their interest to receive an MBMS transmission to the RNC. In case a large number of UEs wishes to receive the service, one has to avoid that all UEs start signaling at the same time, in order to prevent network overload. Therefore, a probabilistic response to the counting indication is used.

When the MBMS counting indication is received, the UE performs a random decision that with a probability equal to the probability factor contained in the counting indication message will generate a counting response from the UE, or with the complementary probability will not generate a counting response. In the latter case the UE continues monitoring the MCCH until the end of the MCCH modification period and if another counting indication is received, repeats the random decision procedure. If the decision is to generate a counting response, the UE uses existing signaling procedures to indicate its presence (RRC connection establishment for idle UEs and Cell Update for UEs in URA-PCH state).

The MBMS counting procedure is terminated if the UE detects that the network has stopped sending the counting indication on the MCCH.

4.3. Broadcast Mode Enhanced with Counting of Users. A special mode in MBMS results from combining counting with the broadcast mode. In the following will use the term “Enhanced Broadcast Mode” (EBM) to refer to this mode. With EBM it is possible to switch to P-t-P bearers if appropriate or to completely stop the transmission in a certain cell if there are no (or not enough) terminals listening to the service. In this way EBM allows a more resource efficient delivery than the vanilla Broadcast Mode which would broadcast a service over a P-t-M bearer even if no terminal is listening.

Figure 7 shows the various situations which could occur if EBM is used. The content channels are delivered over the usual MBMS Broadcast bearers through the Core Network to all RNCs. Each RNC then makes for each cell it controls an intelligent decision, based on the outcome of the counting procedure, whether a particular content channel should be transmitted or not and if it should be transmitted whether a P-t-P or a P-t-M bearer is more efficient.

Note that EBM can be regarded as an optimization in the radio access network. It is fully transparent to higher protocol layers to which EBM appears just like the ordinary MBMS Broadcast mode.

4.4. Multicast Mode. In Multicast mode a terminals indicate interest to receive data of an MBMS Multicast service by “joining” the specific multicast group. The network keeps this “joining” state with the mobility management context in the Gateway GPRS Support Node (GGSN), Serving GPRS Support Node (SGSN) and Radio Network Controllers...
(RNCs). When the terminal moves from one area to another, the “joining” state for all joined services is also transferred to the new serving nodes.

5. Streaming and Download over MBMS

Whereas the MBMS bearer service addresses MBMS transmission procedures below the IP layer, the MBMS user services addresses service layer protocols and procedures above the IP layer.

The MBMS user service is a toolbox, which includes a streaming and a download delivery method. These delivery methods do not differ between or depend on the MBMS bearer services. Figure 8 shows that the MBMS User Services “Download” and “Streaming” sit on top of the MBMS Bearers Services which were explained in the previous section. Both the “Download” and the “Streaming” user service deliver media data encoded in various formats which will be explained in more detail in Section 6.

The MBMS download delivery method is intended to increase the efficiency of file distributions, including messaging services such as MMS. The download delivery method allows the error-free transmission of files via the unidirectional MBMS Bearer Services. The files are “downloaded” and stored in the local file system of the mobile phone. The streaming delivery method is intended for continuous receptions and play-out like in mobile TV applications.

Figure 9 shows the protocol stacks which are used for MBMS. The left hand side depicts that part of the protocol stack which requires an IP unicast bearer. The right hand side shows that part of the protocol stack which was designed for use over multicast/broadcast bearers and thus builds upon UDP. Since UDP packets can also be sent over unicast bearers, the right hand side of the protocol stack can also be implemented on top of a unicast bearer.

It can be seen that Service Announcements and other Metadata can be delivered both over unicast and broadcast/multicast connections. This means that a client can for instance download service announcement related information from a web page or it receives the information via a broadcast/multicast bearer similar to how ESG information today is delivered in broadcast-only services such as DVB-H. Unicast and broadcast/multicast delivery of Service

---

**Figure 7: MBMS Broadcast Mode with “Counting” in UMTS.**

**Figure 8: MBMS Bearer and User Services.**
Announcement information can also be combined. For the Associated-Delivery procedures, certain procedures such as point-to-point file repair and reception reporting require a unicast connection whereas other procedures such as point-to-multipoint file repair (e.g., multicasting of missing packets) can be executed over a broadcast/multicast bearer. Similar holds for the Security functions. Some of them such as Registration and MSK distribution require a unicast connection whereas others such as MTK distribution work over broadcast/multicast bearers. The actual transfer of media data, that is, streaming and download, was designed for broadcast/multicast bearers but as it was said already above, it can also be done over a unicast bearer.

5.1. Download Delivery Method. The download delivery method is intended for file distribution services, which store the received data locally in a terminal. It can be used to deliver arbitrary files from one source to many receivers efficiently. Existing content-to-person MMS services, which deliver short video clips related to live events like a soccer match via MMS, will greatly benefit from this feature. Today, those services rely on point-to-point connections for MMS delivery. In the future the existing MMS sub-system can be easily interfaced with a BM-SC which then distributes the clip via MBMS download. The principle of an MBMS download delivery is shown in Figure 10.

The files are delivered during the MBMS data transfer phase (see Phase 1 in Figure 10). The MBMS transmission bearer is activated with the MBMS Session Start message. This message triggers the paging process in the radio access network, which inform the MBMS receivers about an upcoming transmission. After the MBMS bearer is successfully established the BM-SC starts sending the actual MBMS download data. Included in the download data is a file containing the File Delivery Table (FDT). The FLUTE [35] protocol is used to send the files via UDP. The FLUTE protocol allows the FEC protection of the files and it uses the IETF FEC framework [36]. After the MBMS data transmission, bearer resources are released via the MBMS Session stop message.

During the MBMS data transfer phase, certain mobile phones may experience packet losses due to fading conditions or handovers. Naturally, full reliability cannot be offered in a pure unidirectional distribution scheme because the packet loss rate can be excessive for some users.

Three packet error recovery schemes are foreseen for the download delivery method. The most important one is the use of FEC coding, which allows recovery of lost packets without any server interaction.

The Raptor forward error correcting code [37] was chosen as a basis for FEC scheme [38]. A broadcast of newly created FEC packets is of benefit for all receivers, which have not successfully reconstructed the original source block. Generally, Raptor codes can handle even large files as one source block. But since mobile phones have a limited amount of fast memory for decoding, a single source block for 3GPP release 6 receivers may only contain up to 4100 kByte of data. Thus, larger files are subdivided into a number of source blocks and the FEC repair symbols are generated for each source block. If an interactive bearer is used for the file repair procedure, the repair data is independently sent to different receivers and can even be tailored to the actual losses of that receiver. On the contrary, if the MBMS bearer is used, the same repair data is sent only once to multiple User Equipments (UEs) and the repair data should be useful for all receivers with losses. Therefore, the rateless property of the Raptor code, for example, the possibility to generate an arbitrary number of FEC redundant symbols out of one source block, is very beneficial for the PTM repair mechanism.

Beside FEC protection, MBMS offers two additional types of file repair procedures: one is using interactive bearers and another one uses MBMS bearers. In case of file repair, the MBMS client waits until the end of the transmission of files or sessions and identifies then the missing data from the MBMS download. Afterwards, it calculates a random back-off time and selects a file repair server randomly out of a list. Then, a repair request message is sent to the selected file repair server at the calculated time. The file repair server responds with a repair response message either containing the requested data, redirecting the client to an MBMS download session, redirecting the client to another server, or alternatively, describing an error case. The BM-SC may also send the repair data on a MBMS bearer (possibly the same MBMS bearer as the original download) as a function of the repair process. The performance of the post-delivery file repair procedures described above has been analyzed in [39, 40].
5.2. Streaming Delivery Method. The streaming delivery method is intended for the continuous reception and play-out of continuous media like video, audio or speech. A typical example of an application using the streaming delivery method is the transfer of live channels in a mobile TV service.

RTP [41] is used as transport protocol for MBMS streaming delivery. RTP provides means for sending real-time or streaming data over UDP (see Figure 9).

As previously mentioned packet losses could occur during the streaming data transfer that would result in distortions of the received video and audio quality at the receiver side. In case of streaming with its real-time constraints file repair procedures are not suitable. Instead, forward error correction (FEC) is used [42].

As for download delivery the Raptor forward error correcting code [37] may be used to increase bearer reliability for MBMS transmissions. The Raptor code belongs to the class of rateless codes and can thus generate an arbitrary number of FEC redundant symbols out of one source block. The FEC stream bundling concept shown in Figure 11 allows the protection of the actual audio/video data together with synchronization information (RTCP) and possibly decryption information. Packets of one or more UDP flows may be used to construct the source blocks for the FEC protection. The FEC redundancy information is transmitted in one separate traffic flow. Since the Raptor code is a systematic FEC code, the receiver can simply ignore the FEC flow, if no transmission errors occur.

The advantage of the FEC stream bundling concept, as shown in Figure 11, is that the FEC efficiency is increased when protecting several data flows together, because the FEC code works on a larger portion of data.

Figure 12 depicts how one or more out of several possible packet flows of different types (Audio, video, text RTP and RTCP flows, MIKEY flow) are sent to the FEC layer for protection.

The source packets are modified to carry the FEC payload ID and a new flow with repair data is generated. The receiver takes the source and repair packets and buffers them to perform, if necessary, the FEC decoding. After appropriate buffering received and recovered source packets are forwarded to the higher layers.

MBMS streaming delivery may also use reception reporting, that is, reports about received parts of the service that are sent back to the server. The reception reporting procedure allows an operator to collect reception statistics and usage patterns of the service from actual end-users.

The MBMS User Service framework is harmonized with OMA BCAST and DVB CBMS specifications. In particular the same set of audio/video codecs is supported. Therefore, for a given rate MBMS is able to deliver the same audiovisual quality as DVB-H.

6. Media Codecs

MBMS supports different media types, among which Speech, Audio, Video and Timed Text are ones supported both for MBMS Download and Streaming delivery. In addition, MBMS defines codecs for synthetic audio, still images, bitmap graphics, vector graphics, and text for use in MBMS Download services only.

In what follows we will summarize the codecs which have been specified for MBMS Download and Streaming services. The complete set of supported codecs can be found in [15].
For speech, one can choose between the AMR decoder for narrow-band speech and the AMR wideband speech decoder (AMR-WB) for wideband speech working at 16 kHz sampling frequency.

For audio coding, Enhanced AACPlus (E-AAC+) and AMR-WB+ codec can be used.

Within 3GPP the two supported audio codecs were tested for different type of audio content at different coding rates. These characterization results are summarized in Tables 2 and 3.

For video coding, MBMS uses H.264/AVC [6, 7]. Since dynamic negotiation of media codecs is not feasible in broadcast/multicast services, H.264/AVC is the only new video codec recommended for MBMS.

The current version of MBMS specifies the Baseline Profile at Level 1.2. In this profile/level combination, video content can be encoded up to QVGA resolution (320 × 240 pixels) and up to 30 fps.

In addition to speech, audio, and video, MBMS also supports “Timed Text”. Timed text is text that is rendered at the terminal, in synchronization with other timed media such as video or audio. Timed text is used for such applications as closed captioning, titling, and other visual annotation of timed media. Timed text for MBMS reuses the Timed Text specification [43] which was specified originally for packet-switched streaming [4].

### 7. Service Layer Components

MBMS defines all components that are necessary to transmit audiovisual content to MBMS capable devices. Yet, MBMS does not include all service layer functionality known from cable or satellite TV networks. Most notably, MBMS does not specify a service guide or program guide that describes the available programs or channels in a user friendly way. Further, interactivity must be closely integrated into the service and its architecture, to allow easy and user-friendly access. Also, service access protection and content protection are favorably closely integrated into service guide and service purchase or subscription mechanisms. Thus, some additional functionality is needed in addition to MBMS for a complete and user friendly service offering.

The Open Mobile Alliance (OMA) has defined such a set of functionalities, called the “Mobile Broadcast Enabler”, or abbreviated BCAST 1.0. In detail, the following functions are defined in BCAST 1.0: service guide, file and stream distribution protocols and associated mechanisms like fire repair and forward error correction (FEC), notification function, service and content protection, service interaction,
service provisioning, terminal provisioning, roaming and mobility support, and specification on back-end interfaces for each function. BCAST 1.0 is a generic set of functions. However, in order to better integrate with broadcasting systems, OMA has defined adaptations to some specific broadcast systems. These include IPDC over DVB-H, 3GPP2 BCMCS and, notably here, 3GPP MBMS. Thus, BCAST 1.0 can be nicely combined with MBMS, giving a combined overall system for mobile TV services.

Two of the functions mentioned before are of special importance and are explained in some more detail: the service guide, and service protection. The BCAST service guide (SG) is a central component of the BCAST enabler. It consists of information describing the service and programs, and has some parts targeted to the end-user, and other parts targeted to the device. The SG is encoded in an XML-based format, and is structured into information units, so-called fragments. There are separate fragments that describe the service and program schedule, the technical details required to “tune in” and receive the service, interactivity, additional information for the user, and purchase and delivery related information. Figure 13 shows the logical structure of the BCAST SG. However, many of the fragments are not mandatory to be used for a specific service. A minimalistic SG that describes a continuously ongoing TV channel would for example only need a “service” fragment (describing the service, e.g., the channel name) and an “access” fragment (describing the technical parameters to receive the service, like codec and IP port information). In general however, more detailed information describing the sequence of programs is desired, making the used SG more complex. SG fragments are bundled and packaged into a specific container, called a service guide delivery unit (SGDU). Further, the totality of all current fragments and their grouping into SGDUs for a service is described in a so-called Service guide delivery descriptor (SGDD). SGDUs can be delivered over broadcast/multicast, where they are typically frequently repeated, or on demand over unicast. Delivery over broadcast takes away resources from the service delivery and is thus only preferred in systems that cannot rely on a unicast channel for interaction. In 3G systems however, it is advantageous to deliver the SG on demand, when a device or its user decides that it needs additional information, for example, in case the information present on the device contains only information for a short time of the future. The service guide is dynamic in nature; on-the-fly changes and updates of the SG and its fragments are possible.

Service protection is another important functionality. In fact, service protection is closely coupled to the overall control over service delivery and end user billing; the value chain participant that controls the service protection usually also sends the bill to the end user. Therefore, service protection does not only have a technical role, it is also closely connected to the business control and business model. In order to be more flexible, BCAST has defined two different solutions for service protection, called profiles: the Smartcard profile and the DRM profile. Both profiles follow however the same principle of a key hierarchy, as shown in Figure 14. Key hierarchy means that the audio and video media streams are not directly encrypted with just one key. Rather, the media streams are encrypted with a short term key, which changes frequently (e.g., every 5 minutes) and is broadcasted alongside with the media streams. The short-term key itself (corresponding to the previously mentioned MTK in MBMS) is encrypted with a long-term key. The long-term key (corresponding to MSK) changes less frequently (e.g., once a day or once a week) and is transported to the device in a point-to-point communication. The long-term key itself is encrypted with a shared secret (corresponding to MUK) that somehow has been agreed between sender and receiver beforehand.

The OMA BCAST Smartcard profile is derived from MBMS security and GBA as described above, but provides some extensions. In the Smartcard profile, the shared secret (Layer 1 (L1) as shown in the figure) is based on a key stored in a SIM or USIM or ISIM and also known in the network. From that stored key, the shared secret used for the communication is derived in a bootstrapping procedure. The long-term keys (L2) and short-term keys (L3) are sent in messages following the MIKEY [44] format. The media streams (L4) are streams transported using encrypted transport protocols. BCAST supports either SRTP [19] or ISMAcryp [45] or IPSec [46].

In the DRM profile, the shared secret (L1) is based on DRM public-key certificates issued by a central authority to both sender and receiver, and which are mutually authenticated in a DRM registration procedure. The long-term keys (L2) are sent in DRM Rights Objects, which is a known container format for keys (and permissions to use content). The short-term keys (L3) are sent in a special BCAST message format. As for the Smartcard profile, the media streams (L4) are streams transported using SRTP [19], ISMAcryp [45] or IPSec [46].

The specification of OMA BCAST was finalized in June 2007. First interoperability tests have been done by the companies participating in the bmcoforum [47] in the same month.

8. Use Case: Mobile TV over MBMS in UMTS

In this section we summarize simulation results from [48] showing the capacity improvements which MBMS will add to a UMTS network. We have assumed an MBMS configuration which uses broadcasting combined with “Counting”, commonly referred to as “Enhanced Broadcast Mode” (see Section 4.3).

We assume a mobile TV service with 20 mobile TV channels, each is delivered at 128 kbps. The channel interest is not uniformly distributed. We assume that Channels 0 & 1 are selected each in 25% of the cases, Channel 2 in 15% and Channel 3 in 10% of the cases. The remaining 16 channels are selected with an equal probability of 1.56%. For simplicity we assume that the channel interest does not change during the simulation period.

The mobile TV usage model is derived from statistics on usage and viewing patterns obtained during the “Finpilot” study [49]. In our simulation we assume that each user
is watching mobile TV for a total of 30 minutes within 12 hours. The session length is exponentially distributed with a mean of 12.6 minutes.

The user may switch between different channels during the mobile TV session. The user watches a channel for an exponentially distributed time with a mean of 3.25 minutes. The user switches off the mobile TV reception as soon as the end of the session is reached.

A cell cluster of 48 cells is used as cellular network simulation topology. We further assume that the MBMS Enhanced Broadcast Mode (see Section 4.3) is used. In this mode, an RNC can decide based on information obtained through counting whether a point-to-point or a point-to-multipoint bearer is more efficient. We further assume that the point-to-point bearer is realized as an HSPA bearer and that the point-to-multipoint broadcast bearers requires twice as much radio resources compared to an HSPA bearer. We further assume that each cell has capacity to serve up to 16 simultaneous users with point-to-point HSPA. When point-to-multipoint transmissions are used, then the available point-to-point capacity is reduced accordingly.
Figure 15: Blocking probability over user density for delivery of 20 TV channels over MBMS in a UMTS network.

The mobile TV subscribers are randomly distributed in the cell cluster. We varied the user density between 60 and 190 users per cell. Note that in urban areas the mobile subscriber density is typically around 300 users. However, the addressable market for mobile TV services today is much lower than 100% since not all users are interested in the service and not all users have TV enabled phones. Market research today indicates an addressable market of 10%–15%. By limiting ourselves to 190 mobile TV subscribers we can still capture a market take-off over the coming years.

The resulting channel usage per cell for different channel selection probabilities is depicted in Figure 15. It can be seen that for 140 users per cell the probability that one of the low interest channels 4 to 19 is watched is very low. Channel 0 (25% channel selection probability) is used on average by 1.45 users per cell, whereas channels 4 to 19 have 0.09 users per cell. The maximal number of simultaneous watcher per channel is of course higher.

In Figure 15 we show blocking probabilities for various network configurations. The case where all 20 channels are broadcasted is not shown since the blocking probability in that case is zero at the expense of excessive radio resource consumption: 20 always on broadcast bearers in every cell.

The case \( N_{FTM} = 16 \) corresponds to unicast only transmission, using HSDPA. From 120 mobile TV subscribers per cell onwards the blocking probability starts to increase and reaches 1% at 160 subscribers per cell. One can see that the blocking probability can be considerably decreased by introducing only a few broadcast bearers. With 2 broadcast bearers \( (N_{FTM} = 2) \) the blocking probability stays well below 0.5% even for 190 mobile TV subscribers per cell.

A user density of 190 users per cell corresponds to the case where roughly 30% of the total subscriber base is subscribed to the mobile TV service.

9. MBMS Evolution

MBMS evolution in 3GPP 3G networks from Release 6 is part of the two main tracks of standard developments, as depicted in Figure 16:

(i) evolution of WCDMA,

(ii) 3G long term evolution of UTRA, called 3G LTE (denoted by E-UTRA in the 3GPP specifications).

The evolution of WCDMA comprises improvements for MBMS for the P-t-M radio bearers, mainly the introduction of so called MBMS single frequency networks (MBSFN), and HSPA evolution that allows for higher capacity using P-t-P radio bearers. In 3GPP Release 7 also support for unicast bearers like Streaming [4, 5] or OMA Push [28] as part of an MBMS User Service has been added. The use case which has driven the integration of unicast bearers into MBMS is access to an MBMS service in the home network from a visited network which does not support the same MBMS services as in the home network or which does not support MBMS at all.

In a parallel track 3GPP has standardised the long term evolution of UTRA, called 3G LTE (denoted by E-UTRA in the 3GPP specifications). LTE is part of 3GPP Release 8 and defined in the 36-series of specifications. Originally Release 8 LTE was planned to include enhanced MBMS and significant standardization effort has been made in this area. However, beginning of 2008 MBMS was shifted out of Release 8 and into Release 9. By December 2009 the specification has been ranked 95% complete by 3GPP.

The LTE MBMS related aspects addressed here capture the status of 3GPP agreements at the time of the shift, as of [18].

The main goals for LTE are [50]:

(i) improved spectrum flexibility,
(ii) increased cell edge and cell average throughput,
(iii) efficient support of MBMS,
(iv) simplified architecture.

The most important innovations in LTE with respect to MBMS are the support of larger carrier bandwidths, use of OFDM modulation in the physical layer and efficient and flexible support for MBSFN.

The MBSFN principle is the same in both evolution tracks and is therefore first introduced hereafter for both in common.

9.1. MBMS Single Frequency Networks. If the same multimedia service shall be provided over geographical areas comprising several cells or even an entire nation, advantages can be taken from the inherent single frequency operation of the WCDMA and LTE, that is, all cells can use the same carrier frequency and the terminal is therefore inherently able to receive the signals from several cells simultaneously [51]. If different cells transmit different information, then the signals interfere in the terminal receiver. Whereas this cannot be avoided for user-individual services, it is possible to avoid intercell interference for broadcast services by sending the same signal from all cells in the service area at the same time. By the elimination of intercell interference, a dramatic increase in spectral efficiency and thereby throughput on the used radio resources is achieved.

This efficient exploitation of the single frequency network nature for MBMS is called MBSFN in 3GPP. MBSFN operation is currently being specified by 3GPP as an
enhancement for WCDMA. MBSFN is also a main driver in the 3GPP work item on the long term evolution (LTE) (see next section).

If an entire carrier frequency is dedicated to broadcasting using MBSFN transmissions, such dedicated MBMS carrier obviously does not need to be matched with an associated uplink carrier. Therefore there is an opportunity to operate dedicated MBMS carriers also in unpaired frequency bands. Among potentially suitable frequency bands there are the following:

(i) UMTS core band: 1900–1920 MHz and 2010–2025 MHz—designated for UMTS-TDD operation that have been licensed but are currently unused in most countries.

(ii) UMTS extension band: the range 2570–2620 MHz is in between the UMTS FDD uplink and downlink. The decision on the usage is up to national administrations in Europe. The upper part of this range that is adjacent to the FDD downlink could be used for MBMS dedicated carriers, without coexistence problems between the MBMS dedicated carriers and FDD downlink.

(iii) Part of the terrestrial TV broadcasting band 470–862 MHz. Due to the transition from analog TV to the more spectrally efficient digital TV, it is widely accepted that part of this band can be vacated from terrestrial fixed TV broadcasting services. This so called “digital dividend” could be used to introduce mobile TV services based on MBMS.

9.2. Evolution of WCDMA

9.2.1. MBSFN. For WCDMA, 3GPP has introduced MBSFNs in Release 7 for FDD and TDD in 2007. In Release 8 this has been further developed into Integrated Mobile Broadcast (IMB) that is designed to operate in unpaired spectrum. IMB principles are identical to those for MBSFN FDD, except that a new synchronization code has been added to resolves potential cell search issues of legacy TDD terminals.

In the conventional WCDMA downlink, all the physical channels of a cell, except for the synchronization channels, are scrambled by cell-specific primary or secondary scrambling codes. While scrambling codes are used to differentiate cells, orthogonal spreading is used to separate multiple code-division multiplexed channels within a cell.

With orthogonal spreading, a receiver does not experience own-cell interference when the base station signal travels through a flat channel. State of the art receivers are capable of removing own-cell interference even in the case of non flat channel, using, for example, linear MMSE chip equalizer or G-Rake techniques. Due to lack of signal orthogonality; however, these receivers do not suppress other-cell interference as effectively.

In contrast, in the MBSFN transmission mode the same scrambling is used for all participating cells. When different base stations use the same waveform to simultaneously send a common set of MBMS content channels, the received signal is then the same as that for a single-source transmitted signal traveling through a heavily dispersive channel, where each path corresponds to a signal path between a base station transmitter and the receiver as illustrated in Figure 17. In this case, the other-cell signals share the same orthogonal properties as the own-cell signals. As a result, an advanced receiver (e.g., zero-forcing equalizer) not only collects the signal energy contributed by the multiple base station signals, but also gets rid of interference arising due to multipaths and transmission from multiple base stations. In order to exploit the resulting very high SINR, Release 7 adds support for 16QAM on the S-CCPCH on which an MTCH is mapped. This boosts the spectral efficiency of MBMS in a WCDMA network significantly.
In order to limit the receiver complexity, the delay spread needs to be limited and therefore the cells in the MBSFN need to transmit the same signal at the same time.

Synchronisation on the order of a few micro seconds is required. The synchronisation of NodeBs connected to one RNC is ensured by the RNC. As a consequence, synchronisation is not ensured across RNCs and MBSFN across RNCs are not supported.

In a proposed application of the MBSFN mode, cells have multiple carriers of which some operate in the MBSFN mode using cell common scrambling and some operate in the normal 3GPP Release 6 mode using cell specific scrambling. Multiple MBMS physical channels can be code-division multiplexed on this dedicated carrier. However, the same channelization code is used in all the base stations to transmit the same MBMS content channel. Furthermore, of course the usual 3GPP Release 6 time multiplexing of MBMS is possible.

Table 4 reports MBSFN spectral efficiency results from [52] obtained from radio network simulations with 2800 m inter site distance, 3 sectors per site, urban area propagation environment, with channel model “Vehicular A” and users moving at 3 km/h. The performance criterion is the spectral efficiency achievable with 95% coverage and 1% BLER.

Results are shown for receivers of 3GPP Type-2 (implementing G-RAKE or equivalent techniques) and 3GPP Type-3 (like Type-2 plus implementing antenna diversity). With Type-3 advanced receivers that can equalize, that is, take advantage of, the signals from the 7 closest NodeBs spectral efficiency in excess of 1 b/s/Hz can be achieved if 16QAM is employed. For less advanced Type-3 receivers or in case the inter site distance is significantly larger, so that only the signals from 3 NodeBs can be equalized, the spectral efficiency is still 0.6 b/s/Hz.

The higher cell capacity also allows for higher bearer data rates, thus the maximum supported bearer data rate has been increased to 512 kbps. As the cell capacity is several times higher than 512 kbps, time sliced transmission of services is enabled, where each service is mapped to one 2 ms subframe in a radio frame. This significantly improves the battery saving in the receiver.

9.2.2. HSPA Evolution. HSPA comprises HSUPA and HSDPA. As has been described in Section 3.2.1, HSDPA can be used for MBMS P-t-P radio bearers. For the HSDPA evolution, 3GPP is standardizing $2 \times 2$ MIMO and an increase in the modulation order from 16QAM to 64QAM for 3GPP Release 7, in order to drastically increase the per user data rates and system capacity.

The MIMO scheme chosen for HSPA evolution employs multiple codeword allowing for transmitting two separately encoded streams to a UE. Hence the data on each stream is separately encoded, modulated and spread. The up to 15 spreading codes (of spreading factor 16), available for HSDPA, are reused over both streams. Before transmitted on the antennas, the modulated and spread signal is spatially weighted (pre-coded) using a unitary transform. The weights are taken from the same codebook as used for closed-loop transmit diversity mode 1. For the link adaptation, the UE reports the number of streams, the spatial weight (pre-coding index) and transmission rate that it prefers. HARQ is operating independent between the streams.

Release 6 HSPA systems support the use of 16QAM in the downlink. For indoor or small-cell system deployments, higher SNRs and thereby 64QAM modulation can be supported. The maximum data rate per user achievable with 64QAM is 21.3 Mbps. While the combination of MIMO and 64QAM is not included in Release 7, it is being considered for future releases.

Section 9.4 presents some performance results for evolved HSDPA, together with results for P-t-P transmissions in LTE.

9.2.3. Flat Architecture. 3GPP has standardized an optional flat architecture for WCDMA in Release 8, as illustrated in Figure 18. The RNC functionality can be integrated into the Node-B and IP multicasting in the core network is supported down to the NodeBs. Functionality and signalling
Table 4: MBMS capacity and spectrum efficiency achieved by SFN operation (using 90% of the cell power) for receivers of Type-2 (implementing G-RAKE or equivalent techniques) and Type-3 (like Type-2 plus implementing antenna diversity).

<table>
<thead>
<tr>
<th>Receiver</th>
<th>Receiver capable of equalizing signals from 3 NodeBs</th>
<th>Receiver capable of equalizing signals from 7 NodeBs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type-2</td>
<td>1.54 Mbps, 0.31 b/s/Hz</td>
<td>2.62 Mbps, 0.53 b/s/Hz</td>
</tr>
<tr>
<td>Type-3</td>
<td>3.01 Mbps, 0.60 b/s/Hz</td>
<td>5.38 Mbps, 1.08 b/s/Hz</td>
</tr>
</tbody>
</table>

via the Iur interface has been added to enable the necessary coordination between NodeBs, in particular synchronisation between NodeBs for MBSFN that in the classical architecture is performed by the RNC.

9.3. MBMS in LTE. MBMS in LTE uses an evolved architecture, in order to support MBSFNs with high flexibility, which was an important design goal of LTE from the start. An MBMS architecture is desired that supports the tight synchronisation of broadcast content transmit times across many cells in the network. Furthermore, for LTE it is desired to support MBSFN transmission and user individual services on the same carrier. The architecture needs to support the coordinated allocation of radio resources within the carrier for MBSFN transmission across all cells participating in the particular MBSFN.

Figure 19 shows possible candidate architecture that has been discussed in 3GPP. The architecture is based on enhancements to the architecture of earlier releases.

The following logical entities are proposed:

(i) E-MBMS GW: entity that is located between the content provider and the eNode Bs. The MBMS GW will be involved in the MBMS session start/setup, and will also participate in the content synchronization for MBMS services using MBSFN (i.e., the mUPE functionality defined in 3GPP). This entity is expected to be part of the Evolved Packet Core. The detailed functionality of this logical entity is in the scope of SA2/CT work, for example, it is FFS if this entity would be split in a user plane and control plane part.

(ii) MCE: entity responsible for coordinating the usage of MBSFN transmission in the LTE RAN. This entity is expected to be part of the LTE RAN. The MCE will be responsible for all the eNode Bs in a given area.

Furthermore it is proposed to introduce the following logical interfaces:

(i) M1 is a logical interface between the MBMS GW and the eNode Bs. The transport on this interface will be based on IP multicast. The MBMS content will most likely need to be transport in some framing or tunneling protocol, in order to support content synchronization and other functionalities. IP multicast signaling will be supported in the TNL layer in order to allow the eNode Bs to join an IP multicast group.

(ii) M2 is a logical control interface between the MCE and the eNode Bs. This interface is used to coordinate the setting up of MBMS service in the eNode Bs for MBSFN operation. It is expected that the MCE will have a master role in the coordination, while the eNode Bs provide feedback to the MCE. It is expected that M3 will have the same signaling transport layer as S1/X2.

It is not precluded that M3 interface can be terminated in eNBs. In this case MCE is considered as being part of eNB. Therefore M2 does not exist in this scenario. This is depicted in Figure 19 which depicts two envisaged deployment alternatives. In the scenario depicted on the left MCE is deployed in a separate node. However, in this case the MBSFN controlled by the integrated MCE can only consist of cells belonging to the eNB hosting the MCE. The scenario on the right MCE is part of the eNBs.

(iii) M3 is a logical control interface between the MBMS GW and the MCE. This interface has similar MBMS functionality as currently specified over Iu (RANAP), meaning that the MCE is expected to take the role of the RNC with regards to setting up MBMS resources, while the MBMS GW takes the role of the SGSN. It is expected that M1 will have the same signaling transport layer as S1/X2.
9.3.1. LTE Downlink Physical Layer. In the following we provide a brief introduction to the E-UTRA downlink physical layer [53]. The EUTRA downlink uses OFDM, because it efficiently supports flexible carrier bandwidth, allows frequency domain scheduling, is resilient to propagation delays which is particularly beneficial for SFN broadcasting configurations, and is well suited for multiple-input multiple-output (MIMO) processing.

The possibility to operate in vastly different spectrum allocations is essential. Different bandwidths are realized by varying the number of subcarriers used for transmission, while the subcarrier spacing remains unchanged. In this way operation in spectrum allocations of 1.25, 2.5, 5, 10, 15, and 20 MHz can be supported. Due to the fine frequency granularity offered by OFDM, a smooth migration of, for example, 2G spectrum is possible. Frequency-division duplex (FDD), time-division duplex (TDD), and combined FDD/TDD, are supported to allow for operation in paired as well as unpaired spectrum.

A subcarrier spacing of 15 kHz is adopted, allowing for simple implementation of dual mode Rel-6/LTE terminals as the same clock frequencies can be used. To minimize delays, the slot duration is selected as short as 0.5 ms, corresponding to seven OFDM symbols. The cyclic prefix length of 4.7 μs is sufficient for handling the delay spread for most unicast scenarios, while only adding modest overhead. Very large cells, up to and exceeding 120 km cell radius, with large amounts of time dispersion are handled by reducing the number of OFDM symbols in a slot by one in order to extend the cyclic prefix to 16.7 μs. Multi-cell broadcast services are supported by transmitting the same information from multiple (synchronized) base stations. To the terminal, the received signal from all base stations will appear as multipath propagation and thus implicitly be exploited by the OFDM receiver.

Figure 20 outlines the more detailed time-domain structure for LTE downlink transmission. Each 1 ms subframe consists of two slots of length $T_{\text{slot}} = 0.5$ ms. Each slot then consists of a number of OFDM symbols.

A subcarrier spacing $\Delta f = 15$ kHz corresponds to a useful symbol time $T_u = 1/\Delta f \approx 66.7$ μs. The overall OFDM symbol time is then the sum of the useful symbol time and the cyclic-prefix length $T_{\text{CP}}$. As illustrated in Figure 20, LTE defines two cyclic-prefix lengths, the normal cyclic prefix and an extended cyclic prefix, corresponding to seven and six OFDM symbols per slot respectively. The exact cyclic-prefix lengths can be obtained from Figure 20. It should be noted that, in case of the normal cyclic prefix, the cyclic-prefix length for the first OFDM symbol of a slot is somewhat larger, compared to the remaining OFDM symbols.

The main use of the extended cyclic prefix is expected to be MBSFN-based multicast/broadcast transmission.

In addition to the 15 kHz subcarrier spacing, a reduced subcarrier spacing $\Delta f_{\text{low}} = 7.5$ kHz is also defined for LTE and specifically targeting MBSFN transmission. The use of the reduced subcarrier spacing also scales the OFDM symbol time by a factor of two, thus providing a twice as long cyclic prefix ($\approx 33.3$ μs).

Taking into account also the downlink time-domain structure, the resource blocks mentioned above consist of 12 subcarriers during a 0.5 ms slot, as illustrated in Figure 21. Each resource block thus consists of $12 \cdot 7 = 84$ resource elements in case of normal cyclic prefix and $12 \cdot 6 = 72$ resource elements in case of extended cyclic prefix.

Each subcarrier of each OFDM symbol can be modulated by Quadrature phase shift keying (QPSK), 16-quadrature amplitude modulation (16-QAM), or 64-QAM modulation schemes.

The downlink share channel (DL-SCH) transport channel uses the physical layer to provide P-t-P radio bearers as well as P-t-M radio bearers with feedback from the UEs. The latter transmission mode can be used for MBMS if the number of users per cell is small, typically less than 10, so that the feedback does not cause a high signaling load. The feedback allows to adapt transmit parameters to the particular radio conditions of the few users and HARQ retransmission.
9.3.2. MBSFN. If a larger number of users of a particular MBMS are present in a cell, broadcast radio transmission in the cell are more suitable, which can be used either in single cell or multi cell transmission mode. In the single cell mode the MBMS service is transmitted to a particular user only from a single cell. Multi-cell transmission is supported by means of the MCH (Multicast Channel) transport channel. A multicell transmission essentially means that the cells transmitting the MBMS service are configured to form an MBSFN (see also Section 9.1). If an MBSFN is established using a particular MCH, then the same MCH information is transmitted time aligned from a group of cells using identical transport formats and identical resource allocations and with identical scrambling (cell-group-specific scrambling, see above). From a UE point-of-view, such multicell MCH transmission will appear as a single MCH transmission over a channel being the aggregation of the channels from the UE to each cell.

The MCH can time multiplexed on a subframe granularity of 1ms with other transport channels like the downlink shared channel (DL-SCH) that is used for single cell point to point transmission.

In order to be able to properly demodulate the multicell MCH transmission, the UE needs an estimate of the aggregated channel of all cells involved in the MBSFN. For this to be possible, specific reference signals (“MCH-specific reference signals”) are needed. The MCH-specific reference signals are identical (identical time/frequency locations and identical reference-signal sequence) for all cells involved in the MBSFN. Based on such reference signals, the UE can directly estimate the aggregated channel of all cells involved in the MBSFN.

The aggregated channel of the cells involved in the MBSFN will typically have a large delay spread due to the differences in the propagation delay as well as residual transmit-timing differences. To handle this, LTE defines 3 different values for the OFDM cyclic prefix: 4.67 $\mu$s, 16.7 $\mu$s and 33 $\mu$s.Signals from eNBs arriving within the CP duration of the UE synchronization point contribute useful signal energy and thereby improve the coverage. Signals
arriving outside the CP contribute interference. Since the CP does not contain user data, its length is a tradeoff between the time fraction available for user data and the CINR value achievable with the desired probability in the MBSFN area. If MBSFN is not used, the shortest CP will be sufficient in almost all propagation environments. If MBSFN consisting of more than a few cells are used, the CP of 16.7 μs should be used. In the case “high power high tower” sites are available or deployments in low frequency bands, good coverage can be achieved with even higher distance between sites and in this case the extended long CP of 33 μs should be used. In order to limit the relative overhead imposed by the CP, the OFDM core symbol duration is also doubled for the configuration with CP = 33 μs, and the subcarrier spacing is cut by half to keep the carrier bandwidth unchanged.

Based on the assumptions made by 3GPP, initial radio network simulations have been carried out. Results are presented in [54]. The key parameters are shown in Table 5.

Figure 22 shows the achieved broadcast spectrum efficiency versus inter-site distance for the case of multi-cell broadcasting. The spectral efficiency is valid for subframes allocated for MBSFN transmission. In Release 9 MBMS at most 6 out of every 10 subframes can be allocated to MBSFN transmission. A large cluster of cells transmits the same broadcast content synchronously thereby achieving signal aggregation gains and avoiding strong intercell interference. It can be seen that the spectrum efficiency requirement of 1 bps/Hz, is achieved for inter-site distances of up to approximately 1850 m inter site distance (ISD) for the 3GPP simulation scenario “Case 1”, that is, in the 2 GHz band and with 20 dB indoor penetration loss, and up to 4700 m for the “Case 4”, that is, in the 900 MHz band and with 10 dB indoor loss [53]. Naturally, the capacity decreases with increasing separation between transmitters as the power per transmitter is assumed to be fixed and the proportion of cells that are so far away that they cause interference rather than contribute useful signal increases.

9.4. PTP Bearer Performance of LTE and HSPA Evolution. In [15] throughput results for E-UTRA based on the agreed 3GPP current assumptions have been published where the radio resource is shared equally amongst all users. Here we are more interested in the capacity for streaming where all users need approximately equal throughput. The raw data that was used for [15] has been processed here to get an estimate of the streaming capacity per cell and 5 MHz as shown in Table 6, where the radio resource split between the users follows from the per user channel quality. The table is based on 95% satisfied users. This estimate is intended to show the technology potential, given, for example, perfect channel estimation, error-free feedback and disregarding transmitter impairments.

From these rates we calculate the number of channels for a given per channel rate. From this and 5% blocking we calculate the corresponding load, and from this and the load per subscriber follows the capacity in numbers of subscribers. For the assumption of 600 potential subscribers per cell (active and inactive) the results are plotted in Figure 23.

It can be seen that with LTE, even for 256 kbps and even if all potential subscribers actually subscribe to the service, each user can use the service on average for 25 minutes per 12 hours.

10. Conclusions

Mobile networks have evolved from voice telephony networks to multimedia delivery networks. Mobile TV services have become quite popular during the past two years. Apart from popular TV channels, the offerings often include special mobile editions with highlights from the weekly TV program, such as series and comedies, delivered as looped channels.

Although today’s technologies such as HSPA provide enough capacity for the introduction phase of new mobile TV services, it is foreseeable that the increasing popularity of
Table 5: Radio network simulation parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectrum allocation</td>
<td>10 MHz</td>
</tr>
<tr>
<td>Base station power</td>
<td>40 W</td>
</tr>
<tr>
<td>Propagation</td>
<td>$L = 35.3 + 37.6 \times \log(d)$, $d =$ distance in meters (includes $P_L = 20$ dB penetration loss)</td>
</tr>
<tr>
<td>Inter-site distance</td>
<td>500 to 4000 m (varied)</td>
</tr>
<tr>
<td>Modulation and coding schemes</td>
<td>Depending on inter-site distance, QPSK, 16QAM, or 64QAM, and turbo coding</td>
</tr>
</tbody>
</table>

Table 6: Streaming capacity per cell and 5 MHz.

<table>
<thead>
<tr>
<th>HSPA R6 1-Rx</th>
<th>HSPA R6 2-Rx</th>
<th>HSPA-evolution</th>
<th>LTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9 Mbps</td>
<td>2.9 Mbps</td>
<td>5.3 Mbps</td>
<td>6.7 Mbps</td>
</tr>
</tbody>
</table>

Figure 23: Estimate of addressable market for future UMTS evolution tracks.

those services might sooner or later lead to capacity bottlenecks. Therefore, 3GPP back in 2002 started the specification of broadcast/multicast services for GSM/UMTS.

MBMS adds broadcast and multicast capabilities to existing GSM/UMTS networks. The MBMS functionalities can be accessed through functions provided by the so-called Broadcast/Multicast Service Centre (BM-SC), a new logical node in the 3GPP architecture. Together with new Point-to-Multipoint radio bearers MBMS enables efficient stream and file delivery to large user groups with various error protection and recovery options. Related application and service layer functionality has been defined by the Open Mobile Alliance (OMA), in the Mobile Broadcast Services 1.0 (BCAST 1.0) specification, which was finalized in 2007.

MBMS has been specified such that broadcast/multicast services can be used together with voice and data services within the same radio carrier. This gives greatest flexibility to cellular operators. The various network configurations MBMS supports range from single multicast/broadcast transmission in a single cell up to a nationwide, single frequency network, broadcasting the same content (e.g., TV channels) across the whole country.

The specification of MBMS in the UMTS radio access network (UTRAN) enables the estimation of the approximate or exact number of UEs in a cell that are currently interested in a particular MBMS content channel. This allows the radio access network to make an optimal decision about if and how a broadcast service should be delivered in a particular cell; via P-t-P or P-t-M radio bearers, or not at all. Thus often the term “Enhanced Broadcast Mode” is used to refer to this particular MBMS configuration.

MBMS supports standardized, state-of-the-art audio and video codecs. The used video codec is H.264. Image resolutions up to QVGA $(320 \times 240$ pixel) and frame rates up to $30$ fps are supported. For audio, the High-Efficiency AAC version 2 codec, and the various AMR codecs (narrow band, wideband, and enhanced wideband) are supported.

Simulation results have shown, that an MBMS enabled UMTS 3GPP Release 6 network is able to deliver 20 mobile TV channels to a mobile TV subscriber base which corresponds to roughly 30% of all subscribers.

In 3GPP Release 7 and 8 the MBMS broadcasting capacity when delivering the same service over larger geographical areas (MBSFN) is drastically improved. If all cells use the same carrier frequency, the terminal is able to receive the signals from several cells simultaneously. By the elimination of intercell interference, a dramatic increase in spectral efficiency and thereby throughput on the used radio resources is achieved. MBSFN is an enhancement available for both, networks based on WCDMA technology and LTE. With LTE it is even possible to time multiplex MBSFN and unicast transmission on the same carrier frequency, thereby providing the best possible integration of high performance broadcasting and individual user service delivery.

References


[50] 3GPP TR 25.913, “Requirements for Evolved UTRA and Evolved UTRAN”.


[53] 3GPP TR 25.814, “Physical Layer Aspects for Evolved UTRA”.