5G Evolution for Multicast and Broadcast Services in 3GPP Release 17

Vinay Kumar Shrivastava, Sangkyu Baek, and Youngkyo Baek

ABSTRACT

5G communications systems are on the evolution path with present focus on enabling advanced features and new service capabilities. Multicast and Broadcast Services (MBS) are being considered as one of the most promising use cases of 5G. As introduced to Release 17 of 3rd Generation Partnership Project (3GPP) 5G standards, MBS work item has been developed over existing 5G framework. MBS standardization is targeted to enhance 5G New Radio and 5G Core Network capabilities for serving reliable, low latency, resource efficient, and massively deployable multicast and broadcast services from both architectural and transmission perspectives. In this article, we present an introduction to MBS standardization and outline newly introduced technical features, and their use cases.

INTRODUCTION

Multicast and broadcast communications have been considered to pave the way for resource efficient transmission to multiple end users which require to receive same contents. Due to this obvious gain with a single transmission for multiple users, 3GPP’s Long Term Evolution (LTE) standards have introduced various features to support the broadcast services over more than a decade. Historically, evolved Multimedia Broadcast and Multicast Service (eMBMS) in LTE Release 9-12 was an initial set of standardized features directed towards dissemination of broadcast contents, as detailed in [1] and [2]. eMBMS is based on Single Frequency Network (SFN), and it utilizes synchronized multi-cell transmissions from many eNBs (evolved NodeBs i.e., base stations in 3GPP network) together providing for over-the-air combining of Multicast-Broadcast SFN (MBSFN) signals, enhancing the reliability and coverage areas for services. MBSFN transmissions are time-interleaved with unicast transmission with pre-assigned and dedicated sub-frames over the radio frame and utilize full system bandwidth.

To overcome the limitations of eMBMS, Single Cell Point-To-Multipoint (SC-PTM) was introduced in 3GPP Release 13 [3]. SC-PTM adopted a more flexible approach where dynamic time and frequency resource utilization (even within a sub-frame) is possible for broadcast services over a small dense geographical region (e.g., hotspot) within a single-cell coverage. The dynamic resource utilization effectively allows integration of broadcast services delivery with unicast physical channels [4].

3GPP Release 14/15 targeted broadcast services with dedicated carrier for MBMS transmission supporting up to 100 percent resource usage for MBMS and a large duration Orthogonal Frequency Division Multiplexing (OFDM) symbol to support wider coverage in order of tens of kilometers. 3GPP Release 16 introduced ‘Terrestrial Broadcast’ targeting Enhanced Television (EN-TV) services over large and static transmission areas with dedicated broadcast infrastructure, for example, High-Power High-Tower (HPHT) deployments [5]. Despite several enhancements across different releases, LTE eMBMS and SC-PTM lacked on reliability and latency aspects for multicast and broadcast services with no mechanism in place for error correction and packet retransmission.

Meanwhile, fifth generation wireless communications technology (5G) has been standardized by 3GPP and commercially launched in 2019. However, 5G did not address multicast and broadcast service in its first phase (3GPP Release 15) [6] and second phase (3GPP Release 16) [7], with initial focus being on enhanced Mobile Broadband (eMBB) and Ultra-Reliable and Low-Latency Communication (URLLC). Consequently, in third phase of 5G (3GPP Release 17), 3GPP started to build functional support of multicast and broadcast services abbreviated as MBS over an existing 5G standards framework. The standardization has been conducted for overall 5G system architecture from both Next Generation Radio Access Network (NG-RAN) and 5G Core Network (5GC) perspectives.

Although a natural adoption and evolution of multicast and broadcast services in 5G standards was expected, it also posed new requirements and challenges for 5G MBS. New emerging multicast services in MBS such as mission-critical delay-sensitive signaling and high-resolution IPTV require to achieve same levels of high reliability and low latency as available with unicast services. This demands for an involved protocol stack design with layers and functionalities reinforcing on the reliability and latency aspects in RAN protocol, physical layer and service continuity. On the contrary, broadcast services in MBS can maintain broadly same requirements like LTE eMBMS and SC-PTM, and therefore, inherit many design features.

Also, unique 5G New Radio (NR) characteristics such as Bandwidth Part (BWP), beamforming, absence of always-on reference signals (i.e., Cell-specific Reference Signal, CRS, in LTE), varied Sub-Carrier Spacing (SCS), non-SFN deployments...
and new transmission aspects significantly influence NR MBS design. In 5GC, wider-area service with using not only new MBS functions but also existing network function is further required.

In this regard, key features of the standardized 5G MBS are as follows:
- Group scheduling mechanism to allow User Equipments (UEs) to receive MBS service including simultaneous operation with unicast reception
- Multicast/broadcast delivery in 5GC, i.e., shared delivery
- Reliability enhancements by dynamic change of multicast service delivery between Point-to-Multipoint (PTM) and Point-to-Point (PTP), Automatic Repeat Request (ARQ), and Hybrid Automatic Repeat Request (HARQ), etc.
- Supporting service continuity and lossless handover
- Reception of broadcast data irrespective of UE’s Radio Resource Control (RRC) states
- MBS over legacy network node, for example, 3GPP Release 15/16 network.

5G MBS discussion in 3GPP Working Groups (WGs) progressed with a targeted completion of functional specification by March 2022. Figure 1 depicts the evolution of 5G MBS from legacy LTE and involved aspects. The remainder of this article provides comprehensive overview of each aspect of 5G MBS from service requirements to functional enhancements.

**NEW SERVICE REQUIREMENTS**

MBS has two different cast types, namely, multicast and broadcast. The multicast is defined as the same service and the same specific content data that are provided simultaneously to a dedicated set of UEs which are authorized in the service coverage. The broadcast is defined as the same service and the same specific content data that are provided simultaneously to all UEs in the service area. Which cast type is actually used depends on actual service type. MBS is targeted to enable diverse services including public safety and mission critical, Vehicle-to-Everything (V2X) applications, transparent IPv4/IPv6 multicast delivery, IPTV, software delivery over wireless, group communications and Internet-of-Things (IoT) applications. Service requirements for varied uses cases of MBS differ in their requirements for reliability, latency, Quality of Service (QoS) handling, service area coverage, service continuity and security aspects. Therefore, it becomes imperative to build a comprehensive mechanism for MBS to address these diverse needs.

**Reliability and latency:** For MBS, high reliability and low latency services (e.g., mission critical delay-sensitive signaling with packet delay budget of 60ms and packet error rate of $10^{-6}$ present extreme service requirements as shown in Table 1 [8]. Therefore, there is a new need for dynam-
initial deployments, service continuity across legacy network nodes is an essential requirement and may involve transition in service delivery methods (e.g., unicast based access for MBS services).

**Security:** To ensure secure communication for certain MBS services (e.g., software delivery, mission critical), a security functionality is required which provides for session security context, access credentials and their maintenance across involved network nodes and UEs. Unlike unicast security with Protocol Data Convergence Protocol (PDCP) layer based mechanism, application/service layer security provisioning transparent to NG-RAN and radio protocols is adopted for MBS, in order to avoid any impact on NG-RAN and provide more deployment flexibility to network operators.

**NETWORK ARCHITECTURE**

Starting with 3GPP Release 17, the network architecture of 5G system (5GS) is enhanced to support multicast and broadcast services, with the design aimed to re-use legacy 5G system as much as possible. However, to support MBS in 5GS, some network functions are newly introduced as part of architectural enhancements for 5G multicast/broadcast services’ work item in Service Architecture (SA) group in 3GPP and are depicted in Fig. 2 [10]. These new network functions are as follows:

- **Multicast Broadcast User Plane Function (MB-UPF)** is an ingress point to 5GS and works as a session anchor to 5GS.
- **Multicast Broadcast Session Management Function (MB-SMF)** manages MBS session and configures a user plane function MB-UPF, based on the policy rules for multicast and broadcast services.
- **Multicast Broadcast Service Function (MBSF)** has service level functionality to interact with AF/AS (Application Function/Application Server) and MB-SMF for MBS session operations. Further, it determines transport parameters and session transport, and control MBSTF if used, which can be implemented in the Network Exposure Function (NEF).
- **Multicast Broadcast Service Transport Function (MBSTF)** has generic packet transport functionalities available to any IP multicast enabled application such as framing, multiple flows, packet encoding, and therefore, works as media anchor for MBS data traffic. Moreover, existing network functions are enhanced for MBS. Session Management Function (SMF) is enhanced to handle UE’s multicast session join or leave and to determine the MBS traffic delivery method to use. Access and Mobility Management Function (AMF) is enriched to select appropriate NG-RAN nodes for broadcast and group notification of multicast session activation towards UEs. User Plane Function (UPF) is enhanced to deliver multicast data to UEs via PDU session. Policy Control Function (PCF) is upgraded to support QoS handling for MBS session.

On top of the enhanced network architecture, two delivery methods are supported to deliver the MBS data traffic from MB-UPF to NG-RAN, namely, 5GC Shared Delivery (SD) method and 5GC Individual Delivery (ID) method.

SD is used to save 5GC resource for MBS data transmission. More specifically, when an MB-UPF has a packet to transmit to multiple UEs receiving data of MBS session, it just sends a single copy of MBS data packet to each NG-RAN node. This single packet is dedicated to a single UE but ultimately shared by multiple UEs. Obviously, only NG-RAN nodes with MBS capability can receive data by SD whereas NG-RAN node which does not support MBS capability, for example a legacy gNB (i.e., base stations in 3GPP 5G-NR network), cannot receive data by SD. SD can be used for both multicast services and broadcast services.

In ID, when an MB-UPF has a packet to transmit to multiple UEs receiving data of MBS session, a single copy of MBS data packet is delivered to the UPF and which, in turn, sends separate copies corresponding to individual UEs. Although ID seems to be less efficient from resource consumption perspective, it is essential for cases where in-NG-RAN node does not have MBS capability of receiving data by SD but data transmission of MBS session is still required. For instance, in a mobility scenario, a UE which is receiving MBS data packets can move to other gNB’s coverage not supporting MBS capability; network switches its delivery method to ID for service continuity. Since ID is destined to a particular UE, it is applied to multicast services only.

When the content provider or AF wants to disseminate MBS services, it may request the 5GS to configure the MBS session via NEF (or MBSF). Information for MBS service reception can be provided to UE via service announcement from AF. Such information consists of service type, MBS ID (MBSTF), MBSTF identity, Temporary Mobile Group Identity (TMGI), IP multicast address and so on.

For multicast service reception, UE joins an associated MBS session identified by the TMGI. Thereafter, transmission resources are established for the multicast data transmission to the serving NG-RAN node and to UE, wherein if SD is used,
A main difference from unicast is to support split MRB where a PDCP entity is linked to one PTP RLC and one PTM RLC where gNB may send MBS packets via either PTP RLC, PTM RLC or both based on the required reliability.

**Figure 3.** Protocol architecture for NR MBS.

The broadcast mode does not require any interaction between UE and gNB. The network does not have any feedback from UE side on transmission status (e.g., ACK/NACK) but it transmits the data only in the best-effort manner. Hence, QoS cannot be guaranteed and only low-QoS services are feasible. Also, the broadcast mode does not mandate RRC state transition.

**Protocol Architecture**

NR MBS supports two delivery modes in RAN, namely, multicast mode and broadcast mode, in order to support multicast services and broadcast services, respectively. Each mode has its own characteristics and target services.

Target services of multicast mode have particular QoS requirements which the network should guarantee as in case of unicast. Therefore, UEs receiving multicast data are required to stay in CONNECTED state and a dedicated RRC signaling provides the radio resource configuration including MBS Radio Bearer (MRB) configuration, physical layer configurations and so on, which can be optimally configured based on interaction between UE and gNB. Moreover, in case that reliable transmission is required for cell edge user with bad channel quality, the PTM transmission can be switched to PTP transmission where PTM Radio Link Layer (RLCL) establishes a dedicated transmission path between gNB and UE, and utilizes ARQ to enhance the performance as in unicast transmission.

When no multicast data arrival is expected, the multicast session can be deactivated and UEs belonging to the multicast group can transit to INACTIVE or IDLE state. These UEs have to re-enter CONNECTED state when the multicast session is about to be re-activated. In this case, group paging with the corresponding TMGI is used to notify these UEs.

On the contrary, broadcast mode can be provided to all UEs within a coverage regardless of RRC states. The broadcast mode is a similar mechanism to SC-PTM in LTE. In order to deliver the broadcast data for UEs that are out of CONNECTED state, radio resource configuration for broadcast mode is periodically transmitted via MBS Control Channel (MCCH) from which UEs apply the received configuration for MBS Traffic Channel (MTCH).

The broadcast mode does not require any interaction between UE and gNB. The network does not have any feedback from UE side on transmission status (e.g., ACK/NACK) but it transmits the data only in the best-effort manner. Hence, QoS cannot be guaranteed and only low-QoS services are feasible. Also, the broadcast mode does not mandate RRC state transition.

**Figure 3.** Protocol architecture for NR MBS.

![Protocol Architecture Diagram](image-url)
For broadcast reception supported in all RRC states, Broadcast CFR is configured for receiving group-common PDCCH and PDSCH, which carries data from MCCH and MTCH logical channels, and is confined within the Initial BWP (or same as Initial BWP).

| Scenario | A | B | C |
|----------|---|---|---|
| RRC State | RRC_IDLE/INACTIVE | RRC_CONNECTED | RRC_CONNECTED |
| Initial BWP | | | |
| Frequency | | | |
| Cast Type | Broadcast Only | Broadcast & Multicast | Multicast Only |
| BWP | - MBS CFR is same or a portion of initial BWP. - Multicast is not supported in RRC_IDLE/INACTIVE. | - MBS CFR for multicast and MBS CFR for broadcast are confined in unicast BWP. | - MBS CFR for broadcast and unicast BWP are not overlapped - Broadcast may not be receivable. |

**Figure 4.** An example for BWP/CFR operation and usage for MBS broadcast and multicast reception.

A UE receiving MBS data, in addition to a common DRX configuration for all unicast reception. MAC manages relevant physical layer schemes for MBS, for example, MBS Semi-Persistent Scheduling (SPS) and HARQ operation. Unlike unicast SPS utilizes CS-RNTI for SPS activation, deactivation and retransmission, MBS SPS is controlled by both unicast signaling by CS-RNTI and multicast signaling by G-CS-RNTI.

**Physical Layer Aspects**

One of the most important NR physical layer characteristics influencing MBS design is BWP operation in which multiple UEs can commonly perform MBS reception including group-common Physical Downlink Control Channel (PDCCH) and Physical Downlink Shared Channel (PDSCH) [13]. MBS defines Multicast CFR and Broadcast CFR to enable reception of multicast and broadcast respectively. Multicast CFR is confined within the UE’s dedicated active unicast BWP to support simultaneous reception of unicast and multicast in the same time slot. Starting PRB (Physical Resource Block), length of PRBs and associated group-common PDCCH and PDSCH configurations specify the Multicast CFR. It is worth noting that SCS and CP (Cyclic Prefix) for CFR are same as that for active unicast BWP in order to avoid BWP switching or additional hardware receiver capability. The multicast mode is suitably provisioned along with unicast reception with regard to BWP, channel measurement, Channel State Information (CSI) reporting. For broadcast reception supported in all RRC states, Broadcast CFR is configured for receiving group-common PDCCH and PDSCH, which carries data from MCCH and MTCH logical channels, and is confined within the Initial BWP (or same as Initial BWP).

Figure 4 illustrates the BWP operation depending on reception scenarios of UE. A UE switches BWP with the change of RRC state between IDLE/INACTIVE and CONNECTED. Moreover, BWP switching may also occur within CONNECTED state. Accordingly, UE can be in one of the scenarios as:

- **Scenario A:** UE is in IDLE/INACTIVE state and can only receive broadcast. Broadcast CFR can be confined within Initial BWP or can be same as Initial BWP.
- **Scenario B:** UE is in CONNECTED state and can receive both broadcast and multicast if both CFRs are confined in active BWP.
- **Scenario C:** UE is in CONNECTED state and can receive multicast with multicast CFR confined in active BWP. UE may not receive...
When UE changes its serving cell due to the mobility (e.g., handover), the support of service continuity is important for users not to experience interruption or performance degradation of the ongoing services. The biggest change of 5G MBS in mobility and service continuity compared to the previous generation is packet-level service continuity with lossless mobility in multicast mode as depicted in Fig. 5. This packet-level service continuity is based on packet-level sequence number (SN) synchronization where the same packet has the same PDCP SN over the area supporting the service continuity.

In unicast, PDCP SN for a signaling radio bearer (SRB) or data radio bearer (DRB) of a UE can be continued during handover, utilizing handover preparation procedure with indication of the current SN status and data forwarding from serving gNB to target gNB [11]. In multicast, multiple UEs over the large service area would simultaneously receive packets of an MRB. Therefore, the packet-level SN coordination among all gNBs of the service area is required at the beginning of MRB initialization. The SN synchronization naturally enables UEs to maintain the configured MRBs during the handover. The handover preparation requires only context transfer of the UE which is about to handover to the target gNB. Unlike unicast, all MBS configurations used after the handover are provided by the target gNB.

Since the connection with the source gNB is released at the handover, the packet loss could occur but this can be recovered through retransmission by the target gNB. This recovery is similar to the unicast handover where the target gNB selectively retransmits packets which are not yet successfully delivered to the UE based on PDCP status report [12]. In other words, service continuity with lossless handover to significantly reduce the interruption time up to the same level as unicast, around a few milliseconds, can be eventually supported for multicast. Note that it was not achievable with LTE eMBMS and SC-PTM.

Handover to gNB not supporting MBS features, for example, involving legacy gNB, is a possible scenario considering backward compatibility. In this case, the ID method is employed instead of the SD method for MBS traffic delivery during the handover so that the serving gNB needs to change the ongoing MRB to unicast DRB in advance for the target node to understand the transferred UE context. In case the target gNB does not establish the moving UE’s ongoing MRBs but support MBS, the service continuity with lossless handover can still be supported by the context transfer and establishment of the MRBs.

In broadcast mode, network does not accurately know which UEs are receiving or interested
in the broadcast service. Hence, the packet-level service continuity of multicast mode is not reused for broadcast mode, and only low-level service continuity without lossless handover can be possible. MCCH provides the broadcast configurations possibly including neighbor cell information which UE can directly apply for cell reselection. Also, when the serving gNB does not provide UE’s interested broadcast services, UE can send an MBS Interest Indication (MII) message to request UE’s interested services. Although these mechanisms have some latency of hundreds of milliseconds, the ongoing broadcast services can be continued with low QoS requirement.

Conclusions and Future Directions

In this article, we provided an overview of 5G MBS, its key technical features and standardization progress in 3GPP Release 17. Diverse service requirements and evolving network architecture aspects for MBS were discussed. Certain stringent MBS service requirements for high reliability and low latency led to involved radio protocol layers and functions for MBS operation. While NR physical layer aspects and unique transmission characteristics influenced MBS design, new approaches for MBS CFR, HARQ ACK/NACK feedback and retransmission were realized. Lastly, mobility and service continuity ensuring lossless and seamless MBS service reception were introduced.

It is anticipated that 3GPP Release 17 version of MBS may not be able to address many aspects due to lack of time, for example, multicast data reception is presently restricted to CONNECTED state UEs only. Further, MBS feature is expected to continue evolving in future. In 2022, 3GPP started to study and standardize Release 18 (a.k.a. 5G-Advanced) MBS targeting certain objectives, namely:

- Support multicast data reception in INACTIVE state to cater large number of UEs and possible enhancement for mission critical services
- Improvement of resource efficiency for MBS reception in RAN sharing scenarios
- RAN signaling enhancements to support sharing for MBS broadcast and unicast reception from same or different network operators
- Support for On-demand multicast MBS session triggered by AF, and efficient resource utilization via 5GC choosing multicast and/or unicast delivery for a certain service
- Support group message delivery for capability-limited devices, including NEF enhancement, coexistence of existing power saving mechanisms and MBS

To conclude, MBS is advancing on an interesting journey of technical progress, standardization and market deployments with 5G-Advanced.

References

[1] D. Lecompte, and F. Gabin, “Evolved Multimedia Broadcast/Multicast Service (eMBMS) in LTE-Advanced: Overview and Rel11 Enhancements,” IEEE Commun. Mag., vol. 50, no. 11, Nov. 2012, pp. 68–74.
[2] J. Kim et al., “Group Communication over LTE: A Radio Access Perspective,” IEEE Commun. Mag., vol. 54, no. 4, Apr. 2016, pp. 16–23.
[3] 3GPP TR 36.890, Study on Single-Cell Point-to-Multipoint Transmission for E-UTRA, v13.0.0, July 2015.
[4] Y. Zhang et al., “MBSFN or SC-PTM: How to Efficiently Multicast/Broadcast,” IEEE Trans. Broadcasting, vol. 67, no. 3, Sept. 2021, pp. 582–911.
[5] J. J. Gimenez et al., “Enhanced TV Delivery with EMBMS: Coverage Evaluation for RoofTop Reception,” Proc. IEEE Int’l Symp. Broadband Multimedia Syst. Broadcast (BMSB), 2018, pp. 1–5.
[6] S. Parkvall et al., “NR: The New 5G Radio Access Technology,” IEEE Commun. Standards Mag., Dec. 2017.
[7] Y. Kim et al., “New Radio (NR) and Its Evolution Toward 5G-Advanced,” IEEE Wireless Commun., vol. 26, no. 3, June 2019, pp. 2–7.
[8] 3GPP TS 23.501, System Architecture for the 5G System (5GS), Stage 2, v17.3.0, Dec. 2021.
[9] 3GPP TR 26.881, Study on Forward Error Correction (FEC) for Mission Critical Services, v15.0.0, Mar. 2018.
[10] 3GPP TS 32.247, Architectural Enhancements for 5G MultiCast Broadcast Services, Stage 2, v17.1.0, Dec. 2021.
[11] E. Dahlman, S. Parkvall, and J. Skold, “5G NR – The Next Generation Wireless Access Technology,” Oxford, U.K.: Academic Press, 2018.
[12] 3GPP TS 38.323, Packet Data Convergence Protocol (PDCP) specification, v16.6.0, Dec. 2021.
[13] K. Takeda et al., “Understanding the Heart of the 5G Air Interface: An Overview of Physical Downlink Control Channel for 5G New Radio,” IEEE Commun. Standards Mag., vol. 4, no. 3, Sept. 2020, pp. 22–29.
[14] X. Lin et al., “5G New Radio: Unveiling the Essentials of the Next Generation Wireless Access Technology,” IEEE Commun. Standards Mag., vol. 3, no. 3, Sept. 2019, pp. 30–37.
[15] 3GPP TS 38.300, NR and NG-RAN Overall Description Stage 2, v16.8.0, Dec. 2021.

Biographies

Vinay Kumar Shrivastava (S’06, M’12) (shrivastava@samsung.com) received his Master’s degree in Telecommunications Systems Engineering from Indian Institute of Technology Kharagpur, India, in 2003 and Bachelor’s degree in Electronics Engineering from Rajiv Gandhi Proudyogiki Vishwavidyalaya, India, in 2001. He joined Samsung Electronics in 2003 and has worked extensively on design and development of 3G/4G/5G modems. He is currently a staff engineer at Samsung Research Institute, Bengaluru, India, and a Samsung 3GPP standardization delegate in RAN2 Working group. His research interests include communications standards, MAC protocols, broadcast communication, device-to-device communications.

Sangkyu Baek (S’06, M’12) (sangkyu.baek@samsung.com) received his B.S., M.S. and Ph.D. degrees in mathematics from Korea University, Seoul, Korea, in 2006, 2008, and 2011, respectively. Since 2012, he has been with Samsung Electronics and working on advanced 5G research and pre-5G standardization. He is currently a principal engineer at Samsung Research, Samsung Electronics, Seoul, Korea, and a Samsung 3GPP standardization delegate in RAN2 Working group. His research interests include communications standards, MAC protocols, broadcast communication, device-to-device communications.

Youngnho Baek (youngnho.baek@samsung.com) received an M.S. degree, in 2002, and finished doctoral course in mathematics from Korea University, Seoul, Korea. He joined Samsung Electronics in 2007 and has worked on 3G/4G/5G standardization especially in WMAN and 3GPP. He is currently a staff engineer at Samsung Research, Samsung Electronics, Seoul, Korea, and a 3GPP standardization delegate in SA2 Working group. His research interests include communications standards, non-public network, broadcast communication, proximity services, and queuing theory.