IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 3, 2018, accepted August 6, 2018, date of publication August 10, 2018, date of current version September 7, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2864790

CASH: Content- and Network-Context-Aware Streaming Over 5G HetNets

MD. JALIL PIRAN[1], (Senior Member, IEEE), S. M. RIAZUL ISLAM[1], (Member, IEEE), AND DOUG YOUNG SUH[2], (Member, IEEE)

[1]Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea
[2]Department of Electronics and Radio Engineering, Kyung Hee University, Seoul 17104, South Korea

Corresponding author: Doug Young Suh (suh@khu.ac.kr)

This work was supported by the Korea Electric Power Corporation under Grant R18XA02.

ABSTRACT Heterogeneity is one of the key features that characterizes the future generation of cellular networks, 5G and beyond. However, streaming high-quality bandwidth-hungry multimedia contents over bandwidth-constrained 5G heterogeneous networks (5G HetNets) involves various significant challenges, including long video start time, video start failures, frequent buffering and stalling, and low quality of experience (QoE). Traditional multimedia streaming technologies, however, do not pay attention to either the available network bandwidth or the interaction between content characteristics and resources. To reduce network strain and improve QoE, we propose "Context-Aware Streaming over 5G HetNets (*CASH*)" that allows us to achieve a tradeoff between content-context and network-context. The proposed CASH fundamentally works in a multi-step process. First, the CASH comes with an integrated architecture that includes a media server, a flow scheduler, and a single radio controller (SRC). The SRC and the user equipment (UE) of interest cooperatively prepare a metadata file that contains the network-context. Second, based on the metadata file, which can be accessed from the SRC in the media preparation server, we analyze and cluster the contents based on the content-context, e.g., the actual bitrate of each scene. The metadata file is then updated by adding the content-context information. Third, the flow scheduler basically controls the flow of the clusters of the contents in the server-push mode and conveys that to the appropriate radio access technology (RAT) conforming to the bitrate of the clusters and bandwidth delivered by RATs. Finally, the UE will aggregate the received packets and will play-back the content. We analytically show the validity of CASH. Also, extensive simulations are performed to demonstrate that CASH offers substantial performance improvements compared with exiting works in terms of peak data rate, latency, users' experiences, and spectral efficiency.

INDEX TERMS Multimedia streaming, context-awareness, resource allocation, carrier aggregation, 5G HetNets, QoE.

I. INTRODUCTION

The phenomenal growth, more than twice each year, of mobile traffic introduced the issue of radio spectrum scarcity as one of the most crucial challenges that the fifth generation of mobile networks (5G) will struggle with. CISCO forecasts a mobile data traffic of 49 Exabytes per month by 2021 [1] and universal mobile telecommunications service (UMTS) predicts that the amount will be 351 Exabytes in 2025 [2]. Providing seamless and flawless mobile video services is one the most critical requirements of 5G, whereas over three-quarters, 78%, of the world's mobile data traffic will be video uploading/downloading by 2021 [1]. Therefore, the influx of such a tremendous flood of multimedia traffic would significantly slow down cellular networks. Moreover, there are many other requirements for the next generation such as increased capacity, improved data rate, decreased latency, reduced cost, and consistent QoE provision [3] that need to be taken into consideration.

It is obvious that the traditional macro-cellular networks are incapable of meeting a plethora of surging traffic demands. Therefore, it entails highly distributed networks to satisfy the anticipated capacity constraint. It was emphasized by Cooper [4], the father of the cellular phone, in his position paper that: *"Our history, along with an understanding of the* *potential of known technologies, demonstrates that spectrum is an asset that cannot be separated from the technology assets that enable it; that these technology assets are not finite; and that, in our robust society, they always scale to demand. That is the genius of our society; our policies should exploit that.”*

Mobile network operators have tried various solutions to tackle the issues that arise in the sophisticated radio environment, such as increasing capacity with new radio spectrum, employing multi-antenna techniques, using new kinds of modulation and coding schemes with more efficiency, etc. [5]. However, applying such techniques alone might not prove to be successful in order to solve the problem, especially in crowded places and cell edges with high performance degradation. In this context, whereas UEs are equipped with multi-RAT selection functionality that enables them to select the best network among the accessible networks [6], HetNets come with a number of special capabilities, such as high-speed connections support, flexible resource allocation, and integration of various RATs [7], [8] that pave the way for the breakneck growth of mobile traffic, under the principle of extreme densification and offloading.

HetNets, with densely deployed small cells interoperating with macro-cells, are expected to increase network coverage, capacity, and quality via employing a set of different infrastructures and technologies [9]. Therein, the cells arranged in order of decreasing and in terms of base station (BS) power are called as macro-, micro-, pico-, and femto-cells. Macro-cells ensure coverage while pico- and micro-cells are deployed for capacity enhancement and femto-cells are organized by the users at home or offices to experience better communications.

A key feature of future HetNets is the increased integration among different RATs. 5G UEs with multiple connectivity can be employed to connect to different RATs such as macro-cell (high speed packet access (HSPA) or enhanced voice-data only (EVDO), 5G standard (mmWave frequencies for instance), LTE-U, 4G LTE/LTE-A/LTE-A Pro, 3G), various kinds of WiFi (802.11g, n, ac), and device-to-device communication (D2D) [10]. This will result in increasing capacity of hot spots with so many users, covering the areas far from macro networks (outdoor in terms of distance and indoor in terms of signal strength), and to offload data from the large macro-cells, consequently increasing bitrates per unit area that leads to improved QoE. However, scheduling to communicate with a number of RATs may lead into some other challenges.

To cope with such issues, carrier aggregation (CA) was introduced in LTE Release (Rel.) 10 ratified by 3^{rd} generation partnership project (3GPP), with backward compatibility to Rel. 8, as the aggregation of multiple component carriers with the aim of increasing the total available bandwidth and hence maximizing bitrates. In CA, up to five component carriers can be aggregated, with the accumulated bandwidth of up to 100MHz [11]. Then, in order to improve the reliability and throughput, another technique called coordinated multi point (CoMP) was introduced in LTE Rel. 11 [12], with three goals: improving coverage, cell edge data rate, and system spectrum efficiency. By applying CoMP, a number of transmitter and receiver elements can be coordinated to serve a user. It means that the user can send and receive data from one transmitter in one sub-frame and from another in the next sub-frame both in downlink and uplink. Consequently, with the help of CoMP in HetNets, multiple macro- and small-cells may cooperate to communicate with a user for one transmission [13]. The entire process is transparent to the UE and there is no need for a special configuration on the UE side.

However, particularly for multimedia services in HetNets, flexible resource allocation techniques are needed to be developed considering the limitations on both sides. On the network side, to obtain optimal throughput, ensuring fairness, and handle dynamic requirements of different data traffics, various metrics are important to be taken into account including spectral, energy and cost efficiency. On the other side, for multimedia services, startup time, video start failures, buffering and stalling, quality of the delivered content, and latency are the factors that are of vital importance.

To overcome the above-stated technical difficulties, this article aims to propose a multimedia streaming technique that allocates the resources according to both multimedia content-context and network-context, coined as *CASH* that is content- and network-context-aware streaming over 5G HetNets. In this technique, we consider explicit requirements of multimedia content and allocate resources accordingly. *Context* is ‘‘any information that can be used to characterize the situation of an entity’’ [14]. Based on the given definition, in this research, content-context refers to the content characteristics, e.g., scene complexity represented by scene bitrates, resolution, quality, etc. On the other side, network-context refers to the network throughput, delay, packet error rate, etc. Our proposed scenario thoroughly considers the characteristics of both the content and the networks in order to maintain a trade-off between the available network bandwidth and bitrate of different clusters of the content, and hence to improve system performance, resource utilization, and QoE. It is worth noting that we introduced the idea of content context-aware resource allocation for the very first time in moving picture experts group (MPEG) 109^{th} standardization meeting [15]–[17].

The proposed *CASH* composed of an integrated architecture that consists of a SRC, a flow scheduler, and a media preparation server. In this system, the SRC acts as the intermediate node between UEs and core network (CN) [43]. The flow scheduler is placed between the media preparation server and CN. Moreover, we develop a QoE-driven strategy, which incorporates content real resource requirements of multiple multimedia applications into real-time network resource management on 5G HetNets.

To do that, we contribute the followings: First, with the help of UEs and SRC, we estimate the number and quality of available channels that different RATs provide to the

UE of interest. This information i.e. network-context, are added to a metadata file. Second, according to the information stored in the metadata file, we analyze and cluster the content in the media preparation server. All information regarding the content and cluster's characteristics i.e. content-context, are added to the metadata file. The metadata file including the network-context and content-context information file is named as *"context analysis file (CAF)"*. Third, we develop a flow scheduler that steers the content clusters to the corresponding RAT to be handed over to the UE. Through the CAF, the scheduler has the knowledge of both network-context (received from the media preparation server) and the content-context (received from the SRC).

We conducted a comprehensive simulation study to evaluate the effectiveness of the proposed framework compared with the "conventional video streaming" and "dynamic adaptive streaming over HTTP (DASH)" in terms of average number of handoff, average delay, average number of quality switching, average number of stalls, average peak signal-to-noise-ratio (PSNR), and average mean opinion score (MOS).

Our proposed *CASH* can be distinguished from the other related scenarios in the literature by its unique merits that can be stated in four aspects as follows:

- Resource utilization: *CASH* needs less bandwidth compared to conventional video streaming technologies since resources are allocated in an intelligent way based on the actual scenes demand.
- QoE improvement: *CASH* allocates higher bitrates to more complex video scenes, and thus improves QoE considerably.
- Compared to DASH, *CASH* needs much less storage whereas DASH makes a copy of the segments with different bitrate, and hence increases storage wastage. Moreover, whereas DASH is client-driven, *CASH* is server-push, which results in a reduced number of requests (message overhead) and reduced media latency.
- Compared to adaptive bitrate (ABR) streaming technologies, the proposed method reduces the bitrate switching, and thus decreases QoE degradation by quality changing frequently.

The rest of this paper is organized as follows. Related works and problem statement are described in Section II. Section III, presents the system and network model. The detailed description of the proposed *CASH* is provided in Section IV. Section V validates the derived results and analyzes the performance behaviors. Finally, Section VI, draws conclusions.

II. RELATED WORKS AND PROBLEM STATEMENT

In HetNets, in order to achieve optimal performance, ensuring fairness, as well as meeting dynamic traffic demands from various applications, flexible resource management is of vital importance. Different metrics need to be considered for resource management, such as quality of service (QoS) requirements, spectrum access, energy consumption, and cost [18], particularly for bandwidth-intensive multimedia services that results in QoE degradation without specially designed resource allocation. There are many metrics that can have a non-negligible impact on QoE, such as long startup time, video start failures, frequent buffering/rebuffering and stalling, play-back errors, low quality video, quality oscillations, and playback delay. Such kinds of annoying issues need to be considered in resource allocation schemes especially for multimedia services.

Typically, every multimedia content composed of different scenes with different complexity in terms of bitrate. For example, the bitrate of a low motion (e.g. news) is relatively very low compared to a high motion video (e.g. automobile racing). Since the bitrate of multimedia contents is based on spatial and temporal complexity of the scenes, conventional video compression with the constant quality results in variable bit rate (VBR) [19], [20]. However, transmission of multimedia contents as a continuous service requires a guaranteed bit rate (GBR) in order to achieve high QoE. The issue associated with this method is that the GBR is expedient for complex scenes with a high bitrate and for simple scenes with low bitrate, the allocated fixed bitrate may not be fully utilized.

The ABR streaming is the current technique for video streaming over the Internet [21]–[23]. In ABR, a video source files is encoded with multiple bitrates. Then every device based on its available bandwidth and resolutions requests for an appropriate bitrate that best fits for it. Although, ABR was proposed to be a prominent solution for buffering and stalling issues and provide seamless video services in a fluctuating network environment, it is dramatically involved with inconsistency in delivered video quality.

Then, in order to improve the ABR performance, DASH was proposed by MPEG, which provides suitable formats for streaming of segmented multimedia over HTTP [24]. In the client-driven DASH paradigm, based on the network fluctuations and CPU capacity, through adjusting buffering and stalling, DASH tries to provide seamless video playback. Clients may choose the video rate that is appropriate for their needs, based on the observed current throughput and feeds, measurement of content round trip time (RTT) between itself and the server, the screen size, and the resolution required through some interactions with the media origin server. DASH is adopted by many standardization and industry bodies, such as Microsoft Smooth Streaming [25], Apple HTTP Live Streaming (HLS) [26], Adobe HTTP Dynamic Streaming [27], Digital Video Broadcasting (DVB) [28], 3GPP [29], Hybrid Broadcast Broadband Television (HbbTV) [30], and many others.

However, DASH is involved with the issue of frequent video quality switching, which would lead to QoE degradation because of network fluctuations, especially when there is a big change in video quality (for instance from 4K to 360p). As a result, users may be unable to enjoy a smooth and constant video service experience. Some other frequent faults caused by DASH are: lack of control by service providers on client behavior, resource expiration announced by the client because of network failure or reconfiguration, ensuring quality switching due to mobile handoff (HO), message overhead due to excess of *HTTP GET* messages that clients need to send whenever any network changes occur, and many others [31].

As a result, yet the efforts in the multimedia streaming area are not coherent with the users' expectations. The technical issues that the current multimedia streaming approaches are struggling such as long start-up time, frequent buffering/rebuffering stalls, play-back errors, and low quality, negatively affect the final QoE of the end-users. Low QoE results in users' dissatisfaction where a survey by Conviva proved that 75% of video viewers will stop watching a video within four minutes if the experience is poor [32]. It is all because of the fact that the trade-off between the explicit requirements of the multimedia contents and the available network resources is not taken into consideration while designing resource allocation approaches.

III. SYSTEM AND NETWORK MODEL

As an instance of 5G HetNets, we considered a time-slotted multi-RAT HetNet setting including M small cells within the coverage area of one macro-cell e.g. gNB (the next generation NodeB), as well as one WiFi access points (APs), as shown in Fig. 1, for the purpose of research. Generally, a 5G HetNets is in the form of an integrated framework accounting different RATs including:

- gNB (next Generation NodeB), powerful BSs that cover large areas, i.e., 5G macro-BS.
- BTS (2G), NB (3G), and eNB (LTE/LTE-Advance/LTE-A Pro).
- HgNB (Home gNB), used for indoor applications, e.g., pico-cells in aircrafts.
- RN (Relay Node), used to extend the coverage of macro-cells at the cell edges and hot-spot areas. It is connected to Donor gNB without a separate backhaul. From the UE viewpoint, the RN will act as a BS, however, Donor gNB sees the relay nodes as a UE.
- DgNB (Donor gNB), as the modified version of evolved universal terrestrial access network (E-UTRAN), shares its radio resources for RNs as well as serves its own UEs.
- RRH (Remote Radio Head), operates as soft relay between UE and BSs through the wireline or wireless front-haul links.
- Satellite Technologies, for wide area coverage in Internet of Things (IoT) applications, e.g., vehicle tracking, or even to offload broadcast and multicast linear TV traffic from 5G [33].

For the purpose of brevity, different BSs are not distinguished but are collectively referred to as the BS when they are called in general throughout this paper.

The cell size is defined according to BSs' power, antenna position (indoor/outdoor), and the area environment [34]. Macro cells are assumed as coverage providers. Pico- and micro-cells are supposed to increase the network capacity in a crowded environment such as shopping malls, subway stations, etc. Femto-cells and WiFi may be used in a smaller network size such as home and offices. Femto-cells differ from pico-cells due to the deployment intention that are more autonomous. These kind of BSs are self-installed by the customers, but pico-cells are implemented by the operators. The radio parameters in femto-cells are locally determined, but in pico-cells it is planned centrally.

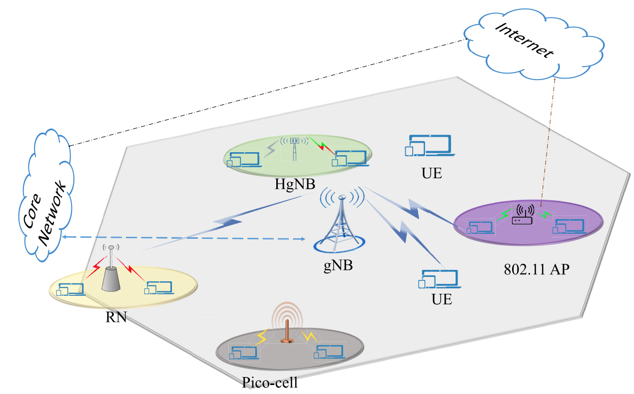

FIGURE 1. A three-tier heterogeneous network consisting of one gNB and multiple RATs.

We assumed that N UEs may associate with a subset of M BSs, operating on non-overlapping frequency bands. Each UE may access a subset of BSs since they have a specific number of RATs. The BSs are indexed by $m \in M$ and the UEs are indexed by $n \in N$. UEs are any device that has wireless communication capability, transceiver. It is assumed that all UEs have the traffic aggregation option.

We follow one of the interesting scenarios that has been considered in LTE Rel. 12, which enables the UEs equipped with multiple radio transceivers to receive control channels and/or data channels from multiple transmitters simultaneously. It is covered by two technologies i.e. the control-plane/user-plane split [35] and dual connectivity (DC) architecture [36]. As illustrated in Fig. 1, the UE located at the edge of the office femto-cell has the capability to receive data from both the femto-cell and macro-cell simultaneously. The split bearer, as a particular form of DC, divides the traffic among the BSs. The macro-BS contains the packet data convergence protocol (PDCP) layer while the small-BS has its own radio link control (RLC), media access control (MAC), and physical (PHY) layers. The packets are sent to the UEs via both macro-BS directly or via the small BS according to the signal strength on the UE side. Macro and small BSs are connected through a non-ideal backhaul link [24].

Another main component of our framework is the media origin server, which is employed to prepare multimedia contents. The server has got several functionalities including transcoding, encoding, clustering, encrypting, etc. It prepares and delivers multimedia contents with great efficiency and superior QoS. The content is encoded using scalable video coding (SVC) [37], which will be composed of a sin-

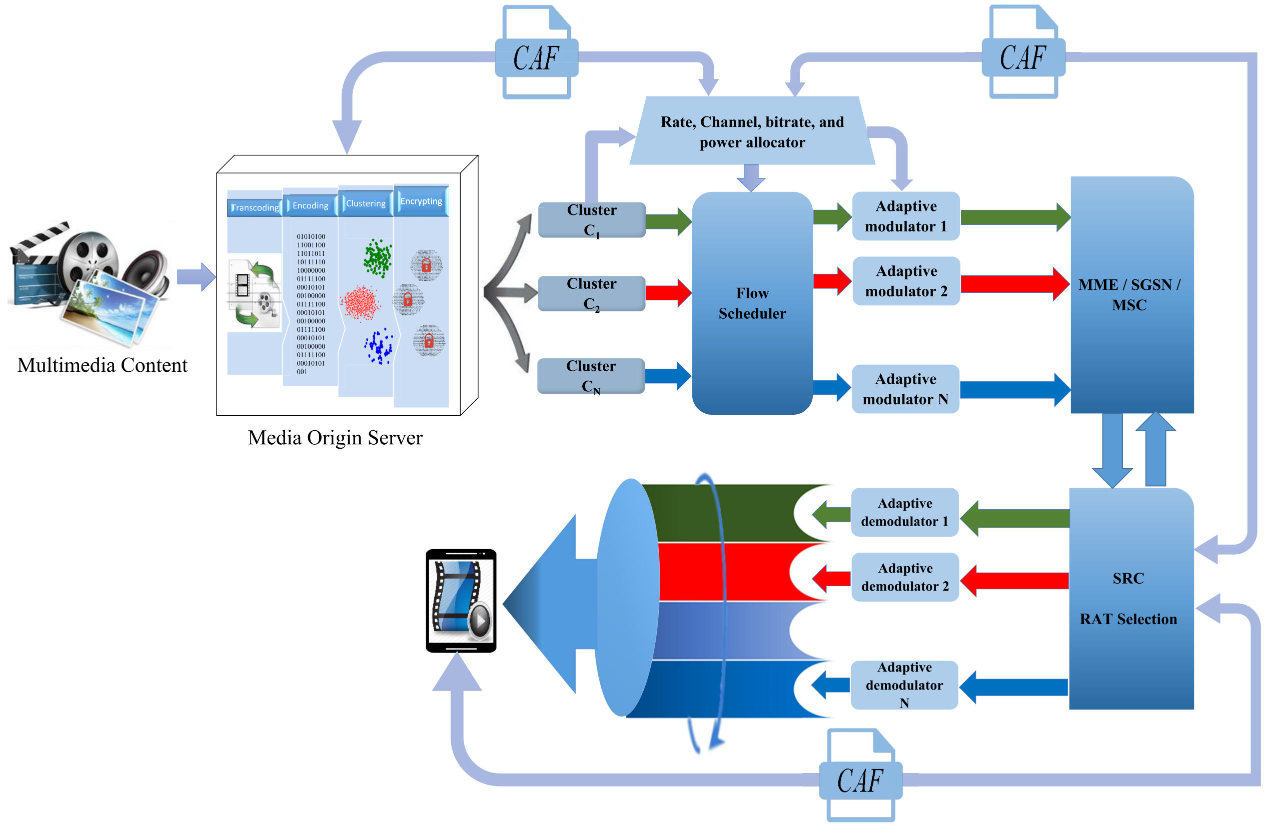

FIGURE 2. **The workflow of the proposed *CASH*.**

gle base layer (BL) and one or several enhancement layers (EL). The BL is compulsory to decode the content on the receiver side [38]. However, the ELs are used to improve the quality of the service. The number of ELs will be chosen according to the quality of the available channels that is indicated in CAF, which is received from the SRC. The SRC as an intermediate network entity is located between UEs and CN.

A flow scheduler is assumed to be another intermediate entity between the media origin server and the SRC. It has got unidirectional or bidirectional and point-to-point communication between the servers and the SRC. It expedites mechanisms for providing services that are content-context-awareness and network-context-awareness. The flow scheduler routes the appropriate cluster of the content to the corresponding RAT. It is done according to the information stored in CAF that is received from the SRC. The role of flow scheduler is similar to server and network assisted DASH (SAND) in DASH, which has been introduced by MPEG in ISO/IEC 23009-5:2017 [31], to address resource allocation, congestion, and QoE signaling. The SRC may not be an independent entity that increases the cost of a network, rather it can be associated with the entities like BSs and mobile switching center (MSCs). In a similar way, the flow scheduler can be implemented on other network entities (i.e. the media origin server).

As stated above, for any communication between UEs and media servers, a CAF is exchanged between the UE and the server via the SRC and flow scheduler. As a manifest file, CAF is composed of both a network-context and content-context; network-context including: average throughput, buffer level, initial playout delay, request/response transactions, shared resources, packet error rate, and content-context including: program information (i.e. title, type), one adaptation set per video stream (i.e. content composition node and data on finding the clusters), format, resolution, quality, location (i.e. URL), and bitrate.

The next section expounds performance analysis of the proposed streaming system.

IV. CASH: CONTEXT-AWARE STREAMING OVER 5G HETNETS

In section II, we discussed the issues associated with the current streaming technologies. In order to overcome those issues, in this section, we explain our solution, *CASH*, which is a video streaming framework that allocates the resources in an intelligent manner based on the actual requirements of the contents and network characteristics.

Fig. 2 illustrates the workflow of the proposed context-aware multimedia streaming over 5G HetNets, *CASH*. The UEs and SRC cooperatively evaluate the accessible channels provided by different RATs and prepare a CAF regarding the number and quality of the channels that are offered by different RATs to the UE. The flow scheduler and SRC exchange CAF to be aware of the network's conditions and the content's characteristics. This information exchange takes place at the PDCP layer in the macro BS [14]. Upon a request by the UE for a specific multimedia content, the corresponding CAF of the UE will be forwarded to the media preparation server where the content is stored. The media origin server is basically made up of multiple functionalities including transcoding, encoding, clustering and encrypting.

Using SVC, the whole content is encoded in one BL and several ELs. The number of ELs are determined according to the information received from the SRC. The better quality channels in the more number of RATs, the more number of ELs.

According to the number and quality of accessible channel provided by different RATs to the UE, the server clusters the content. Then, the server passes the clustered content to the flow scheduler. The adaptive modulator based on the number of bits allocated to each cluster chooses the modulation scheme and also, accordingly, the level of transmission power is adjusted. Then, the SRC will steer the cluster to the appropriate RAT. The received packets from different RATs are stored in the corresponding receiving buffer. Every group of pictures (GoP) period of a specific cluster is decoded according to the source block number information that is stored in the corresponding payload ID. In the case of late arrival packets with previous decoded GoP of different clusters, they will be discarded. According to the information in the network abstraction layer unit (NALU) header [39], the packets are queued in the decoder buffer. If the BL of the corresponding ELs of a specific cluster is not received on time, all the ELs will be dropped. At the end, the player would start playing back the safe and in-order received packets.

As stated earlier, our aim is to utilize the network resources in the most efficient manner and to provide the best feasible service and improving QoE. In doing so, the proposed *CASH* consists of two phases:

1) Phase I: Flexible RAT selection based on network-context
2) Phase II: Content preparation based on content-context

In the next subsections we describe each phase of *CASH* and related functionalities in details.

A. PHASE I: FLEXIBLE RAT SELECTION BASED ON NETWORK-CONTEXT

In this subsection, we explain and formulate different functionalities of the first phase of *CASH* in order to discover and select the best quality channels that are offered to UEs by the accessible RATs.

With the rapid growth and evolution of the mobile broadband, at the time of 5G appearance, the network will be much denser and complex where, each site will be multi-layer/multi-band/multi-mode. Different modes are stated as general packet ratio service (GSM)/UMTS/LTE-A Pro/5G/WiFi, while different layers are low and high frequency macro coverage layer, micro capacity layer, pico and femto layers, as well as WiFi hotspots. Such coexistence of several networks' modes enables the UEs to select the best RAT according to their capabilities. The physical interpretation of RATs' capabilities (i.e. radio resource such as time slots and code sequence), is based on the specific technologies implementation of the radio interface regardless of the used multiple access technology [40]. Hence, RAT capabilities may be stated in terms of effective or equivalent bandwidth that is provided to the UEs as well as delay, packet error rate, request-response transactions, etc.

TABLE 1. List of symbols.

Symbol	Definition
$P_{R_{m,n}}$	Signal power strength received by UE n from BS m
r	Distance between B Sm and UE n
$P_{T_{m,n}}$	Transmitted Power from BS m to UE n
$\mathcal{G}_{m,n}$	Channel gain from the BS m to the UE n
γ^2	Noise power
$\mathcal{P}_I$	Total interference power signal made by the adjacent cells
$\mathcal{SINR}$	Signal-to-noise ratio
$\mathcal{C}_n$	Capacity yielded by UE n
W	Bandwidth
$\mathcal{E}_m$	Level of service efficiency of BS m
$\mathcal{S}_n$	SE requirement of UE n
$\mathcal{N}_m$	Number of UEs associated with BS m
$\mathcal{Q}_S$	PSNR on the server side
$\mathcal{B}$	Total bitrate of the content
b_{BL}	Bitrate of the base layer
b_{EL}	Bitrate of the enhancement layer(s)
Θ	$R-D$ model parameter
b_{BL}	Ratio of packet loss in BL
$\mathcal{Q}_{UE}$	PSNR of the reconstructed video on the UE side
$\mathcal{Q}_{BL}$	PSNR of the base layer
$\mathcal{M}$	MOS
$\mathcal{F}_r$	frame rate
$\mathcal{T}_r$	transmission rate
$\mathcal{E}_r$	packet error rate
η	modulation
σ	coding scheme

Moreover, since UEs can be aware of channel conditions, user experience, and user device features [41], it is in a unique position in enabling the multi-RAT system. In this situation, the SRC in cooperation with UEs will be able to prepare a CAF that includes almost the exact information about the available channels for the UEs in the coverage areas of the BSs. This leads to several advantages such as: improving resource utilization, providing consistent service, and simplifying multi-RAT interoperability and network management processes.

The UEs collect local information about each accessible RAT based on different factors such as spectral efficiency, data-rate, delay, signaling load, etc. BSs' transmitted power and traffic load have a non-negligible impact on the data-rate received by UEs. The signal power strength received by UE

n at distance r from BS m is expressed as:

$$P_{R_{m,n}}(r) = P_{T_{m,n}}.\mathcal{G}_{m,n}(r), \quad (1)$$

where $P_{T_{m,n}}$ is the transmitted power from BS m to UE n, and $\mathcal{G}_{m,n}$ is the channel gain. If the total interference power signal made by the adjacent cells is:

$$\mathcal{P}_I = \sum_{k \in M, k \neq m} \mathcal{P}_{I_{k,n}}(r), \quad (2)$$

then, the signal-to-noise ratio (SINR) is calculated as:

$$\mathcal{SINR}_n = \frac{P_{R_{m,n}}(r)}{\gamma^2 + \mathcal{P}_I}, \quad (3)$$

where γ^2 is the noise power. Utilizing the Shannon's formula, the capacity yielded by an UE with bandwidth W is:

$$\mathcal{C}_n = \mathcal{W}. \log_2(1 + \mathcal{SINR}_n). \quad (4)$$

In the UE perspective, the best RAT is the one that provides the best services in terms of spectral efficiency (SE), latency and signaling load [42]. The level of power transmission of a BS and traffic load is the effective metrics on the UEs' received bitrate. In this context, the accessible BSs can be ranked according to the level of service efficiency that they provide to the UE as:

$$\mathcal{E}_m = \frac{\mathcal{S}_n.\mathcal{N}_m}{P_{T_m}}, \quad (5)$$

where $\mathcal{S}_n$ is the demanded UE SE n as the ratio of the total bitrate received by UE n to the BW measured in $[bps/Hz]$, and P_{T_m} is the total power of BS m that is providing service to the number of UEs, $\mathcal{N}_m$.

According to the stated significances of the multi-RAT architecture, it fits well into the *CASH* scenario whereas the technology steers the traffic in a handled manner to various RATs. Moreover, it supports multiple connections with joint traffic splitting and aggregation for several RATs well.

Traditionally each RAT is end-to-end coupled, which leads to inefficient radio resource utilization. From the signaling point of view, system interoperability between various RATs is handled by the SRC, which is a unified controller network element for unified radio resource and traffic control, as illustrated in Fig. 3. The SRC is integrated with multiple functionalities including radio resource management, joint mobility management, and single non-access stratum (NAS). In *CASH*, the SRC plays an important role, whereas it is the intermediate entity between UEs and CN entities. The SRC and CN use unified interfaces as well as NAS.

In the case of handoff, the SRC changes the spectrum band of the RAT with the same NAS, thus in the CN's point of view, HO is not observable, and thus has no signaling delay. For implementation, the SRC can be implemented independently or it can be associated with RATs or other entities e.g. mobile management entity (MME), mobile switching station (MSC), or it can even be implemented on different devices in a distributed manner. The SRC acts as the coordinating entity between different RATs and has an interface with the network

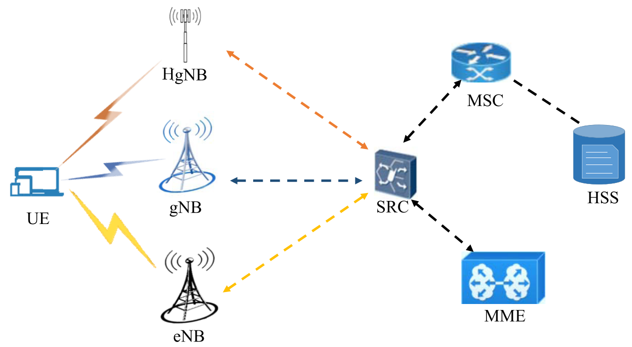

FIGURE 3. UEs have the ability to communicate with several RATs.

controllers of all the RAT involved, as shown in Fig. 3, the UE and SRC may exchange NAS layer 3 (NAS/L3) signaling of 5G network through a gNB and exchange NAS/L3 signaling of LTE via an eNB. Such signaling is called as cross-RAT signaling inter-networking [43]. It enables UEs to connect to an optimal tier of a network in each of the supported RATs and inform the user about the best RAT according to its quality, congestion, and cost [14].

In summary, the RAT selection procedure in *CASH* is as follows:

1) The UE of interest camps on a BS and exchanges NAS/L3 signaling relating to an MSC with an SRC. Means, the UE continuously checks the logical channels that are offered by BSs via a broadcasting control channel (BCCH) and a common control channel (CCCH).
2) The SRC manages radio resources of each RAT and forwards the UE's request to an MSC.
3) The MSC and the UE exchange NAS signaling to accomplish the process.
4) The MSC indicates SRC to assign a channel for the UE.
5) The SRC indicates the UE about a BS radio resource assignment information.
6) The UE collects and sends its local information to the SRC including:
   - Channel status across various RATs.
   - Spectral efficiency

Real-time

   - traffic condition
   - QoS of the received services from different RATs
   - Energy consumptions
7) The SRC prepares a metadata file including network-context and sends to the media preparation server.
8) The media server analysis and extracts the content's characteristics information and then appends the information to the CAF.
9) The media server clusters the content according CAF data.
10) The CAF is shared to the UE, SRC and the flow scheduler.
11) The SRC assigns UE-RAT association.

12) The UE maintains a channel to BSs based on the indicated resource data and asks for each cluster through corresponding RAT.
13) The flow scheduler steers the clusters to the appropriate RAT and starts streaming the content.
14) the UE starts traffic aggregation and plays back the content.

B. PHASE II: CONTENT PREPARATION BASED ON CONTENT-CONTEXT

In this subsection, we explain content preparation on the server side based on the content-context. At the end of this subsection, we also provide quality evaluation methodology that we have used to measure the performance of *CASH*.

SVC-coded video contents can be communicated over multi-RAT HetNets whereas radio channels are jointly managed by the SRC to reduce network strain. SVC provides multi-layer and scalable video bitstreams including one non-scalable BL and several ELs. On the client side, even though some portions of the encoded bitstream are removed i.e. ELs, still valid bit streams can be formed by decoding BL only. It means, ELs are provided to enhance the temporal resolution (frame rate in frame per second unit, fps), the spatial resolution (screen size), or the quality in terms of signal-to-noise ratio (SNR) of the content, however BL is the mandatory portion of the bit stream that must be received safely by the client [44]. In our scenario, the media preparation server using SVC encodes the content into one BL and several ELs. The number of ELs is decided based on the number and quality of the available channels indicated in CAF.

The HTTP server stores multimedia contents and delivers the segments of the content. Recently, video streaming over HTTP as an over-the-top (OTT) is getting plenty of attention because of several reasons such as; (1) the today Internet infrastructure is in that infrastructure evolution stage to support HTTP in the most efficient way, (2) HTTP is a firewall friendly approach in nature, (3) HTTP streaming schemes are client-driven in which the client's capabilities are taken into account [24].

In *CASH*, the server splits the content into some segments in the form of chunks that can be delivered in one or several files. Moreover, for each content, a CAF describes a manifest of the multimedia content, the number and actual bitrate of clusters of the content, and other characteristics. The server then opts for the number and bitrate of the clusters according to the number and bandwidth of the channels that different RATs provide to the UE.

The CAF then will be sent to the UE. Through this metadata file, the UE will be aware of the content characteristics e.g. type, resolution, number of clusters, the actual bitrate required to receive each cluster, timing of each cluster (start time, duration, and end time), etc. Using this information, the UE requests each cluster of the content via the appropriate RAT and starts receiving the video by fetching the segments by HTTP GET requests. This is done by *server-push* instead of client-driven. In the so-called *server-push* method, in order to decrease request-related overhead [45], the server starts a special frame called *push-promise*. At the time of receiving this frame, the UE does not send any other request for those content chunks until the responses are pushed to the UE completely. Then, the UE retrieves the response from its browser cache directly. This mechanism will reduce streaming latency as well.

To evaluate the efficiency of the multimedia streaming techniques, traditionally the quality of delivered content was evaluated in terms of PSNR or distortion rate as a QoS measurement scale. PSNR can be stated as the average of the corresponding assessments over all the frames [46]. The drawback of this measurement technique is that it does not consider the visual masking phenomenon. To alleviate this issue and increase the confidence, in this paper, we consider both PSNR and MOS.

PSNR evaluation [47] can be done in the form of distortion modeling as a continuous function of the stream rate or discrete values according to the number of safely received SVC layers. Thus, on the server side, PSNR as a linear function of bitrate is calculated as:

$$\mathcal{Q}_S = \Theta \cdot (\mathcal{B} - b_{BL}) + \mathcal{Q}_{BL} = \Theta . \sum_{i=0}^{N} b_{EL}^{i} + \mathcal{Q}_{BL}, \quad (6)$$

where $\mathcal{B}$ is the total bitrate of the content, which is equal to sum of the bitrate of the BL (b_{BL}) plus sum of the bitrate of the ELs (b_{EL}) (there may be several ELs or even no EL), and (Θ) is the rate-distortion ($R - D$) model parameter, which is selected based on the spatial-temporal features of the content and the codec. On the UE side, based on the packet loss ratio in BL, b_{BL}, PSNR of the received content is:

$$\mathcal{Q}_{UE} = \Theta . \left(b_{BL} - b_{BL} . \sum_{i=0}^{N} b_{EL}^{i} \right) + \mathcal{Q}_{BL}. \quad (7)$$

In wireless networks, because of the network fluctuations channels condition does not remain the same over time. Hence, packet loss cannot be prevented thoroughly and it is one of the common issues. Therefore, to obtain MOS various factors need to be considered such as frame rate $\mathcal{F}_r$, transmission rate $\mathcal{T}_r$, packet error rate $\mathcal{E}_r$ because of handoff and poor channel quality, modulation η, and coding scheme σ:

$$\mathcal{M} = \frac{\alpha_1 + \alpha_2 \mathcal{F}_r + \alpha_3 (\ln \mathcal{T}_r)}{1 + \alpha_4 \mathcal{E}_r + \alpha_5 (\mathcal{E}_r)^2}, \quad (8)$$

where $\mathcal{E}_r = \frac{1}{1+e^{\eta(SINR-\sigma)}}$, and the coefficients $a_1 - a_5$ are derived by a non-linear regression of the prediction model with a collection of MOS values as in [48].

V. SIMULATION AND RESULTS

We considered many different video streams to evaluate the performance of *CASH* over an instance of 5G HetNets, however in this paper just we report the results about three video contents. Table 2, lists simulation parameters used in this paper. The upper part of the table consists content-context of three video streams, "Blue Planet", "Sony Demo", and

TABLE 2. Simulation parameters.

Characteristics of the video streams (content-context)			
Parameter	Blue Planet	Sony Demo	Transporter2
Resolution	1920x1088	352x288	1920x1088
Frame Rate [fps]	24	30	24
GoP Pattern	G16B3	G16B15	G16B3
Coarse Grain Scalability ELs	4	4	4
Compression Ratio	71.48	16.60	23.37
No. of Frames	9984	17664	9984
Avg. Frame Size [KiB]	78.53	8.9	130.95
Avg. PSNR [dB]	52.49	46.95	47.15
PSNR Standard Deviation [dB]	16.075	4.43	11.08
Quantization Parameter (I-frame, P-frame, B-frame)	(I16,P16, B19)	(I48, PN/A, B48)	(I48, P48, B51)
Network settings (network-context)			
Network Layout	500m gNB layer Inter-Site Distance with 3 small cells per gNB		
Transmission Power [dBm]	gNB: 46, Pico1: 30, Pico2:20, WiFi: 18		
Sub-frame Duration	1ms (11data+3 control)		
Modulation and Coding schemes	QPSK(1/5-3/4), 16-QAM(2/5-5/6), 64-QAM (3/5-9/10)		
Antenna Gain [dBi]	gNB: 14, Pico1:5, Pico2:4, WiFi:3, UE:0		
Delay [ms]	gNB:1, Pico1: 0.015, Pico2: 4, WiFi:3		
Frequency [GHz]	gNB:6, Pico1:28, Pico2:1.8, WiFi: 2.4		
Bandwidth [MHz]	gNB:30, Pico1:40, Pico2:10, WiFi:20		

"Transporter 2". The lower part of the table, lists the network settings i.e. network context. We considered two small cells covered by a macro cell as well as one WiFi access point (AP). For simulation, we have used various coding and simulation tools including FFmpeg [49], MPEG JSVM, EvalVid [50], and NS2.

As an instance, Fig. 4 shows the bitrate of "Blue Plant" content with dramatic levels of motion activity and variations between the complexity of the scenes. Not only for this stream, but for the other contents as well, determination of GBR requires considering a trade-off between the required bandwidth and quality. The higher GBR results in the higher quality, however it squanders precious network resources, especially for low complex scenes. On the other hand, the lower GBR results in lower required bandwidth, though lower quality especially for high complex scenes.

In order to overcome the above-stated issues (i.e. wastage of network resource and QoE degradation), we categorize the content scenes based on their complexity and allocate a channel based on their actual requirements. Fig. 5 shows the scene clustering for Blue Planet content. The content is categorized into four clusters. The red cluster represents the simplest scenes while the blue cluster shows the most complex scenes having the most bitrate. On the server side,

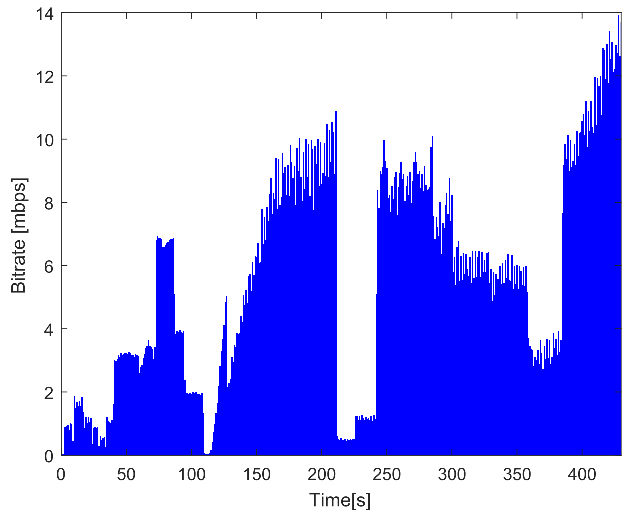

FIGURE 4. VBR of blue planet stream.

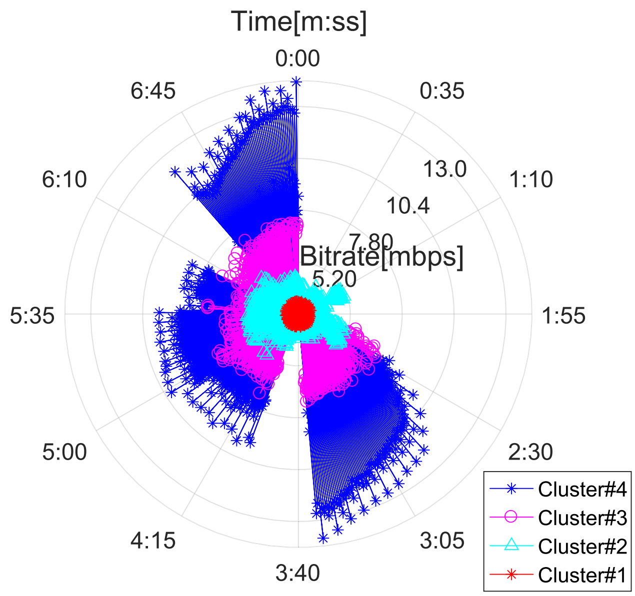

FIGURE 5. Content clustering based on content-context.

a cluster ID is assigned to the chunks that belong to the corresponding cluster.

Fig. 6 presents *CASH* performance compared with conventional streaming and dynamic adaptive streaming i.e. MPEG-DASH, along six dimensions: *average number of HOs*, *average delay*, *average quality switching*, *average stalls*, *average PSNR*, *average MOS*. We supposed that conventional streaming and DASH are working in the same network with no vertical HO and they are operating in a single network. The average number of vertical HOs in the proposed *CASH* is high whereas it is working in the multi-RAT mode, as such for each cluster the UEs needs to handover to another BS vertically. The number of HOs is proportional with dramatic levels of

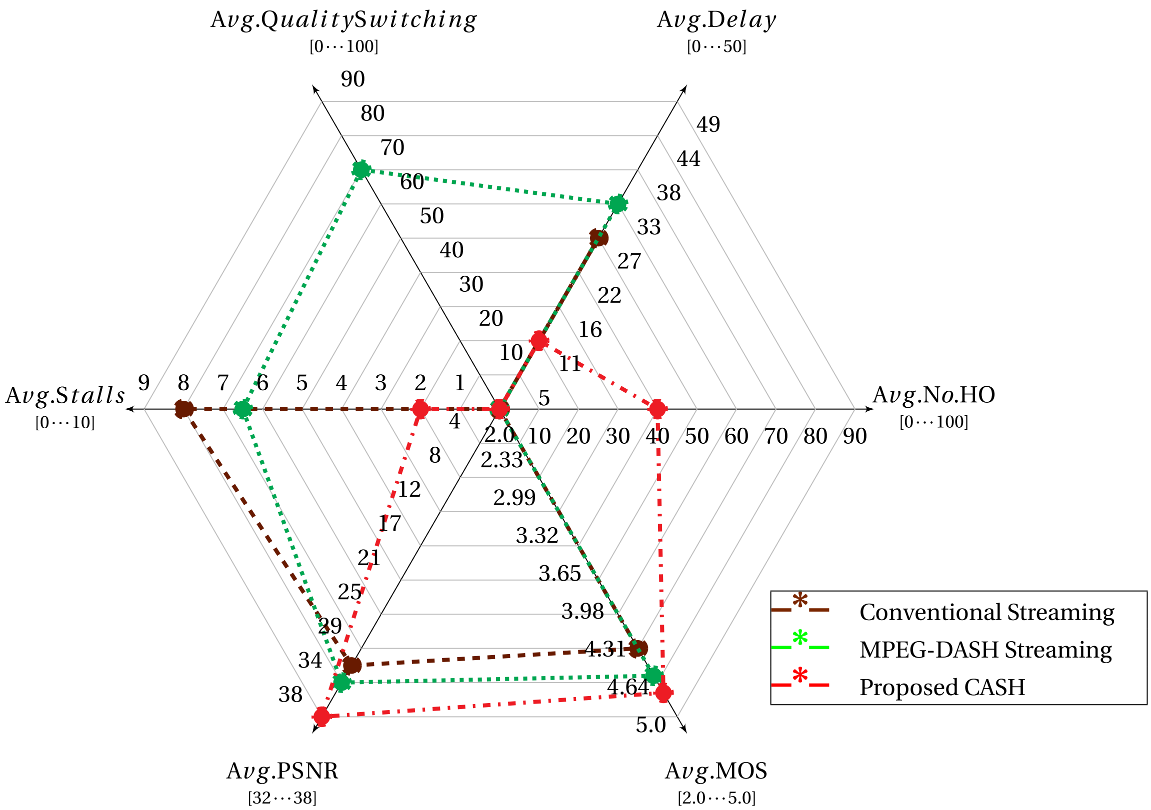

FIGURE 6. CASH performance compared with "Conventional streaming" and "DASH" for streaming of "Blue Planet" content.

motion activity in the content, the more motion difference between the content's scenes, the more number of distinct cluster, and hence the more number of HOs. In order to come up with the issue of vertical HOs and guarantee QoE in *CASH*, we used our earlier QoE-driven handoff management presented in [46]. As stated above, opposite to DASH, *CASH* is a server-push system and hence low message overhead because of the less number of HTTP GET requests result in reduced content latency. DASH has the highest number of quality switching. It is due to the nature of quality adaptation in DASH, in which it needs frequent quality switching according to the network fluctuations. *CASH* has the lowest number of stalls compared to DASH and conventional streaming because of concurrent transmission of different clusters. From the figure, a correlation is salient between MOS and the other metrics i.e. delay, quality switching, stalls, and PSNR. The conventional streaming shows the lowest MOS. Quality switching in DASH results in QoE degradation especially when video quality changes dramatically. However, due to the method of context-aware bitrate allocation of *CASH* where high quality channels are assigned for complex scenes and low quality channels are reserved for simple scenes, there is no QoE degradation in *CASH*, as such most possible MOS can be obtained by *CASH* compared to the others.

Fig. 7 shows the saved bandwidth by the *CASH* for three different video streams e.g. "Blue Planet", "Sony Demo", and "Transporter 2". It demonstrates the performance of the *CASH* well in terms of resource utilization. *CASH* needs less bandwidth considerably because it allocated bandwidth in an intelligent way based on the actual segments requirements. Whereas the enhanced video quality, more than 35dB, is out of human perception capability, either CBR and GBR selection to accommodate the maximum bitrate will result in the wastage of excess bitrate. Moreover, because *CASH* works in server-push mode and not client-pull, the message overhead is considerably reduced.

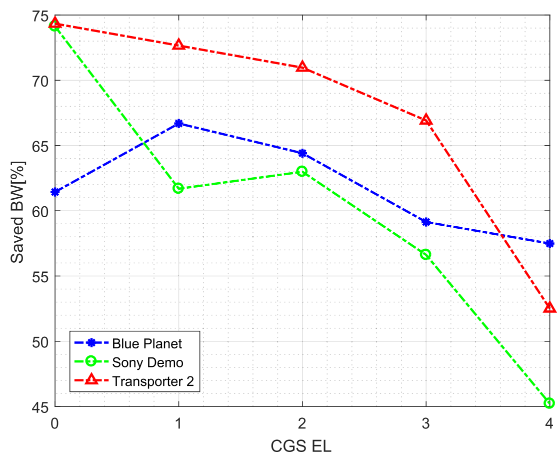

FIGURE 7. Saved bandwidth by CASH.

VI. CONCLUSION

This article presents the *CASH* a context-aware multimedia streaming scheme that enables high quality video streaming over 5G HetNets by maintaining a trade-off between the available network bandwidth and the actual bitrate of different scenes of video content. The *CASH* offers much better QoE performance in terms of PSNR and MOS compared to the conventional video streaming approaches and DASH, since it allocates the available resources based on content-context. In addition, the *CASH* enables us to achieve an enhanced network utilization in terms of saved bandwidth because the RAT of interest is nominated based on network-context. Based on a theoretical analysis and extensive simulation results, the paper concludes that the integrated framework of content- and network-context-aware resource allocation is an effective and smart approach for multimedia streaming over 5G HetNets.

REFERENCES

[1] *CISCO Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021*, CISCO, San Jose, CA, USA, 2017.

[2] I. Godlovitch, N. van Gorp, P. Hausemer, I. Vassileva, R. Womersley, and O. Batura, "Support for the preparation of the impact assessment accompanying the review of the regulatory framework for e-communications," Eur. Commission, Brussels, Belgium, Tech. Rep., 2015. [Online]. Available: https://publications.europa.eu/en/publication-detail/-/publication/2984b37b-9aa6-11e868c-01aa75ed71a1/language-en, doi: 10.2759/5577.

[3] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, "Realizing the tactile Internet: Haptic communications over next generation 5G cellular networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 82–89, Apr. 2017.

[4] M. Cooper, "The Myth of spectrum scarcity," DYNA, Solana Beach, CA, USA, Tech. Rep., 2010. [Online]. Available: http://dynallc.com/wp-content/uploads/2012/12/themythofspectrumscarcity.pdf and http://dynallc.com

[5] S. Islam, M. Zeng, O. A. Dobre, and K.-S. Kwak, "Resource allocation for downlink NOMA systems: Key techniques and open issues," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 40–47, Aug. 2018.

[6] N. Himayat *et al.*, "Network selection in a heterogeneous network," U.S. Patent 9 872 197, Jan. 16, 2018.

[7] H. Zhang, N. Liu, X. Chu, K. Long, A. Aghvami, and V. C. M. Leung, "Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 138–145, Aug. 2017.

[8] T. Han *et al.*, "On downlink NOMA in heterogeneous networks with non-uniform small cell deployment," *IEEE Access*, vol. 6, pp. 31099–31109, 2018.

[9] T. Zhou, N. Jiang, Z. Liu, and C. Li, "Joint cell activation and selection for green communications in ultra-dense heterogeneous networks," *IEEE Access*, vol. 6, pp. 1894–1904, 2018.

[10] P. Wang, W. Song, D. Niyato, and Y. Xiao, "QoS-aware cell association in 5G heterogeneous networks with massive MIMO," *IEEE Netw.*, vol. 29, no. 6, pp. 76–82, Nov./Dec. 2015.

[11] H. S. Kamath, S. B. Mendon, and S. G. Nayak, "Carrier aggregation in LTE-A," *IUP J. Telecommun.*, vol. 9, no. 3, pp. 7–14, 2017.

[12] "LTE aggregation & unlicensed spectrum," 5G Americas, Bellevue, WA, USA, White Paper, Nov. 2015.

[13] Z. Liu, M. Dong, H. Zhou, X. Wang, Y. Ji, and Y. Tanaka, "User-oriented intercell interference coordination in heterogeneous networks (HetNets)," in *Resource Allocation in Next-Generation Broadband Wireless Access Networks*. Hershey, PA, USA: ICI Global, 2017, pp. 145–164.

[14] M. Niaz, "5G mobile communication: Technology enablers & interferencemitigation methods white paper," L&T Technol. Services, Gujarat, India, Tech. Rep., 2017. [Online]. Available: https://pdfs.semanticscholar.org/b4cd/53e8e122a14a33cb2bfdd33114e8b84a71c0.pdf

[15] M. J. Piran and D. Y. Suh, *DASH SAND CE: RBR as a DANE Parameter for RBR Media*, document ISO/IEC JTC1/SC29/WG11 m34318, MPEG 109th Meeting, Sapporo, Japan, Jul. 2014.

[16] D. Y. Suh, Y. W. Cho, M. J. Piran, K. M. Park, and Y. W. So, *MMT SubADC CE: Evidence for SubADC*, document ISO/IEC JTC1/SC29/WG11 m34315, MPEG 109th Meeting, Sapporo, Japan, Jul. 2014.

[17] M. J. Piran and D. Y. Suh, *DASH SAND CE: Scene Clustering, as a New Parameter for DANE*, document ISO/IEC JTC1/SC29/WG11 m35212, MPEG 110th Meeting, Strasbourg, France, Oct. 2014.

[18] B. Xu, Y. Chen, J. R. Carrión, and T. Zhang, "Resource allocation in energy-cooperation enabled two-tier NOMA hetnets toward green 5G," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2758–2770, Dec. 2017.

[19] J. Wu, B. Cheng, C. Yuen, Y. Shang, and J. Chen, "Distortion-aware concurrent multipath transfer for mobile video streaming in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 4, pp. 688–701, Apr. 2015.

[20] P. Pancha and M. El Zarki, "MPEG coding for variable bit rate video transmission," *IEEE Commun. Mag.*, vol. 32, no. 5, pp. 54–66, May 1994.

[21] H. Zhu, Y. Cao, W. Wang, B. Liu, and T. Jiang, "QoE-aware resource allocation for adaptive device-to-device video streaming," *IEEE Netw.*, vol. 29, no. 6, pp. 6–12, Nov. 2015.

[22] A. Scherkus and A. J. Colwell, "Adaptive video streaming over a content delivery network," U.S. Patent 9 060 207, Jun. 16, 2015.

[23] S. Issa, O. O. Khalifa, and T. S. Gunawan, "A dynamic bandwidth schemes base on adaptive video streaming," *Transylvanian Rev.*, vol. 1, no. 8, 2017.

[24] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over the Internet," *IEEE Multimedia*, vol. 18, no. 4, pp. 62–67, Apr. 2011.

[25] A. Zambelli, "IIS smooth streaming technical overview," Microsoft Corp., Washington, DC, USA, Tech. Rep. XPOO2620446, 2009. [Online]. Available: http://www.bogotobogo.com/VideoStreaming/Files/iis8/IIS_Smooth_Streaming_Technical_Overview.pdf

[26] A. Mukherjee, "Macro-small cell grouping in dual connectivity LTE-B networks with non-ideal backhaul," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 2520–2525.

[27] *Adobe HTTP Dynamic Streaming*. Accessed: Apr. 20, 2018. [Online]. Available: http://www.adobe.com/products/httpdynamicstreaming/

[28] L. B. Yuste, "Time and timing within MPEG standards," in *MediaSync*. Cham, Switzerland: Springer, 2018, pp. 411–450.

[29] O. Oyman, "Methods for switching between a MBMS download and an HTTP-based delivery of DASH formatted content over an IMS network," U.S. Patent 9 887 852, Feb. 6, 2018.

[30] M. O. van Deventer, M. Probst, and C. Ziegler, "Media synchronisation for television services through HbbTV," in *MediaSync*. Cham, Switzerland: Springer, 2018, pp. 505–544.

[31] *Information Technology—Dynamic Adaptive Streaming Over HTTP (DASH)—Part 5: Server and Network Assisted DASH (SAND)*, Standard ISO/IEC 23009-5:2017, ISO, Geneva, Switzerland, May 2017.

[32] *The Real-Time Measurement and Intelligence Platform for Streaming TV*. Accessed: Jul. 1, 2018. [Online]. Available: https://www.conviva.com

[33] M. Nekovee and R. Rudd. (2017). "5G spectrum sharing." [Online]. Available: https://arxiv.org/abs/1708.03772

[34] Y.-C. Wang and S. Lee, "Small-cell planning in LTE HetNet to improve energy efficiency," *Int. J. Commun. Syst.*, vol. 31, no. 5, p. e3492, 2018.

[35] *Coordinated Multi-Point Operation for LTE Physical Layer Aspects (Release 11)*, document 36.819 V11.1.0, Dec. 2011.

[36] Z. Yu, X. Zhou, Z. Yu, D. Zhang, and C.-Y. Chin, "An OSGi-based infrastructure for context-aware multimedia services," *IEEE Commun. Mag.*, vol. 44, no. 10, pp. 136–142, Oct. 2006.

[37] G. S. Sandeep, B. S. Kumar, G. S. Mamatha, and D. J. Deepak, "A survey of optimization of rate-distortion techniques for H.264 scalable video coding," in *Proc. Int. Adv. Soft Comput., Intell. Syst. Appl.*, 2018, pp. 227–235.

[38] A. Heindel, E. Wige, and A. Kaup, "Low-complexity enhancement layer compression for scalable lossless video coding based on HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 8, pp. 1749–1760, Aug. 2016.

[39] Y.-S. Yu and C.-H. Ke, "Genetic algorithm-based routing method for enhanced video delivery over software defined networks," *Int. J. Commun. Syst.*, vol. 31, no. 1, p. e3391, 2018.

[40] O. E. Falowo and H. A. Chan, "Multiple-RAT selection for reducing call blocking/dropping probability in cooperative heterogeneous wireless networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 1–12, 2012. [Online]. Available: https://link.springer.com/content/pdf/10.1186%2F1687-1499-2012-221.pdf

[41] Q. Ye, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "Distributed resource allocation in device-to-device enhanced cellular networks," *IEEE Trans. Commun.*, vol. 63, no. 2, pp. 441–454, Feb. 2015.
[42] P. Schulz *et al.*, "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 70–78, Feb. 2017.
[43] M. Fang, Y. Chen, B. Tan, and P. Xing, "Communication method, user equipment, and single radio controller," U.S. Patent 9 706 462, Jul. 11, 2017.
[44] P. Seeling and M. Reisslein, "Video transport evaluation with H.264 video traces," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 1142–1165, 4th Quart., 2012.
[45] D. V. Nguyen, H. T. Le, P. N. Nam, A. T. Pham, and T. C. Thang, "Adaptation method for video streaming over HTTP/2," *IEICE Commun. Exp.*, vol. 5, no. 3, pp. 69–73, 2016.
[46] M. J. Piran, N. H. Tran, D. Y. Suh, J. B. Song, C. S. Hong, and Z. Han, "QoE-driven channel allocation and handoff management for seamless multimedia in cognitive 5G cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6569–6585, Jul. 2017.
[47] G. Lu, X. Zhang, L. Chen, and Z. Gao, "Novel integration of frame rate up conversion and HEVC coding based on rate-distortion optimization," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 678–691, Feb. 2018.
[48] A. Khan, L. Sun, E. Jammeh, and E. Ifeachor, "Quality of experience-driven adaptation scheme for video applications over wireless networks," *IET Commun.*, vol. 4, no. 11, pp. 1337–1347, Jul. 2010.
[49] FFmpeg. Accessed: Jun. 30, 2018. [Online]. Available: https://www.ffmpeg.org and https://trac.ffmpeg.org/wiki/Encode/MPEG-4
[50] C.-H. Ke, "myEvalSVC: An integrated simulation framework for evaluation of H.264/SVC transmission," *TIIS*, vol. 6, no. 1, pp. 379–394, 2012.

MD. JALIL PIRAN (M'10–SM'16) received the Ph.D. degree in electronics and radio engineering from Kyung Hee University, South Korea, in 2016. He was a Post-Doctoral Research Fellow in resource management and quality of experience in 5G cellular networks and Internet of Things (IoT) with the Networking Laboratory, Kyung Hee University. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Sejong University, Seoul, South Korea. He has published a substantial number of technical papers in well-known international journals and conferences in research fields of resource allocation and management in 5G mobile and wireless communication, HetNet, and IoT, multimedia communication, streaming, adaptation and QoE, and cognitive radio networks. In the worldwide communities, he has been an Active Delegate from South Korea in the Moving Picture Experts Group since 2013 and an Active Member of the International Association of Advanced Materials since 2017. He received the IAAM Scientist Medal of the year 2017 for notable and outstanding research in new age technology and innovation, Stockholm, Sweden. He has been recognized as the Outstanding Emerging Researcher by the Iranian Ministry of Science, Technology, and Research in 2017. His Ph.D. dissertation has been selected as the Dissertation of the Year 2016 by the Iranian Academic Center for Education, Culture, and Research in electrical and communications engineering.

S. M. RIAZUL ISLAM received the Ph.D. degree in information and communication engineering from Inha University, South Korea, in 2012. From 2005 to 2014, he was with the University of Dhaka, Bangladesh, as an Assistant Professor and a Lecturer at the Department of Electrical and Electronic Engineering. From 2014 to 2017, he was with the Wireless Communications Research Centre, Inha University, as a Post-Doctoral Fellow. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Sejong University, South Korea. His research interests include wireless communications, Internet of Things, wireless health, and machine learning.

DOUG YOUNG SUH (S'89–M'90) received the B.Sc. degree from the Department of Nuclear Engineering, Seoul University, South Korea, in 1980, and the M.Sc. and Ph.D. degrees from the Department of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 1986 and 1990, respectively. In 1990, he joined the Korea Academy of Industry and Technology and conducted research on HDTV until 1992. Since 1992, he has been a Professor with the College of Electronics and Information Engineering, Kyung Hee University, South Korea. His research interests include networked video and computer game. He has been working as a Korean delegate for ISO/IEC MPEG since 1996.