IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 6, 2021, accepted March 19, 2021, date of publication April 5, 2021, date of current version April 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3070814

Multicast at Edge: An Edge Network Architecture for Service-Less Crowdsourced Live Video Multicast

KAMRAN ZAHOOR[1], KASHIF BILAL[2], AIMAN ERBAD[3], (Senior Member, IEEE), AMR MOHAMED[2], (Senior Member, IEEE), AND MOHSEN GUIZANI[2], (Fellow, IEEE)

[1]Department of Computer Science, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22010, Pakistan
[2]Department of Computer Science and Engineering, College of Engineering, Qatar University, Doha, Qatar
[3]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

Corresponding author: Aiman Erbad (aerbad@hbku.edu.qa)

This work was supported by the Qatar National Research Fund (a member of Qatar Foundation) under Grant NPRP 8-519-1-108, and Grant NPRP 13S-0205-200265.

ABSTRACT Using smartphones, tablets, and other portable/handheld devices, we have become more reliant on the video streaming services for entertainment and remote work. Mobile data traffic has grown eighteen folds over the past five years accounting for the majority of IP traffic. YouTube Live, Facebook Live, Twitch, DouYu and other streaming as well as video conferencing services have increased in popularity so at any given moment they serve thousands of live video streams to millions of users. The Enhanced Multimedia Broadcast Multimedia Service (eMBMS) is the standard multicast protocol for 5G networks. Cellular multicast has gained considerable attention to efficiently utilize the limited spectrum to transmit multimedia content to cellular sites with co-located viewers, lowering the cost, and maximizing the Quality of Experience (QoE). However, popular live video content providers use unicast mode for live video delivery and have limited support in the eMBMS service-oriented network architecture. In this paper, we propose an overlay network architecture to augment eMBMS to address the limitations of the standard eMBMS architecture and enable service-less multicast for crowdsourced live video providers. We propose a Virtual Network Function (VNF) service that identifies potential multicast scenarios based on user requests for a live video within a confined area. The VNF Application Server collects information, validates a potential multicast scenario, and initiates an ad-hoc multicast service on the fly. We use a real-world dataset of Facebook Live videos to evaluate the proposed architecture. The simulation results depict considerable advantages in terms of cost, efficiency, and Quality of Experience (QoE). Our results show that the proposed architecture provides significant benefits in bandwidth saving at the backhaul, transit, and RAN links.

INDEX TERMS MBMS, multicast, video streaming, wireless communication.

I. INTRODUCTION

Using smartphones, tablets, and other portable/handheld devices, we have become more reliant on video streaming services for entertainment and remote work. Mobile data traffic has grown eighteen folds over the past five years accounting for around 63% of total IP traffic [1]. Moreover, it is expected to continue increasing at an annual rate of 47% within the coming years. The majority of the mobile data traffic (around 80%) consists of video content. By the year 2021 [2] live Internet video traffic is anticipated to be at least 13% of total IP traffic. Despite the aforementioned statistics, the COVID-19 pandemic also has led to spike in the use of online conferencing for work-from-home and schooling purposes. Different online meeting and online large events platforms, i.e., Zoom, Microsoft Teams, etc. have experienced unprecedented traffic growth at their networks. According to a report, Microsoft Teams has witnessed 775% spike in MS Teams usage in Italy due to COVID-19.[1]

In live videocast, multiple users within a confined vicinity may request the same content. As the school and university

The associate editor coordinating the review of this manuscript and approving it for publication was Barbara Masini.

[1]https://www.geekwire.com/2020/microsoft-cloud-services-sees-775-spike-regions-affected-social-distancing-mandates/

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

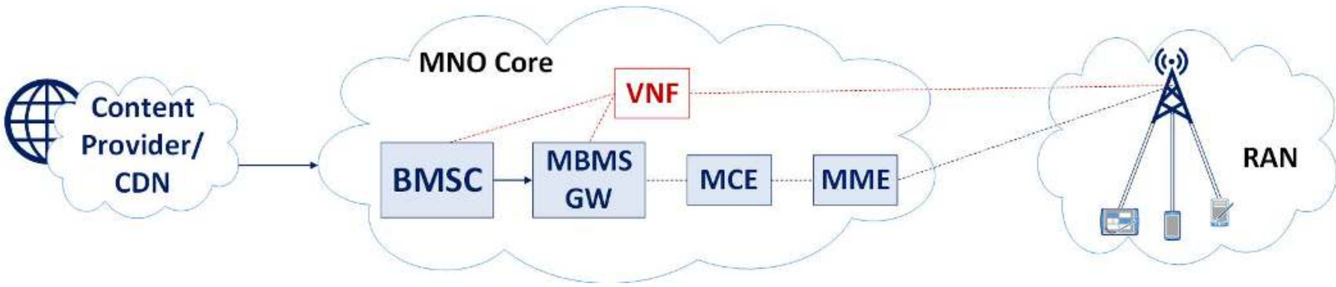

FIGURE 1. Multicast at Edge - A bird-eye view.

exams are being conducted in online fashion these days or as more people work from home, we will have more video streams that is consumed simultaneously by co-located users. Moreover, YouTube Live and Facebook Live, two major sources of live content provide several thousands of live video streams to millions of users at any given moment. This number is expected to grow further due to the COVID-19 outbreak and the new normal beyond the pandemic. Real-world live video traces show that numerous users within the same vicinity request the same content [3]. A separate flow in the unicast mode for each user is streamed from the nearest cache/storage repository, i.e., Content Delivery Network (CDN). This delivery of the same content to an area considerably impacts the network efficiency, cost, Quality of Service (QoS) of the service provider, and Quality of Experience (QoE) of the viewers.

To fairly address the escalating demands for video traffic, multicast in the cellular Radio Access Network (RAN) is a viable option [4]. In legacy, as well as modern Long Term Evolution - Advanced (LTE-A) networks, live video streaming is performed in unicast mode, in general. Popular video content providers, such as Facebook and YouTube use unicast transmission for live video delivery to their users. However, unicast transmissions, specifically in a scenario, where multiple users from a limited area are watching the same stream/channel put unnecessary load on transit, backhaul, and access networks by wasting scarce resources at each tier, increase the delivery cost, and decrease the viewers' QoE [5]. With an increase in the streaming users in a RAN, desired radio resources grow linearly, even when the users request the identical video content. On the contrary, the multicast transmission offers a resource efficient and scalable approach to transmit live videos to several users simultaneously. Multicast can lead to capital-savings for the content providers in terms of fetching less data from the CDN. High QoE and efficient resource utilization can be achieved because of the resources saved due to multicast at transit, backhaul, and RAN levels.

The 4G/5G cellular networks have already standardized the support for broadcast/multicast services. Multimedia Broadcast Multicast Service (MBMS) and Broadcast Multicast Service (BCMSC) are the standard services that are employed by Mobile Network Operators (MNO) to deploy broadcast/multicast architectures. Telstra, AT&T, China Unicom, Verizon Wireless, and Korea Telecom have already field tested the eMBMS services successfully and have been planning to launch eMBMS [6]–[8]. Although MBMS architecture was announced by the 3rd Generation Partnership Project (3GPP) in 2011 for LTE-A, still the telecom operators around the world have not commercialized the eMBMS services. However, such multicast transmissions are only carried out for the predefined user services, e.g., TV, video conferencing, file download, live streaming offered by the Multicast Service Providers (MSP) [9], [10]. Previous research [11]–[13] carried out different optimizations in the wireless multicast services pertaining to the enhancement of the video quality, bandwidth, and energy consumption of user equipment and the network. However, all of these studies presented their works assuming a service-based architecture, where the content provider announces and initiates a service, which is not applicable on crowdsourced live video providers, such as Facebook Live, YouTube Live, and Twitch, where thousands of people deliver the live content without announcing a separate service for each videocast. Therefore, eMBMS multicast cannot be used for such crowdsourced live video providers.

Crowdsourced live video streaming is considered an emerging video streaming format, and Information and Communication Technologies (ICT) giants like Facebook and YouTube are putting considerable efforts in this growing area. The existing eMBMS architecture already supports the broadcast/multicast services at the edge network including 3D and multi-view videos. The trend of live video streaming has gone viral within recent years. Every 1 out of 5 videos is a live video on Facebook. Some of the major examples illustrating the emergence and popularity of live videos include: exploding the watermelon video, which reached over 800,000 simultaneous viewers at Facebook Live in 2016. This video got over 100 requests per second on the spike. The 2015 super bowl encountered 114 million live viewers with an average of 2.36 million users on the live stream. On Twitch TV, 840,000 simultaneous viewers watched a stream in 2015 [14]. We collected a dataset of live videos broadcasted at the Facebook Live from January 2018 to September 2018. The dataset contained the locations of the viewers watching specific live streams. Our study depicted that several users were watching the same live video within close vicinity, presenting a strong motivation to use multicast instead of unicast within the RAN. However, the eMBMS architecture cannot be used to multicast Facebook Live streams, as it works only with the service-based infrastructure. *What if we want to multicast the live streams from service-less crowdsourced live stream provider*? In this paper, we propose an architecture to answer the key question, how to initiate an on-the-fly service-less multicast resource bearer from a crowdsourced live video streaming server, which is not a Multicast Service Provider?

In this paper, we present an overlay multicast network architecture using the Mobile Edge Computing (MEC) and network function virtualization strategies for an ad-hoc multicast service, as depicted in figure 1. We present a VNF - Application Server (VNF-AS) working as a proxy entity on account of both the content provider and the user equipment. There are certain parameters, such as service ID, multicast IP address, session time, etc. that are provided by the user equipment and the multicast service provider to announce a multicast service by the eMBMS architecture. The proposed VNF-AS, on behalf of multicast service provider and user equipment, provides the required parameters to the eMBMS architecture to establish the multicast service. The VNF-AS performs as a proxy entity during the eMBMS phases as the content provider in our case is not a multicast service provider and the user equipment is not subscribed to any multicast service. The details about how the VNF module performs all these aforementioned tasks are mentioned in the design section. The VNF module exploits the multicast transmission rather than the unicast and sends fewer uplink requests for specific content releasing the burden from backhaul, transit, and radio access links. We evaluate the performance of our proposed solution by using a real-world Facebook Live dataset collected between January and September 2018. Results show that our proposed architecture considerably reduces the cost, network consumption, and efficiently utilizes the RAN resources.

Our contributions in this paper include the following:

- Proposed VNF - Application Server-based network architecture to augment eMBMS and enable on-the-fly service-less video multicast without proposing any architectural changes in the standard eMBMS.
- Developed algorithms to request monitoring, detection of potential multicast scenarios, communication between user equipment and VNF-AS, and communication between BMSC and VNF-AS.
- Minimized bandwidth utilization, RAN resources, and cost by reducing the number of streams per content.
- Performed extensive evaluation with a real-world dataset based simulation to depict the validity and efficiency of the proposed model.

The remainder of this paper is organized as follows. We present the related work in Section III. Section IV contains the system design along with the overview of eMBMS architecture and our detailed proposed solution. We present our simulation configuration, evaluation methods, and the detailed results of our extensive simulations in Section V. Finally, we conclude our paper in Section VI.

II. BACKGROUND AND RELATED WORK

eMBMS architecture has been discussed in the literature in general for performance evaluation of broadcast and multicast techniques. This evaluation has been conducted for various aspects of LTE multicast that include, network slicing for multicasting traffic [15], MBMS-based architectures [13], [16] along with effective cooperative strategies [17]–[20]. These strategies include scheduling techniques [21], [22], caching and multicast techniques [23]–[26], and physical layer analysis [27], [28]. Some authors examined the MBMS in specific application scenarios, such as the cooperation of disjoint networks, intelligent transportation systems, and video transmission.

Kamran et. al. in [29] discussed the need of service-less multicast video delivery in 5G cellular networks. Authors discussed the possible challenges in enabling a non-subscription-based on-the-fly multicast service and a promising solution to those challenges. The proposed solution includes a combination of NFV and MEC to augment the eMBMS architecture. This study presents a high level architectural design to instantiate the ad-hoc multicast bearer while optimizing the traffic load on the backhaul, fronthaul, and RAN links.

Qin *et al.* presented a BMSC design in [30] that is most similar to our work. A new structure of the BM-SC is proposed to help IPTV transmission in LTE networks. This paper presented a wireless multicast solution named as Wi-Live exploiting the Software Defined Network (SDN) technique to adaptively detect the potential multicast scenarios in the wireless network. Wi-Live enables a multicast service adaptively and groups the users attached to the same access point. Authors claim that this wireless multicast solution targets the Internet infrastructure, which supports unicast only. The authors used commodity computers as the gateways and eNBs and depicted the transfer of media files over the Ethernet. This study did not mention any part of the solution at the RAN level. Furthermore, the authors did not mention, which cellular technology this solution targets specifically. The eMBMS is standardized by 3GPP for broadcast/multicast services but Wi-Live does not mention anything about eMBMS and how the multicast services will be initiated. Lastly, Wi-Live demands changes in the user equipment to inform the user equipment about multicast services and join them explicitly. But, authors did not mention what are the changes required to support the Wi-Live.

Fuente *et al.* in [31] proposed a joint multicast/unicast scheduling model. This solution exploits the dynamic optimization of the LTE frames to increase the throughput in an MBSFN service area. This scheduling solution joins the unicast and multicast transmissions to achieve a threshold bit rate for a specific multicast. The algorithm selects an optimum MCS index and the optimum number of multicast subframes for each LTE frame. Another MBSFN related

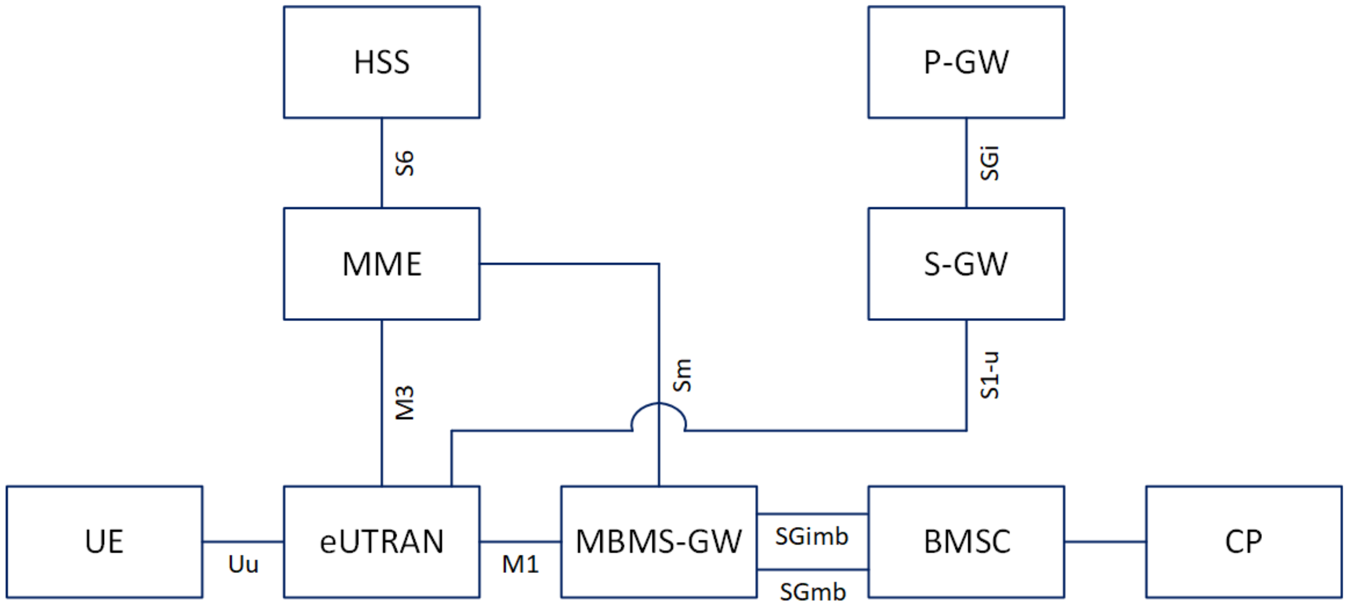

FIGURE 2. Architecture of eMBMS.

study is presented in [32] that is focused on the dynamic deployment of MBSFN service areas. This work studies creation of MBSFN service area and its dependence on the cell size based on the Signal to Interference and Noise Ratio (SINR). This method optimizes the efficiency of broadcasting/multicasting techniques by creating the MBSFN areas dynamically for broadcast/multicast events. A group-based multicast method is introduced in this study to provide efficient broadcast/multicast services to all the users in the specific region based on different MCSs and radio channel quality.

In [33], Damera and Babu proposed an MBS based architecture that offers multiple content to viewers at the same time. The proposed scheme aims to increase QoS of the user, business revenues for the service provider, and provide efficient network resources utilization. To schedule radio resources, the Multicast Coordination Entity (MCE) exploits a scheduling algorithm for each cell taking into account different attributes, i.e., type of user and service, QoS, etc. As far as service architectures are concerned, [34] evaluates the IP Multimedia System (IMS) and LTE framework system. The purpose of this work is to provide an IMS based architecture to increase the efficiency of video conferencing and to reduce the operational cost. The major contribution of this work is to provide an alternative IMS-based architecture that exploits the LTE base stations capacity to increase the video conferencing efficiency in each cell.

Most of the studies in the literature, except [29] neither proposed a RAN level solution nor developed a novel adaptive multicast scheme. Whereas, our proposed solution is an adaptive multicast service that works at the RAN level. Wi-Live is an architecture that is most relevant to our proposed work that uses the commodity computers and Ethernet network to simulate a wireless network multicasting architecture. Whereas, we performed our simulations in a properly configured 4G cellular network. Most of the works, including Wi-Live, do not mention sufficient details about eMBMS that is a standard broadcast/multicast architecture. On the other hand, we considered all the necessary parameters of an eMBMS service to validate our proposed architecture within the eMBMS. Some of the works, specifically Wi-Live, demand changes in the user equipment to start a multicast service but our proposed architecture do not need any changes in the user equipment or the eMBMS architecture.

III. SYSTEM DESIGN

In the design section, we elaborate the 3GPP standard eMBMS architecture and highlight the key requirements to initiate a live video multicast session, the solution to fulfill those requirements, and method to start a service-less live video multicast transmission using eMBMS delivery.

A. eMBMS OVERVIEW

The 3GPP TS 23.246,[2] i.e., release 15, discuss the overall aspects of eMBMS services for 5G networks. The eMBMS, as defined by the 3GPP, is a Point to Multipoint (PTM) service for delivering specific content from one source to numerous destinations. Transmission of identical data to multiple destinations consents the sharing of radio and core network resources that increases the network efficiency. eMBMS supports two transmission modes, i.e., broadcast and multicast mode. The eMBMS uses a common channel for both the broadcast and multicast transmissions, irrespective of the number of destinations. This common channel can be

[2] ftp://ftp.3gpp.org/specs/archive/23series/23.246/

TABLE 1. eMBMS related acronyms.

BMSC	Broadcast Multicast Service Center
BCMSC	Broadcast Multicast Service
CDN	Content Delivery Network
CQI	Channel Quality Indicators
CLVP	Crowdsourced Live Video Provider
eMBMS	evolved Multimedia Broadcast Multicast Services
EPC	Evolved Packet Core
EPS	Evolved Packet System
eUTRAN	evolved UMTS Terrestrial Radio Access Network
MBMS	Multimedia Broadcast Multicast Service
MBSFN	Multicast Broadcast Single Frequency Network
MCE	Multicast Coordination Entity
MCS	Modulation and Coding Scheme
MEC	Mobile Edge Computing
MME	Mobility Management Entity
MNO	Mobile Network Operator
MooD	Multicast operation on Demand
MRB	Multicast Radio Bearer
MSP	Multicast Service Provider
MBMS-GW	Multimedia Broadcast Multicast Service - Gateway
RAN	Radio Access Network
SC-PTM	Single Cell - Point To Multipoint
TMGI	Temporary Mobile Group Identifier
QoE	Quality of Experience
QoS	Quality of Service
VNF-AS	Virtual Network Function - Application Server

the Physical Broadcast Channel (PBCH) for broadcast or the Physical Multicast Channel (PMCH) for multicast. The existing 3GPP architecture accommodates some new capabilities and functional entities to realize the eMBMS, as shown in figure 2.

The eMBMS architecture involves four basic entities to carry out broadcast/multicast services. First entity is Broadcast Multicast Service Center (BMSC) that is the source of multicast content within the core network. The responsibilities of BMSC include authentication of MSP, i.e., if a content provider offers a service?, authorization and authentication of user equipment, i.e., if the user is subscribed to the service provided by the service provider?, authorization and initiation of broadcast/multicast services, i.e., creating a user service offered by the MSP, creating a bearer service offered by the BMSC, and to coordinate with other eMBMS entities to establish the bearer and user services. Second, the eMBMS-Gateway (MBMS-GW) performs the eMBMS data delivery received from BMSC to eNBs and service regions in SC-PTM and MBSFN, respectively during the multicast session. The handovers across cells are handled by the Mobility Management Entity (MME) within the service area that is the third entity within the core network. Lastly, the Multicast/Multicell Coordination Entity (MCE) handles the control signaling between the eMBMS core and the eNBs.

In the current eMBMS scenario, cellular network multicast can only be carried out by receiving the pre-announced content from the content provider and delivered only to the subscribed viewers. To multicast a video, the eMBMS requires various actions performed at content provider and user equipment level. First, the content provider must be an MSP with its user services available for its subscribers. The content provider needs to contact eMBMS for authentication and authorization. The user equipment needs to be subscribed to a content provider for a specific service. Then eMBMS needs to instantiate its multicast bearer services to deliver the multicast user services. At the user equipment end, some services related changes are also required as mentioned in [4]. The user equipment should be able to receive the System Information Blocks (SIB) related to eMBMS, especially SIB-13, SIB-15, and SIB-16. Moreover, user equipment needs to decode the SIBs correctly to receive multicast, support Multicast operation on Demand (MooD), and support unicast to multicast and multicast to unicast switching.

Network function virtualization [35] technology has emerged as a promising solution to deal with the ever-increasing demands for a variety of proprietary hardware in present telecom networks. As in recent times, the applications are retained by dynamically configurable cloud environments, virtualized network functions allow the networks to be swift and responsive towards the requirements of the services running over it [36].

The eMBMS bearer services are used to transmit the broadcast/multicast content. The bearer services are provided by the MNO. However, before establishing the bearer service, the BMSC within the MNO core and the user equipment must be aware of the eMBMS service through the service announcement. The following steps are performed in eMBMS for the announcement of user services, activation of a multicast service, and joining of users.

1) The MSP contacts the BMSC for the announcement of user service, e.g., live coverage of a football match, containing the service-related information.
2) BMSC announces the user service to all the service areas mentioned in the service announcement. According to TS.23.246, the service announcement can be performed through SMS, WAP, Push, Cell Broadcast, etc.
3) The interested users activate the general-purpose Packet Data Protocol (PDP), sends MBMS Service Request, and sends IGMP (v4) Join message to the BMSC with the following information. IMSI: International Mobile Subscriber Identifier; IMSI = MCC + MNC + MSIN MSIN: Mobile Station Identification Number APN: Access Point Name
4) BMSC sends the same data to the MSP requesting the MSP for the authorization and authentication of the user as a subscribed user with the mentioned service.
5) MSP responds to the user subscription query with a positive acknowledgment message.
6) BMSC requests the user equipment to activate the PMM-Connected mode and the user equipment Context, as shown below. The user equipment Context is saved in both the user equipment and BMSC after the user equipment linking phase.
7) user equipment sends the Activate MBMS Context Request to BMSC with the following information. NSAPI: Network Service Access Point Identifier.

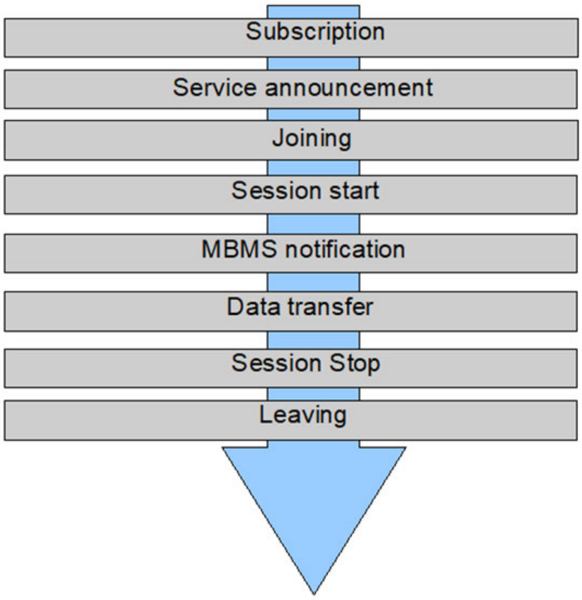

FIGURE 3. Phases of eMBMS.

The MBMS Context contains following information at the BMSC's end. Service Area, TMGI, QoS parameters, List of cells participating in the multicast service, List of downstream nodes.

8) BMSC interrogates the MSP for the availability of the service in the mentioned Service Area by sending the following information. RAI: Routing Area Identity MSP sends the acknowledgment to BMSC as MBMS Authorization Response.
9) MSP sends the acknowledgment to BMSC as MBMS Authorization Response.
10) BMSC sends MBMS Registration Response to the user equipment, accepts the Activate MBMS Context Request of user equipment, and allocates TMGI.

Steps 1, 5, and 9 are performed by the MSP, steps 2, 4, 6, 8, and 10 are performed by the BMSC, and steps 3 and 7 are performed by the user equipment to complete the service announcement, multicast service creation, and user equipment joining. The Session Start, Data Delivery, and the rest of the procedures are mentioned in TS.23.246 in detail. According to TS.23.246, the eMBMS services are provisioned by performing the following phases sequentially as shown in figure 3.

- Subscription: The users subscribe to an MBMS service provided by a multicast service provider and may get charged for the service.
- Service Announcement: Through this phase, all the available user services are notified to the subscribed users mentioning the type of service and time of transmission.
- Joining: In the joining phase, the multicast user groups are formed and the users show their willingness for the reception of an eMBMS Bearer Service.
- Session Start: This point in time marks the start of a multicast session for the subscribed users.
- MBMS Notification: The users are notified with a session start time showing that the data transfer has started.
- Data Transfer: The actual eMBMS data delivery is carried out in this phase on an MRB.
- Session Stop: This phase specifies the end of the broadcast/multicast session and data delivery is stopped. The MNO bearer resources will be released after this phase.
- Leaving: The user no more wishes to receive the eMBMS data and leaves the multicast session.

Most importantly, to initiate a multicast session using eMBMS delivery, there are several requirements, which must be met.

1) In the subscription phase, the user equipment must be subscribed to a pre-defined service offered by an MSP. The provider may charge the users for the services. The BMSC must interrogate the MSP for the user equipment authentication and authorization for the services to be announced.
2) In the Service Announcement phase, the BMSC must announce the Services ID only to the subscribed users. The service ID must be provided by the service provider, which should be an MSP.
3) In the joining phase, the BMSC must interrogate the user equipment for several user equipment related parameters i.e., MSISDN, Cell ID. Then the BMSC decides the Service Area for the users and must allocate a TMGI accordingly. By now, the user is added to a multicast group with an IP multicast address. This IP multicast address is notified to all the intended users before the joining of the service.

In general, the Crowdsourced Live Video Providers (CLVP), such as YouTube or Facebook do not meet the aforementioned requirements because: **(a)** CLVPs are not multicast service providers, **(b)** the viewers of a live video are not required to subscribe, **(c)** CLVPs generally do not announce the service, **(d)** MSPs generally stream a single video at a specific time, e.g., some game event, concert, or similar stream, whereas, thousands of streams are broadcasted by CLVP. To stream the video using multicast, CLVP has to announce separate service for all live videos, which is infeasible at MSP level, and **(e)** CLVP streamer do not follow a strict schedule, i.e., a streamer may start streaming a video at any time for any duration. Therefore, the basic requirements to use the eMBMS architecture for multicast are not met by the CLVPs. In such a scenario, how can a CLVP video viewed by multiple viewers in close vicinity be delivered using multicast? The ensuing section will answer the question.

B. PROPOSED SOLUTION

As discussed in the previous section, service-less multicast for CLVPs is not possible using the standard eMBMS architecture. In this section, we propose a VNF-MEC-based

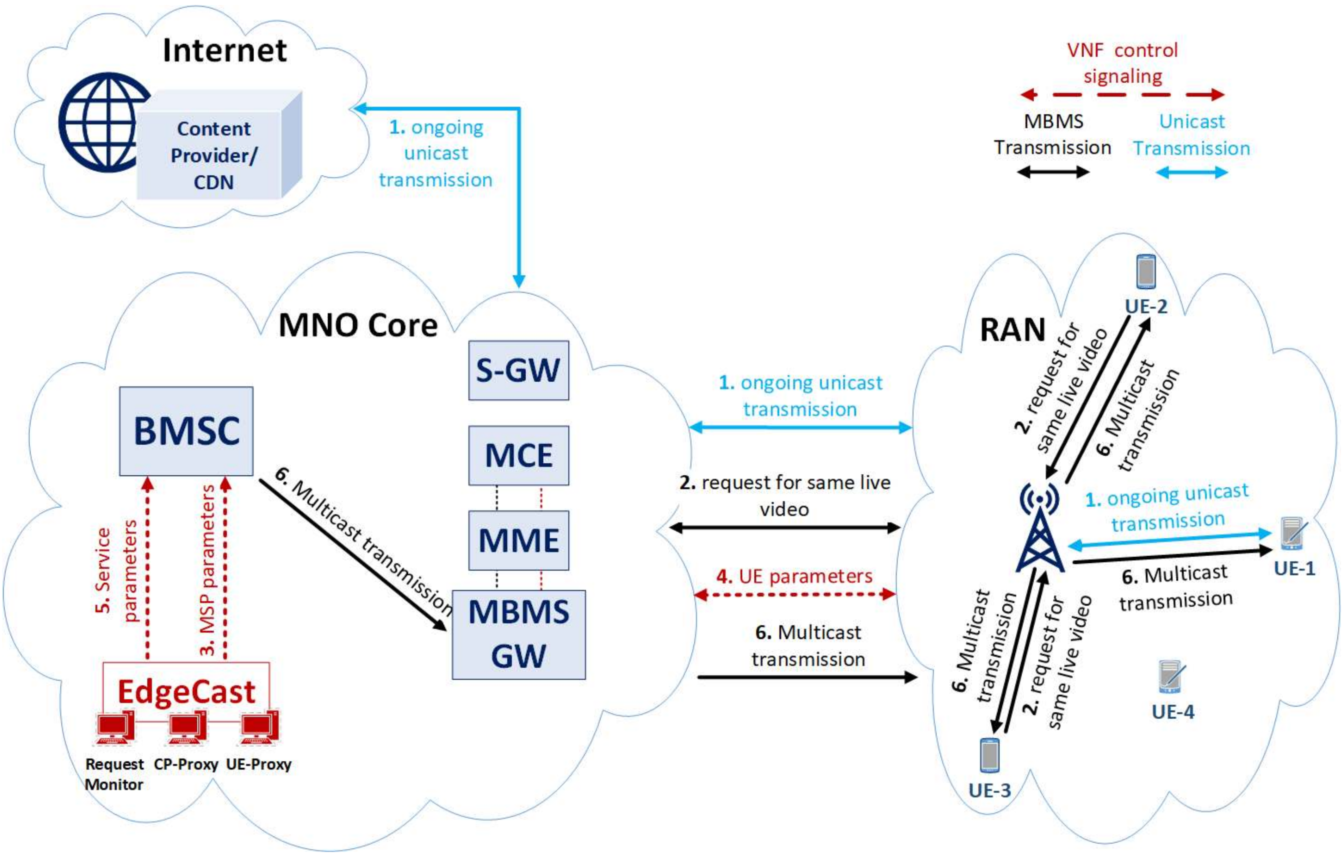

FIGURE 4. Multicast at Edge architecture.

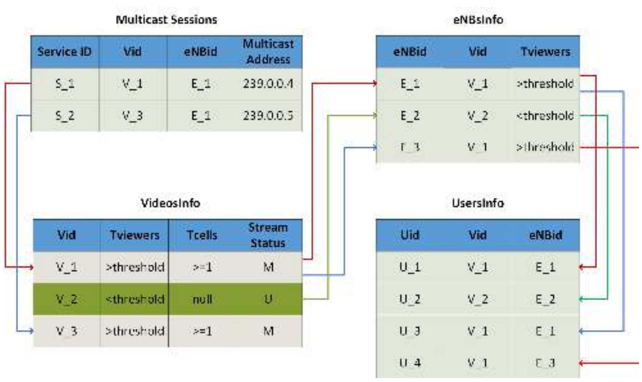

FIGURE 5. Services-Videos-Cells-Users mapping.

overlay network architecture to augment the standard eMBMS architecture. The proposed solution supports multicast for service-less and non-subscription-based live videos from the CLVPs.

Our proposed solution can be plugged in with the standard eMBMS architecture at core network level, with no or minimal changes in the standard eMBMS architecture (to improve the efficiency and viewers QoE). The primary requirements in the eMBMS include announcement and subscription of multicast service, from the content provider and user equipment's perspectives, respectively. We propose EdgeCast, to initiate service on the fly, based on the current demand for a specific video within the area covered by the core network. The EdgeCast imitates the actions required by the content provider and user equipment's to initiate the service, i.e., the EdgeCast acts on behalf of the content provider to announce a multicast service for the BMSC. Similarly, the EdgeCast communicates with BMSC on behalf of user equipment's for the subscription and authentication. When all requirements for an eMBMS multicast are met, the video is multicast using the standard eMBMS. An overview of how the EdgeCast converts an ongoing unicast to multicast is presented in figure 4.

The EdgeCast, being a core network entity keeps monitoring all the ongoing video streams. If, a user requests a live video stream that is already being transmitted to another user within the same cell. The EdgeCast performs the required steps, i.e., communicating with the BMSC on behalf of user equipment and MSP, to create a multicast bearer for that particular live stream in that particular cell or service area. In a nutshell, the EdgeCast creates a multicast service on the fly, subscribes to the participating user equipments, creates multicast groups, and enables eMBMS to multicast the live video stream. The detailed design and working of EdgeCast is discussed below.

The EdgeCast is comprised of three major modules: (a) Request monitoring and multicast/unicast session initiation module, (b) content provider proxy module (CP-Proxy), and (c) user equipment proxy module (UE-Proxy).

The request monitoring module monitors all the incoming requests for the live video and manages the initiation and end of the multicast sessions. The working of the request monitoring module is depicted in Algorithm 1 and Algorithm 2. Algorithm 1 accounts for the users' session start and end. The request monitoring module manages all active users' information in multiple tables, i.e., UsersInfo, VideosInfo, eNBsInfo, and MulticastSessions as depicted in figure 5. UsersInfo table contains the User ID (U_id) e.g., the mobile number, eNB ID (eNB_id) of the user, and Video ID (V_id) being watched. VideosInfo table contains the required information related to all of the active videos being watched within the service area covered by the MNO core, such as Video ID (V_id), total viewers (Tviewers), total cells (TCells), and status of the video, i.e., U for unicast or M for multicast. eNBsInfo table contains information of viewers for a specific video within one eNB coverage area, such as Video ID (V_id), eNB ID (eNB_id), and total viewers (Tviewers). MulticastSessions table contains the information related to ongoing multicast sessions, such as Service Id (dynamically generated on behalf of content provider), Video ID (V_id), multicast address, and the eNB ID (eNB_id) for which the service has been created. The relationship mapping among the tables is shown in figure 5.

Algorithm 1 Request Monitoring Module (Potential Multicast Scenario Detection)

1: **Input:** User ID, Video ID, eNB ID
2: lookup Video ID AND eNB ID in MulticastSessions table
3: **if** found **then**
4: increment Tviewers in Videosinfo table by 1
5: increment Tviewers in eNBsinfo table by 1
6: add User ID, Video ID, eNB ID in Usersinfo table
7: **else**
8: lookup Video ID in VideosInfo table
9: **if** found **then**
10: Generate Service_ID
11: add Service_ID in MulticastSessions table
12: add Video ID in MulticastSessions table
13: add MulticastAddress in MulticastSessions table
14: add eNB ID in MulticastSessions table
15: set status = M
16: increment Tviewers in Videosinfo table by 1
17: increment Tviewers in eNBsinfo table by 1
18: increment Tcells in Videosinfo table by 1
19: **else**
20: Set status = U
21: add User ID, Video ID, eNB ID in Usersinfo table
22: add User ID, Video ID, eNB ID in Videosinfo table
23: add User ID, Video ID, eNB ID in eNBsinfo table
24: **end if**
25: **end if**

Multiple viewers watch a live video stream from the content provider at the same time. The stream monitor module analyzes every video request routing through the core network. On arrival of the request, the monitor module looks into the MulticastSessions table for a collective entry of V_id and eNB_id (line 2, algorithm 1). If the V_id and eNB_id are available, the number of total viewers in Videosinfo and eNBsinfo is incremented by 1, the V_id is updated with the currently requested video, otherwise, a new record is inserted, comprised of U_id, V_id, and eNB_id in Usersinfo, Videosinfo, eNBsinfo tables (line 4-6, algorithm 1). The V_id is searched in the VideosInfo table (line 8, algorithm 1), if found, search eNB_id in the eNBsInfo table. If the eNB_id is found, the number of viewers is incremented by 1 in Videosinfo and eNBsinfo tables, the number of cells is incremented by 1 in Videosinfo table, generate a Service_ID, add Service_ID, V_id, eNB_id, and MulticastAddress in MulticastSessions table and set the status of V_id as multicast in VideosInfo table (line 10-18, algorithm 1). If the eNB_id is not found, a new entry is created in UsersInfo, VideosInfo, and eNBsInfo tables and set status as unicast in VideosInfo table (line 20-22, algorithm 1). The process is repeated in reverse order, when a viewer leaves a video session (line 3-5, algorithm 2), i.e., user details are removed from the UsersInfo, total viewers are decremented in VideosInfo (or video record is deleted), viewers are decremented in eNBsInfo (line 9-10, algorithm 2), and algorithm 2 is called for possible unicast switching (line 11-17, algorithm 2).

Algorithm 2 Request Monitoring Module (User Leaves the Video Session)

1: **Input:** User ID, Video ID, eNB ID
2: lookup eNB ID in eNBsInfo table
3: **if** Tviewers $< \tau_1$ **then**
4: lookup Video ID in VideosInfo table
5: **if** Tviewers $< \tau_2$ **then**
6: delete Service_ID from MulticastSessions
7: **end if**
8: **else**
9: decrement Tviewers in eNBsInfo table by 1
10: decrement Tviewers in VideosInfo table by 1
11: lookup eNB ID in eNBsInfo table
12: **if** Tviewers = 0 **then**
13: delete eNB ID from eNBsInfo table
14: **end if**
15: lookup Video ID in VideosInfo table
16: **if** Tviewers = 0 OR Tcells = 0 **then**
17: delete Video ID from VideosInfo table
18: **end if**
19: **end if**

The key purpose of the monitoring module is to decide whether the video should be delivered as unicast or multicast to the viewers. Algorithm 3 illustrates the process of multicast initiation and termination. If the number of users watching the same video increase the threshold τ_1 (line 3, algorithm 3), request monitor module initiates a multicast session (line 4-9, algorithm 3). The multicast is initiated by requesting the content provider proxy module to act on behalf of the MSP and the user equipment proxy module to act on behalf of the user equipment (line 10-16, algorithm 3).

The parameters τ_1 and τ_2 are important configurable parameters that have huge impact on users' QoE. The value of τ_1 can be set by the MNO and can be changed at any time according to the needs and goals of the MNO to achieve higher QoE, cost benefits, and/or lower core network traffic burden. Whereas the value of τ_2 is updated at each new entry in the VideosInfo table. The value of τ_1 defines the number of users watching the same live stream at which the MNO wants to check for a possible multicast scenario. However, the value of τ_2 specifies that how many users are currently watching the same live video stream. The relation between τ_1 and τ_2 can be seen as the value of τ_2 should always be greater than or equal to τ_1 in order to initiate a multicast session. This means that a higher ratio between τ_2 and τ_1, results in a higher multicast session benefit, i.e., τ_2 / $\tau_1 \geq 1$.

Algorithm 3 Request Monitoring Module (Multicast Resource Bearer Establishment)

1: **Input:** User ID, Video ID, eNB ID
2: lookup eNB ID in eNBsInfo table
3: **if** Tviewers $> \tau_1$ **then**
4: lookup Video ID in VideosInfo table
5: **if** Tviewers $> \tau_2$ **then**
6: add Service_ID in MulticastSessions table
7: add Video ID in MulticastSessions table
8: add eNB ID in MulticastSessions table
9: Multicast_IP_Address in MulticastSessions table
10: call user equipment proxy( )
11: **if** Cell_ID, Media_Description received **then**
12: call content provider proxy (Service_ID)
13: **end if**
14: **if** Service_Description_ACK received **then**
15: Send TMGI to content provider proxy
16: Multicast_IP_Address to content provider proxy
17: **end if**
18: **end if**
19: **end if**

The user equipment proxy module and the content provider proxy module are responsible to imitate the responsibilities of user equipment and the MSP. The steps related to MSP, i.e., steps 1,5, and 9 and the steps related to user equipment, i.e., steps 3 and 7 mentioned in the eMBMS overview section need to be performed by the user equipment proxy and content provider proxy respectively.

Once a potential multicast scenario is found, the EdgeCast needs to create a Multicast Service. For this task, requirements 1, i.e., authorization of MSP and user equipment as a subscribed user and requirement 2, i.e., service ID provision by the MSP presented in the previous section need to be fulfilled. In eMBMS, the sequential steps, presented in the previous section, i.e., the announcement of user service, activation of a multicast service, and joining of users need to be performed.

First, the content provider proxy module of the EdgeCast system contacts the BMSC on behalf of the MSP, over the xMB interface and provides a service announcement.

Algorithm 4 User Equipment Proxy

1: **Input:** Potential_Multicast_Situation_Signal
2: **if** Potential_Multicast_Situation_Signal received **then**
3: Send SIB_1 to user equipment with SystemInfoValueTag = TRUE
4: Wait ModificationPeriod = 8ms
5: Send SIB_2 to user equipment
6: Wait ModificationPeriod = 8ms
7: Send SIB_13 inquiry to user equipment for MSISDN
8: Send SIB_13 inquiry to user equipment for Cell_ID
9: Send SIB_13 inquiry to user equipment for Media_Description
10: Wait for SIB_13 response from user equipment
11: Send MSISDN to Monitor
12: Send Cell_ID to Monitor
13: Send Media_Description to Monitor
14: **end if**
15: **if** request received from content provider proxy **then**
16: Send MSISDN to content provider proxy
17: Send Cell_ID to content provider proxy
18: Send Media_Description to content provider proxy
19: **end if**
20: **if** Service_Announcement Received **then**
21: Decode Service_ID, Session_Description
22: Send Session_Acknowledgment
23: **end if**

The BMSC sends the service announcement in the user equipment's cell, as mentioned by the MSP. The user equipment proxy sends SIB_1 to user equipment with SystemInfoValueTag = TRUE, on the Uu interface (line 3, algorithm 4). After waiting for the modification period of 8ms (line 4, algorithm 4), the user equipment proxy sends Send SIB_2 to user equipment stating that network conditions have changed and wait another 8ms modification period (line 5-6, algorithm 4). Then, user equipment proxy Sends SIB_13 inquiries to user equipment, for MSISDN, Cell ID, NSAPI, and Media Description (line 7-9, algorithm 4). Upon reception of the SIB-13 response from the user equipment, the user equipment proxy module sends MSISDN, Cell ID, and Media Description to the Monitor module (line 11-13, algorithm 4). Following, the user equipment proxy generates the IMSI, i.e., IMSI = MSISDN + MCC + MNC, where MCC and MNC are fixed. Further the user equipment proxy activates the PDP instance on behalf of the user equipment and sends the MBMS Service Request and IGMP Join message to the BMSC over the MB2 interface. user equipment

Algorithm 5 Content Provider Proxy

```
1:  Input: Service_ID
2:  if Service_ID received then
3:      Request MSISDN to user equipment proxy
4:      Request Cell_ID to user equipment proxy
5:      Request Media_Description to user equipment proxy
6:      Receive MSISDN to user equipment proxy
7:      Receive Cell_ID to user equipment proxy
8:      Receive Media_Description to user equipment proxy
9:      Send Service_ID to BMSC
10:     Send Media_Description to BMSC
11:     Send Session_Description to BMSC
12:     Send Type_of_Service to BMSC
13: end if
14: if Multicast_IP received then
15:     Set TMGI = Service_ID + MCC + MNC
16:     TMGI, Multicast_IP to BMSC
17: end if
```

proxy mentions the IMSI, Cell ID as APN, Service ID, and Multicast IP address in the IGMP Join request.

Then the BMSC sends the request received by the user equipment proxy to the content provider proxy on the xMB interface (as the content provider proxy is the MSP in this case) for the user authorization. The content provider proxy sends the positive ACK message to the BMSC in response. The BMSC asks the user equipment proxy (on behalf of user equipment) to activate the PMM-Connected mode and the user equipment Context. The user equipment proxy saves the user equipment Context (IMSI, Multicast IP Address, APN Name, TMGI) on behalf of the user equipment.

Further, the user equipment proxy sends Activate MBMS Context Request to BMSC, over MB2 interface, with Multicast IP Address, APN Name, and NSAPI.

Upon reception of Activate MBMS Context Request by the user equipment proxy, the BMSC sends the IMSI, MSISDN, and RAI to content provider proxy (on behalf of MSP), over xMB interface, to check the availability of the service in the cell (line 2-12, algorithm 5). The content provider proxy sends the MBMS Authorization Response to the BMSC. The BMSC sends the MBMS Registration Response to user equipment proxy (on behalf of user equipment).

The monitor module adds the user in a multicast group and allocates the TMGI and Multicast IP address to the group (line 14-16, algorithm 3). The user equipment proxy sends SIB-1 to user equipment with SystemInfoValueTag = TRUE. The user equipment proxy requests the user equipment to activate PMM Connected Mode. The user equipment proxy sends SIB-13 to user equipment with a new media description mentioning the multicast IP address (line 20-22, algorithm 4). Finally, the BMSC starts delivering the eMBMS data on the multicast IP address.

The functionalities of user equipment proxy and content provider proxy are elaborated in algorithm 4 and algorithm 5 respectively. The communication between monitor, user equipment proxy, content provider proxy, user equipment, and BMSC is depicted in figure 6 in sequential steps.

TABLE 2. Dataset and simulation parameters.

Total Live Video Streams	750,000+
Total Number of Broadcasters	170,000+
Total Categories of Videos	942
Total Number of Fetch Cycles	22
Average Fetch Cycle Period	180 seconds
Total Number of Simulation eNBs	8
Average Viewer Count per eNB	129
Length of the Simulation Video	4608 seconds
Average Video Representation	480p

IV. EVALUATION

A. DATA SET

We used real-world live video traces taken from Facebook to analyze the performance of our system. These traces were generated using the Facebook LiveMap API [3]. This dataset is generated for two periods for January and February. Period 1 is from January 28th, 2018 to February 3rd, period 2 is from February 9th, 2018 to February 16th, 2018. This dataset presents a collection of over 750,000 live video streams from Facebook Live with a fetch period of 3 minutes. A total of 942 different categories of videos are included in this dataset with around 170,000 different broadcasters. The categories of videos include various genres such as Sports, Music, News, etc. It has been seen that around 17% of the total viewers of a specific live video are located within the 20Km vicinity of the broadcaster and 23% in her 50Km diameter. It can be concluded confidently that more than 23% viewers watch the same live video while being within the same MNO core.

We preferred to use this dataset due to several reasons. First, this dataset is created for Facebook live videos. As it is evident that Facebook Live is the major contributor to the total number of live videos globally. Therefore, it is an optimum approach to study such a case that affects the real-world networks on a larger scale. Second, a realistic count of the number of live viewers is presented in this dataset along with the total active users at each fetch time. Third, the start and the end time of the live videos are given in the dataset making it easier to find the total active time of the live broadcast. Lastly and most importantly, the locations of the viewers and the broadcaster are also given. We used these locations to generate a real-world cellular network scenario and calculated the realistic Channel Quality Indicators (CQI) and data rates accordingly considering an LTE-A network.

B. SIMULATION CONFIGURATION

We configured evalvid server i.e., a media streaming server developed in NS3[3] and developed an eMBMS architecture that imitated all the functionalities of required eMBMS

[3]https://www.nsnam.org/documentation/

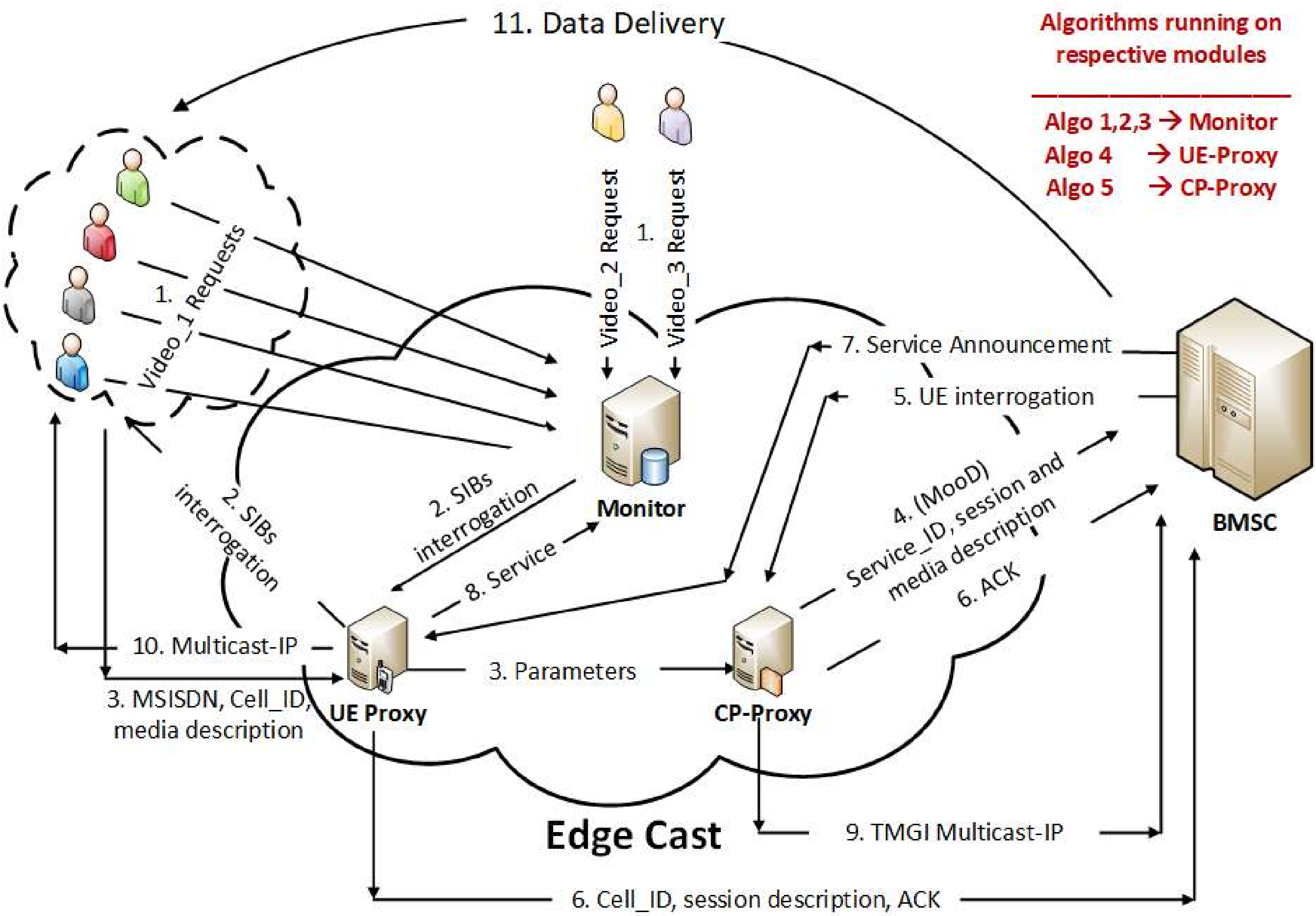

FIGURE 6. Communication between different modules of EdgeCast with other entities.

entities such as BMSC, MBMS-GW, eNBs, and the user equipment in the NS3 simulator. We carefully considered the overheads incurred in a cellular network to establish the multicast sessions. We simulated the live video traces taken from the city of Bangkok, Thailand, and the City of Sao Paulo, Brazil with their respective parameters. We chose Bangkok and Sao Paulo as the target areas because they show the considerable clustering of viewers of the same video within close vicinity. This scenario lets us imitate an LTE network as shown in figure 7. We plot the locations of eNBs and the users on the Microsoft Bing Live Map API.

We configured the MBSFN areas consisting of 8 eNBs at most, as per 3GPP specifications. We simulated a video of 4608 seconds with an average viewer count of 129 for a single eNB for that specific video within the Bangkok city. A 480p representation video is considered in the simulation. The reason to select 480p is that the representation must be of an average quality so that various users with different channel qualities may receive it successfully. Three are a total of 22 fetch periods. In each fetch period, the CQI or the data rates of the respective users remain the same. Most of the users happen to be watching the live video streams between the Bangkok Noi and Pathum Wan districts in Bangkok City, figure 7 (C) and within the Santa Ifigenia in Sao Paulo City, figure 7 (D).

C. RESULTS

We performed comprehensive simulations to analyze the network load and resources used on the transit, backhaul, and RAN links. We compared the obtained results with the legacy unicast methods for live video transmission. The results and comparison with unicast are presented for two cases (a) SC-PTM and (b) MBSFN. In general, users around the globe value their privacy a lot by keeping their location tracking off. The dataset used in our experimentation shows some distinctive facts about the users' locations. It contains the total number of users with their location off, i.e., privacy protection on, and the users with their location on. We show the comparison of actual viewers of a video and the users with their location on in figure 8. Figure 8 shows the results for Sao Paulo city where 10985 viewers watched the video but only 2118 viewers had their locations on. Bangkok city also shows the same results. Our results and analysis are based on the users with their locations ON. However, the information

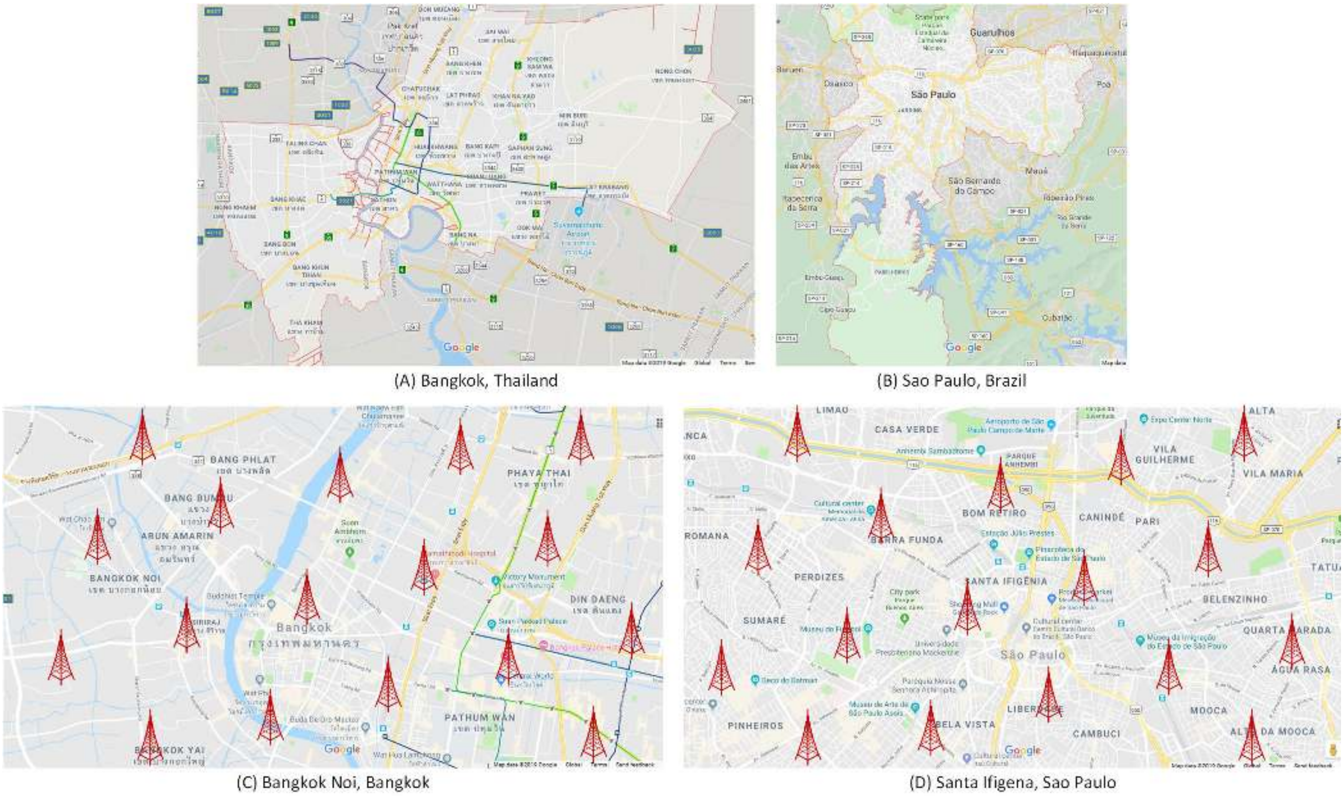

FIGURE 7. **Cellular Network deployment over the Region of Interest.**

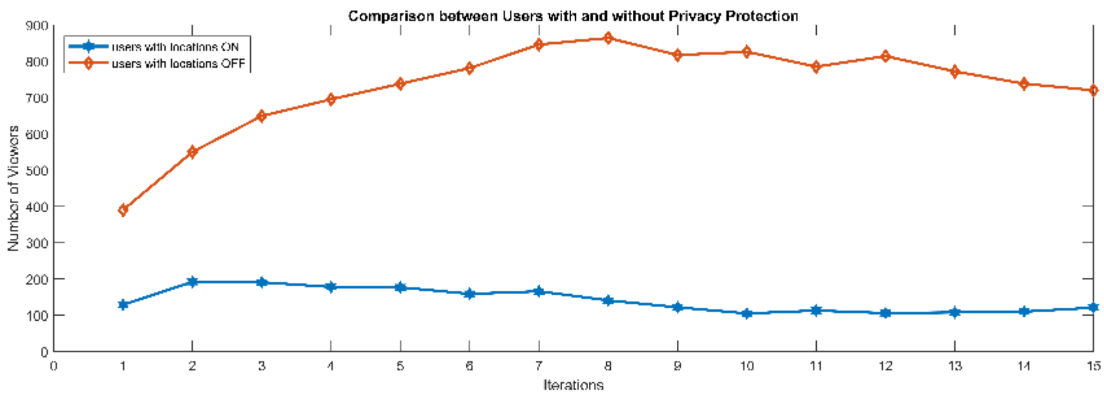

FIGURE 8. **Total viewers vs viewers with location ON (Sao Paulo City, Brazil).**

about the locations of all the users would show a lot of better results.

We start the analysis of our results with the comparison of load on the transit link, i.e., total data transmitted by the CDN in both unicast and multicast modes. The data transmitted by the eNB (backhaul link) is the same as the CDN. Therefore, we only show the results for data transmitted by CDN. In this case, we consider the MBSFN and SCPTM as one scheme, i.e., multicast because the CQI selection at RAN level does not affect the bitrate of the video. The CDN transmits as many streams towards the MNO core as many requests are received by it. We calculated the cumulative data transmitted by the CDN (both unicast and multicast modes) for 22 fetch cycles as follows;

$$\sum_{i=0}^{n}\sum_{j=0}^{m} D_j \times t_j \tag{1}$$

where,

m = number of active users
n = number of active eNBs
D = total data received by a user
t = total active time of the user

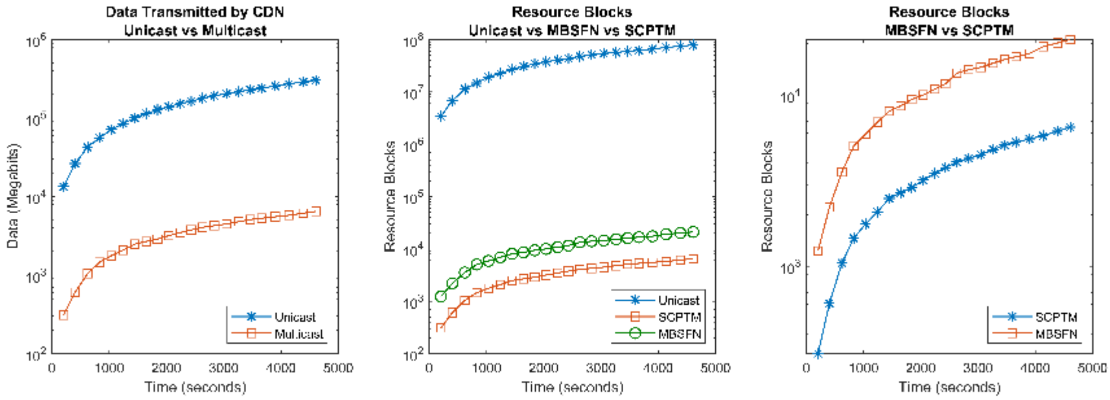

FIGURE 9. Cumulative Data transmitted by CDN and Resource Blocks assigned by eNBs (Bangkok).

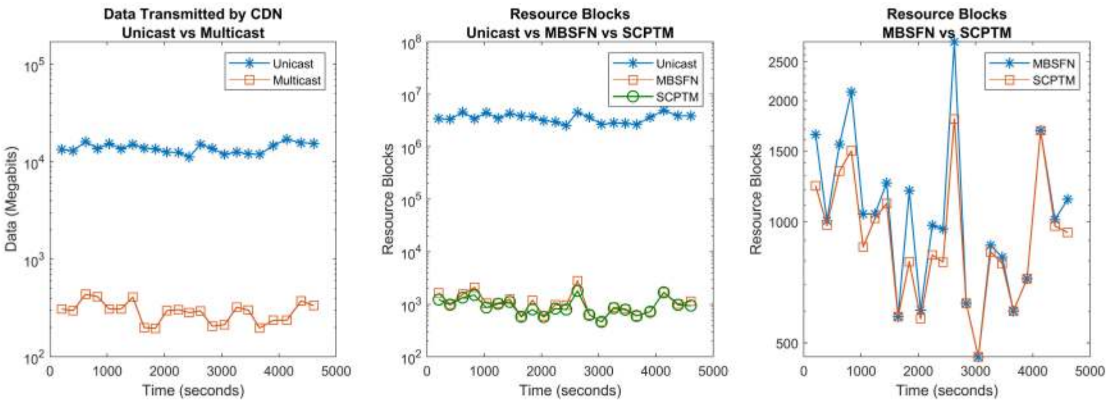

FIGURE 10. Data transmitted by CDN and Resource Blocks assigned by eNBs (Bangkok).

Figure 9 is a log-scaled graph of the cumulative data transmitted (Megabits) over time (seconds) to all the requesting users and the total number of resource blocks assigned by the eNBs during the live video streaming. Figure 10 shows the plots for data transmitted (Megabits) over time (seconds) and the number of resource blocks assigned by the eNBs for each fetch cycle. Resource Block is the smallest unit of data in LTE-A. In LTE-A The data is transmitted in the form of Orthogonal Frequency Division Multiple Access (OFDMA) symbols, which carry several data bits. The number of data bits in one symbol depends on the Modulation and Coding Scheme (MCS) plus the coding rate. The MCS is measured in CQI levels i.e., QPSK, 16-QAM, 64-QAM, and 256-QAM. The CR is measured in indices from 0 (worst) to 31 (best). The MCS and CR, on a combined level, decide how many data bits can be transferred in an OFDMA symbol. When a user's CQI is worst, the eNB has to assign a more number of resource blocks to transmit a specific amount of data as compared to when the user's CQI is good. Moreover, the number of resource blocks also depends on the number of streams flowing through an eNB. More number of streams will incur more usage of resource blocks.

In OFDMA, one resource block is composed of one slot, i.e., 0.5 milliseconds long. One slot is a grid of carries 12 subcarriers each of 15kHz in frequency with 7 Resource Elements (RE) or symbols in each subcarrier. It means one resource block is composed of 84 Res. 2 slots make 1 subframe, i.e., 1 millisecond long. 10 subframes compose one frame, i.e., 10 milliseconds in time. Finally, one second in OFDMA transmission carries a total of 168000 symbols or 2000 resource blocks. We used the following calculations to calculate the total number of resource blocks assigned by each eNB during the 4608 seconds of simulation time.

$$\sum_{i=0}^{n}\sum_{j=0}^{m}\frac{D_j \times t_j}{d_j} \times 2000 \tag{2}$$

where,

$m =$ number of active users
$n =$ number of active eNBs
$D =$ total data received by a user
$t =$ total active time of the user
$d =$ data rate of the user (depends on CQI)

In unicast, the CDN transmitted a total of 303940 Megabytes and in multicast 6485 Megabytes over 4608 seconds. This shows that a huge amount of network load at the transit link is reduced by switching to the multicast transmission mode. The same is the case with total resource blocks used by all the participating eNBs throughout the live video streaming. These results show that a massive amount of resource blocks are saved at the RAN link by exploiting the multicasting technique.

Figures 11 and 12 depict the results (data transmitted by CDN, resource blocks assigned by eNBs) for the city of Sao Paulo, Brazil. These graphs depict the same results as the Bangkok city except that the MBSFN and SCPTM show identical results since almost all the users lie within the same cluster. Therefore, there is a negligible difference in the CQIs of the users for both the schemes. The wireless bandwidth consumption is significantly decreased in Multicast at Edge as the system replaces multiple unicast transmissions by a single multicast transmission.

The bandwidth saving can be calculated through the following formula;

$$B_{multicast} = \left(N_{users} - 1\right) \times BR_{videos} \quad (3)$$

where,

$B_{multicast}$ = the bandwidth saved in multicast mode

N_{users} = the total number of users in each fetch time, receiving the same live video

BR_{video} = the average bitrate of the live video

We calculated the bandwidth consumption at the backhaul link i.e., from core network to the RAN for the period of the live video streaming session. Figures 13 and 14 show the results of bandwidth used by all the participating eNBs through each iteration for Bangkok and Sao Paulo respectively. The bandwidth consumption depends on the bitrate of the streaming video that is 0.5 Mbps, in this case, and the CQIs of the users receiving the stream. In general, the bandwidth consumption of the multicast session must be consistent throughout the streaming time once the multicast bearer is established regardless of the number of active users. Whereas, there is a small variation in the bandwidth consumption of multicast sessions among different fetch cycles, which is due to the selection of new CQI value for each user after every iteration. Results clearly show that the bandwidth consumption in unicast transmission mode is enormously higher than the unicast mode as each user is assigned a different unicast bearer. Therefore, when a new user requests the same stream in unicast mode, the available bandwidth is shared among all the active users degrading the user QoE.

Various QoE metrics have been proposed to quantify the viewers' QoE such that [37]–[40]. He et al. in [39] offered a viewer satisfaction score based on the total number of representations a user can receive. In this work, we also consider the same metric/QoE score with one difference that is the authors in [39] considered 5 different representations whereas we considered 6 representations i.e., 144p, 240p, 360p, 480p, 720p, and 1080p.

Let S_r be the QoE score of a user based on the maximum quality she can receive and the received quality. We calculate a particular user's QoE score based on the user's receivable quality and the received quality of the video. For example, if a user receives a 720p representation whereas she has enough available bandwidth to receive 1080p representation, then the QoE score is less. Let $l = |R|$ represents the all possible

TABLE 3. QoE score evaluation.

Receivable Video	Received video representation					
	1080	720	480	360	240	144
1080	1	0.83	0.67	0.50	0.33	0.17
720	-	1	0.83	0.67	0.50	0.33
480	-	-	1	0.83	0.67	0.50
360	-	-	-	1	0.83	0.67
240	-	-	-	-	1	0.83
144	-	-	-	-	-	1

representations set, i.e., $l \leq 6$ in this case, $a = \frac{1}{l}$ represents the loss in the QoE score while receiving a low-quality representation, l_r is the QoE score achieved by the user, and l_u is the maximum achievable QoE score of the user, then the loss in the QoE score is:

$$d_u = |l_u - l_r| \times a \quad (4)$$

The user's achieved QoE score can be then formulated as:

$$q_u = q_{max} - d_u \quad (5)$$

where q_{max} is the maximum receivable quality/QoE score i.e., 1. The satisfaction level or QoE score of the user can then be obtained as:

$$S = U_{max} - q_u \quad (6)$$

where U_{max} represents the highest satisfaction/QoE score i.e., 1.

To calculate the achievable QoE score on the real-time data, we considered the eNB with the highest number of attached users. Furthermore, we considered the iteration number 13 from all the 22 iterations because, in this iteration, a maximum number of users appeared with their locations on, i.e., 150 users. We dispersed these users within the cell area using the normal distribution and assigned them the CQI values based on the distance from the eNB. We calculated total data received by each user and the total number of RBs assigned to the user depending on the CQI level. As mentioned earlier, a lower CQI level demands a greater number of resource blocks affecting the network operation cost. For example, a user with 1,2, or 3 CQI can receive 6 bits/RB whereas a user with 15 CQI will receive 36 bits/RB. We then calculated total available RBs at the given eNB from the real-world data, i.e., 29361 RBs for iteration number 13.

In general, if the network cost needs to be reduced then QoE of the user could get affected. To calculate the QoE, we conducted multiple experiments considering various scenarios. When considering the 150 users in a cell, regardless of the selection of multicast mode, i.e., MBSFN or SC-PTM, the QoE in multicast mode increases tremendously in comparison to unicast. We calculated the aggregated QoE of all the 150 users in both unicast and multicast modes that appear to be 0.47 for unicast and 1 for the multicast. However, the QoE does not need to always be enhanced in the multicast mode. In some cases, the QoE of the users decreases when multicasting the video as the number of RBs is limited. For example, when we considered only one user per CQI level,

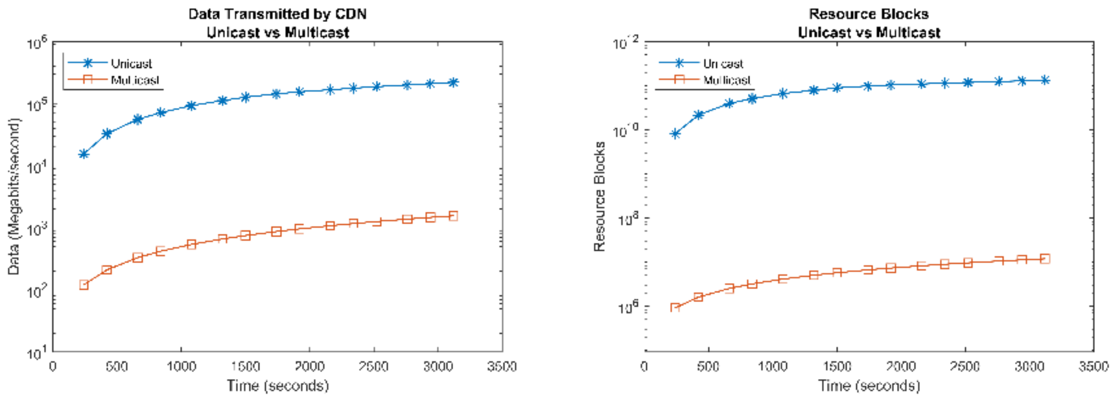

FIGURE 11. Cumulative Data transmitted by CDN and Resource Blocks assigned by eNBs (Sao Paulo).

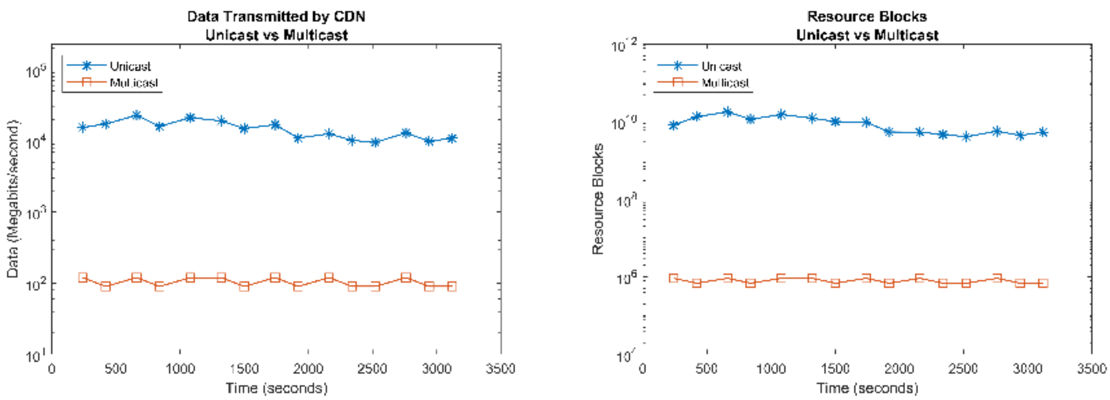

FIGURE 12. Data transmitted by CDN and Resource Blocks assigned by eNBs (Sao Paulo).

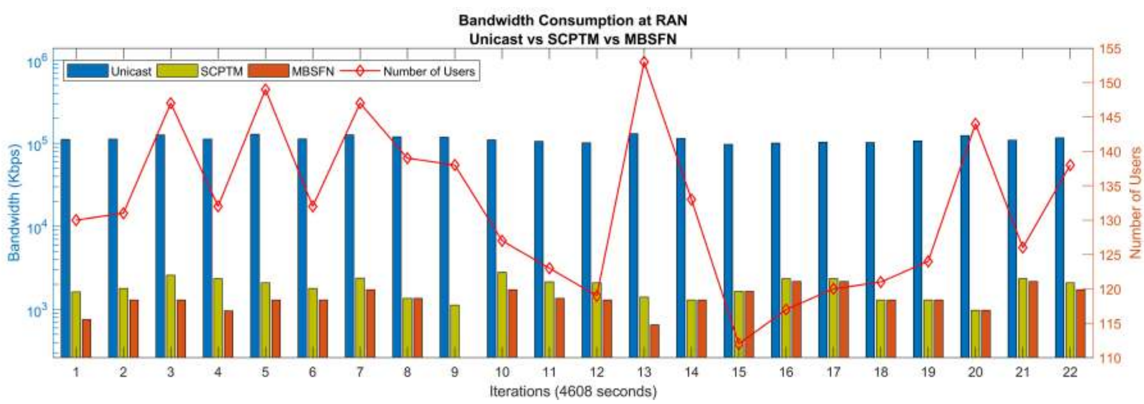

FIGURE 13. Bandwidth Consumption at backhaul link (Bangkok).

then the QoE of all the users is decreased except the user with the lowest CQI. In multicast mode, the video representation to be transmitted is selected based on the lowest CQI. If we wish to send better representation to a user with a lower CQI then a greater number of RBs will be used to do so. This results in higher network cost as well as the compromise in RBs assigned to other users that affect the QoE of all other users.

After the experiments, it is seen that when there is only one user per CQI level, the collective QoE of all the users in unicast mode is 0.55 whereas in multicast mode collective QoE is 0.17. This results in huge QoE degradation in the case of multicast mode when there are fewer users in the cell with different CQI levels. Conclusively, it is evident that there is a minimum number of users in a cell that decides when to start the multicast session to achieve a better QoE

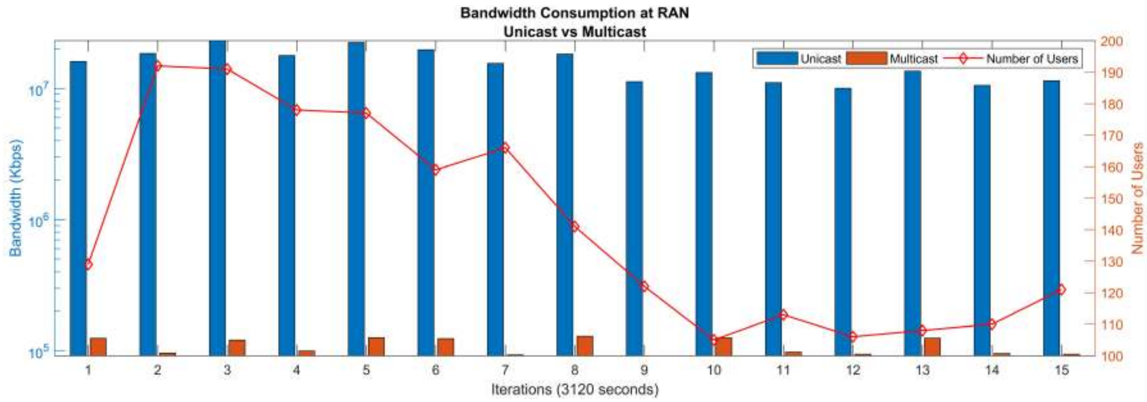

FIGURE 14. Bandwidth Consumption at backhaul link (Sao Paulo).

score than the unicast mode as well keeping the network cost low. We performed numerous experiments to calculate the collective and individual QoE of users by varying the number of users in each experiment. We deduced that increasing the number of users with a higher CQI did not affect/enhance the QoE. Since the video representation to be transmitted is based on the user with the lowest CQI level in the cell at the time. Therefore, to achieve better QoE with a given number of RBs, there should be a minimum number of users with the lowest CQI, say n, which indicates the threshold to switch to the multicast mode. In our experiments, we concluded that for 29361 RBs the minimum number of users with lowest CQI must be greater than or equal to 3 to enhance the collective QoE of all the users, as shown in figure 13. Figure 15 shows by increasing the number of users with the lowest CQI level gradually decreases the QoE in unicast mode, as the number of RBs is limited. Whereas, the QoE is considerably increased in multicast mode when increasing the number of users. While the ongoing academic and industrial network related experiments pave the path for faster and more efficient network infrastructure. The number of users per cell is expected to reduce further when mm-wave cellular cells (with very short coverage range per cell) are deployed in the future. Moreover, the emerging technologies, such as crowdsourced-based Internet of Vehicle (IoV) transmission [41], edge router multicasting [42], and denser cellular networks, such as femtocells, mm-wave, etc will aid the proposed mechanism to perform even better. This is due to the fact that in future generation cellular networks, there will be low number of users within a smaller cellular cell. Resulting in a lower diversity in CQI values and greater signal strength due to the smaller distance between the eNB and the user equipment. Therefore, it is most likely that most of users may be able to receive a higher bitrate version of the streamed video because of being in close proximity of the eNB/Cellular Access Point. Also, we aim to extend this work in developing an approach of collaborating multiple neighboring cellular cells using an MBSFN-SC-PTM hybrid approach to achieve a higher QoE considering the mm-wave and femtocells cellular networks. Furthermore, we performed all these experiments on the data fetched from the real-world dataset. In the future, we aim to develop an optimization model to enhance QoE in multicast mode when the number of available RBs, total number of users, and number of users in each CQI level are given.

To compare the cost of fetching the total data from the CDN in the case of both the unicast and the multicast, we calculated total data transmitted by the CDN to the relevant eNBS from the aforementioned real-world traces. We considered the total transmission time of all the videos for 4608 seconds streamed through all the participating eNBs. In the case of the unicast, total data sent by the CDN is approximately 296 GBs and 215 GBs for Bangkok and Sao Paulo cities respectively. We calculated the cost in USD/GB as per the Amazon CloudFront Policy.[4] We have taken costs for Singapore, South Korea, Taiwan, Hong Kong, & Philippines region for Bangkok city and cost for the South America region for Sao Paulo city. Figure 16 shows the cost comparison of unicast and multicast for two given regions (Bangkok and Sao Paulo). The result shows a considerable cost saving when the multicast transmission is used instead of unicast, when possible. The cost to fetch the data from the CDN in the case of unicast and multicast is \$41 and \$0.887 in the Bangkok region and \$23 and \$0.167 in Sao Paulo region respectively.

D. STORAGE OVERHEAD

The storage overhead of a legacy tree-based multicast scheme is $O(G \times T)$, where G is the size of the multicast group and the T is the size of the tree. The storage overhead of the Multicast at Edge is $O(N_{cells})$ where N_{cells} is the number of total cells participating in the multicast sessions. Although there is the storage overhead in Multicast at Edge, this overhead is simpler than the storage overhead of the legacy multicast scheme. The $O(G \times T)$ is the product of all the multicast trees and the users in it. Whereas, the $O(N_{cells})$ is the sum of all the users in the multicast group making it simpler than the other.

E. COMPUTATION OVERHEADS

Multicast at the edge, achieves the data, bandwidth, and resource blocks saving at transit, backhaul, and RAN links

[4]https://aws.amazon.com/CloudFront/pricing/

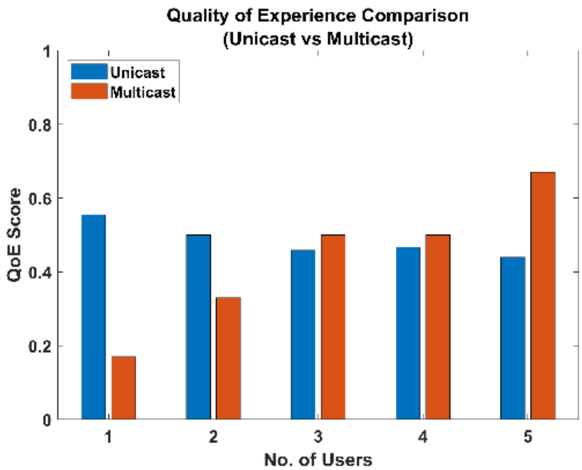

FIGURE 15. QoE Comparison between Unicast and Multicast.

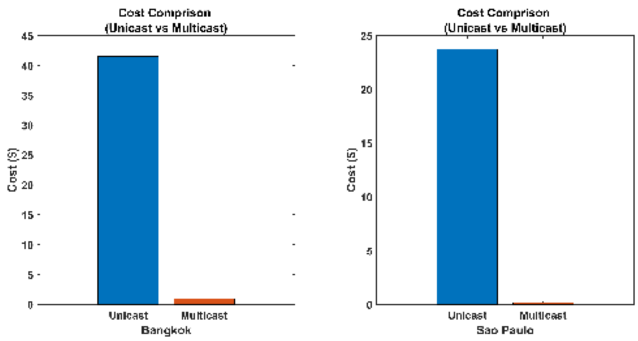

FIGURE 16. Cost Comparison between Unicast and Multicast.

respectively at the expense of the pre stall time required to switch the unicast transmission to the multicast mode. This pre stall time includes the time required to search the potential wireless multicast scenario, the time required to update the four-tables-mapping, and the time required to set up the multicast resource bearer. As the Multicast at Edge is considered to work continuously within the core network of the LTE environment. Therefore, we estimate that there is incremental computation overhead when a new request for a video is received or when the status of an ongoing live video is changed. The storage structure of the Multicast at Edge, i.e., four-tables-mapping is really simple that can easily be implemented like a hashmap dictionary. Therefore, in general, the time complexity for looking up the specific video in the tables is $O(log(m))$, where the m is the number of videos and the time complexity for the cell search is $O(log(n))$), where n is the number of cells with multicast status. The average time complexity to update a multicast group is $O(L_{group})$, where the L_{group} is the length of the group and time complexity of updating the tables is $O(1)$.

F. MRB ESTABLISHMENT OVERHEAD

For the MRB establishment, the main algorithm of the Multicast at Edge sends multiple SIB messages to user equipment and collaborates with the BMSC and content provider for information exchange. In general, a single SIB type 1-2 message needs 5.71 milliseconds to be transmitted and decoded by a CAT-3 user equipment. Multicast at edge sends at least 4 SIB messages to the user equipment, already discussed in the design section. As the Multicast at Edge is considered to be a core network entity and act as a part of the modified BMSC. Therefore, the time required for the communication between the Controller and the BMSC is negligible. This MRB establishment overhead occurs only once for a live video when a second user requests for the same live video. Once the MRB is established, the latter users will not have to wait for the MRB establishment. For the users, after the second user, the pre stall time will be the same as the playout buffer filling time.

V. CONCLUSION

In this paper, we propose EdgeCast, an adaptive multicast solution based on NFV to reduce the data transmitted by CDN (transit link load), bandwidth usage (backhaul link load), resource blocks usage (RAN link load) of live video streaming in the wireless cellular network. EdgeCast adaptively sets up multicast groups according to potential wireless multicast scenarios. The EdgeCast is deployed within the MNO core to monitor all the live video streams and produce multicast streams. Then a measurement study based on the real data collected from the Facebook LiveMap platform is conducted to demonstrate the feasibility and the potential of our work. Towards the end, we estimate the performance improvement of deploying EdgeCast in the 5G cellular network based on the real data of a live stream in Bangkok, Thailand. The results show that EdgeCast could save a huge amount of network resources on transit, backhaul, and RAN links by exploiting the multicasting technique.

ACKNOWLEDGMENT

Open Access funding provided by the Qatar National Library. The statements made herein are solely the responsibility of the author(s).

REFERENCES

[1] I. SANDVINE. (2016). *Global Internet Phenomena Report. 2016*. [Online]. Available: https://www.sandvine.com/hubfs/downloads/archive/2016-global-internet-phenomena-report-latin-america-and-north-america.pdf

[2] *Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper*, document 958959758, V. N. I. Cisco, 2016.

[3] E. Baccour, A. Erbad, K. Bilal, A. Mohamed, M. Guizani, and M. Hamdi, "FacebookVideoLive18: A live video streaming dataset for streams metadata and online viewers locations," in *Proc. IEEE Int. Conf. Informat., IoT, Enabling Technol. (ICIoT)*, Feb. 2020, pp. 476–483.

[4] Expway. (2016). *LTE Broadcast—Lessons Learned From Trials and Early Deployments*. [Online]. Available: https://www.expway.com/wp-content/uploads/2016/12/LTE-Broadcast-Paper-Final-30-November-2016.pdf

[5] K. Bilal, E. Baccour, A. Erbad, A. Mohamed, and M. Guizani, "Collaborative joint caching and transcoding in mobile edge networks," *J. Netw. Comput. Appl.*, vol. 136, pp. 86–99, Jun. 2019.

[6] Ericsson. (Aug. 2017). *Gigabit Network Launched in China*. [Online]. Available: https://www.ericsson.com/en/press-releases/2017/8/gigabit-network-launchedin-china

[7] Verizon, Verizon News Archives. *LTE Multicast: The Future of Mobile Content Delivery*. Accessed: Mar. 1, 2021. [Online]. Available: https://www.verizon.com/about/news/vzw/2014/01/lte-multicast-verizon-power-house>

[8] *LTE Broadcast (EMBMS) Market Update, 2016-2021 White Paper*, GSA, Washington, DC, USA, Nov. 2017.

[9] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. New York, NY, USA: Academic, 2013.
[10] M. Gruber and D. Zeller, ''Multimedia broadcast multicast service: New transmission schemes and related challenges,'' *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 176–181, Dec. 2011.
[11] Y.-F. Liu and W. Yu, ''Wireless multicast for cloud radio access network with heterogeneous backhaul,'' in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, Oct. 2017, pp. 531–535.
[12] J. Yoon, H. Zhang, S. Banerjee, and S. Rangarajan, ''MuVi: A multicast video delivery scheme for 4G cellular networks,'' in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw. (Mobicom)*, Aug. 2012, pp. 209–220.
[13] T. Yamada and T. Nishimura, ''Possibility of a new type of wireless network: Connected small cells with 3G-LTE signaling primitives and PON for the backhaul,'' in *Proc. 16th Int. Telecommun. Netw. Strategy Planning Symp. (Networks)*, Sep. 2014, pp. 1–6.
[14] M. Laterman, M. Arlitt, and C. Williamson, ''A campus-level view of Netflix and Twitch: Characterization and performance implications,'' in *Proc. Int. Symp. Perform. Eval. Comput. Telecommun. Syst. (SPECTS)*, Jul. 2017, pp. 1–8.
[15] Y. Qin, Q. Xia, Z. Xu, P. Zhou, A. Galis, O. F. Rana, J. Ren, and G. Wu, ''Enabling multicast slices in edge networks,'' *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8485–8501, Sep. 2020.
[16] T. Jiang, W. Xiang, H.-H. Chen, and Q. Ni, ''Multicast broadcast services support in OFDMA-based WiMAX systems [advances in mobile multimedia],'' *IEEE Commun. Mag.*, vol. 45, no. 8, pp. 78–86, Aug. 2007.
[17] Y. Cao, T. Jiang, X. Chen, and J. Zhang, ''Social-aware video multicast based on device-to-device communications,'' *IEEE Trans. Mobile Comput.*, vol. 15, no. 6, pp. 1528–1539, Jun. 2016.
[18] R. Chandra, S. Karanth, T. Moscibroda, V. Navda, J. Padhye, R. Ramjee, and L. Ravindranath, ''DirCast: A practical and efficient Wi-Fi multicast system,'' in *Proc. 17th IEEE Int. Conf. Netw. Protocols*, Oct. 2009, pp. 161–170.
[19] X. Zhang, R. Yu, Y. Zhang, Y. Gao, M. Im, L. Cuthbert, and W. Wang, ''Energy-efficient multimedia transmissions through base station cooperation over heterogeneous cellular networks exploiting user behavior,'' *IEEE Wireless Commun.*, vol. 21, no. 4, pp. 54–61, Aug. 2014.
[20] Y. Zhou, H. Liu, Z. Pan, L. Tian, J. Shi, and G. Yang, ''Two-stage cooperative multicast transmission with optimized power consumption and guaranteed coverage,'' *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 274–284, Feb. 2014.
[21] J. Cano and L. Almeida, ''Feedback management for scaling clients in streaming multicast,'' in *Proc. 30th Annu. ACM Symp. Appl. Comput.*, Apr. 2015, pp. 669–671.
[22] S. Pizzi, M. Condoluci, G. Araniti, A. Molinaro, A. Iera, and G.-M. Muntean, ''A unified approach for efficient delivery of unicast and multicast wireless video services,'' *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8063–8076, Dec. 2016.
[23] K. Poularakis, G. Iosifidis, V. Sourlas, and L. Tassiulas, ''Exploiting caching and multicast for 5G wireless networks,'' *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2995–3007, Apr. 2016.
[24] M. Tao, E. Chen, H. Zhou, and W. Yu, ''Content-centric sparse multicast beamforming for cache-enabled cloud RAN,'' *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118–6131, Sep. 2016.
[25] B. Zhou, Y. Cui, and M. Tao, ''Stochastic content-centric multicast scheduling for cache-enabled heterogeneous cellular networks,'' *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6284–6297, Sep. 2016.
[26] Y. Zhou, F. R. Yu, J. Chen, and Y. Kuo, ''Video transcoding, caching, and multicast for heterogeneous networks over wireless network virtualization,'' *IEEE Commun. Lett.*, vol. 22, no. 1, pp. 141–144, Jan. 2018.
[27] S. Almowuena, M. M. Rahman, C.-H. Hsu, A. A. Hassan, and M. Hefeeda, ''Energy-aware and bandwidth-efficient hybrid video streaming over mobile networks,'' *IEEE Trans. Multimedia*, vol. 18, no. 1, pp. 102–115, Jan. 2016.
[28] J. Chen, M. Chiang, J. Erman, G. Li, K. K. Ramakrishnan, and R. K. Sinha, ''Fair and optimal resource allocation for LTE multicast (eMBMS): Group partitioning and dynamics,'' in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 1266–1274.
[29] K. Zahoor, K. Bilal, A. Erbad, and A. Mohamed, ''Service-less video multicast in 5G: Enablers and challenges,'' *IEEE Netw.*, vol. 34, no. 3, pp. 270–276, May 2020.
[30] T. Qin, H. Li, Q. Wu, N. Liu, and F. Li, ''Wi-live, adaptive scenario-aware wireless multicast solution for living streaming based on software defined networks,'' in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 48–53.
[31] A. de la Fuente, A. G. Armada, and R. P. Leal, ''Joint multicast/unicast scheduling with dynamic optimization for lte multicast service,'' in *Proc. 20th Eur. Wireless Conf.*, May 2014, pp. 1–6.
[32] A. de la Fuente Iglesias, R. P. Leal, and A. G. Armada, ''Performance analysis of eMBMS in LTE: Dynamic MBSFN areas,'' in *Proc. OPNETWORK*, Aug. 2013, pp. 1–6. [Online]. Available: https://e-archivo.uc3m.es/handle/10016/21212?show=full
[33] B. Damera and P. C. Babu, ''Broadcast the user chosen content over LTE in a simplified approach using evolved multimedia broadcast multicast service (eMBMS) with optimized MCE scheduling algorithm,'' in *Proc. Int. Conf. Intell. Comput. Appl.*, Mar. 2014, pp. 90–94.
[34] T. A. Le, H. Nguyen, and N. Crespi, ''IMS-based distributed multimedia conferencing service for LTE,'' in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2012, pp. 3210–3214.
[35] M. Liyanage, A. Gurtov, and M. Ylianttila, *Software Defined Mobile Networks (SDMN): Beyond LTE Network Architecture*. Hoboken, NJ, USA: Wiley, 2015.
[36] J. Zhang, Z. Wang, N. Ma, T. Huang, and Y. Liu, ''Enabling efficient service function chaining by integrating nfv and sdn: Architecture, challenges and opportunities,'' *IEEE Netw.*, vol. 32, no. 99, pp. 1–8, Nov./Dec. 2018.
[37] K. Bilal and A. Erbad, ''Impact of multiple video representations in live streaming: A cost, bandwidth, and QoE analysis,'' in *Proc. IEEE Int. Conf. Cloud Eng. (ICE)*, Apr. 2017, pp. 88–94.
[38] L. Toni, R. Aparicio-Pardo, G. Simon, A. Blanc, and P. Frossard, ''Optimal set of video representations in adaptive streaming,'' in *Proc. 5th ACM Multimedia Syst. Conf. (MMSys)*, Mar. 2014, pp. 271–282.
[39] Q. He, J. Liu, C. Wang, and B. Li, ''Coping with heterogeneous video contributors and viewers in crowdsourced live streaming: A cloud-based approach,'' *IEEE Trans. Multimedia*, vol. 18, no. 5, pp. 916–928, May 2016.
[40] K. Bilal, A. Erbad, and M. Hefeeda, ''QoE-aware distributed cloud-based live streaming of multisourced multiview videos,'' *J. Netw. Comput. Appl.*, vol. 120, pp. 130–144, Oct. 2018.
[41] B. Masini, A. Bazzi, and A. Zanella, ''Vehicular visible light networks for urban mobile crowd sensing,'' *Sensors*, vol. 18, no. 4, p. 1177, Apr. 2018.
[42] B. Yang and P. Mohapatra, ''Edge router multicasting with MPLS traffic engineering,'' in *Proc. 10th IEEE Int. Conf. Netw. (ICON)*, Aug. 2002, pp. 43–48.

KAMRAN ZAHOOR received the M.S. degree in computer science from COMSATS University Islamabad, Abbottabad Campus, Pakistan. He is currently a Visiting Faculty with COMSATS University Islamabad, Abbottabad Campus. His research interests include crowd-sourced multimedia, multimedia broadcast multicast services (eMBMS), and next-generation cellular networks.

KASHIF BILAL received the Ph.D. degree from North Dakota State University, USA. He is currently a Postdoctoral Researcher with Qatar University, Qatar. He is also an Assistant Professor with the COMSATS Institute of Information Technology, Pakistan. His research interests include cloud computing, energy-efficient high-speed networks, and crowd-sourced multimedia. He was named as a CoE Student Researcher of the Year, in 2014, based on his research contributions during his Ph.D. studies at North Dakota State University.

AIMAN ERBAD (Senior Member, IEEE) received the B.Sc. degree in computer engineering from the University of Washington, in 2004, the Master of Computer Science degree in embedded systems and robotics from the University of Essex, U.K., in 2005, and the Ph.D. degree in computer science from The University of British Columbia, Canada, in 2012. He is currently an Associate Professor with the College of Science and Engineering, Hamad Bin Khalifa University (HBKU). His research interests include cloud computing, edge computing, the IoT, private and secure networks, and multimedia systems. He received the Platinum Award from H. H. The Emir Sheikh Tamim Bin Hamad Al Thani at the Education Excellence Day 2013 (Ph.D. category). He received the 2020 Best Research Paper Award from Computer Communications, the IWCMC 2019 Best Paper Award, and the IEEE CCWC 2017 Best Paper Award. He has served as a Guest Editor for IEEE NETWORK. He serves as an Editor for the *KSII Transactions on Internet and Information Systems* and the *International Journal of Sensor Networks (IJSNet)*.

AMR MOHAMED (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical and computer engineering from The University of British Columbia, Vancouver, Canada, in 2001 and 2006, respectively. He has worked as the Advisory IT Specialist of the IBM Innovation Centre, Vancouver, from 1998 to 2007, taking a leadership role in systems development for vertical industries. He is currently a Professor with the College of Engineering, Qatar University. He has more than 25 years of experience in wireless networking research and industrial systems development. He has authored or coauthored more than 200 refereed journal and conference papers, textbooks, and book chapters in reputable international journals, and conferences. His research interests include wireless networking, and edge computing for the IoT applications. He holds three awards from IBM Canada for his achievements and leadership and four Best Paper Awards from IEEE conferences. He has served as the Technical Program Committee (TPC) Co-Chair for many IEEE conferences and workshops. He has been serving as the Technical Editor for two international journals.

MOHSEN GUIZANI (Fellow, IEEE) received the B.S. (Hons.) and M.S. degrees in electrical engineering and the M.S. and Ph.D. degrees in computer engineering from Syracuse University, Syracuse, NY, USA, in 1984, 1986, 1987, and 1990, respectively. He is currently a Professor with the Computer Science and Engineering Department, Qatar University, Qatar. Previously, he has served in different academic and administrative positions at the University of Idaho, Western Michigan University, the University of West Florida, the University of Missouri–Kansas City, the University of Colorado at Boulder, and Syracuse University. He is the author of nine books and more than 600 publications in refereed journals and conferences. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He is a Senior Member of ACM. He has served as a member, the Chair, and the General Chair for a number of international conferences. Throughout his career, he received three Teaching Awards and four Research Awards. He was a recipient of the 2017 IEEE Communications Society Wireless Technical Committee (WTC) Recognition Award, the 2018 AdHoc Technical Committee Recognition Award for his contribution to outstanding research in wireless communications and Ad-Hoc Sensor Networks, and the 2019 IEEE Communications and Information Security Technical Recognition (CISTC) Award for outstanding contributions to the technological advancement of security. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the Chair of the TAOS Technical Committee. He is currently the Editor-in-Chief of *IEEE Network Magazine*, serves on the editorial boards for several international technical journals, and the Founder and the Editor-in-Chief for *Wireless Communications and Mobile Computing* journal (Wiley). He guest edited a number of special issues in IEEE journals and magazines. He has served as the IEEE Computer Society Distinguished Speaker and is currently the IEEE ComSoc Distinguished Lecturer.