A Unified Approach for Efficient Delivery of Unicast and Multicast Wireless Video Services

S. Pizzi*, M. Condoluci†, G. Araniti*, A. Molinaro*, A. Iera*, and G.-M. Muntean‡

*DIIES Dept., University Mediterranea of Reggio Calabria, Italy,
e-mail: {*sara.pizzi, araniti, antonella.molinaro, antonio.iera*}*@unirc.it*
†Department of Informatics, King's College London, UK,
email: *massimo.condoluci@kcl.ac.uk*
‡School of Electronic Engineering, Dublin City University, Ireland,
email: *gabriel.muntean@dcu.ie*

Abstract—Recently, mobile multimedia services have represented an increasingly large source of revenue for the telecommunication industry. Subscribers are often interested in simultaneously receiving the same data flow and, hence, they fuel the growing demand of multicast multimedia services. Therefore, the design of efficient radio resource management strategies to jointly handle multicast and more traditional unicast traffic and to increase user satisfaction is of primary importance for the successful deployment of future mobile networks. This paper proposes an efficient radio resource management framework to handle unicast and multicast multi-layer video services. The proposed *Virtual Unified Group* (VUG) approach makes use of a channel-aware subgrouping principle to provide fair throughput to both unicast and multicast users. The idea is to assign unicast subscribers to a *virtual* group thus allowing them to compete for network resources on an equal footing with multicast users. Simulations highlight the benefits that *VUG* provides in a Long Term Evolution (LTE) network scenario, under different traffic load conditions, in terms of throughput and fairness in the distribution of resources to unicast and multicast users.

Index Terms—Unicast, Multicast, Resource Allocation, Subgrouping, LTE, OFDMA

I. INTRODUCTION

Future mobile networks will have to provide effective support to the increasing demand of diverse wireless multimedia services with a multicast/broadcast nature, including mobile TV, IP radio broadcasting, and video streaming so as to maintain high levels of quality for these services. It is expected that video will account for over two-thirds of the mobile data traffic of the world by 2018 [1]; thus, Telco operators are called to manage massive number of requests for multicast video services alongside traditional unicast traffic, which includes voice, video, web browsing, file transfers, and so on. On the one hand, this will be a big source of revenue for service providers and network operators; on the other hand, it becomes important to ensure that network resources are efficiently allocated to secure high user satisfaction and efficient spectrum utilization while coping with radio channel impairments.

That the joint management of unicast and multicast traffic is a key design issue of future mobile networks is demonstrated by the intense research activity conducted by standardization (e.g., the 3rd Generation Partnership Project, 3GPP) and academic bodies. Recently, also the 5G-PPP (5G Infrastructure Public Private Partnership) committee has indicated the full integration between emerging mobile group-based (i.e., multicasting) and more traditional unicast communications as one of the major research priorities in next-to-come 5G networks [2]. The main challenge in achieving this targeted goal is in terms of resource allocation, due to the use of different transmission modes for unicast and multicast services. Indeed, in case of unicast services, data is delivered to a single user; so the modulation and coding scheme (MCS) is selected according to *individual* channel conditions between base station (BS) and user equipment (UE). Differently, in the multicast case, transmission parameters are selected by the BS on a *per-group* basis. Multicast group members could measure different radio link qualities and support different MCSs: poor links need robust MCSs to face hostile propagation conditions, while users with good channels could receive data at higher bit rates. However, the multicast group member with the worst channel condition imposes the MCS selection for the entire group, and thus it becomes the bottleneck of the whole group; this severely affects the transmission efficiency. In general, favoring multicast traffic permits a provider to achieve a high overall system capacity (measured as the number of users simultaneously served by the BS). This is at the expense of an unfair bandwidth distribution between multicast and unicast traffic with negative consequences on unicast users' degree of satisfaction. Vice-versa, by advantaging unicast user requests, the network capacity could significantly decrease, which is undesirable. This unbalance is due to the fact that multicast group(s) and unicast users have different sizes. It is clear that for the best balance between service quality and efficiency, there is a need for a combined unicast and multicast traffic management solution. This paper proposes a novel heuristic radio resource management policy, called *Virtual Unified Group* (VUG), which achieves fair resource distribution between unicast and multicast traffic. The proposed *VUG* approach is to be deployed at the level of BS and can be efficiently utilized for the management of multi-layer video streaming services. The main idea behind *VUG* is to clusterize unicast destinations into one or more *virtual groups* that are managed by the BS like multicast groups. Unicast users can be interested in different data, but they are *virtually* assigned to a group to compete for resource distribution on an

equal footing with multicast traffic. A subgrouping strategy is then applied to all groups (both real and virtual) that aggregate subsets of users with similar channel quality levels.

This paper extends an early version of the proposed *VUG* approach in [3] by: *(i)* introducing subgrouping techniques for the finer management of the channel state information (CSI) granularity in the groups, *(ii)* presenting the proposed idea through an algorithmic approach and presenting the details of the *VUG* operation in a step-wise manner, and *(iii)* assessing the performance of *VUG* in LTE systems under different traffic load conditions in terms of throughput and fairness in the distribution of resources to unicast and multicast users.

The remainder of the paper is organized as it follows. In section II, an overview of the research activity related to unicast and multicast traffic management and resource assignment is presented. The proposed *VUG* policy for concurrently handling unicast and multicast video services is described in details in section III. The simulation-based testing and result analysis are discussed in section IV, and the conclusive remarks are given in the last section.

II. RELATED WORK

This section presents a non-exhaustive review of current proposals in the literature that are related to the proposed *VUG* resource management policy. These address: *(i)* multicast support in broadband wireless access networks, *(ii)* multiple multicast group management, and *(iii)* mixed unicast and multicast traffic delivery.

A. Multicast Support in Broadband Wireless Access Networks

Various research groups have focused their attention on providing solutions to support multicast content delivery in wireless networks. For instance, the work described in [4] enables multiple destinations to be served by the same multicast group through a point-to-multipoint (PMP) transmission. This approach represents an effective solution to efficiently exploit the broadcast nature of the radio channel. However, the main issue relevant to PMP approaches is that the link adaptation procedures, i.e., the selection of transmission parameters such as MCS, are performed on a *per-group basis* [5]. As a consequence, the mere presence of cell-edge users forces the BS to use more robust MCSs to guarantee an error-free reception to all multicast receivers; therefore, the performance of the UEs with a high channel quality, which could potentially attain very high data rates, is dramatically decreased. This adversely affects the overall system spectral efficiency, due to the use of less efficient MCSs. The Conventional Multicast Scheme (CMS) [6] implements the link adaptation procedures by adopting a conservative approach, which assigns the group data rate based on the user that experiences the *worst* channel conditions. Although this approach guarantees the highest fairness (i.e., all multicast members are served at the same data rate), it introduces severe spectrum management inefficiencies, which is not fully exploited with a consequent poor performance in terms of spectral efficiency.

Several solutions have been proposed to overcome the limitations of the CMS approach. For example, an opportunistic scheme introduced in [7] selects for transmission only the "best" subset of multicast members, to maximize the Quality of Service (QoS) levels. A similar approach is followed in [8], where a cooperative transmission model composed of two steps is presented. In the first phase, the BS multicasts data at a high rate by exploiting an opportunistic scheme, whereas in the second phase users with good channel conditions relay the received data to other users. Although opportunistic multicasting introduces several improvements for high quality users, it may require the adoption of additional data coding, i.e, rateless codes, to make the service work properly (due to the fact that the subset of receivers selected could change slot by slot). This approach also introduces several performance issues to the cell-edge subscribers which will experience poor throughput and increased delivery latency [7]. A different solution designed for concurrently leveraging the benefits of unicast (i.e., low energy consumption) and multicast (i.e., reduced network load) is foreseen in [9], where the authors propose an hybrid approach that assigns the available blocks to individual videos, decides whether to use multicast or unicast, and determines the MCSs of individual blocks, in order to maximize the overall energy saving while guaranteeing smooth playout. Other approaches like random linear network coding [10] have been proposed with the aim of optimizing the transmission scheme and minimizing the number of broadcast packets on each downlink channel.

An interesting approach that aims to enhance the radio channel utilization in multicast environments by exploiting multi-user diversity is *subgrouping* [11], [12], [13]. The main idea is to split multicast users into several subgroups, where each subgroup collects the subscribers that experience similar channel quality levels. The goal of subgrouping is to improve both system capacity and session quality since the negative effects caused by the presence of cell-edge users is reduced and the whole multicast group is served in each scheduling frame with more efficient MCSs compared to CMS.

B. Multiple Multicast Group Management

Several works in the literature address the problem of the efficient delivery of multimedia streaming content to different multicast groups simultaneously when served by a given BS [14], [15], [16], [17]. Such works consider multi-layer services where each video stream (or program) is encoded into different layers. The focus is the selection, for any resource within a given scheduling frame, of which layer of which multicast group to transmit at what MCS. In order to reach this goal, the number of receivers per layer has to be dynamically adjusted according to the experienced channel conditions and the available network bandwidth, so as to maximize a given utility function, which is rather complex. The authors of [14] focus on finding a utility-based resource allocation scheme for layer-encoded IPTV multicast streaming services over IEEE 802.16 WiMAX networks. Besides, they prove that the problem of finding the best subset of users to serve per each layer for each multicast group is NP-hard, and propose an approach that can run in polynomial time. The works reported in [15] and [16] follow an approach similar to that

described in [14], by proposing near-optimal algorithms for subgroup management. In particular, authors in [15] present a fast greedy algorithm aimed at achieving a proportional fair resource allocation, which is probably within a constant approximation of the optimal solution (based on a metric that reflects video quality as perceived by the user according to the data rate experienced by the multicast members). The idea at the basis of [16] is to maximize the total system utility (defined as a generic non-negative and non-decreasing function of the received rate) through a two-step dynamic programming algorithm. With the aim to maximize the total system utility of all video multicast sessions, it is assumed that the base layer should not be received by all users, i.e., some video sessions could be dropped or some users could decode any layer. Finally, [17] extends the referred problem by defining an algorithm for transmitting multicast data in bursts to save the energy of mobile receivers with promising results.

The main issue related to these works is the presence of real-time services, rather than the more realistic case wherein such applications co-exist with unicast services; this dictates for novel solutions to simultaneously manage group- and user-based link adaptation procedures.

C. Mixed Unicast and Multicast Traffic Delivery

The future mobile communication systems are expected to manage both multicast and unicast traffic [2]. However, the design of efficient algorithms for the delivery of mixed unicast and multicast traffic is still an open research field. The authors of [18] assess the joint performance of unicast and multicast services in LTE. In particular, they consider the transmission of both streaming and file delivery services when the Multicast and Broadcast Single Frequency Network (MBSFN) operational mode is selected to deliver streaming services. Results show that, for streaming services, increasing the MBSFN area size enhances the coverage and/or allows employing a higher MCS level at the expenses of a lower flexibility when adapting the content to different geographical areas. For file delivery services, increasing the MCS level reduces the file download time. However, in case of too-large MCS values, retransmissions may be required following potential errors that deteriorate the system performance. Once unicast and multicast users are jointly considered, the inclusion of the Enhanced Multimedia Broadcast Multicast Service (E-MBMS) in the same carrier affects the performance of unicast users. The performance of both unicast and multicast transmission schemes in terms of system capacity, worst average channel user capacity, and outage probability are evaluated and compared in [19] in different cell environments. Furthermore, the authors introduce a transmission scheme in a mixed-traffic environment. The proposed approach takes into account Signal-to-Noise Ratio (SNR) threshold values only, not strongly correlated with system capacity, and no minimum performance guarantees are assured to unicast or multicast services. A comparison between unicast and multicast transmissions is presented also in [20], where the authors have determined switching thresholds, as a function of the number of users per cell, to switch between unicast and multicast modality for download or streaming services.

When focusing on the literature dealing with joint management of unicast and multicast traffic, three main trends can be identified: *Unicast Maximization* (UM), *Equal Sharing* (ES), and *Equal Competition* (EC). The first trend prioritizes unicast traffic; the second one aims to equally share the available resources between unicast and multicast users; and the third one aims to maximize the number of conveyed bits to improve the spectrum utilization. Example of proposals which belong to the *UM* philosophy in multi-carrier Orthogonal Frequency Division Multiplexing (OFDM) systems are given in [21] and [22]. The main idea behind these solutions is to guarantee a minimal required data rate to all multicast users, based on the CMS logic, and then to assign the remaining resources to unicast users in order to maximize the throughput. As a consequence, an increased throughput is offered to unicast destinations at the expense of multicast receivers, which experience significant limitations in QoS. An example of *ES* strategy instantiation is available in [23], which presents a power-saving scheduling algorithm that manages mixed unicast and multicast traffic. According to the *ES* philosophy, multicast and unicast services equally share the network capacity; the resources to assign to the two types of traffic (unicast and multicast) are statically split into two equal sets. This policy prevents any traffic class to utilize the resources assigned to the other class. This seems fair, but its main inefficiency is due to the static bandwidth assignment that cannot adapt to the dynamic load variation of both unicast and multicast users, thus resulting in inefficient spectrum utilization and possible negative effects on the quality of one of the traffic classes. Finally, a technique that follows the *EC* logic and aims to guarantee the minimum data rate of both unicast and multicast flows is proposed in [24]. The extra resources are assigned according to a maximum throughput scheme, whiffch iteratively selects the service that conveys the highest number of bits to the destinations. Another example of the *EC* philosophy can be found in [35], which proposes a proportional fairness problem where unicast and multicast users equally compete to get the available resources (i.e., the available resources are assigned to either unicast or multicast users according to which ones are able to maximize the sum of logarithmic data rates). With the aim to maximize the system throughput, this technique intrinsically gives higher priority to multicast services, since multiple destinations can be satisfied with a single transmission. Therefore, multicast users have higher probability to be served with respect to unicast users, which may suffer from poor throughput performance.

The issues identified in relation to the existing solutions suggest the need for an efficient strategy to jointly manage multicast and unicast traffic, such as *VUG*, presented in section III. In addition, section IV introduces a new fairness index that captures the extent to which unicast and multicast users are equally satisfied, which is employed in the performance evaluation illustrated in section IV.

Even though a considerable body of work has focused on the support multicast services in broadband wireless access network, only a few works provide solutions for the concurrent management of both unicast and multicast traffic. This

topic is very important since these two kinds of traffic will definitely co-exist in future mobile communication systems, and their different characteristics must be taken into account to guarantee a fair resource distribution between them. For these reasons, we focus our attention on the design of an efficient strategy to jointly manage multicast and unicast traffic, such as *VUG*, presented in section III. In addition, section IV introduces a new fairness index, which captures the extent to which unicast and multicast users are equally satisfied and is employed during performance evaluation in section IV.

III. VUG - The Proposed Radio Resource Allocation Scheme

A single-BS coverage area (hereinafter referred to as a "cell") is considered, in which both unicast and multicast users receive real-time video streaming, e.g., IPTV, Internet video streaming, or live telecast. In our reference scenario, multi-layer video coding is employed. A video stream is split into a base layer and more enhancement layers [14]. Users can make use of the received video content once they get the base layer and they benefit from a quality of the video that increases with the number of layers they receive. Similarly to [15], this paper assumes that the video layers are synchronized and that data is grouped on a per-layer basis. In doing so, the bits relevant to a given video layer are managed by the packet scheduler as a single data unit; accordingly, the data unit corresponding to a given layer is scheduled only if the units associated to the preceding layers have already been scheduled.

In this scenario, the *VUG* algorithm for radio resource management is proposed to efficiently assign the downlink radio resources in the cell to both unicast and multicast users. The basic goal of this heuristic algorithm is to guarantee the delivery of the base layer to all the users (unicast and multicast) in the cell, while also trying to deliver to them as many additional layers as possible. We operate on a resource allocation frame basis. The number of video layers to be potentially scheduled is received by the base station as an input from the application layer. By taking into account multiple resource allocation frames, we could consider that the base station has a set of layers potentially delivered to the involved user(s), and this set is chosen by the source video server according to the buffering capabilities of the devices as well as the feedback received from the network in terms of delivered video layers. As a consequence, if the network informs the application that only the first layer has been delivered for several consecutive time windows, then the application could choose to avoid providing the further enhancement layers.

Deployed at the level of BS, the main idea behind *VUG* is to organize unicast users into one or more *virtual groups*[1] that are managed by the BS in a manner similar to the multicast groups. Groups are named "virtual" since their unicast members are not necessarily interested in the same type of data, like the multicast group members are. Since multicast users can be located in any position within the cell, unicast users are randomly associated to virtual groups in order to guarantee the

presence of both users with good and bad channel conditions in each virtual group. Instead, unicast users are grouped together based on similar user perceived channel quality in virtual subgroups; thus, users belonging to the same virtual subgroup homogeneous from the delivery conditions point of view. This guarantees a similar heterogeneous environment in terms of channel state inside both multicast and virtual groups. Furthermore, unicast users will likely request video with different bit rates. This further heterogeneity is taken into account during the enhancement layer allocation phase. This *VUG*-based resource management strategy allows unicast users to compete for network resource allocation with the multicast groups on an equal footing, which is highly beneficial for a fair distribution of resources, as it will be validated through simulations. The behavior of *VUG* is strictly related to the information that the BS gathers from all the UEs regarding the measured quality on the downlink. The scheduler located at the BS is, then, in charge of selecting the downlink resources to assign to each traffic flow with the related MCS. Specifically, it dynamically chooses the MCS most suitable to the currently perceived channel quality on each BS-to-UE link.

Summarizing, the main operations performed by the proposed *VUG* policy (i.e., organization of unicast users in virtual groups, and subgroups formation inside both multicast and virtual groups) are illustrated in Fig. 1.
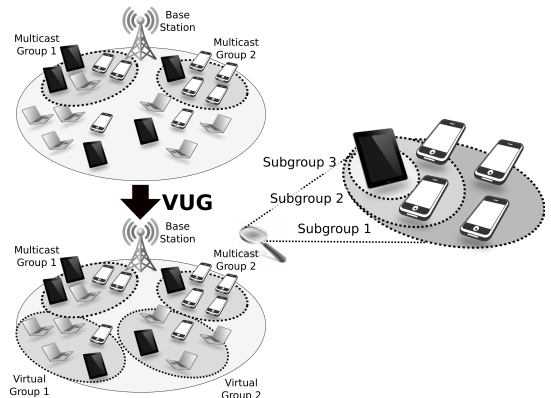

Fig. 1. Operation of the *VUG* technique.

In the remainder of the section, first the system model and the main assumptions are presented. Then, the proposed *VUG* policy for unicast and multicast traffic delivery is described.

A. VUG System Model and Assumptions

Let N be the number of downlink resources available at the BS every scheduling frame lasting T ms, $\mathcal{U}$ be the set of users served by the BS, and C be the number of available MCS levels in the system. The users belonging to $\mathcal{U}$ are split by the BS into different groups. Let $\mathcal{G}^{mul}$ be the set of multicast groups and $\mathcal{G}^{uni}$ the set of virtual groups in the cell, we define $\mathcal{U}_g^{mul}$ the user set in the multicast group $g \in \mathcal{G}^{mul}$ and $\mathcal{U}_g^{uni}$ the user set in the virtual group $g \in \mathcal{G}^{uni}$.

Let $m_u \in \{1, 2, \ldots, C\}$ be the index of the maximum MCS[2] sustained by the generic user u; $c(\cdot)$ is the function

---

[1]The number of activated virtual groups is a system parameter tunable by the network operator, as explained in the following section.

[2]By convention, we consider the lowest index associated to the most robust MCS, and the highest index associated to the MCS with the highest data rate.

that maps a MCS to the respective transport block size (i.e., in case of LTE, this is the number of bytes per resource unit).

The variable $L_{s,u} > 0$ represents the number of layers required for the multi-layer transmission of the video service s selected by user u. The case $L_{s,u} = 1$ models services that do not use scalable coding techniques [25]; these applications are referred to as *single-layer*.

Let $d_{s,u,l}$ denote the number of bits required for the transmission of layer l (with $l = 1, 2, \ldots, L_{s,u}$) of the video service s selected by user u. Note that, in the case of users belonging to a multicast group, the variables $L_{s,u}$ and $d_{s,u,l}$ are the same for all users u in the considered group since they request the same video service s.

All notations used in this paper are summarized in Table I.

TABLE I
LIST OF NOTATIONS

T	Frame duration
N	Number of available resources
C	Maximum number of MCS levels
$\mathcal{U}$	User set
$\mathcal{G}^{mul}$	Multicast group set
$\mathcal{G}^{uni}$	Virtual group set
$\mathcal{U}_g^{mul}$	User set of multicast group g
$\mathcal{U}_g^{uni}$	User set of virtual group g
m_u	Index of the maximum MCS supported by the user u
$c(\cdot)$	MCS-transport block size mapping function
$L_{s,u}$	Number of layers required for the transmission of video s to user u
$d_{s,u,l}$	Number of bits required for the transmission of layer l of video s to user u
$\mathcal{S}_{g,m}^{mul}$	User set for the subgroup relevant to the m-th MCS of the multicast group g
$\mathcal{S}_{g,m}^{uni}$	User set for the subgroup relevant to the m-th MCS of the virtual group g
$r_{g,m}^{mul}$	Number of resources assigned to subgroup related to the m-th MCS of multicast group g
$r_{g,m}^{uni}$	Number of resources assigned to subgroup related to the m-th MCS of virtual group g
$b_{g,m}^{mul}$	Number of bits required by subgroup related to the m-th MCS of multicast group g
$b_{g,m}^{uni}$	Number of bits required by subgroup related to the m-th MCS of virtual group g
a_u^{uni}	Amount of resources already allocated to user u

B. VUG Policy

This subsection describes the details of *VUG* in terms of the following five steps: *(i)* channel condition monitoring for both unicast and multicast users, *(ii)* grouping of users in unicast and multicast groups, respectively, *(iii)* subgroup formation based on similar BS-to-UE channel conditions, *(iv)* video base layer allocation to users such as minimum service level is ensured, and *(v)* enhancement layer allocation to enable the users with better delivery channel conditions to avail from higher video service quality. All the above listed steps are executed every scheduling frame in order to keep into account the variable channel quality and users requests.

Step 1 - Channel state collection

The BS collects the CSIs from each user u located in the cell and computes the related MCS $m_u, \forall u \in \mathcal{U}$. This step is repeated every scheduling frame with the aim to adapt the transmission parameters to the radio channel variations

experienced by the served users due to the mobility speed. To reduce the signaling load in case of massive multicast groups, more sophisticated channel state reporting schemes could be conceived that exchange, for instance, the update of channel state information when the terminal observes a channel variation for more than a pre-defined number of scheduling frames only [26]; this aspect is not considered in this manuscript and is left out for future work.

Step 2 - Groups formation

In this step, the BS splits the user set $\mathcal{U}$ into different groups. In particular, all the users interested in a given multicast service will join the same group $g \in \mathcal{G}^{mul}$, and form the user set $\mathcal{U}_g^{mul}$. As already mentioned, the users from the same multicast group access the same video content. We point out that a multicast group is created if it is composed of at least two members. If the number of members in a multicast group drops to one then the user joins a virtual group. Analogously, unicast users are grouped into *virtual groups*. In a virtual group $g \in \mathcal{G}^{uni}$, each user $u \in \mathcal{U}_g^{uni}$ may ask for a different video service. Since it is likely that users in a multicast group will experience heterogeneous channel conditions, the unicast users are randomly associated to virtual groups. This approach guarantees the presence of both users with good and bad channel conditions in each group.

The BS decides the number and the size of virtual groups to activate in its cell by considering the fairness in the treatment of unicast and multicast users as the main design parameter. In general, a *variable number of virtual groups* can be generated according to the traffic conditions and the degree of fairness that the network operator targets to provide. In [3] we analyzed the performance of different schemes for virtual groups creation. Obtained results showed that the best choice is setting the virtual group(s) size not greater than the average size of the multicast groups in the cell. Furthermore, we demonstrated that higher (lower) values of the virtual group(s) size correspond to greater (smaller) advantages to the unicast traffic. Based on these results, we adopted this choice in the paper.

Step 3 - Subgroups formation

In this phase the BS computes the admissible subgroup configurations. A subgroup collects users that have similar channel qualities. The number of subgroups that can be formed is bounded by the number of MCSs supported by the system.

Let $\mathcal{S}_{g,m}^{mul} \subseteq \mathcal{U}_g^{mul}$ indicate the subset of users from the multicast group $g \in \mathcal{G}^{mul}$ that support the m-th MCS level, and $\mathcal{S}_{g,m}^{uni} \subseteq \mathcal{U}_g^{uni}$ be the subset of users from the virtual group $g \in \mathcal{G}^{uni}$ that support the m-th MCS level (lines 1-3 of the pseudo-code in Table II). The subgroups are created to include users which can support individual m_u greater or equal with the m-th MCS level associated with that subgroup, as follows:

$$\mathcal{S}_{g,m}^{mul} = \{u \in \mathcal{U}_g^{mul} | m_u \geq m\} \ \forall g \in \mathcal{G}^{mul}$$
$$\mathcal{S}_{g,m}^{uni} = \{u \in \mathcal{U}_g^{uni} | m_u \geq m\} \ \forall g \in \mathcal{G}^{uni} \quad (1)$$

In the case of a multicast subgroup, all its members will be served with the MCS m that is the lowest among the MCSs supported by the subgroup members. Multicast subgroups can be efficiently used with multi-layer video coding techniques by

allowing subgroup members to receive an enhancement layer only after the previous layers have been received. Successive enhancement layers are in fact transmitted with a progressively increasing data rate, i.e., by using a less robust MCS. For example, subgroup users supporting 64-QAM modulation, also support 16-QAM, so they will receive the 64 QAM-transmitted enhancement layer only after they have already received the 16 QAM-transmitted layer.

Unlike the multicast subgroup members, each member of a virtual subgroup will be served by considering the individually supported MCS m_u.

Step 4 - Video base layer allocation

This allocation assumes that the minimum video service must be assured to all users in the cell and, therefore, each member of any group (virtual and multicast) has assigned at least the base layer of the requested video service. We assume that the system performs an admission control task which is aimed at guaranteeing that the base station has enough resources to serve at least the base layer for all the scheduled services. Note that in this phase, only the first enabled subgroup is considered (i.e., the one served with the lowest MCS among those supported in the original group), as it comprises all the group members. In particular, $\mathcal{S}_{g,m}^{mul} = \mathcal{U}_g^{mul}$, with $m = \min\limits_{u \in \mathcal{U}_g^{mul}} \{m_u\}$, and $\mathcal{S}_{g,m}^{uni} = \mathcal{U}_g^{uni}$, with $m = \min\limits_{u \in \mathcal{U}_g^{uni}} \{m_u\}$.

VUG determines the amount of resources required to transmit the base layer of the requested video service s to each multicast ($r_{g,m}^{mul}$) and virtual ($r_{g,m}^{uni}$) group g, as in Eq. (2):

$$
\begin{aligned}
r_{g,m}^{mul} &= \lceil d_{s,u,1}/c(m) \rceil, \text{ if } g \in \mathcal{G}^{mul} \\
r_{g,m}^{uni} &= \sum_{u \in \mathcal{S}_{g,m}^{uni}} \lceil d_{s,u,1}/c(m_u) \rceil, \text{ if } g \in \mathcal{G}^{uni}
\end{aligned}
\tag{2}
$$

where $m = \min\limits_{u \in \mathcal{S}_{g,m}^{mul}} \{m_u\}$, $m_u \in \{1,2,\ldots,C\}$ is the maximum MCS sustained by user u in the virtual group, $d_{s,u,1}$ is the number of bits required for the transmission of the base layer of the video service s requested by user u, and $c(\cdot)$ is the transport block size (i.e., number of bits per resource unit) for a given MCS.

The amount of resource units for the multicast group is calculated on the basis of the number of bits for the base layer of the requested video service $d_{s,u,1}$ (the same for all) and the transport block size related to the MCS of the user in the worst channel conditions.

Differently, the resources for a virtual group are the sum of the resources assigned to each single unicast user in the group; this amount is referred to as a_u^{uni} in the following text. The resources assigned to a single user depend on the number of bits for the base layer of the video service requested by the user $u \in \mathcal{S}_{g,m}^{uni}$ and the transport block size related to the MCS of that user (m_u). Lines 4-5 of the pseudo-code in Table II illustrate this step.

Step 5 - Enhancement layers allocation

Following the allocation of the base layer to all the users in the cell, if still some bandwidth is available then the BS iteratively determines the multicast or virtual subgroup to serve

and the amount of resources to assign to the selected subgroup. The policy implemented for the eligible subgroup selection is up to the network provider, since *VUG* is independent from the strategy utilized. However, this paper assumes that the BS selects the subgroup that maximizes the aggregated data rate for the system, thus increasing its efficiency. The decision is taken, layer by layer, iteratively until no more resources are available in the current scheduling frame.

For each set of users in a multicast subgroup $\mathcal{S}_{g,m}^{mul} \subseteq \mathcal{U}_g^{mul}$ or in a virtual subgroup $\mathcal{S}_{g,m}^{uni} \subseteq \mathcal{U}_g^{uni}$, the BS first computes the requested number of bits $b_{g,m}$ as in Eq. (3), then selects the subgroup to serve and assigns the resource to it:

$$
\begin{aligned}
b_{g,m}^{mul} &= \sum_{l=1}^{L_{s,u}} d_{s,u,l} - \sum_{m=1}^{C} r_{g,m}^{mul} c(m), \text{ if } g \in \mathcal{G}^{mul} \\
b_{g,m}^{uni} &= \frac{1}{|\mathcal{S}_{g,m}^{uni}|} \sum_{u \in \mathcal{S}_{g,m}^{uni}} \left(\sum_{l=1}^{L_{s,u}} d_{s,u,l} - a_u^{uni} c(m_u) \right), \text{ if } g \in \mathcal{G}^{uni}
\end{aligned}
\tag{3}
$$

For a multicast subgroup, the requested number of bits is computed as the difference between the total number of bits needed to deliver all the video stream's layers and the number of bits already allocated (i.e., $\sum_{m=1}^{C} r_{g,m}^{mul} c(m)$).

In the case of a virtual subgroup, the requested number of bits is computed as the *average* number of bits required to deliver the remaining requested video streams to the virtual subgroup members. For each unicast user, the requested data is dependent on the amount of resources already allocated a_u^{uni}. The operation of averaging the data rate of unicast users $\left(\frac{1}{|\mathcal{S}_{g,m}^{uni}|} \right)$ is introduced to assimilate the operation of the virtual subgroup to a multicast subgroup. In such a way, a unified requested data rate for the entire subgroup can be considered.

Once the per-subgroup required resources are determined, the algorithm selects the multicast subgroup g^* operating at MCS m^* and the virtual subgroup $\hat{g}$ operating at MCS $\hat{m}$ that maximizes the system throughput as follows:

$$
\begin{aligned}
g^*, m^* &= \underset{g \in \mathcal{G}^{mul}, m \in \{1,2,\ldots,C\}}{\arg\max} \{b_{g,m}^{mul} \cdot |\mathcal{S}_{g,m}^{mul}|\}), \text{ if } g \in \mathcal{G}^{mul} \\
\hat{g}, \hat{m} &= \underset{g \in \mathcal{G}^{uni}, m \in \{1,2,\ldots,C\}}{\arg\max} \{b_{g,m}^{uni} \cdot |\mathcal{S}_{g,m}^{uni}|\}), \text{ if } g \in \mathcal{G}^{uni}
\end{aligned}
\tag{4}
$$

The resources allocated to the selected subgroup (g^* or $\hat{g}$) are calculated as in Eq. (5):

$$
\begin{aligned}
r_{g^*,m^*}^{mul} &= \lceil b_{g^*,m^*}^{mul}/c(m^*) \rceil, \text{ if } \{b_{g^*,m^*}^{mul} \cdot |\mathcal{S}_{g^*,m^*}^{mul}|\}) \geq \{b_{\hat{g},\hat{m}}^{uni} \cdot |\mathcal{S}_{\hat{g},\hat{m}}^{uni}|\}) \\
r_{\hat{g},\hat{m}}^{uni} &= \lceil b_{\hat{g},\hat{m}}^{uni}/c(\hat{m}) \rceil, \text{ if } \{b_{g^*,m^*}^{mul} \cdot |\mathcal{S}_{g^*,m^*}^{mul}|\}) < \{b_{\hat{g},\hat{m}}^{uni} \cdot |\mathcal{S}_{\hat{g},\hat{m}}^{uni}|\})
\end{aligned}
\tag{5}
$$

If a multicast subgroup is selected, then the algorithm simply assigns the amount of resources required to deliver the requested data rate to the selected subgroup, and deletes the served group from the multicast group set if it is already served with the maximum sustained data rate (i.e., all layers are scheduled).

In case a virtual subgroup is selected, the scheduled resources must be later distributed among the unicast users in the subgroup. Different strategies can be implemented, this representing a further flexibility feature of *VUG*. This paper

TABLE II
VUG: Subgroup selection

1: Let $n_g^{uni} = \min_{u \in \mathcal{U}_g^{uni}} \{m_u\}$ ▷ Subgroups formation

2: Let $n_g^{mul} = \min_{u \in \mathcal{U}_g^{mul}} \{m_u\}$

3: Create admissible subgroup configurations $\mathcal{S}_{g,m}^{mul}$ and $\mathcal{S}_{g,m}^{uni}$

4: Compute $r_{g,m}^{mul}$, $r_{g,m}^{uni}$, and a_u^{uni} according to (2) ▷ Base layer allocation

5: $N = N - \left(\sum_{g \in \mathcal{G}^{uni}} r_{g,n_g^{uni}}^{uni} + \sum_{g \in \mathcal{G}^{mul}} r_{g,n_g^{mul}}^{mul} \right)$

6: **while** $N > 0 \vee (\mathcal{G}^{mul} \neq \emptyset \wedge \mathcal{G}^{uni} \neq \emptyset)$ **do**   ▷ Enhancement layers allocation

7:     **for all** $g \in \mathcal{G}^{uni}$ **do**

8:        **for all** $m = n_g^{uni}, n_g^{uni} + 1, \ldots, C$ **do**

9:           Compute $b_{g,m}^{uni}$ according to (3)

10:        **end for**

11:     **end for**

12:     **for all** $g \in \mathcal{G}^{mul}$ **do**

13:        **for all** $m = n_g^{mul}, n_g^{mul} + 1, \ldots, C$ **do**

14:           Compute $b_{g,m}^{mul}$ according to (3)

15:        **end for**

16:     **end for**

17:     Select $\mathcal{S}_{\hat{g},m}^{uni}$ according to (4)

18:     Select $\mathcal{S}_{g^*,m^*}^{mul}$ according to (4)

19:     **if** $b_{g^*,m^*}^{mul} |\mathcal{S}_{g^*,m^*}^{mul}| > b_{\hat{g},m}^{uni} |\mathcal{S}_{\hat{g},m}^{uni}|$ **then**

20:        $n_{g^*}^{mul} = m^*$

21:        $r_{g^*,m^*}^{mul} = r_{g^*,m^*}^{mul} + \lceil b_{g^*,m^*}^{mul}/c(m^*) \rceil$

22:        $N = N - \lceil b_{g^*,m^*}^{mul}/c(m^*) \rceil$

23:        **if** $\sum_{i=1}^{C} r_{g^*,i}^{mul} c(i) \geq \sum_{l=1}^{L_{g^*}^{mul}} d_{g^*,l}^{mul}$ **then**

24:           $\mathcal{G} = \mathcal{G} \setminus \{g^*\}$

25:        **end if**

26:     **else**

27:        $n_{\hat{g}}^{uni} = \hat{m}$

28:        $r_{\hat{g},\hat{m}}^{uni} = r_{\hat{g},\hat{m}}^{uni} + \lceil b_{\hat{g},\hat{m}}^{uni}/c(\hat{m}) \rceil$

29:        $N = N - \lceil b_{\hat{g},\hat{m}}^{uni}/c(\hat{m}) \rceil$

30:        Perform Unicast User Selection and update a_u^{uni}

31:     **end if**

32: **end while**

---

focuses on a maximum throughput approach, so that the BS iteratively selects the user with the best channel condition, as described in Eq. (6):

$$
\begin{cases}
\hat{u} = \arg\min_{u \in \mathcal{S}_{\hat{g},\hat{m}}} \left\{ \left(\sum_{l=1}^{L_{s,u}} d_{s,u,l} - a_u^{uni} c(m_u) \right) / c(m_u) \right\} \\
s.t. \quad u = \arg\max_{u \in \mathcal{S}_{\hat{g},\hat{m}}} \{m_u\}
\end{cases}
\tag{6}
$$

Once the unicast user $\hat{u}$ is selected, the base station assigns the resources to such a user; in case the available resources are higher than those requested by user $\hat{u}$, then the resource assignment is iterated until all the resources for the selected virtual subgroup are made available. As a consequence, the complexity burden for this step can be defined as $O(nr)$, where n is the number of users in the virtual subgroup (which is thus bounded by the overall number of unicast users in the cell) and r is the number of resources allocated to the virtual subgroup. Alternatively to the throughput maximization criterion we implemented, alternative solution taking into account user fairness could be utilized in case the network operator aims at achieving a fair resource distribution among unicast users. This is not the scope of this paper.

Lines 6-32 of the pseudo-code in Table II describe the operation of this step.

IV. Simulation-based Testing and Result Analysis

A. Testing setup

The performance of the proposed *VUG* radio resource management solution has been tested by using Matlab simulations over an LTE network centrally managed by an eNodeB (the LTE's BS) that offers connection-oriented services to pedestrian UEs. LTE guarantees low latency, increased system capacity, and improved spectral efficiency compared to previous cellular network releases; furthermore, being designed to efficiently support multicast transmissions in both the core and the radio access network by implementing the MBMS [27], LTE is currently the most promising wireless system able to support high-quality group-oriented services.

Scheduling procedures are performed by the eNodeB that handles resource allocation in both the time and frequency domain. The LTE's resource unit is the Resource Block (RB), which corresponds to 180 kHz and 0.5 ms in the frequency and time domains, respectively. Every Transmission Time Interval (TTI), lasting 1 ms, the scheduler assigns the resources units on a RB-pair basis. Link adaptation is performed according to the Channel Quality Indicator (CQI), which represents the maximum MCS supported by a UE according to the experienced Signal-to-Interference-plus-Noise Ratio (SINR) value. In particular, LTE defines 15 CQI levels.

The performance evaluation is based on the guidelines defined in [28]. We have considered the cell deployment as defined by 3GPP and ITU-R (i.e., macro-cell case A with 19 neighboring cells, where the cell at the center of the scenario is the cell under consideration for simulation results while adjacent cells act as source of interference with background traffic). Channel quality for each terminal is evaluated in terms of the SINR measured over each sub-carrier when path loss, slow and fast fading affect the signal reception. The Exponential Effective SIR Mapping (EESM) [29] is used to obtain the effective SINR value and the related CQI which ensures a BLER target value smaller than 10%. We assume that unicast users are uniformly distributed across the cells and move according to the Levy-walk mobility model [30]; multicast UEs are distributed in limited areas as in a typical on-campus environment with SLAW mobility [31].

B. Models and assessment metrics

The Matlab simulations have evaluated *VUG* in comparison with the three approaches described in section II: Unicast Maximization (*UM*), Equal Sharing (*ES*), and Equal Competition (*EC*). *UM* guarantees the minimal data rate to unicast and multicast users and then assigns the remaining resources *only* to unicast users. In *ES*, multicast and unicast users equally share the network capacity, and the bandwidth sharing is fixed statically in advance. *EC* guarantees the minimum data rate to unicast and multicast users and assigns the residual resources based on a maximum throughput scheme, which iteratively selects the service (unicast or multicast) that conveys the highest number of bits to the destination. Each simulation

TABLE III
MAIN SIMULATION PARAMETERS

Parameter	Value
Distance attenuation	128.1+37.6*log(d), d [km]
Shadow fading	Log-normal,0 mean, $\sigma = 8$ [dB]
Shadowing Correlation distance	50 m [28]
Fast Fading	ITU-R PedB (extended for OFDM)
Carrier frequency	2 GHz
Cell layout	3GPP Macro-cell case #1, Hexagonal grid, 19 cell sites, 3 sectors per site [28]
Inter Site Distance	500 m
RB size	12 sub-carriers, 0.5 ms
Sub-carrier spacing	15 kHz
Data/Control OFDM symbols	11/3
Number of RBs	100
CQI scheme	Wideband
BLER target	10%
TTI	1 ms
Scheduling frame	10 ms
EUTRA UE	Antenna gain 0 dBi, Noise Figure 9 dB [28]
EUTRA Node-B	Antenna gain 14 dBi, Noise Figure 5 dB [28]
eNodeB transmit power	43 dBm [28]
MIMO Configuration	1 Tx, 2 Rx
Thermal Noise	-174 dBm/Hz
Traffic type	Layered video [121-564] kbps with AL-FEC as in [17]

run has been repeated several times to get 95% confidence intervals in the presented results.

The analyzed performance parameters are:

- **Average Throughput**: is the achieved bit rate (per user or per cell), accounting for the packets successfully delivered (to the target UEs or all UEs) during the simulation time.
- **Inter-Class Throughput Distribution (ICTD)**: evaluates the fairness in the throughput achieved by unicast and multicast traffic; it is calculated as in the following:

$$ICTD = 1 - \frac{t_{avg}^{uni}}{t_{avg}^{uni} + t_{avg}^{mul}} \qquad (7)$$

where t_{avg}^{uni} and t_{avg}^{mul} represent the average value of the ratio between the experienced data rate and the maximum sustained data rate of all unicast and multicast users, respectively. This index takes values in the $[0, 1]$ range and for the case of perfect fairness it is equal to 0.5. ICTD is higher than 0.5 in the case when the multicast traffic benefits from higher throughput values, and is lower than 0.5 when the unicast traffic benefits from the throughput distribution. The rationale behind the definition of the ICTD parameter is to show which class of services (multicast or unicast) is receiving more throughput than the other. This kind of information cannot be obtained with either GDI [12] or Jain's fairness index [36] parameters.
- **Percentage of assigned resources**: shows how network resources (i.e., RBs) are split between unicast and multicast services.
- **Spectral efficiency (SE) [bps/Hz]**: represents the ratio between the users' received rates and the whole amount of channel bandwidth exploited by the eNodeB to schedule the transmissions to unicast/video users.

- **Percentage of UEs per video layer**: is the percentage of UEs in the cell (both unicast and multicast) which receives a given video layer. This metric allows to clearly assess the video quality experienced by the video users.

C. The impact of unicast traffic load

In this section, we evaluate the performance of the addressed policies by varying the load of unicast users in the cell. One multicast group composed of 20 UEs is considered and the number of unicast users is varied from 10 to 50.

We first present the analysis in terms of ICTD (see Fig. 2). This metric clearly testifies that *UM* and *ES* give higher priority to unicast and multicast services, respectively. Up to the presence of 15 unicast UEs, the *EC* policy reaches values close to 0.5; then *EC* starts to give a higher priority to multicast streams. Finally, our proposed subgroup-based *VUG* provides the performance closest to 0.5; this validates our design objective of achieving a fair treatment of unicast and multicast traffic by means of creating virtual groups which have similar features (number of users, amount of requested resources, etc.) and this involves a fair competition among unicast and multicast users to get the system resources.

We highlight that the desired behavior shown by the proposed *VUG* in terms of ICTD, as will be shown in the following simulation results, can be achieved at the cost of other performance parameters.
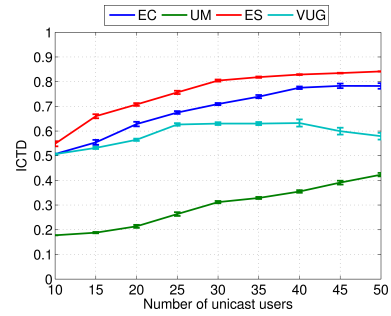

Fig. 2. ICTD by varying the number of unicast UEs.

In this subsection, we also analyze the impact of subgrouping in *VUG* by considering two versions: *VUG* which benefits from the subgroup formation feature and *VUG (w/o subgroups)* which does not create subgroups.

The comparison in terms of average cell throughput (i.e., the mean throughput of all users served by the eNodeB) is shown in Fig. 3. The use of *VUG* with subgrouping results in a performance increase with respect to the case of *VUG (w/o subgroups)* by about 30% on average. Up to 25 unicast users, the impact of subgrouping is not evident; under higher unicast load, subgrouping shows to be effective since it allows to efficiently manage the spectrum, though a large portion of resources are needed to allocate the base layer to unicast subscribers. When comparing *VUG* with the other policies, it can be noticed that our approach achieves the highest throughput under a load up to 35 unicast UEs; then its performance decreases and the mismatch w.r.t. *EC* and *ES* (the highest performing policies in case of heavy unicast

traffic load) is of about 23% in the heavy load scenario with 50 unicast UEs. The reason behind this behavior is the fact that the *VUG* algorithm is intended to achieve fairness in the satisfaction of unicast and multicast users. As a consequence, when the number of unicast users is high, the overall system throughput is reduced with respect to *EC* and *ES* because these policies give higher priority to multicast traffic, although the higher load in the cell is due to unicast UEs. Note that this apparent performance decrease does not involve an inefficient use of system resources, as it will be clearly shown in the following subsection.

The output of further tests are illustrated in Fig. 4, which plots the percentage of UEs per video layer under an increasing number of unicast users (set to 10, 30, and 50 in Fig. 4(a), Fig. 4(b), and Fig. 4(c), respectively). These figures allow to evaluate the average video quality delivered to both unicast and multicast users. It emerges that *EC* guarantees good video quality (i.e., up to layer four) to a large portion of users, and similar results are obtained by *VUG* both in scenario (a) and (b). We highlight that, although *ES* delivers an higher number of enhancement layers in scenario (c) w.r.t. *VUG*, it is not able to guarantee the base layer to all users. In all cases, the use of subgrouping achieves an increase in video quality w.r.t. the case without subgroups.

Above analyses testify to the benefits in terms of throughput and video quality introduced by the use of subgrouping in *VUG*. For this reason, the remainder of this section will consider only *VUG* with the subgroup-formation capability.
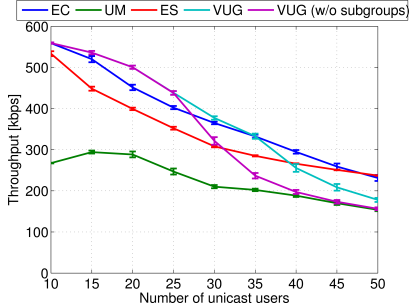
Fig. 3.   Average cell throughput by varying the number of unicast UEs.

Fig. 5 plots the throughput performance achieved by unicast (Fig. 5(a)) and multicast (Fig. 5(b)) UEs. As expected, when considering the average unicast throughput in Fig. 5(a), the *UM* policy guarantees the best throughput performance, while *ES* obtains the poorest throughput values. Noteworthy, *VUG* has a performance close to *UM*. Fig. 5(b) shows the average user throughput experienced by the multicast members. The obtained results are similar to Fig. 5(a); this time in favor of the solutions that support multicast transmissions. Indeed, *ES* and *EC* are the best performing schemes while *UM* achieves the worst performance (i.e., only the base layer is allocated to multicast UEs). However, *VUG* offers the same throughput as *EC* and *ES* in the cases with low number of unicast users. These results testify that *VUG* prevents wasting of system resource due to multicast service (as for *EC* and *ES* schemes) in favour of unicast traffic, without drastically affecting the performance of multicast UEs (as for the *UM* scheme).

In Fig. 6 the percentage of assigned resources is depicted in the three scenarios analyzed in Fig. 4. The poor spectrum utilization of *ES*, which is unable to exploit the whole available spectrum, clearly emerges. The reason is that *ES* equally splits the available spectrum between unicast and multicast traffic, without taking into account the amount of resources required by each type of traffic (as an example, in the analyzed scenarios multicast services require only a limited amount of the whole system resources). All the considered policies assign only a small portion of RBs to multicast streams (from 5 up to 25%); in the case of 10 unicast UEs, *EC*, *UM* and *VUG* exploit about 65% of the available spectrum for unicast services while the *ES* policy reserves a percentage equal to 50% of the system resources to unicast traffic. When the unicast load increases, all policies (except *ES*) exploit the whole available scheduling resources. Note that in case of heavy unicast load, *VUG* reduces the number of RBs assigned to multicast UEs to improve the portion of resources for unicast traffic.

In Fig. 7 the performance in terms of Spectral Efficiency is illustrated. *ES* suffers from the poorest performance. This is due to the fact that *ES* does not efficiently exploit the bandwidth, i.e., only a limited portion of the available resources are scheduled by the eNodeB; thus, both throughput and spectrum efficiency are reduced. Furthermore, as the number of unicast users increases, the efficiency of *ES* decreases while the performance of *EC* remains constant. This is due to the fact that *ES* always reserves the same amount of resources to unicast traffic regardless of the unicast traffic load. Thus, the reserved resources may not be sufficient when the unicast traffic load increases, and this causes a decrease in the spectral efficiency. Differently, *UM* and *VUG* outperform the other policies and achieve very similar performance; for both algorithms, the spectral efficiency exhibits an increasing trend when the unicast traffic load grows, thanks to the resources reserved to unicast UEs. As a consequence, *VUG* achieves very promising spectral efficiency results, which are not adversely affected by an increase of unicast load in the cell.

The analyses above demonstrate the benefits of our proposed *VUG* which is able to achieve fair inter-class satisfaction and to increase the throughput of unicast UEs without significantly affecting multicast performance and overall spectral efficiency.

D. Effect of the number of multicast groups

This subsection evaluates how much the number of multicast groups supported for simultaneous video content delivery in the same cell impact the performance. Testing conditions similar to the previous study are set. A fixed number of unicast users (i.e., 20 UEs) is considered and the number of active multicast flows is varied between 1 and 8. Each multicast group is composed of 20 members. The performance of the compared solutions is assessed in terms of ICTD and throughput of both unicast and multicast users, and the results are plotted in Fig. 8 and Fig. 9, respectively.

By looking at Fig. 8, it clearly emerges that *VUG* obtains the best performance and largely outperforms other approaches. This result testifies to *VUG*'s effectiveness in guaranteeing throughput fairness between unicast and multicast traffic. *UM*
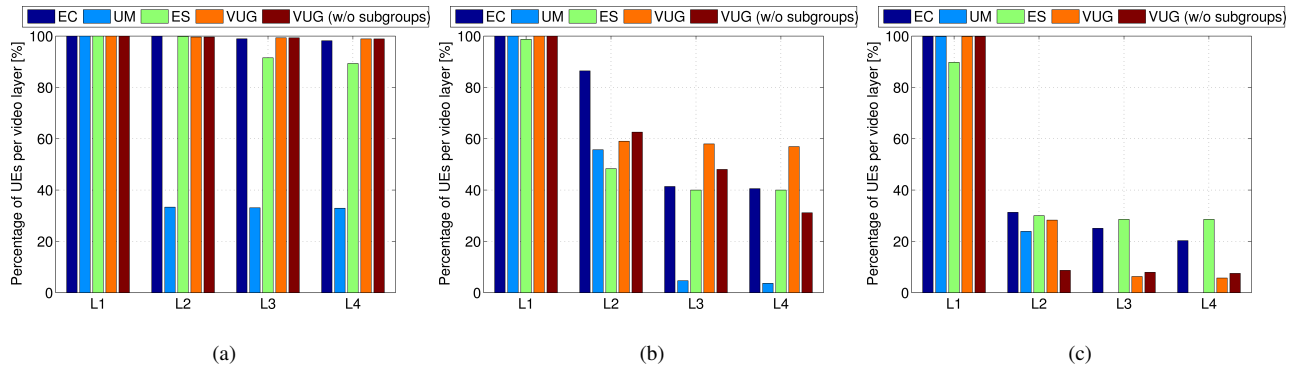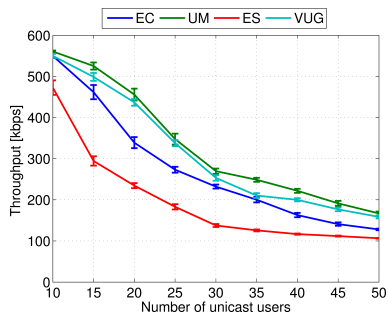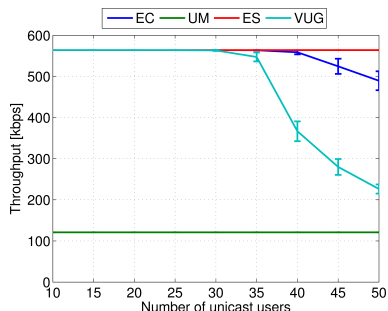
Fig. 4. Percentage of users per video layer in the scenario with (a) 10, (b) 30, and (c) 50 unicast UEs and one multicast group composed by 20 members.

Fig. 5. Throughput of (a) unicast, and (b) multicast users by varying the number of unicast UEs.

and *EC* policies manifest the same trend as in the previous analysis, i.e., higher priority to unicast and multicast streams, respectively. Finally, *ES* gives higher priority to unicast traffic in case of several multicast groups in the cell, as the portion of RBs reserved to multicast traffic (i.e., 50%) is not satisfactory to adequately schedule video flows to all multicast groups.

The throughput achieved by unicast users is illustrated in Fig. 9(a). It can be noted that, although it achieves higher throughput results, the *ES* policy is not affected by the number of active multicast groups, since a constant amount of bandwidth is reserved to unicast traffic. Additionally, all the other algorithms suffer from a decrease in the throughput perfor-

mance. In particular, *UM*, *VUG* and *EC* throughputs decrease down to 267, 161 and 121 kbps, respectively. Fig. 9(b) shows the average user throughput experienced by the multicast members. It can be observed that *UM* achieves the worst performance (i.e., only the base layer is allocated to multicast UEs). Differently, all the other approaches show a decrease in the throughput values when the number of multicast groups grows. In particular, *EC* is the best performing scheme. However, *VUG* offers a throughput close to that of *EC*.

E. The effect of non-video unicast traffic

In this section we analyze the behavior of the proposed *VUG* approach when a varying portion of non-video unicast traffic is managed by the eNodeB. In this simulation scenario, we assume that the network load is equal to 20 unicast users and 20 multicast members joined in one multicast group. We focus on two different cases. In the first case, non-video unicast traffic is taken into consideration in the operations of *VUG*. In the second case, the eNodeB manages the traffic as it follows: (i) assigns the base layers to both unicast and multicast video users, thus non-video traffic is scheduled with a maximum throughput approach; (ii) allocates enhancement layers to video flows by exploiting the unicast/multicast policies addressed in this paper.

In the first analyzed scenario, we assume that the eNodeB manages a growing amount of bursty unicast traffic and, to this aim, it allocates a portion of resources to non-video unicast UEs (not taken into account by *VUG*), which varies during the simulation time; the average amount of resources for background unicast users varies up to 50%.

The performance is evaluated in terms of ICTD (Fig. 10). The proposed *VUG* is not affected by the presence of unicast background traffic (i.e., the ICTD is constantly close to 0.5). Differently, the performance of all other policies varies at the expense of unicast traffic, i.e., the ICTD increases. These results indicate that, if a portion of resources is scheduled for non-video services, other addressed policies give a higher priority to multicast streams while *VUG* guarantees a fair resource distribution between unicast and multicast traffic.

In the second analyzed scenario, we varied the percentage of video and non-video UEs in a total of 20 unicast users.
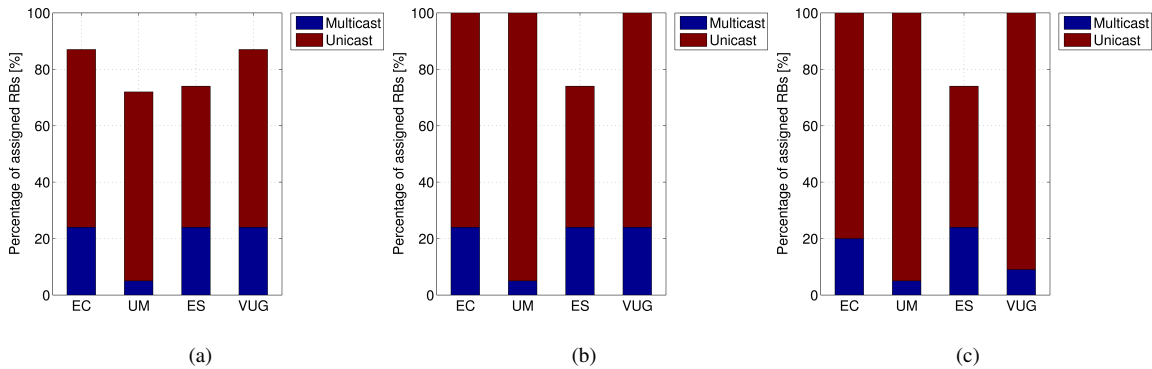
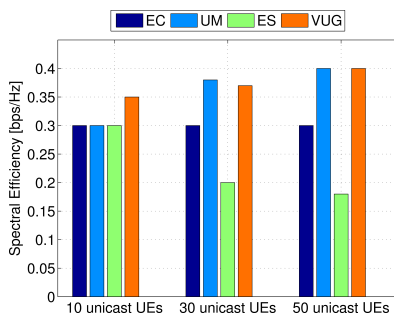Fig. 6.   Percentage of assigned resources with (a) 10, (b) 30, and (c) 50 unicast UEs and one multicast group with 20 members.

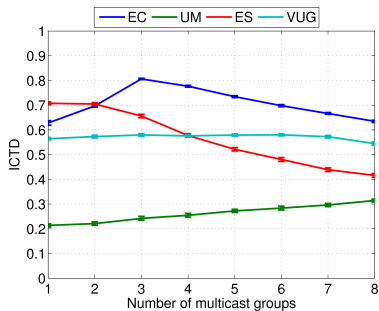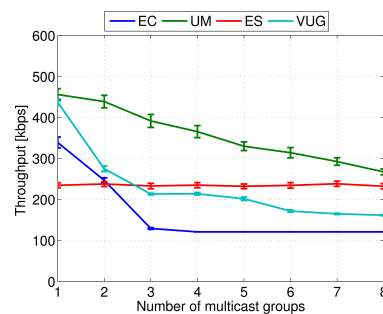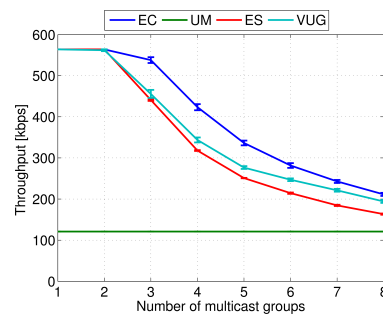Fig. 7.   Spectral efficiency by varying the number of unicast UEs.

Fig. 8.   ICTD by varying the number of multicast groups.

(a)

(b)

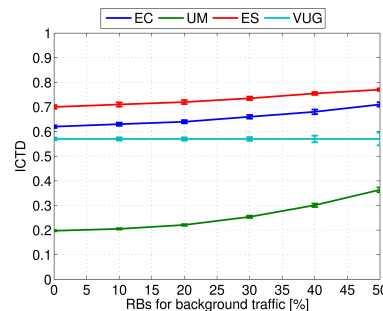Fig. 9.   Throughput of (a) unicast, and (b) multicast users by varying the number of multicast groups.

Fig. 10.   ICTD by varying the portion of resources allocated to non-video background traffic.

The latter traffic is modeled by using the bursty traffic model (FTP model 1, Poisson distributed user arrival [32], file size of 2 Mbytes), and its throughput during the file download is measured. We first focus on the quality experienced by unicast and multicast video streams, measured by taking into account the average % of UEs per video layer (Fig. 11). One can note that the presence of bursty traffic does not affect the video quality. Although this result can be expected for the benchmark policies addressed in this work, it testifies the effectiveness of *VUG* which, being natively designed to support video flows, can be easily enhanced to jointly work with non-video traffic without causing performance degradation. In Fig. 12, we plot the average throughput of bursty UEs. Again, we can note that the performance of bursty UEs does not meaningfully vary under increasing non-video unicast flows.
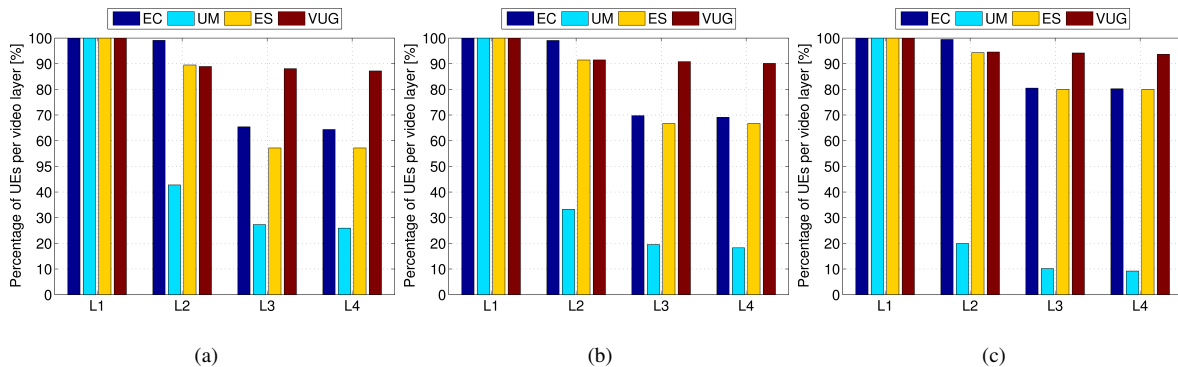
Fig. 11. Bursty traffic analysis: Percentage of users per video layer with (a) 5, (b) 10, and (c) 15 bursty unicast UEs.
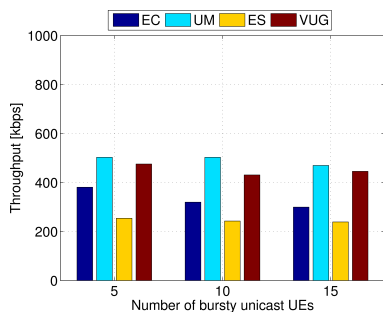

Fig. 12. Bursty traffic analysis: throughput of bursty unicast UEs.

F. Comparative discussion

The simulations results provided in this section highlighted the benefits introduced by the proposed *VUG* strategy and Table IV summarizes these simulation results.

The *EC* strategy gives intrinsically a higher priority (and throughput) to multicast traffic at the expense of a reduced throughput for unicast users. On the contrary, the *UM* approach is tailored to maximize the throughput of unicast users, while guaranteeing only the base layer allocation to multicast flows. The *ES* policy uniformly splits the available resources between unicast and multicast traffic, thus it achieves adequate multicast/unicast throughput only in particular multicast/unicast load scenarios. Our proposed *VUG* strategy is able to allocate dynamically the resources to guarantee an increase in the throughput experienced by unicast users without a severe deterioration of the performance of multicast members. The above mentioned behavior is supported by the performance in terms of ICTD. Indeed, while for other approaches ICTD is close to 0.5 (i.e., equal level of satisfaction for both unicast and multicast flows), in case of a huge cell load only (i.e., the load in the cell limits the number of available resources for enhancement layer allocation), *VUG* always guarantees the best ICTD performance; this confirms the effectiveness of the strategy proposed in this paper.

As a final remark, we underline that the improvements achieved by *VUG* and summarized above do not require an increase in the overhead, since no additional message exchange is required to create the virtual group. Nonetheless, a slight additional processing is required at the eNodeB to distribute resources among unicast users in the virtual group.

V. CONCLUSIONS AND FUTURE WORKS

This paper addressed the problem of fairly distributing radio resources between unicast and multicast multimedia services in next-generation mobile networks. It proposed the *VUG (Virtual Unified Group)* approach, an innovative radio resource management framework that organizes unicast users into *virtual groups* that assemble destinations depending on the user perceived channel quality levels. These virtual groups are managed by the base station in analogy to the multicast groups by employing a subgrouping strategy. The performance of the proposed *VUG* was assessed in comparison with alternative approaches in terms of both throughput and fairness in the satisfaction of unicast and multicast users. The obtained results showed the benefit of employing subgrouping, and demonstrated the effectiveness of *VUG* in fairly distributing system resources among multicast and unicast users, for diverse scenarios with variable load of unicast, multicast and additional background traffic. Future works will be finalized to the improvement of the video quality estimation accuracy by using models such as those proposed by [33] and [34] for different types of content.

REFERENCES

[1] Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018, White Paper, Cisco, 2014.
[2] 5G-PPP, "Advanced 5G Network Infrastructure for the Future Internet," Nov. 2013.
[3] S. Pizzi, M. Condoluci, G. Araniti, A. Molinaro, and A. Iera, "A Novel Approach for Unicast and Multicast Traffic Management in Wireless Networks," in *IEEE Vehicular Technology Conf. (VTC)*, 2015, pp. 1-5.
[4] A. Richard, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for OFDMA-based systems: A survey," *IEEE Communications Surveys and Tutorials*, 15:1, pp. 240-254, 2013.
[5] C. Suh, and J. Mo, "Resource Allocation for Multicast Services in Multicarrier Wireless Communications," *IEEE Transactions on Wireless Communications*, 7:1, pp. 27-31, 2008.

TABLE IV
SUMMARY OF PERFORMANCE FOR OUR PROPOSED VUG STRATEGY.

Metric	EC	UM	ES	VUG
ICTD	Higher satisfaction for multicast users w.r.t. unicast users	-Higher satisfaction for unicast users -Close to 0.5 in case of limited resources (only base layer to both unicast and multicast traffics)	-Higher satisfaction for multicast users (in case of few multicast groups) -Higher satisfaction for unicast users (in case of several multicast groups)	Close to 0.5 in all scenarios and for all unicast/multicast configurations
Throughput	-High for multicast users -Low for unicast users	-High for unicast users -Lowest throughput (only base layer) for multicast users	-High for multicast users (only in case of few multicast flows) -Low for unicast users	-High unicast throughput (close to that of UM) -Very good multicast throughput (equal to those of EC and ES) in case of few unicast users
Spectral Efficiency	Independent from unicast traffic load	High	Low and strongly dependent from the unicast traffic load	High

[6] J. Liu, W. Chen, Z. Cao, and K. Letaief, "Dynamic power and sub-carrier allocation for OFDMA-based wireless multicast systems," in *IEEE International Conference on Communications (ICC)*, 2008.

[7] T. P. Low, M. O. Pun, Y. W. P. Hong, and C. C. J. Kuo, "Optimized opportunistic multicast scheduling (OMS) over wireless cellular networks," *IEEE Transactions on Wireless Communications*, 9:2, 2010.

[8] F. Hou, L.X. Cai, Pin-Han Ho, X. Shen, and J. Zhang, "A cooperative multicast scheduling scheme for multimedia services in IEEE 802.16 networks," *IEEE Transactions on Wireless Communications*, 8:3, pp. 1508-1519, 2009.

[9] S. Almowuena, M. M. Rahman, C.-H. Hsu, A. A. Hassan, and M. Hefeeda, "Energy-Aware and Bandwidth-Efficient Hybrid Video Streaming Over Mobile Networks," *IEEE Transactions on Multimedia*, 18:1, pp. 102-115, 2015.

[10] A. Tassi, I. Chatzigeorgiou, and D. Vukobratovic, "Resource-Allocation Frameworks for Network-Coded Layered Multimedia Multicast Services," *IEEE Journal on Selected Areas in Communications*, 33:2, pp. 141-155, 2015.

[11] G. Araniti, V. Scordamaglia, M. Condoluci, A. Molinaro, and A. Iera, "Efficient Frequency Domain Packet Scheduler for Point-to-Multipoint Transmissions in LTE Networks," in *IEEE International Conference on Communications (ICC)*, pp. 4405-4409, 2012.

[12] G. Araniti, M. Condoluci, L. Militano, and A. Iera, "Adaptive resource allocation to multicast services in LTE systems," *IEEE Transactions on Broadcasting*, 59:4, pp. 658 664, 2013.

[13] C. K. Tan, T. C. Chuah, and S. W. Tan, "Adaptive multicast scheme for OFDMA-based multicast wireless systems," *Electronics Letters*, 47:9, pp. 570-572, 2011.

[14] W.-H. Kuo, T. Liu, and W. Liao, "Utility-Based Resource Allocation for Layer-Encoded IPTV Multicast in IEEE 802.16 (WiMAX) Wireless Networks," in *IEEE International Conference on Communications (ICC)*, 2007, pp. 1754-1759.

[15] S. Deb, S. Jaiswal, and K. Nagaraj, "Real-Time Video Multicast in WiMAX Networks," in *IEEE Conference on Computer Communications (INFOCOM)*, 2008, pp. 1579-1587.

[16] P. Li, H. Zhang, B. Zhao, and S. Rangarajan, "Scalable Video Multicast in Multi-carrier Wireless Data Systems," in *IEEE International Conference on Network Protocols (ICNP)*, 2009, pp. 141-150.

[17] S. Sharangi, R. Krishnamurti, and M. Hefeeda, "Energy-Efficient Multicasting of Scalable Video Streams Over WiMAX Networks," *IEEE Transactions on Multimedia*, 13:1, pp. 102-115, 2011.

[18] J. F. Monserrat, J. Calabuig, A. Fernandez-Aguilella, and D. Gomez-Barquero, "Joint Delivery of Unicast and E-MBMS Services in LTE Networks," *IEEE Trans. on Broadcasting*, 58:2, pp. 157-167, 2012.

[19] S. Y. Baek, D. Y.-J. Hong, and D.K. Sung, "Adaptive Transmission Scheme for Mixed Multicast and Unicast Traffic in Cellular Systems," *IEEE Trans. on Vehicular Technology*, 58:6, pp. 2899-2908, 2009.

[20] M. Konrad, W. Koch, and J. Huschke, "Throughput Based Switching Threshold for Multicast and Unicast in Cellular OFDM Networks," in *IEEE Vehicular Technology Conference (VTC)*, 2007.

[21] H. Seo, S. Kwack, and B. G. Lee, "Channel Structuring and Subchannel Allocation for Efficient Multicast and Unicast Services Integration in Wireless OFDM Systems," in *IEEE Globecom*, 2007.

[22] J. Shen, Na Yi, Bo Wu, W. Jiang, and H. Xiang, "A greedy-based resource allocation algorithm for multicast and unicast services in OFDM system," in *WCSP*, 2009.

[23] C. Wu, X. Sun, and T. Zhang, "A power-saving scheduling algorithm for mixed multicast and unicast traffic in MBSFN," in *ComComAp*, 2012.

[24] H. Deng, X. Tao, and J. Lu, "Qos-aware resource allocation for mixed multicast and unicast traffic in OFDMA networks," *EURASIP Journal on Wireless Communications and Networking*, 2012.

[25] D. Munaretto, D. Jurca, and J. Widmer, "Scalable video broadcast in cellular networks: impact on QoS and network resources," in *IEEE ISCC*, 2010.

[26] Y. Cai, S. Lu, L. Zhang, C. Wang, P. Skov, Z. He, and Kai Niu, "Reduced Feedback Schemes for LTE MBMS," in *IEEE Vehicular Technology Conference (VTC-Spring)*, 2009, pp. 1-5.

[27] 3GPP, "General aspects and principles for interfaces supporting multimedia broadcast multicast service (MBMS) within E-UTRAN," TS 36.440, Rel. 11, Sep. 2012.

[28] 3GPP, "Physical layer aspect for evolved universal terrestrial radio access (UTRA)," Tech. Rep. TR 25.814, Rel. 7, Oct. 2006.

[29] C. Mehlfuhrer et al., "Simulating the long term evolution physical layer," in *EUSIPCO*, 2009, pp. 1471-1478.

[30] R. Injong et al., "On the levy-walk nature of human mobility," *IEEE/ACM Transactions on Networking (TON)*, pp. 630-643, 2011.

[31] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "SLAW: mobility model for human walks," in *IEEE Int. Conf. Comput. Commun.*, 2009, pp. 855-863.

[32] 3GPP, TR 36.814, Further advancements for E-UTRA physical layer aspects, V9.0.0, Mar. 2010.

[33] S.-B. Lee, G.-M. Muntean, and A. F. Smeaton, "Performance-aware replication of distributed pre-recorded IPTV content," *IEEE Transactions on Broadcasting*, vol. 55, no. 2, pp. 516-526, 2009.

[34] Y. Han, Z. Yuan, G.-M. Muntean, "An Objective Metric for Non-Intrusive 3D Video Quality Assessment," in *IEEE Int. Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2014.

[35] 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 Conference on Computer Communications (INFOCOM' 15)*, 2015, pp. 1266-1274.

[36] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," DEC res. rep.TR-301, 1984.