EMB: Efficient Multimedia Broadcast in Multi-tier Mobile Networks

EMB: Efficient Multimedia Broadcast in Multi-tier Mobile Networks

Chetna Singhal, *Member, IEEE*, Carla Fabiana Chiasserini, *Fellow, IEEE*, and Claudio Ettore Casetti, *Senior Member, IEEE*

***Abstract*—Multimedia broadcast and multicast services (MBMS) in mobile networks has been widely addressed, however an investigation of such a technology in emerging, multi-tier, scenarios is still lacking. Notably, user clustering and resource allocation are extremely challenging in multi-tier networks, and imperative to maximize system capacity and improve quality of user-experience (QoE) in MBMS. Thus, in this paper we propose a clustering and resource allocation approach, named EMB, which specifically addresses heterogeneous networks and accounts for the fact that multimedia content is adaptively encoded into scalable layers depending on the QoE requirements and channel conditions of the heterogeneous users. Importantly, we prove that our clustering algorithm yields Pareto efficient broadcasting areas, multimedia encoding parameters, and resource allocation, in a way that is also fair to the users. Furthermore, numerical results obtained under realistic conditions and using real-world video content, show that the proposed EMB results in lower churn count (i.e., higher number of served users), higher throughput, and increased QoE, while using fewer network resources.**

***Index Terms*—Multimedia broadcast, QoE, Multi-tier architecture, clustering, resource allocation.**

I. Introduction

Mobile video demands are predicted to grow to 38 exabytes per month in mobile networks by 2021 [1], thus generating most of the expected mobile traffic growth. Video data also have much higher bit rates than other mobile content types. It follows that next-generation network services need dynamic radio resource allocation to support flexible multicast and broadcast of high resolution videos [2]. Digital Television (DTV) over wireless networks is one of the key applications that is becoming commonplace, wherein cellular Base Stations (BSs) broadcast multimedia content to stationary and mobile users on their heterogeneous devices [2]–[4].

In parallel, conventional cellular systems are evolving towards increasingly heterogeneous networks, which will play a big role in 5G networks and beyond. Such networks include hierarchical multi-tier deployments [5], [6], as shown in Fig. 1, which increase capacity and coverage by enabling dense reuse of the spectrum and improving link quality. In urban areas, macrocells provide coverage distance of more than 500 m, especially apt to vehicular users, while microcells and picocells provide essential coverage in capacity hungry locations. Macro-, micro-, and pico-cells are all connected to the network through operator-owned backhaul; conversely, femto access points (FAP), covering areas of about 10-50 m in a house, utilize residential backhaul links (digital subscriber line (DSL) or cable) and are privately owned.

C. Singhal is with Department of Electronics and Electrical Commun., Indian Institute of Technology Kharagpur, India (email: chetna.iitd@gmail.com). C. Chiasserini and C. E. Casetti are with Department of Electronics and Communication, Politecnico di Torino, Italy

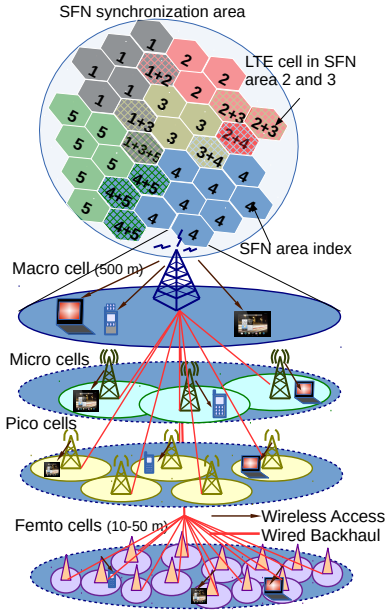


Fig. 1. Multimedia broadcast scenario in a multi-tier LTE-A network.

As far as video services are concerned, on the one hand, significant progress has been made in video coding techniques. To support the demand for high-definition (HD) multimedia content by DTV subscribers in wireless environments, a successor of MPEG-4 advanced video coding/H.264 standard, called ISO/IEC 23008-2 MPEG-H Part 2, ITU-T H.265, or, alternatively, High Efficiency Video Coding (HEVC) has been jointly developed by ISO/IEC MPEG and ITU-T VCEG [7]–[9]. On the other hand, the LTE-A standard specifies multicast/broadcast single frequency network (SFN), configured for providing enhanced Multimedia Broadcast and Multicast Service (eMBMS). The radio resources are adaptively shared between unicast and eMBMS, and multi-cell transmission is enabled for the latter in order to increase the quality of the received signal [10]. Specifically, as depicted in Fig. 1, in a multi-tier LTE-A network, the LTE cells form eMBMS areas that operate as a single frequency network (SFN). An SFN synchronization area consists of several LTE cells (macro). Within each LTE cell, there are several micro-, pico-, and femtocells [11], [12] that participate in broadcasting the multimedia content to the heterogeneous users. However, the resource allocation (carrier, power, modulation, and coding scheme) for each of these tiers (micro, pico, and femto) needs to be efficiently carried out in order to mitigate inter- and intra-tier

interference. As an additional challenge, the broadcast content is to be served to the heterogeneous users that vary in terms of their QoE requirements (namely, heterogeneous display resolution), channel conditions, and premium/regular service subscription. The SFN areas are therefore formed based on several factors, including the program requested by the users, the SINR they experience, the type of User Equipment (UE) they own, as well as their macrocell association).

To address such a complex problem, we propose a TV broadcast framework, called Efficient Multimedia Broadcast (EMB), which first clusters users into hierarchical Pareto-efficient groups, where each group is served with the layered multimedia content by the macro, micro, pico, and femto base stations (BSs). Then the LTE-A resources are assigned to each BS, in every tier of the architecture, for each video layer, by using a matching algorithm. In particular, while accounting for content popularity, user device type and radio connectivity, and the availability of scalable video coding, the EMB scheme uses a multi-criteria optimization approach and makes decisions about optimal *(i)* video encoding parameters and *(ii)* creation of broadcasting areas (SFNs). Importantly, we find that the EMB framework:

(a) reduces the churn rate,
(b) improves the QoE and reduces the perceptual distortion,
(c) increase throughput, and
(d) reduces the amount of resources used for providing the service while ensuring a high system-level fairness.

The rest of the paper is organized as follows. Section II discusses related works and further highlights the novelty of our contribution. Section III presents the network scenario and the system components, and it formalizes the problem we address. Section IV describes the EMB framework and proves that it yields a fair, Pareto-efficient solution. A Pareto-efficient solution defines a system state such that a deviation from it (to make a particular criterion better off) makes another criterion worse off. Then Section V provides details on the simulation scenario and methodology, and presents the key performance results. Finally, Section VI draws our conclusions.

II. Related Work

The joint video team of ITU-T VCEG and the ISO/IEC MPEG has standardized the scalable video coding (SVC) [13] as an extension of H.264/AVC, [7], which achieves a rate-distortion performance comparable to H.264/AVC and has the same visual perception quality with at most 10% higher bit rate [14]. HEVC is a next generation standard with up to 50% reduced bit rate compared to the existing standards, while providing an equivalent perceptual quality [15]. Scalable high-efficiency video coding (SHVC) is a scalable extension of HEVC, which helps HD content delivery over band-limited wireless channel to the heterogeneous UEs [16].

The scalability is in terms of spatial resolution, frame rate, and quantization level. The content is in the form of video layers, with the base layer being the most important and essential content that ensures the delivery of a minimum acceptable video quality. The enhancement layers improve the decoded video quality when received in addition to the base layer. SHVC scalable video is primarily used for adaptive multimedia services [16], [17].

LTE eMBMS resource allocation in order to maximize proportional fair utility of users with heterogeneous channel conditions has been discussed in [18]. [19] presented a formulation for SFN formation, aiming at maximizing the total system throughput as well as a heuristic solution. Network selection and transmission mode for seamless multimedia broadcast in heterogeneous network is discussed in [20]. However, the schemes in [19], [20] do not account for video coding or heterogeneous display capabilities.

[21] proposed static clustering deployment of eNBs in LTE system in order to balance downlink spectral and energy efficiency. Clustering-based load balancing in LTE networks has been given in [22]. Dynamic SFN area creation to optimize multicast transmission efficiency has been studied in [23]. Clustering has been used to discover evolving communities in dynamic networks [24], resource allocation in wireless mesh networks [25], resource management in mobile ad-hoc networks [26], and resource allocation in cellular OFDMA networks [27]. Resource allocation in two-tier LTE het-nets with carrier aggregation has been discussed in [28]. Femto BS placement in enterprise environment considering co-channel interference with macro cell has been discussed in [29]. Multi-cluster scheduling for uplink in LTE-A with carrier aggregation was studied in [30].

[31], [32] analyzed the intrinsic statistical properties and [33] proposed a modified Poisson distribution for modeling of video demands and popularity in multimedia systems. [34], [35] leveraged the user viewing behavior for placement and caching strategy, thereby reducing network bandwidth and storage requirements.

Resource management considering the social and wireless context of users in wireless small cell networks using matching game has been discussed in [36], [37]. Weight based matching for dynamic resource allocation is discussed in [38]. [39] has discussed a two-phase strategy for solving large scale matching problems. Scalable algorithm for maximum matching in random networks has been discussed in [40].

[41] discusses audience driven TV services on 4G LTE networks using system-level scheduling. Radio resource allocation for maximizing served user count for Mobile TV in LTE networks is discussed in [42]. While, radio resource allocation for layered video multicast to heterogeneous groups of users has been discussed in [43]. [44] discussed dynamic formation of SFNs based on the multimedia requests from users. Small cell resource allocation to enhance the multicast throughput and users experience over SFN was discussed in [45]. [46] discussed subgrouping technique for delivering scalable video multicast services in LTE network. Subgrouping technique [46] and small cell resource allocation [45] for scalable video multicast over SFN in LTE network is referred to as 'Layered MBMS' scheme. We compare our proposed EMB scheme with respect to 'Single-tier conventional' scheme [41]and 'Layered MBMS' scheme [45], [46].

However, user-centric multi-criteria clustering to dynamically define SFN areas, adaptive multimedia encoding, and optimized LTE eMBMS resource allocation, in multi-tier LTE-

A het-nets is novel and has not been discussed in literature by far. Overall, clustering along with matching theory (game) gives us a promising solution to dynamically define SFN areas, form Pareto efficient user groups, and manage resources for multi-tier LTE-A het-net.

III. Logical Architecture and System Components

Here we introduce the logical architecture and components of the system under study, including the multimedia layered broadcast, the user demand, and the model we adopt for the data rate and user QoE.

A. Logical architecture

Fig. 2 shows the logical architecture of our proposed efficient multimedia broadcast solution in multi-tier LTE-A networks. The content provider sends to the eMBMS gateway the TV content (prerecorded broadcast program or live video), which is adaptively encoded by the gateway based on the network load and resource constraints. The MBMS coordination entity (MCE) is where our EMB clustering and matching algorithms run periodically in each broadcasting cycle (160 ms [47]). The cycle duration is based on the multicast/broadcast control channel modification period that is conveyed in the system information block (sib2) [48], [49]. This ensures that the MCE periodically performs resource allocation to efficiently provide TV broadcast service to the mobile users (MUs). Then the control and SHVC layer information is sent to the BSs (macro, micro, pico, and femto), based on the clusters, TV content, and SHVC layers to be served to the heterogeneous users. The UE receives the TV content based on the device information, user preference, and service request, previously sent to the network system. The respective SHVC layers of the TV program are sent by the eMBMS gateway to the BSs, which then broadcast them to their users.

An SFN area consists of macrocell BSs that broadcast the base layer of the SHVC content and ensure that the users experience a minimum acceptable quality level, denoted by *Fair*. The micro, pico, and femto BSs associated with a macrocell BS, broadcast higher layers of the corresponding content to deliver higher quality video to users.

We assume that the micro, pico, and femto cells are nested and can be located anywhere across the macrocell coverage area. Hence, to avoid excessive interference between the tier-levels, we assume that different resource blocks (within same frequency band) are used by the different tier-levels for multimedia broadcast. All the micro, pico, and femto BSs, at a particular tier-level and associated with the a particular macro BS, use the same set of resource blocks for content broadcast. All the smaller cells in the coverage area of a macrocell transmit higher layers of programs whose base layer is being broadcast by the eNB (in the macrocell). Hence, resource block reuse (that is being used by eNB) is prevented in a micro, pico, or femto cell, else it would cause interference at the receiver that is receiving SVC layers from the multiple tier levels. Also, the overall throughput (thereby the spectral efficiency) of the system is enhanced due to the higher data rates being supported by the smaller cells (for broadcasting higher SVC layers), using the given frequency band.

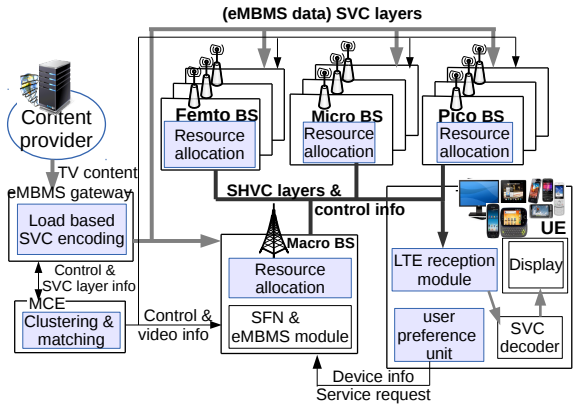


Fig. 2. Logical architecture of the proposed EMB framework.

B. Modeling the user-demand

Video access/demand in multimedia server systems and TV content popularity in IPTV systems are typically modeled using a Zipf-like distribution [32], [35]. In such a system, some videos (or TV contents) are more popular than others, with the content popularity distribution being inversely proportional to the content rank. The multimedia viewing pattern is diurnal with a prime time peak [50], and the video requests are represented using a twenty-four hour period and fixed probability density function. Users request a TV program at prime time with probability 0.4, which drops gradually to 0.1 over the next twelve hours and, again, increases gradually to 0.4 in twenty-four hours. The user-requested TV-content is sampled according to the Zipf-like distribution mentioned above.

A modified Poisson distribution is instead used to model the number of users in the system (system load), such that it caters to peak hours and special events with significantly high system load (e.g., sporting events). $\mathbb{P}(\Delta, \lambda, t)$ is the probability that the system load is $\nu(t)$, at time t, in a system having an arrival rate λ, and maximum number of users arriving per unit of time as Δ. The modified Poisson distribution is defined as: $\mathbb{P}(\Delta, \lambda, t) = \frac{e^{-\lambda}\lambda^{\Delta-\nu(t)}}{(\Delta-\nu(t))!}$.

C. SHVC video QoE and rate model

In order to broadcast the multimedia content to heterogeneous users, the video content is encoded into SHVC layers. The SHVC spatio-temporal scalability grid used for broadcasting is as shown in Fig. 3. There, SHVC layers are indexed by the (s, f) pair, where s is the spatial resolution level and f is the frame rate level. Let $\mathcal{P}$ be the catalogue of TV programs, with cardinality P. The SHVC content consists of a base layer ($l = 1$) and enhancement layers ($1 < l \leq L$) that provide an incremental improvement in QoE when a higher layer is decoded along with all its lower layers. For a program $p \in \mathcal{P}$, if a total of L_p SHVC layers are transmitted, then users need to successfully receive all SHVC layers till l in order to reconstruct l ($l \leq L_p$) layers [16]. Additionally, video may be encoded using a different quantization level q (equivalent to higher quantization step size resulting in higher compression but lower quality), which yields a tradeoff between bit rate

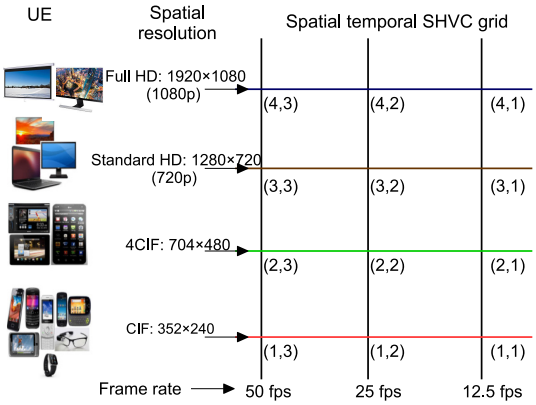


Fig. 3. SHVC spatial and temporal scalable layer grid, and devices categorized in terms of spatial resolution

TABLE I
SUBJECTIVE VIDEO QUALITY (I.E., QOE) CORRESPONDING TO PARAMETRIC VIDEO QUALITY MEASURE $Q(q,f,s)$ AND MOS.

MOS	$Q(q,f,s)$	Quality level (QoE)
1	0	Bad
2	$(0.0-0.25]$	Poor
3	$(0.25-0.5]$	Fair
4	$(0.5-0.75]$	Good
5	$(0.75-1.0]$	Excellent

and quality (the higher the q, the poorer the quality but the lower the bit rate requirement).

The overall video quality is assessed using a parametric function $0 \leq Q(q,f,s) \leq 1$ that approximates the Mean Opinion Score (MOS), a subjective measure indicating the user QoE. $Q(q,f,s)$ has a direct relationship with the MOS [51]: MOS $= 4 \times Q(q,f,s) + 1$. The numerical correspondence among $Q(q,f,s)$, MOS, and QoE is listed in Table I.

The parameters for the quality model are specific to a video and are based on its inherent features. In this work, we will use the quality parametric model defined in [51], according to which, for a given spatial resolution s, $Q(q,f,s)$ is a function of the quantization level q and frame rate f, as follows:

$$Q(q,f,s) = \hat{Q} \cdot Q_f(f) \cdot Q_q(q) \tag{1}$$

where $\hat{Q}$ is the top quality level of video received by the user when it is encoded at minimum quantization level $\hat{q}$ and at the highest frame rate $\hat{f}$. For normalization, we consider $\hat{Q}$ to be equal to 1. Clearly, Q_f is an increasing function of f, while Q_q is an exponentially decreasing parametric function of q ($\hat{q} \leq q \leq \check{q}$, $\check{q}$ is the maximum quantization level). Further details can be found in [51].

We use a similar parametric model for the bit rate, derived from [52], where the bit rate is expressed as a function of quantization level q, frame rate f, and resolution s:

$$R(q,f,s) = \hat{R} \cdot R_f(f) \cdot R_q(q) \cdot R_s(s) \tag{2}$$

with $\hat{R}$ being the maximum bit rate of the video sequence with minimum quantization level $\hat{q}$, maximum frame rate $\hat{f}$, and maximum spatial resolution $s_{\max}$. We remark that the higher the SHVC layer, the larger the resolution value s or the frame rate f, hence the rate requirement.

D. Problem formulation

In a 5G heterogeneous network with N users, we consider that a synchronization area can cover a maximum of B macro-cells. Within each synchronization area there are multiple SFN areas, whose number, denoted by M, must not exceed $M_{\max} = 256$ [53]. All the BSs within a synchronization area have to be time-synchronized. $\mathcal{P}$ is the set of distinct multimedia content that can be requested by the users in the system. However, each SFN area m serves a disjoint subset of content, denoted by $\mathcal{P}_m \subseteq \mathcal{P}$.

The SFN is formed by macrocells only: once the SFN is formed and the set of programs to be broadcast within it has been determined, then we also know which programs will be broadcast by the micro, pico, and femto BSs. We allocate separate resources blocks from the same LTE band at each tier-level for content broadcast, in order to minimize inter-tier interference. We verify that enough resources are available to satisfy quality constraints, necessarily at the macrocell level, as it ensures acceptable QoE (by broadcasting the SVC base layer).

We define an indicator function $x_{k,m}$ that takes 1 when the BS of macro-cell k is associated with the SFN area m, and 0 otherwise. We define also the indicator function a_k that takes 1 if the BS of macro-cell k is associated with at least one SFN area (i.e., $\sum_{m=1}^{M} x_{k,m} > 0$), and 0 else. The maximum synchronization area size, Z, is then given by:

$$\sum_{k=1}^{B} a_k \leq Z \tag{3}$$

$\mathcal{B}_m$ represents the set of BSs (macro-cell) that belong to SFN area m, i.e.,

$$\mathcal{B}_m = \{k : x_{k,m} = 1 \ \forall \ 1 \leq k \leq B\}, \ \forall \ 1 \leq m \leq M, \tag{4}$$

Clearly, the following constraint on SFN area formation holds: no two SFNs within a synchronization area can entirely overlap and be absolutely identical in terms of the BSs participating in multimedia broadcast:

$$|\mathcal{B}_m \cap \mathcal{B}_n| \leq \min\{|\mathcal{B}_m|, |\mathcal{B}_n|\}, \ \forall \ 1 \leq m,n \leq M, \tag{5}$$

Each user i is associated with a particular macro-cell BS, say k; we denote this case by the indicator function $y_{i,k}$. Also, user i is associated with a BS at tier-level v ($v = 0, 1, 2,$ and 3, for macro, micro, pico, and femto-cell levels, respectively), which is denoted by the indicator function $\tilde{y}_{i,v}$. Each user is interested and requests for only one content π_i, $\pi_i \in \mathcal{P}$, for viewing, at a given instance of time. This framework can be extended for serving users requesting more than one content by serving each request independently. The user SINR is denoted by γ_i.

Let β_k represent the sum-total of resource (time-frequency) blocks assigned for multimedia broadcast by BSs at all tier-levels associated with the macro-cell k. This is upper bounded by $\beta_{\max}$, i.e., the maximum number of resource blocks (e.g., 60% in LTE-A Rel. 9-13 [53] and up to 100% in LTE-A Rel. 14 [47]) that can be allocated for multimedia broadcast by a macro-cell. A cell assigns $\rho_{p,m}^{(v)}$ resources towards the

multimedia broadcast of content $p \in \mathcal{P}_m$ in SFN area m at tier-level v. $\mathbb{1}_{p,m}$ is the indicator function denoting that content p is being broadcast in SFN area m.

The macrocell BS broadcasts the base layer and the smaller cell (micro, pico, and femto) BSs broadcast the higher (enhancement) layers of the SHVC content. The resources are allocated based on the content being broadcast at each tier-level. Using (2), the resources at tier-level v have to deliver rate $R_{p,m}^{(v)}$ for program p in SFN area m, which is defined as:

$$R_{p,m}^{(v)} = \begin{cases} R(q_{p,m}, f_0, s_0), & \text{if } v = 0 \\ R(q_{p,m}, f_v, s_v) - R_{p,m}^{(v-1)}, & \text{if } v > 0. \end{cases} \tag{6}$$

The number of resource blocks being used at tier-level v to broadcast the SHVC layers of content p is represented as $\rho_{p,m}^{(v)}$. The delivery of the base layer (ensuring that the minimum quality level of the content gets delivered to every user) by the macrocell BS is ensured by $\rho_{p,m}^{(0)}$. The higher SHVC layers (providing higher quality) are broadcast by the other tier-level BSs. The factors that govern $\rho_{p,m}^{(v)}$ include the minimum SINR amongst all users requesting content p that are connected to BS at tier-level v, and also the bit rate of the SHVC layers being broadcast at tier-level v, $R_{p,m}^{(v)}$ (given in (6)). The quantization level for encoding content p in SFN area m is represented as $q_{p,m}$. Corresponding to the SHVC layers being broadcast by BS at tier-level v, f_v is the frame rate and s_v is the spatial resolution.

Within SFN area m, amongst all the users that are requesting content p, the minimum SINR at tier-level v is denoted by $\gamma_{p,m,v}^{\min}$. We consider that the cells can allocate one out of a set of MCS levels: the MCS $\Omega_{p,m,v}$ governs the amount of resources used for broadcast of content p in the SFN area m at tier-level v. The SINR threshold associated with the MCS level $\widetilde{\Omega}$ is given by $\gamma^{th}(\widetilde{\Omega})$. The minimum SINR user that gets served at any tier level is governed by the SINR threshold for the lowest MCS level. This strategy at the macrocell level guarantees that the base layer (minimum acceptable QoE) is broadcast to all such users ($\gamma_i > \gamma^{th}(\min \widetilde{\Omega})$). The bit rate per unit resource block is denoted by $r(\gamma_{p,m,v}^{\min})$ and it is equal to the product of frequency bandwidth of each resource block and the spectral efficiency corresponding to MCS level $\Omega_{p,m,v}$. The system selects a higher MCS (resulting in a higher rate per resource block) in case users have a higher SINR. Thus, β_k and $\rho_{p,m}$ can be written as:

$$\beta_k = \sum_{v=0}^{3} \sum_{m=1}^{M} \sum_{p=1}^{P} \rho_{p,m}^{(v)} \cdot x_{k,m} \tag{7}$$

$$\beta_k \leq \beta_{\max}, \forall\ 1 \leq k \leq B \tag{8}$$

$$\mathbb{1}_{p,m} = \begin{cases} 1, & \text{if } \exists\ i,\ y_{i,k} = 1,\ x_{k,m} = 1,\ \pi_i = p, 1 \leq i \leq N \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

$$\gamma_{p,m,v}^{\min} = \min\{\gamma_i : y_{i,k} = 1, \tilde{y}_{i,v} = 1, x_{k,m} = 1, \pi_i = p, 1 \leq i \leq N\} \tag{10}$$

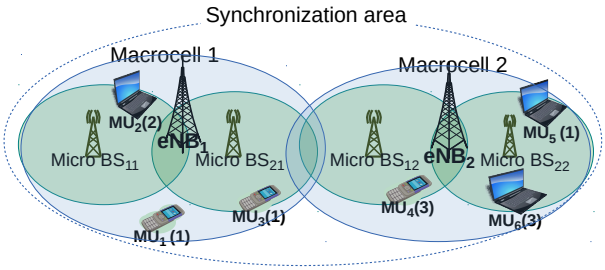


Fig. 4. Multi-tier network illustration with $B = 2, P = 3$, and $v = 2$

$$\rho_{p,m}^{(v)} = \begin{cases} \frac{R_{p,m}^{(v)}}{r(\gamma_{p,m,v}^{\min})}, & \text{if } \mathbb{1}_{p,m} = 1 \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

$$\Omega_{p,m,v} = \max\{\widetilde{\Omega} : \gamma^{th}(\widetilde{\Omega}) \geq \gamma_{p,m,v}^{\min}, \forall\ \widetilde{\Omega}\} \tag{12}$$

The video bit rate and quality (MOS), both decrease with increase in q for all f and s levels of the scalability grid. We need to ensure that even for the lowest-SINR user requesting the content p in SFN area m, the QoE (given in (1)) for the received base layer content is at an acceptable level. This is done by broadcasting the base layer (frame rate f_0 and spatial resolution s_0) at the macrocell level. Based on (1), (2), and (11), the constraint on QoE for the lowest-SINR user at the macrocell level is:

$$Q_w(\rho_{p,m}) = \hat{Q} \cdot Q_f(f_0) \cdot Q_q\left(R_q^{-1}\left(\frac{\rho_{p,m}^{(0)} \cdot r(\gamma_{p,m,0}^{\min})}{\hat{R} \cdot R_f(f_0) \cdot R_s(s_0)}\right)\right), \tag{13}$$

$$Q_w(\rho_{p,m}) \geq 0.25$$

An example of multi-tier network is depicted in Fig. 4. It has two tiers ($v = 2$), three programs ($P = 3$), two eNBs in the synchronization area ($B = 2$), and three SFN areas ($M = 3$). Each MU i requesting program p_i is represented as $\text{MU}_i(p_i)$. In this scenario, the indicator functions have the following values: $X = (x_{k,m}) \in \mathbb{R}^{2\times 3} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$,

$$Y = (y_{i,k}) \in \mathbb{R}^{6\times 2} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}^T,$$

$$\tilde{Y} = \tilde{y}_{i,v} \in \mathbb{R}^{6\times 2} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T,$$

$$\mathbb{1} = \mathbb{1}_{p,m} \in \mathbb{R}^{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The overall objective is to minimize the heterogeneous network resources required for efficient multimedia broadcast given the constraints on available resources, heterogeneous users' QoE requirements, and synchronization area size. Our decision variables are therefore the β_k, which are functions of the $\rho_{p,m}^{(v)}$; the objective function is as follows:

$$\min \sum_{k=1}^{B} \beta_k \tag{14}$$

TABLE II
NOTATION

Symbol	Significance
$\mathbb{1}_{p,m}$	Program p broadcast in SFN area m
$x_{k,m}$	BS k associated with SFN area m
$y_{i,k}$	User i associated with BS k
$\tilde{y}_{i,v}$	User i associated with BS at tier-level v
$\rho_{p,m}^{(v)}$	Resource allocated tier-level v in m for program p
$q_{p,m}$	Quantization level for program p broadcast in m
$\Omega_{p,m,v}$	MCS allocated for program p broadcast in m at tier-level v
β_k	Resources allocated by BS k for multimedia broadcast
$\gamma_{p,m,v}^{\min}$	min SNR amongst users requesting p in m at tier-level v

The above is subject to the linear constraints:

$$\beta_k \leq \beta_{\max}, \forall k \in [1, B] \tag{15}$$

$$\rho_{p,m} \geq 0, \ \forall p \in [1, P], m \in [1, 2^B - 1] \tag{16}$$

and the non-linear constraints, derived from (5)–(13):

$$\hat{Q} \cdot Q_f(f_v) \cdot e^{\left(-\frac{\xi_p}{\hat{q}} \cdot \left(\left(\frac{\hat{q} \cdot \hat{R} \cdot R_f(f_v) \cdot R_s(s_v)}{\sum_{u=0}^{v} \rho_{p,m}^{(u)} \cdot r(\gamma_{p,m,u}^{\min})} \right)^{(1/\delta_p)} - 1 \right)} \geq \overline{Q}_v, \quad \forall \ p \in [1, P], v \in [0, 3] m \in [1, 2^B - 1] \tag{17}$$

In the above expression, ξ_p and δ_p are video specific parameters in $Q_q(q)$ (QoE model) and $R_q(q)$ (rate model), respectively, for SHVC content p, while $\overline{Q}_v$ is the quality-level ensured by BS of tier-level v, where $Q_0 = 0.25$. The objective function given by (14) minimizes the resource allocation by the BSs for multimedia broadcast. The optimization problem is subject to the maximum resource constraint (given by (15)) and acceptable quality levels (given by (17)).

Since, the above problem is non-linear and mixed integer, solving it to the optimum is impractical for scenarios with $B > 3$ and $P > 3$. Thus, below we develop an efficient, yet tractable, solution to the problem that, as proved later, turns out to be Pareto efficient and fair to the users.

IV. The EMB framework

Fig. 5 shows the block diagram of our EMB solution for efficient resource allocation in multi-tier network (all symbols used in Fig. 5 are summarized in Table II). The proposed scheme primarily adopts multi-criteria clustering and nested matching algorithm for SFN area formation, adaptive multimedia encoding, and eMBMS resource allocation. The SFN area formation is performed in every broadcasting cycle (160 ms) while the adaptive multimedia encoding and eMBMS resource allocation is performed more frequently (every 20 ms). This ensures that the EMB system adapts to channel condition variations experienced by mobile users.

A. SFN area formation, and resource allocation

The SFN areas are formed using multi-criteria clustering of users (premium and regular) based on the their content demands, experienced channel conditions, and UE capabilities for various TV programs. The clustering problem is defined as: $(\phi, \Gamma_1, \Gamma_2, \Gamma_3)$, where ϕ is the set of feasible clusters and Γ_1, Γ_2, and Γ_3 are the three criterion functions (given in (18a) and defined by Definitions 1, 2, and 3, respectively).

$$\Gamma_i : \phi \rightarrow \mathbb{R} \tag{18a}$$

$$\Gamma_i(S) < \Gamma_i(S'), \Gamma_j(S) \leq \Gamma_j(S'), \forall S' \in \phi, \ i, j \in [1, 3], i \neq j \tag{18b}$$

$$\mathcal{S} = \{S \in \phi : \text{(18b) is satisfied}\} \tag{18c}$$

In the above expression, S is a dominant solution, which is Pareto efficient if (18b) is satisfied with strict inequality for at least one criterion. $\mathcal{S}$ is the set of minimal elements that satisfy the dominance relation, defined by (18c).

Three criterion functions are used for EMB solution: total resource allocation for video broadcast (SHVC layers encoded with quantization level: $q_{p,m}$ ($1 \leq p \leq P$, $1 \leq m \leq M$)) is Γ_1; perceptual video distortion metric is Γ_2; and bias metric is Γ_3. The range of each criterion Γ_1, Γ_2, and Γ_3, is $[0, 1]$.

Definition 1. *The fraction of resources (resource blocks) assigned for multimedia broadcast, Γ_1, is defined as:*

$$\Gamma_1 = \frac{\sum_{v=0}^{3} \sum_{m=1}^{M} \sum_{p=1}^{P} \frac{\rho_{p,m}^{(v)}}{\beta_{\max}} \cdot \mathbb{1}_{p,m}}{\sum_{m=1}^{M} \sum_{p=1}^{P} \mathbb{1}_{p,m}}. \tag{19}$$

Definition 2. *Perceptual video distortion metric, Γ_2, indicates the overall perceptual distortion for unacceptable QoE levels of the broadcast content in the system. It is defined as:*

$$\Gamma_2 = \frac{\sum_{m=1}^{M} \sum_{p=1}^{P} D_{p,m}}{\sum_{m=1}^{M} \sum_{p=1}^{P} \mathbb{1}_{p,m}} \tag{20}$$

$$D_{p,m} = \begin{cases} 1, \ \textit{if } \mathbb{1}_{p,m} = 1, \ Q(q_{p,m}, f_v, s_v) < \overline{Q}_v, \forall \ v \in [0, 3] \\ 0, \quad \textit{otherwise.} \end{cases} \tag{21}$$

Definition 3. *Contents, $p \in [1, P]$, are the contenders competing for the network resources. Bias metric, Γ_3, indicates the overall performance of the clustering algorithm in terms of system's fairness. It is derived from the Jain's fairness measure [54] for resource allocation and is defined as follows:*

$$\widetilde{Q}_p = \min_{1 \leq m \leq M, \mathbb{1}_{p,m} = 1} \{Q(q_{p,m}, f_{L_{p,m}}, s_{L_{p,m}})\} \tag{22}$$

$$\Gamma_3 = 1 - \frac{\left(\sum_{p=1}^{P} \widetilde{Q}_p \right)^2}{P \cdot \sum_{p=1}^{P} \widetilde{Q}_p^2} \tag{23}$$

where, $L_{p,m}$ is the highest SHVC layer of p broadcast in m.

As detailed below, our clustering method forms SFNs areas that are Pareto-efficient for the above multi-criteria clustering problem, using a modified local optimization (relocation) algorithm.

1) Modified relocation algorithm: The multi-criteria clustering and SFN area formation algorithm (see Alg. 1) is used to cluster the N users in the geographical area. It is based on the following information related to users: the experienced SINR (γ_i) that can be easily derived from the reported CQI, UE type

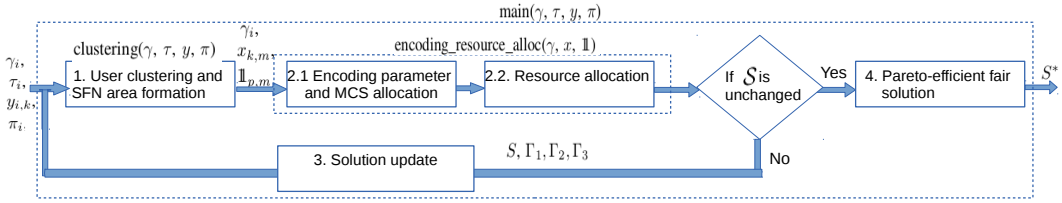


Fig. 5. The EMB solution for efficient resource allocation in multi-tier networks.

(τ_i), requested program (π_i), and the macrocell to which the user is associated (indicated by $y_{i,k}$). UEs are clustered into groups; each cluster consists of users with similar channel conditions, device type, and location. User clusters are then translated into SFN areas, $M \leq 2^B - 1$.

Algorithm 1 User clustering and SFN area formation

Input: γ_i, τ_i, $y_{i,k}$, π_i, $1 \leq i \leq N$, $1 \leq k \leq B$
Function: clustering(γ, τ, y, π)
1) **Determine initial clustering** $\mathcal{C} = \{\mathcal{U}_k | 1 \leq k \leq M\}$
1a) Select the first centroid c_1 uniformly at random from set of users: $1 \leq i \leq N$.
1b) Each user i is associated to $b_i = k$, given $y_{i,k} = 1$.
1c) **for each** $i = 1$ to N
Compute $d^2(i, c_1) = (\gamma_i - \gamma_{c_1})^2 + (b_i - b_{c_1})^2 + +(\tau_i - \tau_{c_1})^2$
1d) Select all other centroids, c_m, $2 \leq m \leq M$:
for each $m = 2$ to M
Select each centroid $c_m = n$ at random with probability $\frac{\min\limits_{1 \leq i < m} d^2(n, c_i)}{\sum\limits_{j \in C_i} d^2(j, c_i)}$, where
C_i is a set of all j closest to $c_i, 1 \leq i \leq m$
1e) Each user is assigned to the cluster with the closest centroid
for each $i = 1$ to N
%Assign user i to cluster $\mathcal{U}_{m^\star}$
$m^\star = \min\limits_m d^2(i, c_m)$
2) **Cluster mapping to form** M **SFN areas**
Initialize $x_{k,m} = 0$, $\mathbb{1}_{p,m} = 0$, $\forall p \in [1, P], m \in [1, M]$
%Associate BSs (maximum 8 [55], [56]) and content to each SFN area
for each $k = 1$ to B
if $x_{k,m^\star} = 0$, $\mathbb{1}_{\pi_i, m^\star} = 0$, $y_{i,k} = 1$, and $\sum\limits_{m=1}^{M} x_{k,m} < 8$ then
$x_{k,m^\star} = 1$
$\mathbb{1}_{\pi_i, m^\star} = 1$
Output: $x_{k,m}$, $\mathbb{1}_{p,m}$, $1 \leq k \leq B$, $1 \leq p \leq P, 1 \leq m \leq M$

2) MAR: Matching Algorithm for Resource allocation in multi-tier mobile networks: Matching theory tractably solves the combinatorial problem of matching players in two distinct sets by suitably characterizing the interaction between heterogeneous nodes (varied types, information, and objectives). It defines preferences for heterogeneous and complex objectives in order to provide stable and optimal solutions using efficient and fast algorithmic implementations. Radio resource allocation in multi-tier networks for scalable video broadcast with heterogeneous users can be posed as a matching problem.

Algorithm 2 MAR: Matching algorithm for resource allocation

Input: γ_i, $x_{k,m}$, $\mathbb{1}_{p,m}$, $1 \leq k \leq B$, $1 \leq i \leq N$
Function: encoding_resource_alloc(γ, x, $\mathbb{1}$)
1) Encoding parameter and MCS allocation for content
1a) Initialize $q_{p,m} = \hat{q}$, if $\mathbb{1}_{p,m} = 1$
1b) Minimum SINR of poorest user in SFN area m requesting program p at level v is $\gamma^{\min}_{p,m,v}$
1c) Number of users in m requesting for p is $N_{p,m}$
1d) Ordered set $\mathcal{T}$, $\mathcal{T} = \{T_i : T_i = (p, m), 1 \leq i \leq P \cdot M, N_{T_1} > N_{T_2}\}$
1e) Initialize $\Gamma_j = 1$, $\forall j \in [1, 3]$, $\mathbb{\Gamma} = 1$, $\rho^{(v)}_{T_i} = 0, \forall i \in [1, P \cdot M]$,$w_1 = w_2 = w_3 = 1/3$
1f) Assign MCS
for each $i = 1$ to $|\mathcal{T}|$
if $\mathbb{1}_{T_i} = 1$, then
for each $v = 0$ to 3
Assign $\Omega_{T_i, v}$ using $\gamma^{\min}_{p,m,v}$ in (12)
2) Encoding parameter and resource allocation
for each $v = 0$ to 3
for each $i = 1$ to $|\mathcal{T}|$
for each $q = \hat{q}$ to $\check{q}$
for each $k = 1$ to B
if $x_{k,m} = 1, \sum\limits_{u=0}^{3} \sum\limits_{m=1}^{M} \sum\limits_{p=1}^{P} \rho^{(u)}_{p,m}\, x_{\mathcal{B}_k, m}\, \mathbb{1}_{p,m} \leq \beta_{\max}$
Assign $\rho^{(v)}_{T_i}$ using (11) for selected q
Update the overall metric $\mathbb{\Gamma} = \sum\limits_{j=1}^{3} w_j \Gamma_j$
if $\mathbb{\Gamma} \geq \sum\limits_{j=1}^{3} w_j \Gamma_j$ then
Update encoding parameter $q_{T_i} = q$
3) Compute the criteria: $\Gamma_1, \Gamma_2, \Gamma_3$
4) Assign solution $S = \{(x_{k,m}, \mathbb{1}_{p,m}, \rho^{(v)}_{p,m}, \Omega_{p,m,v}, q_{p,m}) | 1 \leq k \leq B, 1 \leq m \leq M, 1 \leq p \leq P, 0 \leq v \leq 3\}$
Output: S, $\Gamma_1, \Gamma_2, \Gamma_3$

Set 1 comprises the scalable video layers of all broadcast content in each SFN area. *Set 2* comprises the constricted set of resource blocks at each hierarchical level of the multi-tier architecture. A one-to-one matching between the two sets is performed with externalities and dynamics. The

Algorithm 3 Multi-criteria optimization algorithm for efficient multimedia broadcast

Input: γ_i, τ_i, $y_{i,k}$, π_i, $1 \leq i \leq N$, $1 \leq k \leq B$

Function: main(γ, τ, y, π)

1) $[\gamma, x, \mathbb{1}]$ =clustering(γ, τ, y, π)
2) $[S, \Gamma_1, \Gamma_2, \Gamma_3]$ =encoding_resource_alloc(γ, x, $\mathbb{1}$)
3) **Solution update**
   - 3a) If $\mathcal{S}$ is an empty set then include S in $\mathcal{S}$.
   - 3b) If there does not exist clustering $S' \in \mathcal{S} : S'$ dominates S, i.e. (18b) holds true for $\Gamma_1, \Gamma_2, \Gamma_3, \mathbb{\Gamma}$, then replace S' with S, i.e. $\mathcal{S} := \mathcal{S} \cup S$ and $\mathcal{S} := \mathcal{S} \setminus \{\tilde{S} \in \mathcal{S} : S \text{ dominates } \tilde{S}\}$.
   - 3c) Repeat steps from 1, till Pareto-efficient clustering set $\mathcal{S}$ remains unchanged.
4) **Pareto-efficient fair solution**, $S^* \in \mathcal{S}, \forall S' \in \mathcal{S}$, $min(\max\limits_{1\leq i\leq 3}(w_i\Gamma_i(S^*))) \leq min(\max\limits_{1\leq i\leq 3}(w_i\Gamma_i(S')))$, $S' \neq S^*$

Output: Pareto-efficient fair solution S^*

externalities are in the form of dependencies between the video layers, while dynamics are in terms of time-dependent user demands. We have developed a Matching Algorithm for Resource allocation (MAR), reported in Alg. 2, that gives a two-sided stable solution. It ensures that the MCS assignment (Step 1f) and resource allocation (Step 2) guarantee the data rate required to transmit the video layers of the program to all the users requesting it. MAR is derived from the basic deferred acceptance matching, which has been implemented using adaptive reverse greedy algorithm.

3) EMB solution using multi-criteria optimization: The EMB scheme finds the Pareto-efficient fair solution, S^*, using the multi-criteria modified optimization algorithm (see Alg. 3) for $(\phi, \Gamma_1, \Gamma_2, \Gamma_3)$. This is equivalent to the optimization problem defined in (14)-(17). Γ_1 ensures that S^* is the efficient resource allocation strategy for BSs, $\beta_k, 1 \leq k \leq B$, that is equivalent to solving the optimization objective given by (14). This is achieved while satisfying constraints (15) and (16). Γ_2 and Γ_3 ensure that S^* satisfies the non-linear constraint given by (17) while maintaining system fairness.

B. Convergence to a fair and Pareto-efficient solution

To show that our scheme yields a Pareto-efficient solution for multi-criteria clustering, we draw on [57], [58], and on the proof of convergence that follows from the discussion on iterative algorithms given in [59].

Proposition 1. *Given a set of coefficients $w_i \geq 0$, $i = 1, 2, 3$, such that, $\sum\limits_{i=1}^{3} w_i = 1$, then the solution to $\min\limits_{S\in\phi} \sum\limits_{i=1}^{3} w_i\Gamma_i(S)$, is a Pareto-efficient solution to the multi-criteria problem $(\phi, \Gamma_1, \Gamma_2, \Gamma_3)$.*

Proof: We first observe that the function, $f(\Gamma_1, \Gamma_2, \Gamma_3) = \min\limits_{S\in\phi} \sum\limits_{i=1}^{3} w_i\Gamma_i$, is strictly increasing on each component of the set $\{(\Gamma_1(S), \Gamma_2(S), \Gamma_3(S)) : S \in \phi\}$, i.e., for each $k = 1, 2, 3$, $\Gamma_k(S) < \Gamma_k(S') \Rightarrow f(\Gamma_j, \Gamma_k(S)) < f(\Gamma_j, \Gamma_k(S')), k, j \in [1, 3], j \neq k$. Thus, the assumption and definition of our proposed solution is in line with [57], as has been discussed in Proposition 5 in the Appendix. Furthermore, based on [57, Ths. 3–4], the clustering solution that minimizes $f(\Gamma_1, \Gamma_2, \Gamma_3)$, is a Pareto-efficient solution to the multi-criteria problem $(\phi, \Gamma_1, \Gamma_2, \Gamma_3)$. □

Proposition 2. *Alg. 3 converges to a Pareto-efficient solution set at a linear rate and its execution is more efficient than an extensive search algorithm.*

Proof: The modified relocation algorithm is similar to an iterative multi-criteria cluster and update algorithm [59] (similarity discussed in the Appendix - Proposition 7), where, at each iteration, the efficient clustering solution replaces the other solutions in the Pareto-set. Therefore, based on* Section 4.3 and Theorem 2 in [59], convergence is ensured for the proposed Alg. 3.

The existence of Pareto-efficient solutions in the output of Alg. 3 follows from Proposition 1 and from Proposition 6 in the Appendix. By using Proposition 1, Alg. 2 populates $\mathcal{S}$ with Pareto-efficient solutions in step 3 of Alg. 3. The solution update steps 3a) and 3b) in Alg. 3 ensures that the solution set $\mathcal{S}$ is never empty.

Let us assume that there is a solution $S' \in \mathcal{S}$ that is not Pareto efficient. This implies that there exists another feasible clustering S that satisfies (18b), i.e., $\Gamma_i(S) \leq \Gamma_i(S')$, $i = 1, 2, 3$, with strict inequality for at least one criteria. In such a case, Alg. 3 in step 3b) will include S and exclude S' from the Pareto solution set $\mathcal{S}$. Since, S dominates S', the algorithm does not give S' as a solution as its output. Hence, the assumption $S' \in \mathcal{S}$ is refuted. Thus, the algorithm ensures that clustering set $\mathcal{S}$ only includes the Pareto-efficient solutions.

Overall, the total number of partitions (possible clustering solutions) of N users into M clusters is given by the Stirling number of second kind, η_N^M.

$$\eta_N^M = \frac{1}{M!}\sum_{i=0}^{M}(-1)^{M-i}\begin{pmatrix} M \\ i \end{pmatrix} i^N \tag{24}$$

However, using the distance metric (d^2) when the users are allocated to a cluster, the number of partitions (clustering solutions) are upper bounded by $\kappa = \begin{pmatrix} N \\ M \end{pmatrix}$. This effectively reduces the number of iterations executed in the multi-criteria optimization algorithm for multimedia broadcast (Alg. 3) over the feasible clusters only, making it more efficient than an extensive search over the η_N^M clustering space. The complexity of the multi-criteria optimization algorithm is $\mathcal{O}(\kappa)$.

The function, $f_t(\Gamma_1, \Gamma_2, \Gamma_3) = \min\limits_{S\in\phi_t} \sum\limits_{i=1}^{3} w_i\Gamma_i$(S), represents the solution to the multi-criteria problem $(\phi_t, \Gamma_1, \Gamma_2, \Gamma_3)$ at iteration t of the function given in Alg. 3. By definition, $0 \leq \Gamma_i \leq 1, \forall i \in \{1, 2, 3\}$. The Pareto efficient solution is given by function f_* achieved at iteration t_*. The solution update in step 3 of Alg. 3 ensures that $f_*(\Gamma_1, \Gamma_2, \Gamma_3) \leq f_{(t+1)}(\Gamma_1, \Gamma_2, \Gamma_3) \leq$

*The Pareto set created using the iterative cluster and Update algorithm is of bounded size, which guarantees convergence.

$f_t(\Gamma_1, \Gamma_2, \Gamma_3) \leq 1$. Since, $\frac{(f_{(t+1)} - f_*)}{(f_t - f_*)} < 1$ for $1 \leq t < t_* \leq \kappa$, the convergence rate is linear. □

Proposition 3. *The clustering and resource allocation solution that is obtained using the proposed algorithm is fair (unbiased) in terms of performance criteria Γ_1, Γ_2, and Γ_3, with respect to each program (content), $p \in [1, P]$.*

Proof: Let us assume that the Pareto-efficient solution of the algorithm is not fair (i.e., biased) towards a particular content p. This implies $\rho_{p,m}^{(v)} >> \rho_{p',m}^{(v)}$, in terms of resource allocation at a tier-level v. In terms of $D_{p,m}$ (defined by (21)), it implies that $D_{p,m} << D_{p',m}$. In terms of $\widetilde{Q}_p$ (defined by (23)), it implies $\widetilde{Q}_p \gg \widetilde{Q}_{p'}$, $p, p' \in [1, P], p \neq p'$, and both programs, p and p', are being broadcast in SFN area m (i.e., $\mathbb{1}_{p,m} = \mathbb{1}_{p',m} = 1$).

The multi-criteria clustering finds the Pareto-efficient solution S such that $\Gamma_i(S) < \Gamma_i(S')$, $\Gamma_j(S) \leq \Gamma_j(S')$, $1 \leq i, j \leq 3, i \neq j$. Γ_1 is minimum when $\rho_{p,m}^{(v)}$ are equal $\forall p \in [1, P]$. Hence, the Pareto-efficient solution cannot be biased, i.e., the biased condition with $\rho_{p,m}^{(v)} \gg \rho_{p',m}^{(v)}$ does not hold for criteria Γ_1. Γ_2 is minimum when $D_{p,m} = 0, \forall p \in [1, P]$. Hence, the biased condition $D_{p,m} \ll D_{p',m}$ does not hold for Pareto-efficient solution with respect to criterion Γ_2. The minimum of Γ_3 is equal to 0 when $\widetilde{Q}_p, \forall p \in [1, P]$ are equal. It follows that the biased condition $\widetilde{Q}_p \gg \widetilde{Q}_{p'}$ is not valid for criterion Γ_3.

The solution that is Pareto efficient excludes the biased cases making the solution essentially fair. Hence, the algorithm maintains fairness among the various contents using an iterative execution. □

Proposition 4. *The solution obtained using the proposed algorithm is fair with respect to each of the multiple criteria Γ_1, Γ_2, and Γ_3.*

Proof: The $\mathcal{S}$ set (after step 3 in Alg. 3) includes the Pareto-efficient clustering solutions that are in accordance with Proposition 1 and with Proposition 6 in the Appendix. These clustering solutions are a subset of the Pareto front [59][ii]. We prove that the solution S^* that minimizes $\max_{1 \leq i \leq 3} (w_i \Gamma_i(S^*))$ is fair with respect to each criterion, by using contradiction. [iii]

Let us assume a Pareto-efficient solution S is biased towards criterion Γ_i, i.e., $\Gamma_i(S) << \Gamma_j(S), 1 \leq i, j \leq 3, i \neq j$. Then, step 4) in Alg. 3 would minimize $w_j \Gamma_j(S)$. Hence, minimization of $w_j \Gamma_j(S)$ in step 4) of Alg. 3 will happen by opting for a Pareto-efficient solution S^* that ensures $\Gamma_i(S^*) \approx \Gamma_j(S^*)$. This would rule out all the Pareto-efficient solutions that are biased towards any one criterion. Consequently, the Pareto-efficient solution S^* that is fair to each criterion will be selected as a result of Step 4) in Alg. 3. □

V. Simulation Framework and Results

Here we introduce the system scenario we simulated and the methodology that we adopted. We then show the results obtained in both a small and a large scale scenario. Importantly, in the small-scale scenario, we compare the performance of our EMB scheme against the optimum, showing an excellent matching between the two.

[iii]The set of Pareto optimal outcomes, $\mathcal{S}$.

A. Test videos and variation of rate and QoE for SHVC levels

SHVC content consists of layers; a subset of the received layers determines the user QoE level. Fig. 3 shows spatial resolution based UE categorization and temporal scalable SHVC layers for each of such UEs. The considered resolution category for the UEs are: Full HD (1080p, i.e., 1920×1080 resolution), Standard HD (720p, i.e., 1280×720 resolution), 4CIF (i.e., 704×480 resolution), and CIF (i.e., 352×240 resolution) . Also, we consider three temporal frame rates, namely, 50, 25, and 12.5 frames per second (fps). To test the overall performance of the EMB framework, the video content was adaptively encoded using SHVC Test Model 4 (SHM 4) software [60]. SHVC already includes the spatial and temporal scalability in the form of layers (shown in Fig. 3).

Three sample HDTV programs are used, with each of the program's video (Town, Tree, and Ducks) having different spatial and temporal variance. Snapshots of the three video sequences along with their spatial perceptual information (SI) and temporal perceptual information (TI) measures are shown in Fig. 6(a). Spatial and temporal variance of a video sequence is obtained using the SI and TI metric, respectively, that are defined in ITU-T P.910 [61]. SI is the maximum standard deviation over all pixels of a Sobel filtered frame in a video sequence. TI is the maximum standard deviation over space of the motion difference feature (difference between the pixel values in successive frames). The spatial resolution and temporal frame rates of these video sequences are according to the spatio-temporal grid shown in Fig. 3.

The variation of rate and QoE of the Town video is shown in Fig. 7, based on encoding using SHM software and subjective video quality assessment using absolute category rating method [61], respectively. The rate and QoE model (discussed in Section III-C) is in accordance with the parametric modelling for SHVC in [62]. Since, the data rate and QoE decrease with increase in q (for all the spatial resolutions and frame rates), our scheme balances the tradeoff between these quantities. This is achieved by obtaining the quantization level $q_{p,m}$ that meets the constraints given by (17).

B. Performance evaluation in a small-scale scenario

We compare the performance of the EMB scheme against the optimum obtained by solving the problem in Section III-D through the solver for constrained nonlinear minimization in optimization tool of Matlab. We consider a small synchronization area consisting of two and three macro-cells (i.e., $B = 2$ and $B = 3$), as shown in Figs. 6(b) and 6(c), respectively. Results are obtained as the number of users in each cell (N_i, $N = \sum_{i=1}^{B} N_i$) increases, and for two and three broadcast contents, i.e., $P = 3$ and $P = 2$. The users are uniformly distributed in the B cells, the SINR is computed based on LTE system parameters, and all the users randomly request content from their BS. Clearly, in each scenario, there are $2^B - 1$ possible SFNs.

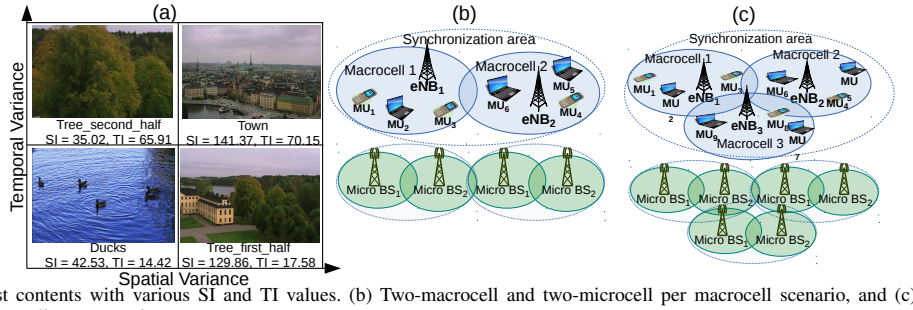


Fig. 6. (a) Broadcast contents with various SI and TI values. (b) Two-macrocell and two-microcell per macrocell scenario, and (c) Three-macrocell and two-microcell per macrocell test scenario.

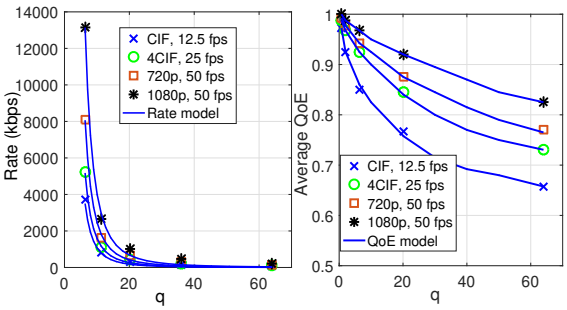
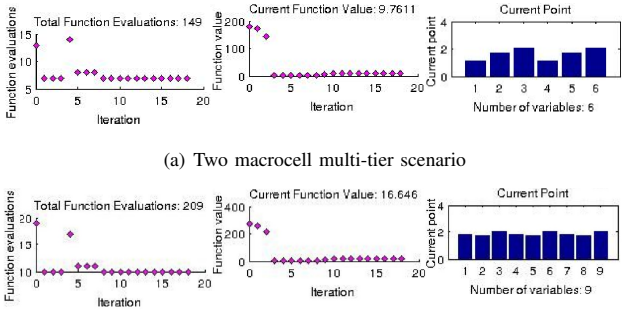


Fig. 7. Rate and QoE of 'Town' SHVC video, for various scalability levels.



(a) Two macrocell multi-tier scenario

(b) Three macrocell multi-tier scenario

Fig. 8. Optimization output for multi-tier scenarios in Figs. 6(b) and (c)

We would like to emphasize that the number of variables in the optimization problem in (14)-(17) is determined by the number of BSs (or cells) and of programs being broadcast by each BS. An increased number of optimization variables considerably increases the convergence and iteration execution time of the optimal solution. As an example, Figures 8(a) and 8(b) show the output of the optimization tool for a two-cell and a three-cell multi-tier scenario (depicted in Figures 6(b) and 6(c)), respectively, while broadcasting three programs to 20 users. The two-cell scenario has six variables, requires 149 function evaluations, and gives 9.76 resource blocks as output; the execution took 1.25 minutes. The three-cell scenario has nine variables, requires 209 function evaluations, and gives 16.64 as output; the execution took 4.18 minutes.

The resource (average number of resource blocks allocated for broadcast) and QoE performance for the two and three-cell scenario is shown in Figs. 9 and 10, respectively, as the number

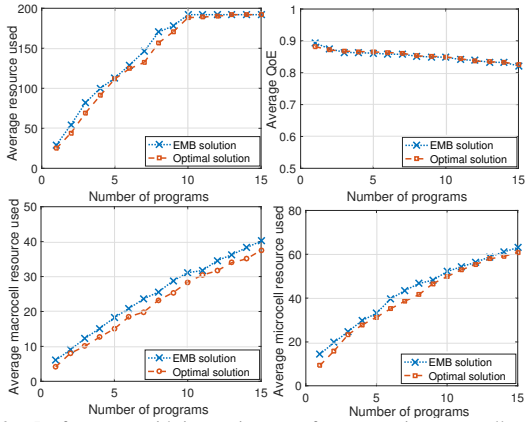


Fig. 9. Performance with increasing no. of programs in a two-cell scenario

of programs increases, and in Figs. 11 and 12, respectively, as the number of users increases. The average QoE (optimal and heuristic) decreases with increase in the number of users per cell and number of programs for both scenarios, while the number of resource blocks (optimal and heuristic) required per cell per content increases with increases. Observe that all the available resource blocks get used for the broadcast service as the number of programs increases (more than 5 and 10 programs in three-cell and two-cell scenario, respectively). This is evident from the saturation in the average resource used shown in Figs. 9 and 10, respectively. As the number of users increases, more resource blocks are required to broadcast programs to the increased number of low SINR users, shown in Figs. 11 and 12. Saturation is reached when all resource blocks are being used for broadcasting content in such scenarios. Importantly, Figs. 9-12 highlight that the EMB solution closely matches the optimum: the root mean square error (RMSE) and Pearson correlation coefficient (PC) between the EMB and the optimal solution are 0.12% and 0.99 (on average), respectively.

C. Performance evaluation in large-scale network scenarios

We now evaluate the performance of our EMB solution for a large multi-tier heterogeneous network with 450 macrocells, depicted in Fig. 13. The simulation parameters are listed in Table III. The wireless channel for the multi-tier LTE-A scenario is modeled with Gaussian fading distribution, log-normal shadowing, and the free-space path loss model [63]–[65]. We have simulated multiple instances ($\geq$ 100) of network

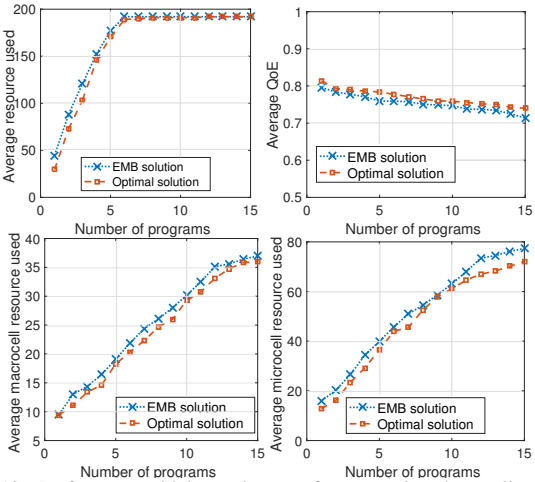


Fig. 10. Performance with increasing no. of programs in a three-cell scenario

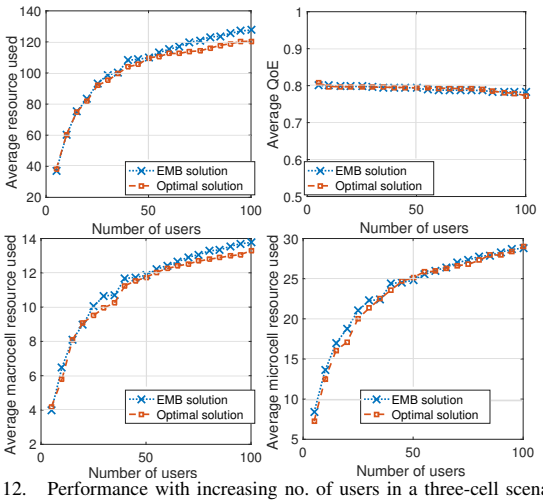


Fig. 12. Performance with increasing no. of users in a three-cell scenario

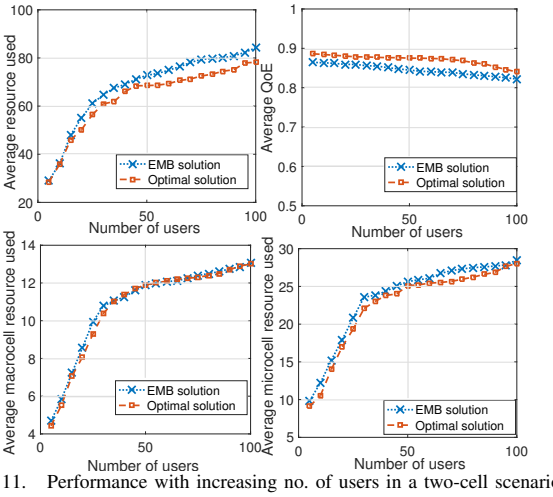


Fig. 11. Performance with increasing no. of users in a two-cell scenario

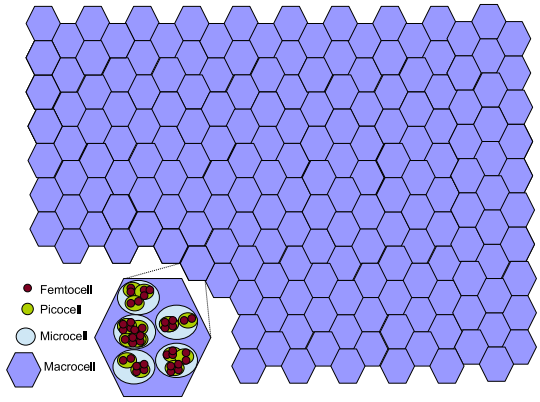


Fig. 13. Large multi-tier heterogeneous network scenario.

scenarios (with confidence interval of 95%) with random deployment of heterogeneous users over the MBMS area. The ratio of users in only macro-cell, macro+micro-cell, and macro+micro+pico or femto-cell coverage is considered to be 20:30:50. The femto and pico-cell coverage in the MBMS area is considered to be 50-75%.

The performance of the proposed EMB framework for an increasing number of UEs per macrocell is presented in Fig. 14 in terms of churn rate, QoE, throughput per macrocell, and criterion $(\Gamma_1, \Gamma_2, \Gamma_3)$. In particular, Fig. 14 shows that the churn rate is less than 7%, QoE greater than 0.75 (above 'Very good' subjective video quality), and fairness (based on Jain's fairness index) is greater than 0.75 with 32.41% (on average) resources used for up to 15 programs broadcast using the EMB framework, and delivered to 1000 users per macrocell.

The performance of the proposed EMB framework is compared to the Single-tier Conventional [41] and the Layered MBMS [45], [46] scheme, in Fig. 15. We have simulated multiple instances ($\geq$ 100) of network scenarios (with confidence interval of 95%) with random deployment of 500-1000 heterogeneous users per macrocell within the MBMS area.

In comparison to the benchmark schemes, with increasing number of programs, EMB results in lower churn rate (on average 80.43% and 65.46%, respectively), higher QoE (on average 99.31% and 45.73%, respectively), and increased throughput (on average 94.22% and 61.98%, respectively) with a reduced proportion of resource usage (on average 65.84%

TABLE III
SIMULATION PARAMETERS

Parameter	Value
Channel bandwidth	10 MHz
Frequency	1.8 GHz
Carrier spacing	15 KHz
Transmission mode	FDD
Number of data carriers	1200
Receiver noise figure	7 dB
Maximum transmitter output power (macrocell)	46 dBm
Maximum transmitter output power (macrocell)	36 dBm
Maximum transmitter output power (femtocell)	12-17 dBm
Maximum transmitter output power (picocell)	20-28 dBm
Transmitter cable and connector loss	2.0 dB
Transmitter power splitter loss	3.0 dB
Transmitter antenna gain	18 dBi
Receiver antenna gain	0 dBi
Additional losses (e.g., building, wall)	14.0-10.0 dB
Receiver noise floor	-97.5 dBm
Receiver sensitivity	-106.4 dBm
Shadowing standard deviation	8 dB
Guard interval	16.67μs

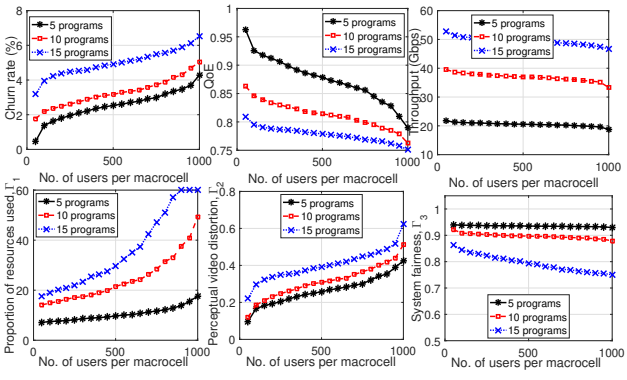


Fig. 14. Performance of EMB with increasing no. of users in a large scenario

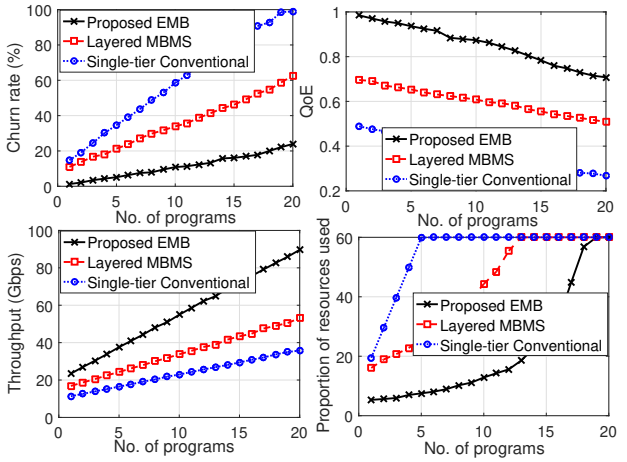


Fig. 15. Performance comparison of EMB to Layered MBMS and Single-tier Conventional, with increasing number of programs in a large scenario.

and 47.20%, respectively).

We have further studied the performance of EMB framework in urban, suburban, and rural scenarios, with respect to the two state-of-the-art schemes. The ratio of users in macro, macro+micro, and macro+micro+pico or femto cell coverage range for urban, suburban, and rural areas are 25:25:50, 25:50:25, and 50:25:25 respectively. The femto and pico-cell coverage for urban, suburban, and rural areas is considered to be 60-75%, 35-50%, and 15-25%, respectively. The corresponding results in terms of churn rate, QoE, throughput, and proportion of resources used, are noted in Table IV. EMB performs better in terms of all performance metrics than the Single-tier Conventional scheme (on average 73.02%, 84.27%, 79.45%, and 52.34%, respectively) and Layered MBMS (on average 56.98%, 34.11%, 46.77%, and 35.67%, respectively), under all the considered scenarios.

VI. Conclusions

We introduced the EMB framework for efficient multimedia broadcast in multi-tier mobile networks. The scheme aims to perform multi-criteria clustering, adaptive encoding of scalable (layered) multimedia content, and efficient resource allocation to provide high-quality video broadcast services to heterogeneous users. The EMB framework includes a multi-criteria clustering algorithm, which yields a Pareto-efficient system setting, and a matching algorithm that efficiently allocates resource blocks and determines the modulation and coding scheme to be used. Our results, obtained in small-scale scenarios, show that EMB closely matches the the optimum (on average, 0.12% RMSE and 0.99 PC). Also, in large-scale scenarios the EMB framework significantly outperforms state-of-the-art schemes such as the Single-tier Conventional and Layered MBMS, in terms of churn rate (by 76.72% and 61.22%, respectively), higher QoE (by 91.79% and 39.92%, respectively), higher throughput (by 86.83% and 54.37%, respectively), and reduced resource requirement (by 59.09% and 41.43%, respectively).

The proposed EMB scheme can be extended to the case of multiple frequency bands and different wireless technologies at the various tier-levels, which represents an interesting direction for future research.

Acknowledgment

This work is supported by the European Commission through the H2020 project 5G-EVE (Project ID 815074), the Indian Department of Science and Technology under the grant no. DST/INSPIRE/04/2015/000793, and the IIT Kharagpur ISIRD grant no. IIT/SRIC/ECE/USW/2016-17.

References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2016-2021," *Cisco White Paper*, Mar. 2017.

[2] 3rd Generation Partnership Project, "Technical specification group services and system aspects; service requirements for the 5G system; stage 1 (rel. 16)," Valbonne, France, Doc. 3GPP TS 22.261 V16.4., Jun. 2018.

[3] Ericsson, "LTE Broadcast: a revenue enabler in the mobile media era," *Ericsson White Paper*, Feb. 2013.

[4] Expway, "LTE Broadcast business cases," *White Paper*, Jun. 2014.

[5] Y. L. Lee, T. C. Chuah, J. Loo, and A. Vinel, "Recent advances in radio resource management for heterogeneous LTE/LTE-A networks," *IEEE Commun. Surveys Tutorials*, vol. 16, no. 4, pp. 2142–2180, Oct. 2014.

[6] Y.-C. Wang and C.-A. Chuang, "Efficient eNB deployment strategy for heterogeneous cells in 4G LTE systems," *Computer Networks*, vol. 79, pp. 297 – 312, Mar. 2015.

[7] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, July 2003.

[8] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang, and T. Wiegand, "High Efficiency Video Coding (HEVC) text specification draft 10," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Geneva, Switzerland, Doc. JCTVC-L1003_v9, Jan. 2013.

[9] "High Efficiency Video Coding HEVC / H.265," 2013. [Online]. Available: http://www.vcodex.com/h265.html

[10] A. Urie, A. N. Rudrapatna, C. Raman, and J. M. Hanriot, "Evolved multimedia broadcast multicast service in LTE: An assessment of system performance under realistic radio network engineering conditions," *Bell Labs Technical Journal*, vol. 18, no. 2, pp. 57–76, Sep. 2013.

[11] I. AlQerm and B. Shihada, "Energy-efficient power allocation in multitier 5G networks using enhanced online learning," *IEEE Trans. on Vehicular Technology*, vol. 66, no. 12, pp. 11 086–11 097, Dec 2017.

[12] X. Ge, S. Tu, G. Mao, C. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Wireless Commun.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.

[13] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[14] *Scalable video coding (SVC)*, 2008 Ed. (ISO/IEC JTC1/SC29/WG11. N9560).

TABLE IV
EMB PERFORMANCE IN DIFFERENT SCENARIOS

Scenario	EMB gain with respect to Layered MBMS				EMB gain with respect to Single-tier Conventional			
	Churn rate	QoE	Throughput	Proportion of resources used	Churn rate	QoE	Throughput	Proportion of resources used
Urban	67.92%	42.01%	62.81%	49.64%	81.26%	98.12%	93.22%	60.42%
Suburban	54.33%	35.89%	45.51%	37.28%	73.55%	83.25%	79.35%	54.37%
Rural	48.71%	24.43%	31.99%	20.08%	64.27%	71.45%	65.79%	42.23%

[15] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[16] J. Nightingale, Q. Wang, and C. Grecos, "Scalable HEVC (SHVC)-based video stream adaptation in wireless networks," in *Proc. IEEE PIMRC*, London, UK, Sep. 2013, pp. 3573–3577.

[17] *Overview of the MPEG-4 Standard*, 2002 Ed. (ISO/IEC JTC1/SC29/WG11 N4668).

[18] 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 *IEEE INFOCOM*, Kowloon, Hong Kong, 2015.

[19] C. Borgiattino, C. Casetti, C. Chiasserini, and F. Malandrino, "Efficient area formation for LTE broadcasting," in *IEEE SECON*, Seattle, WA, June 2015.

[20] Y.-h. Xu, C.-O. Chow, M.-L. Tham, and H. Ishii, "An enhanced framework for providing multimedia broadcast/multicast service over heterogeneous networks," *Journal of Zhejiang University SCIENCE C*, vol. 15, no. 1, pp. 63–80, Jan 2014.

[21] E. Katranaras, M. Imran, and M. Dianati, "Energy-aware clustering for multi-cell joint transmission in LTE networks," in *IEEE ICC*, Budapest, Hungary, June 2013, pp. 419–424.

[22] O. Altrad and S. Muhaidat, "Load balancing based on clustering methods for LTE networks," *Cyber Journals: J. of Selected Areas in Telecommun.*, vol. 3, no. 2, pp. 1–6, Feb. 2013.

[23] A. de la Fuente Iglesias, R. P. Leal, and A. G. Armada, "Performance analysis of eMBMS in LTE: Dynamic MBSFN areas," in *OPNETWORK*, Washington DC, Aug. 2013, pp. 1–6.

[24] F. Folino and C. Pizzuti, "A multiobjective and evolutionary clustering method for dynamic networks," in *Intl. Conf. Advances in Social Netw. Analysis and Mining (ASONAM)*, Odense, Denmark, Aug. 2010, pp. 256–263.

[25] H. T. Cheng and W. Zhuang, "Pareto optimal resource management for wireless mesh networks with QoS assurance: Joint node clustering and subcarrier allocation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1573–1583, Mar. 2009.

[26] A. Bentaleb, A. Boubetra, and S. Harous, "Survey of clustering schemes in mobile ad hoc networks," *Commun. and Netw. J.*, vol. 5, no. 2, pp. 8–14, May 2013.

[27] A. Abdelnasser, E. Hossain, and D. I. Kim, "Clustering and resource allocation for dense femtocells in a two-tier cellular OFDMA network," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1628–1641, Mar. 2014.

[28] Z. Limani, C.-F. Chiasserini, and G. Dell'Aera, "Interference-aware resource scheduling in LTE HetNets with carrier aggregation support," in *IEEE ICC*, London, UK, June 2015, pp. 3137–3142.

[29] V. Sathya, A. Ramamurthy, and B. Tamma, "Joint placement and power control of LTE femto base stations in enterprise environments," in *Intl. Conf. Computing, Netw. and Commun. (ICNC)*, Garden Grove, CA, Feb 2015, pp. 1029–1033.

[30] H. Wang, H. Nguyen, C. Rosa, and K. Pedersen, "Uplink multi-cluster scheduling with MU-MIMO for LTE-advanced with carrier aggregation," in *IEEE WCNC*, Shanghai, China, Apr. 2012, pp. 1202–1206.

[31] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "Analyzing the video popularity characteristics of large-scale user generated content systems," *IEEE/ACM Trans. on Networking*, vol. 17, no. 5, pp. 1357–1370, Oct 2009.

[32] T. Qiu, Z. Ge, S. Lee, J. Wang, Q. Zhao, and J. Xu, "Modeling channel popularity dynamics in a large IPTV system," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 37, no. 1, pp. 275–286, Jun. 2009.

[33] H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng, "Understanding user behavior in large-scale video-on-demand systems," *SIGOPS Oper. Syst. Rev.*, vol. 40, no. 4, pp. 333–344, Apr. 2006.

[34] K. W. Hwang, D. Applegate, A. Archer, V. Gopalakrishnan, S. Lee, V. Misra, K. K. Ramakrishnan, and D. F. Swayne, "Leveraging video viewing patterns for optimal content placement," in *NETWORKING 2012*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 44–58.

[35] T. R. G. Nair and P. Jayarekha, "A rank based replacement policy for multimedia server cache using zipf-like law," *J. Comput.*, vol. 2, no. 3, Mar. 2010.

[36] Y. Gu, W. Saad, M. Bennis, M. Debbah, and Z. Han, "Matching theory for future wireless networks: fundamentals and applications," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 52–59, May 2015.

[37] O. Semiari, W. Saad, S. Valentin, M. Bennis, and H. V. Poor, "Context-aware small cell networks: How social metrics improve wireless resource allocation," *CoRR*, vol. abs/1505.04220, 2015.

[38] C. Karemera and J. Ngubiri, "Complexity of the resource allocation/matching problem with weight based ceilings," in *Algorithms and Architectures for Parallel Processing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 245–254.

[39] D. Applegate and W. J. Cook, "Solving large-scale matching problems," in *Proc. DIMACS Workshop Network Flows And Matching*, New Jersey, USA, Oct. 1991, pp. 557–576.

[40] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis, "Optimality of fast-matching algorithms for random networks with applications to structural controllability," *IEEE Trans. on Control of Network Systems*, vol. 4, no. 4, pp. 770–780, Dec. 2017.

[41] C. P. Lau, A. Alabbasi, and B. Shihada, "An efficient live TV scheduling system for 4G LTE broadcast," *IEEE Sys. J.*, vol. 11, no. 4, pp. 2737–2748, Dec 2017.

[42] A. Shokair, M. Crussiere, J. Helard, Y. Nasser, and O. Bazzi, "Mobile TV directed resource allocation scheme for LTE networks," in *Proc. Inter. Symp. WPMC*, Bali, Indonesia, Dec. 2017, pp. 241–246.

[43] A. Tassi, C. Khirallah, D. Vukobratovi, F. Chiti, J. S. Thompson, and R. Fantacci, "Resource allocation strategies for network-coded video broadcasting services over LTE-advanced," *IEEE Trans. on Vehicular Technology*, vol. 64, no. 5, pp. 2186–2192, May 2015.

[44] S. Almowuena and M. Hefeeda, "Mobile video streaming over dynamic single-frequency networks," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 12, no. 81, pp. 1–26, Nov. 2016.

[45] Y. Liu, R. Hou, K. Lui, and H. Li, "A multicast transmission scheme in small cell networks with wireless backhaul," in *Proc. IEEE GLOBECOM*, Singapore, Dec. 2017, pp. 1–6.

[46] M. Condoluci, G. Araniti, A. Molinaro, and A. Iera, "Multicast resource allocation enhanced by channel state feedbacks for multiple scalable video coding streams in LTE networks," *IEEE Trans. on Vehicular Technology*, vol. 65, no. 5, pp. 2907–2921, May 2016.

[47] 3rd Generation Partnership Project, *3GPP TS 36.331: Evolved universal terrestrial radio access (E-UTRA); Radio Resource Control (RRC); Protocol specification (Release 14)*, Sophia Antipolis, France, Mar. 2019.

[48] ——, "Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification (3GPP TS 36.331 version 13.12.0 release 13))," Sophia Antipolis, France, Tech. Rep. ETSI TS 136 331 V13.12.0 (2019-01)., Jan. 2019.

[49] L. Christodoulou, O. Abdul-Hameed, A. M. Kondoz, and J. Calic, "Adaptive subframe allocation for next generation multimedia delivery over hybrid LTE unicast broadcast," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 540–551, Sep. 2016.

[50] D. Han, D. Andersen, M. Kaminsky, D. Papagiannaki, and S. Seshan, "Hulu in the neighborhood," in *Proc. IEEE Int. Conf. on Commun. Systems and Networks*, Bangalore, India, Jan. 2011, pp. 1–10.

[51] Y. Wang, Z. Ma, and Y.-F. Qu, "Modeling rate and perceptual quality of scalable video as functions of quantization and frame rate and its application in scalable video adaptation," in *Intl. Packet Video Wksp.*, Seattle, WA, May 2009.

[52] S. Azad, W. Song, and D. Tjondronegoro, "Bit rate modeling of scalable videos using quantization parameter, frame rate and spatial resolution," in *IEEE ICASSP*, Dallas, TX, Mar. 2010.

[53] *3GPP TS 36.331 Evolved universal terrestrial radio access (E-UTRA) (Release 13)*, 2015.

[54] R. Jain, D. M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *Technical Report, Digital Equipment Corporation*, vol. TR-301, 1984.

[55] EBU Operating Eurovision and Euroradio, "Delivery of broadcast content over LTE networks," Geneva, Tech. Rep. TR 027, Jul. 2014.
[56] 3rd Generation Partnership Project, "LTE; evolved universal terrestrial radio access network (E-UTRAN); M2 application protocol (M2AP) (3GPP TS 36.443 version 11.2.0 release 11))," Sophia Antipolis, France, Tech. Rep. ETSI TS 136 443 V11.2.0 (2013-04), Apr. 2013.
[57] F. A. and B. V., "Direct multicriteria clustering algorithms," *J. of Classification*, vol. 9, no. 1, pp. 43–61, 1992.
[58] S. Ayramo and T. Karkkainen, "Introduction to partitioning-based clustering methods with a robust example," *Reports of the Dept. of Math. Inf. Tech.*, vol. 1, 2006.
[59] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multiobjective optimization," *Evol. Comput.*, vol. 10, no. 3, pp. 263–282, Sep. 2002.
[60] J. Chen, J. Boyce, Y. Ye, and M. M. Hannuksela, "SHVC Test Model 4 (SHM 4)," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Geneva, Switzerland, Doc. JCTVC-O1007, Oct. 2013.
[61] *Subjective video quality assessment methods for multimedia applications*, ITU-T Recommendation, P.910, Apr. 2008.
[62] C. Singhal and S. De, "UE-TV: User-centric energy-efficient HDTV broadcast over lte and wi-fi," *IEEE Trans. on Mobile Computing*, vol. pp., no. 1, pp. 1–15, 2018.
[63] S. Sesia, I. Toufik, and M. Baker, *LTE UMTS Long Term Evolution, From Theory To Practice*. John Wiley and Sons, Ltd., 2009.
[64] J. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of LTE networks," in *Proc. IEEE VTC Spring*, May 2010, pp. 1–5.
[65] G. Araniti, M. Condoluci, L. Militano, and A. Iera, "Adaptive resource allocation to multicast services in LTE systems," *IEEE Trans. Broadcasting*, vol. 59, no. 4, pp. 658–664, Dec. 2013.

APPENDIX

Proposition 5. *Definitions and assumptions of our heuristic solution are in coherence with those in [57]–[59], given as follows:*

1) *Multicriteria clustering problem is modeled as multi-criteria optimization problem [57]*
2) *Iterative clustering algorithm is used for solving the discrete multi-criteria optimization problem [57], [58]*

Proof: The multi-criteria optimization problem for the criteria: $\Gamma_1, \Gamma_2, \Gamma_3$ is solved using heuristic multi-criteria clustering algorithm, as given in Algorithms 1 and 2. A single criterion, $\min_{S\in\phi} \sum_{i=1}^{3} w_i \Gamma_i(S)$, is derived from the multi-criteria and is used to obtain Pareto efficient solutions, as mentioned in Alg. 2. Modified relocation iterative clustering algorithm (Alg. 3) has been used for obtaining pareto-efficient solution of the multi-criteria optimization problem. □

Proposition 6. *By rearranging the results in [57], we can prove the following. If, $S^* \in \phi$ is the unique optimal clustering solution to criterion Γ_i, $i = 1, 2,$ or, 3. Then S^* is also a Pareto-efficient solution of the multi-criteria optimization problem* $(\phi, \Gamma_1, \Gamma_2, \Gamma_3)$.

Proof: We use contradiction to prove this. Let us assume that $S^* \in \phi$ is not a Pareto-efficient solution of the multi-criteria optimization problem $(\phi, \Gamma_1, \Gamma_2, \Gamma_3)$. Instead, there exists a solution S $(S \neq S^*)$, i.e. Pareto efficient, such that $\Gamma_1(S) < \Gamma_1(S^*)$ and it holds that $\Gamma_2(S) \leq \Gamma_2(S^*)$ and $\Gamma_3(S) \leq \Gamma_3(S^*)$. Since, S^* is a unique minimal solution for the criterion Γ_1. Therefore, $\Gamma_1(S) = \Gamma_1(S^*)$, holds true. This results in the contradiction and hence proves that $S^* \in \phi$ indeed is the Pareto-efficient solution of $(\phi, \Gamma_1, \Gamma_2, \Gamma_3)$. Same rationale holds true for other criterion Γ_2 and Γ_3. □

Proposition 7. *The modified relocation algorithm is similar to the iterative multi-criteria cluster and update algorithm [59].*

Proof: An iterative multi-criteria cluster and update is essentially an iterative multi-objective search algorithm. It comprises of two functions: **generate** cluster S and **update** pareto efficient solution set $\mathcal{S}$ if S dominates $S', S' \in \mathcal{S}$. The representation of the algorithm is shown in Fig. 16.

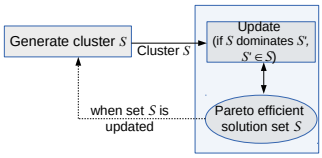


Fig. 16. Representation of iterative multi-objective search algorithm for multi-criteria Pareto-efficient clustering

The proposed algorithm in this paper has the following properties that are same as in [59]:

1) Single clustering instance is considered at each iteration.
2) A finite memory solution set, $\mathcal{S}$, is maintained that stores only the dominant solution till a specified iteration.
3) $\mathcal{S}$ is independent set of the iteration count. It consists of the representative subset of the best clustering solutions (pareto efficient) received till any given iteration.

Hence, the modified relocation algorithm proposed in this paper is similar to the iterative algorithm given in [59] and the proofs in [59] can be applied to our algorithm. □

BIOGRAPHIES

Chetna Singhal (S'13-M'15) received B.Eng. in electronics and telecommunications from University of Pune, India, in 2008, MTech and PhD from Indian Institute of Technology (IIT) Delhi, in 2010 and 2015, respectively. She has worked in IBM Software Lab as software engineer in 2010. She is currently an assistant professor with the Department of Electronics and Electrical Communication Engineering, IIT Kharagpur, since 2015. Her research interests are in next generation heterogeneous wireless networks, with emphasis on cross-layer optimization, adaptive multimedia services, energy efficiency, and resource allocation.

Carla Fabiana Chiasserini (M'98, SM'09, F'18) received her Ph.D. from Politecnico di Torino, Italy, in 2000. She worked as a visiting researcher at UCSD in 1998–2003, and as a Visiting Professor at Monash University in 2012 and 2016. She is currently a Professor with the Department of Electronic Engineering and Telecommunications at Politecnico di Torino. Her research interests include architectures, protocols, and performance analysis of wireless networks. Dr. Chiasserini has published over 300 papers in prestigious journals and leading international conferences.

Claudio Casetti (M'05) received his PhD in Electronic Engineering from the same institution in 1997. He has been a visiting researcher at the University of California, Los Angeles and at the Monash University. He is an Associate Professor at the Dipartimento di Elettronica of Politecnico di Torino. He has coauthored more than 200 journal and conference papers in the fields of networking and holds three patents. His interests focus on 5G, vehicular networks, and mobile services.