Enabling Multicast and Broadcast in the 5G Core for Converged Fixed and Mobile Networks

Tuan Tran, David Navrátil, Peter Sanders, Jon Hart, Roman Odarchenko, Carlos Barjau, Baruch Altman, Christophe Burdinat, and David Gomez-Barquero

***Abstract*—The fifth generation of cellular technologies, 5G, promises to deliver a wide variety of services. However, the first release of 5G (3GPP Rel-15) did not consider multicast capabilities in the 5G system including both access and core networks. This article fills the gap by describing solutions for 5G mobile core network architecture incorporating multicast and broadcast capabilities. Two innovative solutions to provision multicast and broadcast, based on 3GPP Rel-15 arechitecture, have been proposed and analyzed. One is 5G architecture friendly with low imprint over Rel-15 while the other is based on the evolution of LTE Broadcast architecture. The architectural solutions feature two new main transmission modes: transparent multicast transport and point-to-multipoint services. The former enables the 5G network as a network pipe supporting multicast traffic providing a cost-effective delivery mode for multimedia while the latter enables multicast and broadcast as a service. The paper also presents a convergent fixed wireline and mobile network architecture based on the proposed mobile network architecture altenatives. The proposed architecture alternatives are suitable for different verticals and applications such as Digital Terrestrial Television, Public Warning, Internet of Things, V2X (vehicle to everything) and mission critical communications (MCC).**

***Index Terms*—5G, 5GC, broadcast, MBMS, multi-access edge computing (MEC), core network, MooD, multicast, multilink, point-to-multipoint.**

I. INTRODUCTION

THE FIRST 3GPP release of 5G technology (Release 15 or Rel-15), also known as 5G NR (New Radio), was completed in 2018. Rel-15 was structured in three phases. The first phase addressed a non-standalone (NSA) version that requires LTE EPC (Evolved Packet Core) for the control plane (CP). It introduced 5G NR to enhance the user plane (UP) performance and efficiency using dual connectivity across the LTE and NR bands. In the second phase, the stand-alone (SA) version of 5G was standardized, including the 5G core network (5GC), that enabled deployments without any LTE infrastructure. The last phase enabled more architecture options for hybrid LTE and 5G NR deployments using the 5GC. It enabled using the 5GC to inter-work with both LTE RAN (Radio Access Network) and NG-RAN (Next Generation RAN), using the NG-RAN for the control plane.

3GPP Rel-15 only supports unicast communications (UC) in the core network (CN) and point-to-point (PTP) transmissions in the RAN. This limitation implies an inefficient service provisioning, and utilization of the network and spectrum resources, when distributing the same data to multiple users and devices. Nevertheless, 3GPP system requirements for the 5G system realize the need for flexible multicast/broadcast services [1], since it is considered as an essential feature for 5G applications in multiple vertical sectors; such as: media and entertainment (M&E), automotive, Internet-of-Things IoT (machine-type communications), and Public Warning (PW) & Safety (mission critical communications - MCC) [2], [7]. Another vertical that also requires point-to-multipoint (PTM) transmissions is Airborne Communications (e.g., drone communications).

5G multicast/broadcast was one of the topics that is under discussion at 3GPP for 5G phase II (Rel-16). Two different tracks were identified: (i) "Terrestrial Broadcast" and (ii) "Mixed Mode Multicasting" [3]. "Terrestrial Broadcast" enables a dedicated downlink-only and broadcast-only network suitable for Digital Terrestrial Television. For the "Terrestrial Broadcast" track, 3GPP uses the LTE-based Rel-14 EnTV (Enhanced TV) work [4] as a basis with Rel-16 enhancements to meet the 5G requirements for multimedia broadcast services [1], [5], [6]. EnTV is based on the enhanced Multicast/Broadcast Multimedia Service (eMBMS), an extension of the LTE core and radio to provide PTM capabilities [51]. Other notable evolutions of eMBMS include [53], [54], [55]. On the other side, "Mixed Mode Multicasting" uses NR as a basis, and allows for dynamic mode switching between unicast PTP and multicast PTM.

This article presents the work done in the H2020 5G-PPP 5G-Xcast research project on mobile core network architecture [9] that introduces multicast and broadcast

Manuscript received December 1, 2019; revised February 17, 2020; accepted February 26, 2020. Date of publication May 22, 2020; date of current version June 5, 2020. This work was supported in part by the European Commission through the 5G-PPP Project 5G-Xcast (H2020-ICT-2016-2 call) under Grant 761498. *(Corresponding author: Tuan Tran.)*

Tuan Tran and Christophe Burdinat are with Enensys, 75007 Paris, France (e-mail: tuan.tran@enensys.com; christophe.burdinat@enensys.com).

David Navrátil is with Nokia Bell-Labs, 02610 Espoo, Finland (e-mail: david.navratil@nokia-bell-labs.com).

Peter Sanders is with One2many, 7411 CL Deventer, The Netherlands (e-mail peter.sanders@one2many.eu).

Jon Hart is with British Telecom, Suffolk IP5 3RE, U.K. (e-mail jonathan.hart@bt.com).

Roman Odarchenko is with Bundleslab, 1117 Budapest, Hungary (e-mail roman@bundleslab.com).

Carlos Barjau and David Gomez-Barquero are with the iTEAM Research Institute, Universitat Politècnica de València, 46022 Valencia, Spain (e-mail: carbare1@iteam.upv.es; dagobar@iteam.upv.es).

Baruch Altman is with LiveU Ltd., Kefar Sava 4442211, Israel (e-mail: baruch@liveu.tv).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2020.2991548

capabilities in the 5GC. The proposed solutions provide PTM capabilities in 5G as built-in delivery features for network optimization, integrating both PTP and PTM modes under one common framework and enabling dynamic use of PTM to maximize deivery network and radio spectrum efficiency [7], [8].

The proposed mobile core network architecture was designed upon the 3GPP 5G network architecture while Rel-15 was being standardized. The design of mobile core network is aligned with 3GPP key design principles and concepts for the 5GC, such as service-based interface between control plane network functions (NFs), function separation and modularization, and the control and user plane separation (CUPS) paradigm introduced in LTE Rel-14 [52], where the control logic is decoupled from the data flow and interconnected using standardized interfaces, providing increased scalability and flexibility. Two different architecture alternatives are proposed. One provides an approach that is 5G architecture friendly. The other provides minimal changes to the eMBMS architecture and specification described for LTE. Both alternatives aim at enabling PTM communication capabilities as a built-in optimization feature of the core network.

The rest of this article is structured as follows. Section II provides a high-level description of eMBMS architecture as a background on how multicast and broadcast communications are currently supported by an LTE system. Section III presents a brief overview of the 3GPP 5G core network architecture in Rel-15. Section IV presents the design principles to construct core network architecture enabling multicast/broadcast capabilities. Section V describes the new functionalities and technologies while Section VI describes the required building blocks, network functions. Sections VII and VIII introduce the proposed mobile core network architectures and analysis, respectively. Section IX describes the application of the proposed architectural alternatives to different verticals and applications. Section X presents the converged core network architectural alternatives based on the proposed mobile core network architectures. Section XI concludes the article and discusses its outcomes to standard organization such as 3GPP.

II. OVERVIEW OF eMBMS ARCHITECTURE

eMBMS is able to provide multicast and broadcast multimedia services through the LTE network, combining unicast with multicast/broadcast data in the same LTE radio frame as specified in [10], [11]. In Fig. 1, the BM-SC (Broadcast Multicast Service Centre) provides functions for MBMS User Service provisioning and delivery to the content provider. It can also serve as an entry point for MBMS data traffic from the MBMS User Services. On the data plane, the MBMS-GW (MBMS Gateway) receives the IP multicast traffic from the BM-SC and forwards it to the eNodeB(s).

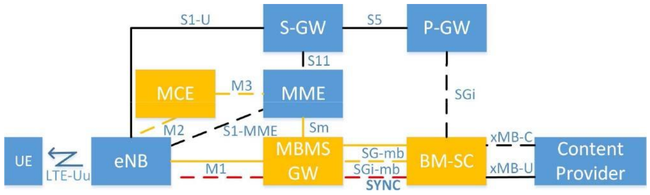

Fig. 1. eMBMS architecture in LTE.

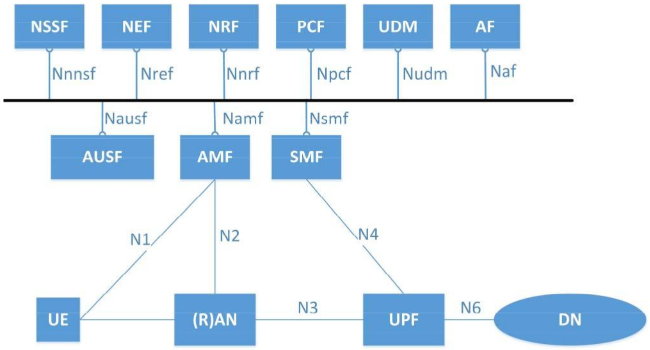

Fig. 2. Non-Roaming 5G System Architecture in service-based representation.

On the control plane, the MBMS-GW sends MBMS Session Control Signaling towards the downstream node (i.e., Mobility Management Entity - MME) which routes the signaling messages to the MCE (Multi-cell/multicast Coordinating Entity) serving the broadcast area. The MCE provides the admission control and the allocation of the radio resources used by all eNodeBs in the target area. The MCE also decides the radio configuration such as modulation and coding scheme for both control and data.

III. 5G ARCHITECTURE IN 3GPP REL-15

The work in 3GPP Rel-15 started with the studies on new services and markets technology enablers for next generation mobile telecommunications. The studies focused on the three domains of 5G: massive IoT, critical communications (known as ultra-reliable low latency communications) and enhanced mobile broadband [12], [13], [14].

The architecture of 5G system has been defined by 3GPP since Rel-15. The architecture model of the CN has been significantly changed compared to the previous generations of 3GPP systems (e.g., LTE, 3G). This model has been developed considering some key design concepts and networking technologies such as modularity, reusability, network function virtualization (NFV), software defined networking (SDN), CUPS, minimizing dependencies between access network (AN) and CN, or application support including local and centralized services. As a result, the 5G architecture model is defined as service-based architecture where its elements are defined as NFs that offer services to other NFs via well-defined interfaces in a common framework. The 5G architecture for non-roaming scenario is shown in Fig. 2 where the service-based interfaces are depicted by a line that interconnects the 5GC NFs [15]. Fig. 3 shows the 5G architecture in a reference point representation that shows direct interactions between NFs. Another feature of the 5G architecture is the separation of compute and storage resource that can be optionally enabled by deploying Unstructured Data Storage Function (UDSF) residing in Unified Data Management (UDM), which offers storage and retrieval of information (e.g., NF stores the context of registered user equipments (UEs)). The separation of compute and

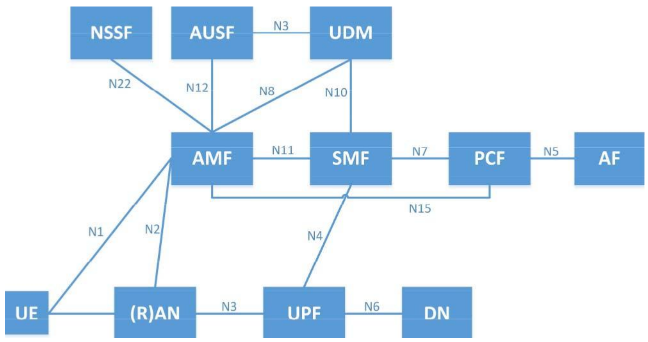

Fig. 3. Non-Roaming 5G System Architecture in reference point representation.

storage resource simplifies the procedure of changing NFs that serve a UE (e.g., Access and Mobility Management Function - AMF), and improves system's resilience and load balancing. The procedures over the 5G architecture are described in [16], [17].

The 5G core network has been designed to operate with different ANs. The key enabler for this operation is the generalized design of the interface between the AN and the CN. In Rel-15, the CN can serve 3GPP and non-3GPP defined ANs which are NG-RAN and untrusted WLAN (Wireless Local Access Network) access, respectively. The architecture allows a UE to be simultaneously connected to the same CN over 3GPP and non-3GPP ANs as described in Section X. In this case, UE is served by a single AMF, which allows for seamless mobility between 3GPP and non-3GPP access networks.

In Fig. 2, a UE could request AMF to establish a Packet Data Unit (PDU) session via N1. The AMF authenticates the UE with Authentication Server Function (AUSF), checks which policies are applicable for this UE with Policy Control Function (PCF), and requests available slices with Network Slice Selection Function (NSSF). Subsequently, AMF requests Session Management Function (SMF) to establish the session and SMF provides a tunnel over which the UE can exchange PDUs with the Data Network (DN) (e.g., the Internet) via User Plane Function (UPF). The UE context is stored in UDM.

An AF is an application function (AF) and XCF (as introduced in Section VII) can be an instantiation of an AF.

Previous generations of mobile systems are typically deployed as a single network. This approach is not suitable for 5G that needs to meet diverse requirements of various verticals. Network slicing is an important concept, a network slice comprises of CN, AN features and functionalities that are necessary to meet diverse requirements of a specific deployment or scenario (e.g., mobile broadband, massive IoT, ultra-reliable low-latency communication). A network slice is then deployed on common hardware and networking infrastructure as a dedicated and customizable network.

IV. DESIGN PRINCIPLES

The mobile network architecture proposed in this article was built based on the 3GPP 5G network whose first release was being standardized. 3GPP TS 23.501 [15] provides some key principles and concepts as follows:

- Separate the UP functions from the CP functions. This principle allows independent scalability, evolution and flexible deployments (e.g., centralized or distributed).
- Modularize the function design (e.g., to enable flexible and efficient network slicing).
- Wherever applicable, define procedures (i.e., the set of interactions between network functions) as services.
- Enable each NF to interact with other NF directly if required.
- Minimize dependencies between the AN and the CN. The architecture is defined with a converged core network with a common AN - CN interface which integrates different access types (e.g., 3GPP and non-3GPP access).
- Support a unified authentication framework.
- Support "stateless" NFs, where the "compute" resource is decoupled from the "storage" resource.
- Support network capability exposure function (via Network Exposure Function - NEF).
- Support concurrent access to local and centralized services. To support low latency services and access to local data networks, UP functions can be deployed close to the access network.
- Support roaming with both home routed traffic as well as local breakout traffic in the visited PLMN (Public Land Mobile Network).

The design principles in this article are aligned with 3GPP direction with additional principles related to multicast and broadcast capabilities from a core network perspective as follows:

- Enabling multicast and broadcast capabilities should require a small footprint on top of the existing unicast architecture.
- Wherever possible, treat multicast and broadcast as an internal optimization tool inside the network operator's domain.
- Consider terrestrial broadcast as a service offered also to UEs without uplink capabilities that can be delivered as a self-containing service by subset of functions of multicast and broadcast architecture.
- Simplify the system setup procedure to keep the system cost marginal. The design aims to develop an efficient system in terms of architecture/protocol simplicity and resource efficiency. Despite simplified procedures, the architecture also should allow flexible session management.
- Focus on the protocols that allows efficient IP multicast.
- Enable caching capabilities inside the network.

Although this article focuses on mobile core network, the network design is viewed from converged network perspective where the fixed and mobile networks are taken into consideration in Section X. The fixed network in this article refers to wireline broadband network (e.g., cable network).

V. 5G-XCAST RELATED TECHNOLOGIES

This section describes the three new 5G functionalities and technologies that the 5G-Xcast project enables: *i)* a new concept of converged autonomous 5G-Xcast multicast

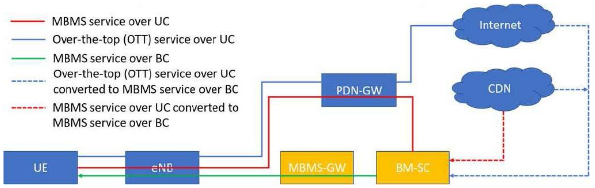

Fig. 4. High-level 3GPP MooD architecture.

operation on demand (“5G-Xcast MooD”) (MBMS operation on Demand), which is applied in a converged network including both fixed and mobile networks; *ii)* the multilink technology to increase the available bandwidth, service reliability, seamless mobility, etc.; *iii)* Multi-access Edge Computing (MEC) to enable new applications and services in 5G such as Augmented Reality and Virtual Reality (AR/VR).

A. Multicast and Unicast Switching

3GPP MooD enables the dynamic establishment of MBMS User Services according to actual consumption in mobile network, in order to offload unicast content delivery while efficiently utilizing network resources when the traffic volume exceeds a certain threshold. 3GPP MooD in LTE networks was standardized in 3GPP Rel-12 specified in 3GPP TS 26.346 [18] and further described in TR 26.849 [19]. In Fig. 4, in unicast delivery, over-the-top (OTT) content from the Internet or the operator CDN (Content Delivery Network) to the UE passes by the PDN-GW (Packet Data Network Gateway). If a certain threshold of UE population receiving the same OTT content is reached (i.e., MBMS is activated by MooD functionality), the BM-SC fetches the OTT content in unicast and converts to MBMS traffic and deliver to the UEs through the MBMS-GW and then eNBs to save network resource usage.

In fixed broadband networks, there is no standardized solution for switching to multicast delivery on demand although proprietary solutions exist. In the 5G-Xcast project, a dynamic delivery mode selection has been enabled where a seamless unicast/multicast/broadcast switching can be performed within both mobile and fixed access networks. Furthermore, the seamless switching can also be performed when the UE switches from fixed to mobile network and vice versa. The reader is referred to [20] that describes the aspirations in this area. This “5G-Xcast MooD” dynamic delivery mode selection is not a single solution but a collection of tools, procedures and best practices applied in various parts of end-to-end system as shown in [20] to enable efficient multicast delivery. The dynamic delivery mode switching comprises two stages. In the first stage, content is provided in a format suitable for multicast including transport, i.e., IP multicast. In the second stage, the (radio) AN optimizes its resources by selecting the best transport mode possible. For example, a RAN such as NG-RAN determines a set of PTP and PTM bearers to deliver the content to the receiving UEs in the most efficient way.

B. Multilink and Multi-Connectivity

This technology uses simultaneous several IP connections by a single UE using multiple radio access points. Multi-connectivity can be used to aggregate bandwidth and coverage seamless utilization of different IP access technologies such as 5G, LTE, and unlicensed technologies such as IEEE 802.11 (Wi-Fi). In this article, a subset of multi-connectivity technologies was considered, which is called multilink (Fig. 5).

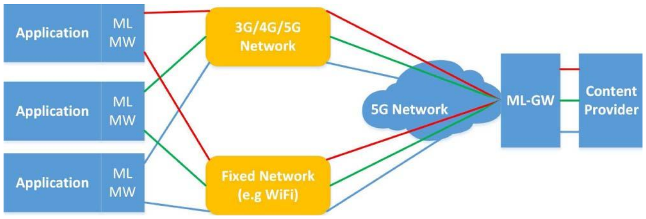

Fig. 5. Example of Multilink.

For example, data aggregation from multiple subscriptions to LTE and Wi-Fi (or directly to the fixed networks) increases the bandwidth available for UE as well as the IP service reliability and availability. A 5G cellular network access is required to maintain the service continuity when a UE is at the cell edge or leaves the Wi-Fi coverage, and vice versa. In these strategies, a ML-GW (Multilink Gateway) is able to reroute in real time some or all of data packets through the different available links. The ML-MW (Multilink middleware) performs the adequate data operation at the UE. The ML-MW communicates with the ML-GW which can be located either at the core network, the publisher site or on the cloud. These two ML entities exchange information about the performance of each link.

The content transmitted from the ML-GW down to the UE is split or duplicated over available links which are possibly from different operators or uses different access technologies, according to their temporal performance. The content is then reassembled at the UE by the ML-MW as a coherent data stream. The delivery is completely agnostic to the content. ML can be used to improve reliability, to provide ancillary content information, increase available bandwidth (e.g., higher video resolution), and/or to perform traffic planning and shaping.

Three different types are described as follows:

- *ML-CP:* additional functionality in the control plane of the mobile core network, which performs the estimation of QoS (Quality of Service) parameters for data transfer via each available link, multilink session setup and release;
- *ML-UP:* additional functionality in the user plane of the mobile core network, which performs data splitting, IP tunnel establishment;
- *ML-MW:* ML middleware functionality in the UE between the Application and the lower transport levels, which performs data combining, signaling (channel quality data transmitting), caching, providing ML session setup request (QoS parameters).

C. Multi-Access Edge Computing

Edge computing refers to a general paradigm of moving computation and storage to the network edge. The edge

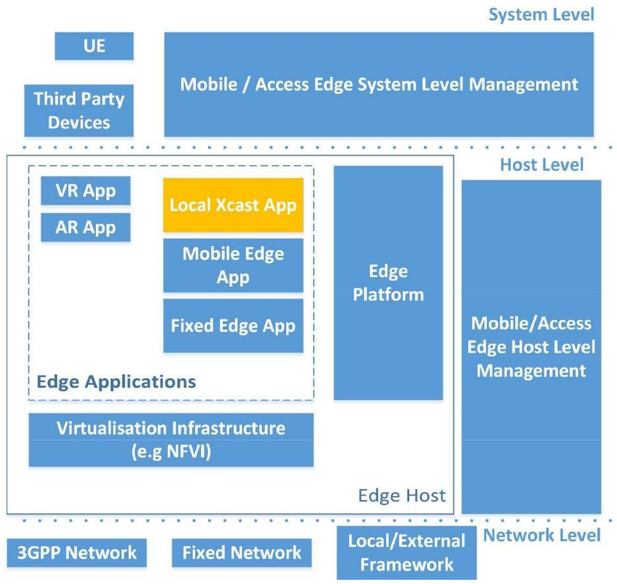

Fig. 6. Mobile edge computing framework adapted to multi-access.

computing paradigm aims at exploring the potential that can be achieved through the convergence of diverse fields such as communication and information technology (IT). Such a convergence enables the development of new applications and services at the edge of the fixed and/or wireless access network. The ETSI multi-access edge computing (MEC) framework presented in [21], being adapted to a converged network is as shown in Fig. 6.

Currently envisioned key use cases for MEC include video analytics, IoT, mass delivery of AR/VR, data caching and optimized local content distribution. MEC could play a key role in hosting the low-latency VR/AR applications which could then be delivered to the end user clients using fixed/mobile access networks. The data caching at the edge can allow MNOs (Mobile Network Operator) to significantly reduce the transport network load, thereby minimizing deployment costs, while improving latency. Multicast IP routing between a cloud application and a user application running on UE becomes more feasible. Thus, the edge computing paradigm allows for a new mobile/converge network architecture with multicast capabilities that can be utilized by any edge application which become a multicast source in the architecture, e.g., as illustrated by the Xcast App in Fig. 6 and Fig. 7.

In Fig. 7 MEC platform takes the place of the AF in the 5G system architecture. The NFs may offer services to the MEC platform, which exposes the services to MEC application via the mobile edge platform application enablement framework including radio network information APIs and others [22], [23], [24], [25]. The MEC applications are located in the local data network. The 5GC performs the traffic steering from the UPF to the local data network of MEC platform via a N6 interface towards MEC application.

VI. BUILDING BLOCKS AND APPROACHES

Two approaches to enable PTM have been proposed in this article: *i)* transparent multicast transport inside the 5G network operator, *ii)* PTM capability as a service.

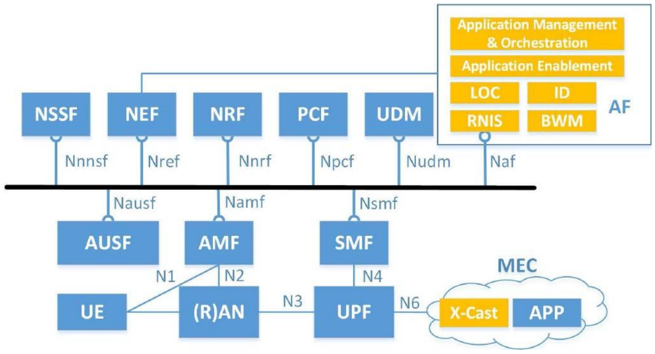

Fig. 7. MEC in 5G system architecture.

A. Transparent Multicast Transport

The 5G network needs to support an elemantary NF (e.g., transport of IP multicast datagram) to deliver multicast data through the network to the UE. This elementary function is enough to satisfy certain use cases such as LTE for mission critical application (e.g., Mission Critical Push-to-Talk (MCPTT) [26]). In MCPTT, eMBMS is used to transport data transparently to the UEs [27]. MCPTT relies on the MB2 interface for the transparent transport of user data via eMBMS [28]. The transport of multicast user data can be further simplified by allowing the 5G network to receive multicast user data directly from the multicast source that can reside inside (e.g., in multi-access edge cloud) or near the 5G network in operator's network.

The transparent multicast transport proposes that the UPF treats the multicast data (e.g., IP multicast datagrams) in the same or very similar way as the unicast data even if multicast data are delivered to the UPF via a tunnel. For example, the UPF does not perform any multicast specific functionalities such as application layer forward error correction (AL-FEC). It is assumed that any protocols or optimization for unidirectional transport may implement the transparent multicast transport capability of the 5G system by the system (e.g., DVB multicast Adaptive Bit Rate (ABR) [30]) or the applications. In this approach, specific functionalities (e.g., reliability from retransmission or AL-FEC) may be performed at the multicast source rather than the 5G system providing such functionalities.

For QoS control [17], the transparent multicast transport could utilize the policy and charging control framework which allows the interaction between PCF and AF. The multicast source taking the role of AF can provide a filter information to identify the service data flows for policy control and/or differentiated charging.

B. Point-to-Multipoint Services

5G system may offer content delivery through a set of services (e.g., file delivery) using the system's PTM data transport capability to a group of users or in a geographical area. Taking xMB interface in Fig. 1 as an example [31], a service corresponds to a content provider's service offering for delivery over a network supporting MBMS to UEs. A PTM service in this article refers to the service defined in [31].

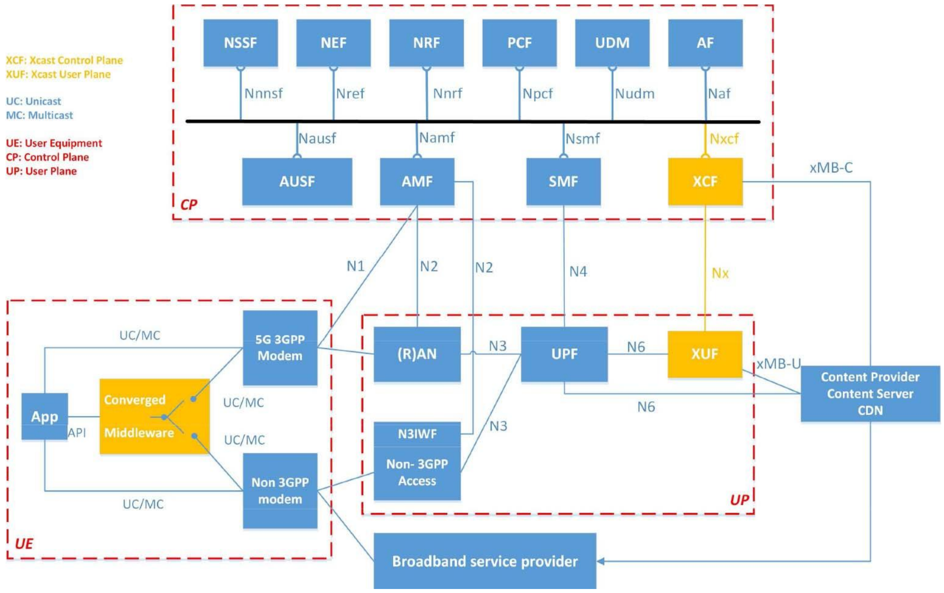

Fig. 8. 5G core network system architecture alternative 1: Transparent multicast transport.

In case of eMBMS, the BM-SC selects a delivery method (e.g., download delivery, streaming delivery, transparent delivery in [18]) including unicast delivery for a session. The BM-SC implements functionalities such as encapsulation and AL-FEC for the download delivery method and the streaming delivery method. The BM-SC can also support the associated delivery procedures, e.g., a file repair procedure.

The xMB reference point in eMBMS could be adopted in 5G system including necessary functionalities and associated delivery procedures, which are discussed in the following sections, although underlying system functions for delivery of multicast and broadcast data may be different.

1) User Data Encapsulation and Reliable Content Delivery: Delivering files or media segments using PTM requires user data encapsulation techniques which differ from unicast delivery due to one-way communication from the source to multiple destinations. Indeed, efficient and reliable data delivery over a unidirectional and lossy channel implies he usage of dedicated data encapsulation protocols (e.g., FLUTE [32]) and associated procedures. Unidirectional delivery protocols are designed to allow the use of AL-FEC. Furthermore, by associating the reliable data delivery with a repair procedure, a fully reliable delivery method can be achieved by allowing the UEs to request the missing data. Besides the M&E vertical, the file distribution service (e.g., file delivery method defined in [18]) is also required by practically use cases of all verticals: V2X (e.g., software update, traffic message, etc.), IoT (e.g., massive software update) and for public safety (MCData file distribution defined in [33], [34]). By using the xMB interface or its evolution in 5G, a content provider can provision the content to a file distribution service. A UE's application receives the content via an API such as File Delivery Application Service API from the MBMS Client [35].

2) Geographical Broadcast: There are applications and services that require or benefit from the broadcast data delivery in a geographical area to users. Unlike the cases when the UEs express explicitly an interest in receiving multicast user data (e.g., by joining IP multicast group) and the network can setup the resources based on this demand, the geographical broadcast requires that the network resources are allocated upon request from the content provider. The required functionality could be provided by a dedicated control network function (e.g., XCF described in Section VII-A) responsible for geographical broadcast management. It's noted that the geographical area could be extended to nation-wide for specific services (e.g., terrestrial broadcast TV).

3) Audience Size Measurement and Metric Reporting: Starting from a given number (or higher) of UEs consuming the same content in a specific geographical area, broadcast delivery is more beneficial than PTP (i.e., unicast) delivery (e.g., in terms of spectrum efficiency). The ability to automatically switch between unicast and broadcast requires the measurement of the audience size. The network side could explicitly ask the UEs to periodically send the reports including various metrics (e.g., Quality of Experience, network quality) or implicitly measures the audience size through different means (e.g., passive measurement).

4) Multicast Offloading: Multicast offloading refers to a capability of multicasting the popular or most requested content (both real time and non-real time). Multicast offloading for real time content is based on audience measurement. The network continuously analyses the consumption reports sent by the UEs to detect the popular live contents and trigger the multicast offloading to optimize the use of radio resource within the areas where the audience size is above a given threshold.

VII. PROPOSED MOBILE CORE NETWORK ARCHITECTURE

Two alternative architectural solutions have been proposed. The NFs specific to multicast or broadcast are shown in Figures 8 and 9. The functionalities of BM-SC and MBMS-GW as they exist in LTE are split into a control plane part

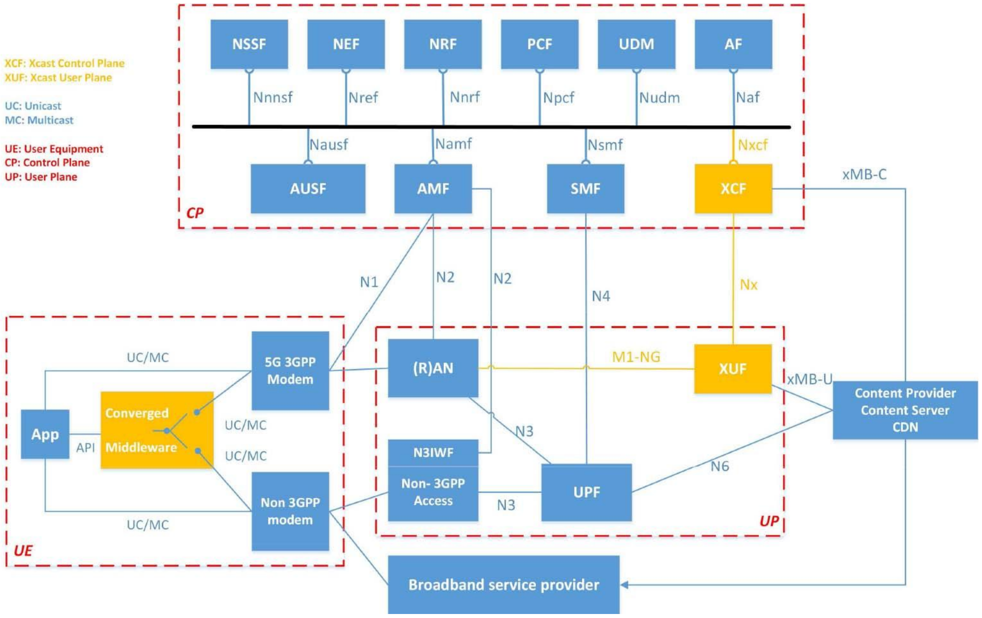

Fig. 9. 5G core network system architecture alternative 2: PTM services.

(XCF or 5G-Xcast Control Plane Network Function) and a user plane part (XUF or 5G-Xcast User Plane Network Function). UE functionalities are the same for both alternatives. Alternative 1 (Alt1) provides a 5G architecture friendly approach while Alternative 2 (Alt2) provides a solution with minimal changes to the eMBMS architecture and specification described for LTE.

A. Alternative 1

In Fig. 8, the XUF interfaces with the content provider via xMB-U interface and with the UPF via N6 reference point. The XUF provides multicast and broadcast specific functionalities for the user plane. The N6 reference point and the functionalities of UPF need to be enhanced to support and handle IP multicast traffic received from the XUF or the content source at the PDU session level, e.g., the UPF needs to support IGMP (Internet Group Management Protocol), MLD (Multicast Listener Discovery) and PIM (Protocol-Independent Multicast) and the IP multicast routing is needed to be supported by the N6 interface for the IP PDU session type.

The UE architecture illustrates an abstract model that consists of an application, a converged middleware, a 5G 3GPP modem and a non-3GPP modem. The modems offer connectivity services through access networks. The connectivity services provide exchange of PDUs between the Content Server and the modems. The role of the converged middleware is to hide the communication complexity to the application. The converged middleware could represent one or several of the following roles.

- Peer entity of XCF and XUF applicable only to the connectivity via 5G 3GPP modem (i.e., similar to MBMS APIs in LTE [35]);
- Multicast Gateway of DVB multicast ABR as described in [30];
- HTTP client (e.g., available as a client library) with the support of multicast QUIC (Quick UDP Internet Connections) [36];
- ML-MW as described in [37].

The XCF represents the peer endpoint to the content provider for the xMB-C reference point, i.e., the control plane part of xMB interface [18], [31]. The XCF functionalities related to xMB-C reference point includes the following:

- Authentication and authorization of XCF for a content provider;
- Authentication and authorization of a content provider for XCF;
- Creation, modification and termination of a service;
- Creation, modification and termination of a session;
- Status notification and query.

The XCF interacts with other network functions for session control and management through service-based interfaces and uses the services offered by them to manage network resources for the xMB session. The following (non-exhaustive list) functionalities are supported in the XCF:

- Network resource management for xMB session using SMF services including the allocation of UPF resources and maintenance of core network tunnels between UPF(s) and (R)AN node(s) and the allocation of (R)AN resources by (R)AN upon SMF request(s) in the geographical area;
- AL-FEC configuration;
- Allocation of reference point for multicast data transport to the UE (i.e., a multicast IP address);
- Session and service announcement;
- Reception of consumption and reception reports about a service;
- File repair management;
- Control multicast (or broadcast) transport availability based on the consumption reporting (i.e., functionality similar to 3GPP MooD in LTE);

- DRM (Digital Right Management) management;
- ML session setup and release upon request from UE;
- Estimation of QoS parameters for data transfer via each available link.

Some of these control functions are similar to the functions provided by BM-SC in LTE. It's noted that the current specification of xMB reference in 3GPP Rel-14 may not fulfil the requirements for 5G multicast/broadcast capabilities and might need to be enhanced.

The XUF represents the peer endpoint for the content provider through the xMB-U reference point, i.e., the user plane part of xMB interface [18], [31]. The XUF functionalities related to xMB-U reference point include *i)* delivery of content to XUF from the content provider, *ii)* retrieval of content by XUF from the content provider. The XUF functionalities are the following (non-exhaustive list):

- Reliable delivery of data over unidirectional transport (e.g., FLUTE);
- AL-FEC to protect content against packet loss.

The XUF sends the multicast IP packets to the UPF over N6 reference point, which in turn will need to be enabled to send the multicast IP packets via N3 tunnel.

B. Alternative 2

Alt2 as shown in Fig. 9 differs from Alt1 in the fact that the XUF interfaces with the RAN directly (via an M1-NG reference point) whereas in Alt1 the XUF interfaces with the RAN via the UPF. Hence, in Alt2 the XUF needs to support generic UPF capabilities (enhanced for multicast or broadcast capabilities), whereas in Alt1 the XUF would only require to support dedicated multicast or broadcast functionalities. However, in Alt1 the UPF would need to support multicast capabilities, which is not required for a UPF in Alt2.

An M1-NG interface is the same as M1 interface in LTE as specified in [38]. It would use GPRS Tunnelling Protocol User Plane (GTP-U) encoding to multicast the data packets to the RAN nodes. The XCF sends tunnel information IP multicast address to the RAN nodes or non-3GPP access nodes. The difference is that M1-NG supports IP multicast while the N3 does not up to 3GPP Rel-16.

VIII. ARCHITECTURE ANALYSIS AND DISCUSSION

Previous section describes two alternative solutions for the mobile core network architecture enabling multicast/broadcast capabilities. Both architectures can deliver the same functionalities and support both transparent multicast transport and PTM services. The architectures are described with the reference to the 5G system architecture [15]. The architectures differ in the set of NFs, the functionalities each NF implements and the reference points between the NFs.

A. Alternative 1

In Alt1, the functionalities needed to support the PTM services are separated from the existing NFs in 5GC specified by 3GPP in Rel-15. This solution leverages the transparent multicast transport introduced to the system by enhancing UPF, SMF, PCF, and possibly other NFs (e.g., UDSF). In addition, XCF and XUF are introduced to offer PTM services. The architecture for Alt1 follows the CUPS design principle in 5G. The architecture also separates the NFs used for offering multicast and broadcast services over xMB interface while the architecture also allows for the transparent multicast transport of data from the Content Server to the UPF using the multicast routing capabilities at the N6 reference point.

The XCF controls the XUF over a new reference point Nx. The XUF can use Nx to notify the XCF about its status and user plane events. An option in which the XUF connects to the SMF via the N4 reference point was considered. However, the N4 reference point between SMF and XUF would be only used for a transparent transport of messages between XCF and XUF. Hence, the introduction of a new reference point is more appropriate.

In this alternative, the UPF scalability potentially depends on the number of the multicast streams to be managed. However, the UPF scalability is an implementation issue which can be addressed in various ways. For instance, one or more UPFs can handle a single multicast stream. A single UPF can handle both unicast and multicast or a single UPF can handle multiple multicast streams.

When a content is delivered to multiple UEs using PTM service offered by the XCF and the XUF, the XUF retrieves or receives the data from the Content Server through the xMB interface. The XUF then performs the encapsulation (e.g., FLUTE) with the protection against packet loss using AL-FEC. The encapsulated data is sent over N6 reference point to the UPF as IP multicast packets. The UPF then forwards the encapsulated data through the N3 reference point to the (R)AN nodes. The (R)AN nodes transmit the encapsulated data to the UEs over the air. The converged middleware decapsulates the encapsulated data with additional AL-FEC decoding if required and delivers the content to the application.

In cases where the XUF functionalities are not desirable (e.g., encapsulation, AL-FEC protection) or alternative solutions to the XUF functionalities are performed at the Content Server, the Content Server sends directly IP multicast traffic to the UPF over the N6 (i.e., network infrastructure between UPF and the Content Server must support multicast routing). XUF and XCF are not needed for such deployments. The deployment scenarios in which the MNO may support the direct injection of IP multicast are the following (see Section VI-A). An example deployment scenario when the direct injection of IP multicast traffic is possible is the case of MEC deployment. Indeed, the MEC server sends the data to the UEs using LADN (Local Area Data Network) [15] which belong to the MNO. In this deployment, alternative solutions to the XUF functionalities may be implemented by a MEC application running on MEC host platform. Another example deployment scenario could utilise Automatic Multicast Tunnelling (AMT) [29].

In this architecture, the N3 reference point supports multicast in order to deliver IP multicast data packets to the appropriate (R)AN nodes. 3GPP in Rel-15 does not support multicast at the N3 reference point.

This alternative has several points that need to be taken into account:

- The synchronization for multi-cell transmission cannot be performed in the XUF (like the SYNC protocol between the BM-SC and the eNodeB in LTE [39]) since the N6 interface does not provide tunnelling functionality (e.g., GTP-U) for the SYNC protocol. However, the synchronization could be done at the RAN, which also avoids the need for the synchronization of the RAN with the core network entities.
- The UPF delivers the same content to multiple gNodeBs and in the case of a large number of gNodeBs (e.g., large geographic coverage) the support of multicast at the N3 reference point and possibly at the N9 reference point seems beneficial.

B. Alternative 2

The Alt2 leverages the LTE eMBMS architecture for offering PTM services. XUF implements a sub-set of UPF functionalities (e.g., packet routing and forwarding, traffic usage reporting, QoS handling, etc.) required for multicast or broadcast capabilities. The rationale is to offer minimized changes to the current eMBMS specification in LTE. Indeed, the architecture in this alternative is similar to the actual eMBMS architecture in LTE where the current BM-SC is split into control plane (e.g., XCF) and user plane (e.g., XUF). In this respect, the XUF in Alt1 is not an equivalent to the XUF in Alt2. The same applies to the XCF and its functionalities, which are different from Alt1. The XCF in the Alt2 is responsible for a session management for PTM services including the maintenance of tunnels and resources. The separation of control and user plane for BM-SC is already described in 3GPP TS 23.285 [40] which aims at the latency reduction for V2X applications.

Alt2 supports both transparent multicast transport and PTM as a service. Indeed, the solution where multicast functionalities are performed at the Content Server (e.g., through the MEC platform) requires the multicast capabilities in the UPF through the N6 reference point as described in Alt1. In addition, multicast or broadcast capabilities as existed in LTE can be supported in the 5G architecture through the XCF/XUF which reflect the BM-SC and MBMS-GW in the LTE architecture. For instance, the XUF could provide SYNC functionality for a very large coverage area up to nation-wide as in LTE.

IX. APPLICATION OF BROADCAST AND MULTICAST TO VERTICALS AND APPLICATIONS

The two proposed architecture alternatives are the baseline to develop the solutions for different verticals and applications such as IoT, PW, Mission Critical Communications, V2X.

A. Group Message Delivery for IoT

3GPP Rel-15 addressed the use of MBMS in LTE for the potential data delivery to a large number of IoT devices. The considered requirements are the support of reliable delivery, acknowledgements on successful reception, eMBMS delivery procedures for devices with limited capabilities (e.g., battery life of 15 years). In 3GPP Rel-16, the first study on the support and evolution of Cellular IoT (CIoT) functionalities for the 5G system was performed and captured in TR 23.724 [42]. In [42], group message delivery can be supported by unicast MT NIDD (Mobile Terminated Non-IP Data Delivery) between the NEF and UE. However, this solution is based on a unicast approach which is not efficient. The proposed architectural alternatives described in this article provide a more efficient solution using a broadcast approach.

In the context of LTE, the communication between an IoT Application hosted by an Application Server (AS) in the external network and the IoT Application in the UE uses services provided by the LTE system, and optionally services provided by a Services Capability Server (SCS) for value-added services. In particular, [43] provides a solution for the group message delivery service is offered by SCEF (Service Capability Exposure Function). The SCS/AS delivers a message through the SCEF using T8 reference point to the LTE EPC with eMBMS enabled and then to a group of IoT UEs.

In the context of 5GC, if the SCS/AS is trusted by 5GC, it can use directly the xMB interface to ask for a group message delivery. Otherwise, it accesses services offered by 5GC via the NEF using the NEF Northbound interface specified in [44]. It is proposed to introduce an API for group message delivery to the NEF Northbound interface similar to APIs for group message delivery via MBMS specified for the interface T8 [45]. The work flow solution based on the proposed architectural alternatives in this article for IoT group message delivery using this new API is described in details in [46]. In addition, the TR 26.850 [47] describes further optimizations at the service layer (e.g., binary data format) for IoT.

B. Public Warning With Multimedia Content

Emergency events are largely unpredictable and therefore the initiative to transmit warning messages needs to lie with the network and UEs in the affected area need to be triggered to start receiving warning messages.

Many network operators have already deployed a Public Warning System based on Cell Broadcast technology which is limited only to text messages. To enable PW message with multimedia content, the UE application could use received Cell Broadcast messages to show interest in receiving (additional) multimedia content through an IGMP, MLD or PIM request. The multimedia content would then be delivered to the UEs using transparent multicast transport as described in Section VI-A. For example, the procedure in which the content is provided over multicast QUIC would be as follows:

- UE receives a Cell Broadcast (CB) message with a URL (e.g., http(s)://eu-alert.org/picture.html/);
- The UE sends the request for the resource to the server;
- The UE receives an advertisement for a multicast QUIC session in the HTTP Alt-srv header;
- The UE joins the multicast group which triggers the resource allocation as per transparent multicast flow;
- The UE receives the IP multicast content.

The work flow solution for PW with multimedia content is described in details in [46]. The PW application on the UE could offer the user a more sophisticated way of presenting multimedia content if such content consists of multiple components, such as text or speech (for the visually impaired) in multiple languages or video in sign language (for the hearing impaired). Those components could be selected by configuration or by ad-hoc selection by the user.

Instead of using Cell Broadcast to trigger UEs to start receiving multimedia content, the paging message could be used to carry an *mbms-Indication*, similar to the *cmas-Indication* and *etws-Indication* [41] that are now used for triggering UE reception via Cell Broadcast technology.

Multimedia warning messages would be submitted by the alert originator via the xMB reference point and be broadcast based on Alt1. The only additional step would be that not only the PTM Configuration data needs to be broadcast to the UEs, but also the MBMS indication in the paging message.

C. Mission Critical Communications

Public safety organizations such as police, firefighters, and ambulances have specific requirements on services offered by communication systems. Such services include push-to-talk, mission critical video and data which are available for private communication between two parties but also for groups. Multicast is a natural technology choice for providing a group communication to ensure service scalability.

In LTE, a group communication service application server (GCS AS) has access to broadcast services offered by the MBMS system via the MB2 interface [48]. The GCS AS receives cell identities of serving cells from UEs in a group and uses this information to determine whether to use MBMS or unicast path to deliver content. It is understandable that the decision based on cell identities is very coarse and may result in suboptimal operation of radio access network. Moreover, service continuity is challenging to achieve. UE mobility causes switching from MBMS bearer to unicast bearer. The switching is initiated by a UE upon detection of bad MBMS bearer condition or that a candidate target cell for reselection provides the service using SC-PTM [9]. It becomes obvious that the GCS AS role in controlling radio resources via switching between multicast and unicast is not optimal and it is caused by the fact that MBMS in LTE was designed for broadcasting.

The transparent multicast transport solves the challenges of multicast for group communication [46]. Alternatively, the PTM services provided by XCF can offer a multicast service and allow the GCS AS to specify a group of UEs to which multicast should be delivered. The XCF and the XUF then uses the transparent multicast transport to deliver the service [46]. The key component of this solution is to enable RAN to determine the transmission mode (unicast, single-cell or multi-cell transmission) for delivering multicast content to UEs [49]. Service continuity is provided using RAN mechanisms such as including multicast flows in handovers or more advanced techniques like RAN multicast area management [49].

The adaptation of the proposed architecture alternatives for mission critical communications leads to the simplification of GCS AS implementation. A GCS AS does not need to implement the control of MBMS and unicast bearers. At the same time, the system architecture allows for a better control of radio resources and service continuity controlled in RAN. The work flow for group communication using multicast in Mission Critical Communications is described in details in [46].

D. Cellular Vehicle-to-Everything

V2X communication refers to communications in many vehicular applications including road safety, traffic efficiency, local services and others [50]. 5G thanks to its low latency and high reliability performance is envisioned to enable more advanced applications such platooning and cooperative driving. In general, a V2X application requires a communication system to provide communication between two stations of the system (e.g., two vehicles or a vehicle and a road side unit), stations in a group (e.g., vehicles in a platoon), and dissemination of information – typically a warning – in a geographical area.

The support for V2X services was introduced to 3GPP system in Rel-14. The 3GPP architecture for V2X services shares some principles with the architecture for mission critical services [40], namely that the role and the functionalities of V2X application server (V2X AS) and GC AS are the same with the regard of MBMS. A V2X UE provides its geographical location to a V2X AS. This information is used to determine whether to switch between unicast and MBMS delivery as described for mission critical services. The same challenges and drawbacks are applicable to group communication in V2X context, e.g., cooperative maneuver such as merging into a platoon. The transparent multicast transport or multicast provided by the PTM services address these challenges in the same way as described for the mission critical services above.

The capability of communication system to disseminate information in a geographical area is critical for V2X services. The geographical broadcast of PTM services offer exactly this critical functionality for V2X systems. The procedure of UE joining a multicast communication group of a platoon using transparent multicast transport is described in details in [46].

X. 5G-XCAST CONVERGED CORE NETWORK ARCHITECTURE

In the past, fixed and mobile networks have been deployed as separate systems delivering independent but often similar services such as broadband, M&E, messaging, voice and video communications, etc. The 5G architecture should seamlessly integrate with fixed and mobile access technologies under a fully converged end-to-end system in order to deliver future service requirements.

As mentioned in the previous section, in addition to M&E services there are a number of other services which could benefit from broadcast/multicast functionality. Some devices which use these services (e.g., IoT devices, Set-top box (STB), etc.) may only be connect via fixed access mechanisms. In addition, devices which are connect via both fixed and mobile (e.g., UE in the home connected to a FN RG (Fixed Network Residensial Gateway) and 5G RAN, STB connected to a 5G RG, etc.), would have a number of advantages from converged connectivity such as the simultaneous use of multiple access

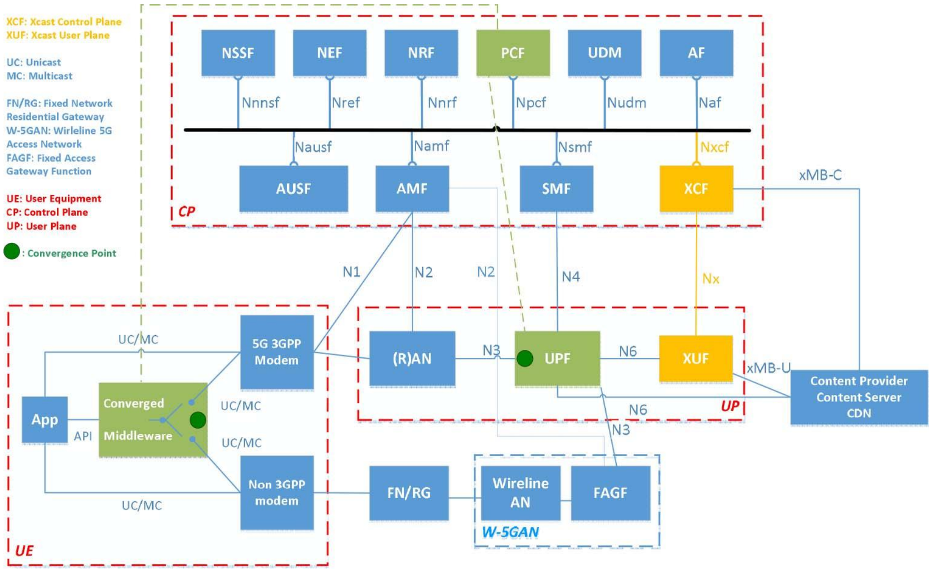

Fig. 10. Converged 5G Fixed and Mobile Network Architectural Option Example: Trusted with UE and UPF as the convergence points.

networks, bandwidth boost, failover with session continuity, fast provisioning, symmetric bandwidth, etc.

Work in standards bodies on the topic of convergence happened in parallel to the 5G-Xcast project. The main activities and work areas on convergence that were tracked during the project were:

- 3GPP TR 26.891 - Media Distribution in 5G;
- 3GPP TR 23.716 - Study on Wireline & Wireless Convergence on 5GS;
- 3GPP TR 23.793 - Study on Access Traffic Steering, Switching and Splitting on 5GS;
- Broadband Forum TR.348 - Hybrid Access Broadband Network Architecture;
- Broadband Forum SD-407 - 5G Fixed and Mobile Convergence Study.

A number of architectural alternatives in regards to convergence was considered, whilst simultaneously aligning the aforementioned options with the standards bodies' activities on the subject of convergence. Though not all alternatives are covered in this article, as an example one such architectural solution is an integrated converged core network approach in which the UE and UPF act as the points of convergence (Fig. 10). This solution has a number of benefits including: good access to core information and policy, common authentication and policy enforcement across networks, potential for lowest cost of ownership and deployment for operators.

The most relevant aspects for the converged architecture shown in Fig. 10 are listed hereafter:

- UE – In order for this architectural solution to be attainable the UE must have converged middleware support (note: in this use case the converged middleware is ML-MW). The UE has simultaneous connectivity with both the NG-RAN and the Fixed Access Network. The latter connection is achieved by the UE through a Wi-Fi connection to the FN RG which in turn is connected to the wireline access network. The UPF also has a logical connection to the PCF.
- Fixed Access Gateway Function (FAGF) – A new entity to be defined by 3GPP acts as an intermediary function between the Fixed Network and Mobile Core. The FAGF has an N3 interface to the UPF.
- UPF – Is responsible for handling all the user plane traffic between the converged core network and the UE. It supports simultaneous connectivity with both the NG-RAN and the Fixed Access Network via its N3 interfaces to the FAGF and NG-RAN. The UPF has N6 interfaces to both the XUF and CDN respectively. The UPF also has an N4 connection to the SMF and a logical connection to the PCF.
- PCF – Has logical connections to both the UE and UPF. The PCF is responsible for instructing the UE and UPF on the functional behavior to perform for any given multilink session, based on Operator policy.

XI. CONCLUSION AND OUTLOOK

This article proposed two different architecture alternatives to enable multicast and broadcast capabilities in 3GPP 5G core network architecture. Both architectures can deliver the same functionalities and support both approaches (transparent multicast transport and PTM services). The architectural alternatives are described with reference to the 5G system architecture and differ in the set of NFs, the functionalities each NF implements and the reference points between the involved NFs. Two new network functions (i.e., XCF and XUF) are introduced to offer multicast and broadcast capabilities. The first alternative provides an approach that is 5G architecture friendly. The second aims at minimizing the changes to the functionalities developed in LTE eMBMS for offering PTM services.

This article also presented how the proposed architecture alternatives applied to different applications and verticals with diverse requirements such as IoT, V2X, PW, MCC. In addition, based on the proposed architecture alternatives for mobile network, this article describes their extension to multiple 5G converged network architecture alternatives for different deployment options.

At the time of writing this article, 3GPP starts working on the architectural enhancements for 5G multicast and broadcast services. The work done in this article intends to become a reference point for standardization work in the 3GPP.

ACKNOWLEDGMENT

The views expressed in this contribution are those of the authors and do not necessarily represent the project.

REFERENCES

[1] *Service Requirements for Next Generation New Services and Markets*, 3GPP Standard TS 22.261, Mar. 2018.

[2] D. Ratkaj and A. Murphy, *Definition of Use Cases, Requirements and KPIs. Deliverable D2.1*, 5G-PPP, Valencia, Spain, 2019.

[3] *Interim Report From Email Discussion on 5G Broadcast Evolution*, 3GPP, Sophia Antipolis, France, Mar. 2018.

[4] K. Flynn. (2019). *Enhanced Television Services Over 3GPP eMBMS*. Accessed: Nov. 16, 2019. [Online]. Available: http://www.3gpp.org/news-events/3gpp-news/1905-embms_r14

[5] "Study on scenarios and requirements for next generation access technologies," 3GPP, Sophia Antipolis, France, Rep. TR 38.913, Jun. 2017.

[6] "Study on LTE-based 5G terrestrial broadcast," 3GPP, Sophia Antipolis, France, Rep. TR 36.776, Jul. 2018.

[7] D. Gomez-Barquero, D. Navratil, S. Appleby, and M. Stagg, "Point-to-multipoint communication enablers for the fifth generation of wireless systems," *IEEE Commun. Stand. Mag.*, vol. 2, no. 1, pp. 53–59, Mar. 2018.

[8] N. Nouvel, *Content Delivery Vision. Deliverable D5.1*, 5G-PPP, Valencia, Spain,, 2017.

[9] T. Tran, *Mobile Core Network. Deliverable D4.1*, 5G-PPP, Valencia, Spain, 2019.

[10] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description; Stage 2*, 3GPP Standard TS 36.300, Apr. 2017.

[11] *Multimedia Broadcast/Multicast Service (MBMS); Architecture and Functional Description*, 3GPP Standard TS 23.246, Sep. 2019.

[12] "Massive Internet of Things," 3GPP, Sophia Antipolis, France, Rep. TR 22.861, Jan. 2016.

[13] "Feasibility study on new services and markets technology enablers for critical communications; stage 1," 3GPP, Sophia Antipolis, France, Rep. TR 22.862, Jan. 2016.

[14] "Feasibility study on new services and markets technology enablers for enhanced mobile broadband; stage 1," 3GPP, Sophia Antipolis, France, Rep. TR 22.863, Jan. 2016.

[15] *System Architecture for the 5G System*, 3GPP Standard TS 23.501, Dec. 2019.

[16] *Procedures for the 5G System*, 3GPP Standard TS 23.502, Dec. 2016.

[17] *Policy and Charging Control Framework for the 5G System; Stage 2*, 3GPP Standard TS 23.503, Sep. 2017.

[18] *Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codecs*, 3GPP Standard TS 26.346, Jan. 2015.

[19] "Multimedia broadcast/multicast service (MBMS) improvements; MBMS operation on demand," 3GPP, Sophia Antipolis, France, Rep. TR 26.849, Jun. 2015.

[20] T. Stevens, "Key technologies for the content distribution framework. deliverable D5.2," 5G-PPP, Valencia, Spain, 2019.

[21] *Mobile Edge Computing (MEC); Framework and Reference Architecture*, ETSI Standard GS MEC 003, Jan. 2019.

[22] *Mobile Edge Computing (MEC); Mobile Edge Platform Application Entablement, V1.1.1*, ETSI Standard GS MEC 011, Jul. 2017.

[23] *Mobile Edge Computing (MEC); Radio Network Information API, V1.1.1*, ETSI Standard GS MEC 012, Jul. 2017.

[24] *Mobile Edge Computing (MEC); Location API, V1.1.1*, ETSI Standard GS MEC 013, Jul. 2017.

[25] *Mobile Edge Computing (MEC); Bandwidth Management API, V1.1.1*, ETSI Standard GS MEC 015, Oct. 2017.

[26] *Functional Architecture and Information Flows to Support Mission Critical Push to Talk (MCPTT); Stage 2*, 3GPP Standard TS 23.379, Jun. 2016.

[27] *LTE; Mission Critical Push To Talk (MCPTT); Media, codecs and Multimedia Broadcast/Multicast Service (MBMS) Enhancements for MCPTT Over LTE*, 3GPP Standard TS 23.379, Apr. 2017.

[28] *Group Communication System Enablers for LTE (GCSE_LTE); MB2 Reference Point; Stage 3*, 3GPP Standard TS 29.468, Sep. 2019.

[29] "Automatic multicast tunneling," IETF, RFC 7450, Feb. 2015.

[30] DVB. (2019). *DVB Releases Reference Architecture for IP Multicast—DVB*. Accessed: Nov. 16, 2019. [Online]. Available: https://www.dvb.org/news/dvb-releases-reference-architecture-for-ip-multicast

[31] *Representational State Transfer Over xMB Reference Point Between Content Provider and BM-SC*, 3GPP Standard TS 29.116, Jan. 2017.

[32] "FLUTE—File delivery over unidirectional transport," IETF, RFC 3926, Oct. 2004.

[33] *Mission Critical Data Over LTE*, 3GPP Standard TS 22.282, Jan. 2016.

[34] *Functional Architecture and Information Flows to Support Mission Critical Data (MCData); Stage 2*, 3GPP Standard TS 23.282, Jun. 2016.

[35] *Multimedia Broadcast/Multicast Service (MBMS); Application Programming Interface and URL*, 3GPP Standard TS 26.347, Jan. 2016.

[36] Datatracker.ietf.org. (2019). *Draft-pardue-quic-http-mcast-05—Hypertext Transfer Protocol (HTTP) Over Multicast QUIC*. Accessed: Nov. 16, 2019. [Online]. Available: https://datatracker.ietf.org/doc/draft-pardue-quic-http-mcast/

[37] R. Odarchenko, B. Altman, R. Aguiar, and Y. Sulema, "Multilink approach for the content delivery in 5G networks," in *Proc. 5th Int. Sci. Pract. Conf. Probl. Infocommun. Sci. Technol.*, 2018, pp. 140–144.

[38] *Evolved Universal Terrestrial Radio Access Network (E-UTRAN); M1 Data Transport*, 3GPP Standard TS 36.445, Aug. 2016.

[39] *MBMS Synchronization Protocol (SYNC)*, 3GPP Standard TS 25.446, Feb. 2016.

[40] *Architecture Enhancements for V2X Services*, 3GPP Standard TS 23.285, Jun. 2016.

[41] "Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification," 3GPP, Sophia Antipolis, France, Rep. T2.36.331, Oct. 2016.

[42] "Study on cellular Internet of Things (CIoT) support and evolution for the 5G system (5GS)," 3GPP, Sophia Antipolis, France, Rep. TR 23.724, Sep. 2017.

[43] *Architecture Enhancements to Facilitate Communications With Packet Data Networks and Applications*, 3GPP Standard TS 23.682, Sep. 2019.

[44] *5G System; Network Exposure Function Northbound APIs; Stage 3*, 3GPP Standard TS 29.522, Mar. 2018.

[45] *T8 Reference Point for Northbound APIs*, 3GPP Standard TS 29.122, Jun. 2017.

[46] B. Altman, *Session Control and Management. Deliverable D4.3*, 5G-PPP, Valencia, Spain, 2019.

[47] "MBMS for Internet of Things (IoT)," 3GPP, Sophia Antipolis, France, Rep. TR 26.850, Jun. 2017.

[48] *Common Functional Architecture to Support Mission Critical Services; Stage 2*, 3GPP Standard TS 23.280, Jun. 2016.

[49] M. Säily and C. Barjau, *RAN Logical Architecture and Interfaces for 5G-XCast. Deliverable D3.3*, 5G-PPP, Valencia, Spain, Feb 2019.

[50] "Intelligent transport systems (ITS); vehicular communications; basic set of applications; definitions, v1.1.1," ETSI, Sophia Antipolis, France, Rep. TR 102 638, Jun. 2009.

[51] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: Overview and rel-11 enhancements," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 68–74, Nov. 2012.

[52] *Architecture Enhancements for Control and User Plane Separation of EPC Nodes*, 3GPP Standard TS 23.214, Mar. 2018.

[53] T.-X. Do, V.-G. Nguyen, and Y. Kim, "SDN-based mobile packet core for multicast and broadcast services," *Wireless Netw.*, vol. 24, pp. 1–14, Dec. 2016.

[54] H. Chantre and N. L. S. da Fonseca, "Reliable broadcasting in 5G NFV-based networks," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 218–224, Mar. 2018.

[55] H. Chantre and N. L. S. da Fonseca, "Redundant placement of virtualized network functions for LTE evolved multimedia broadcast multicast services," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2017, pp. 1–7.