Resource Allocation Frameworks for Network-coded Layered Multimedia Multicast Services

Andrea Tassi, Ioannis Chatzigeorgiou and Dejan Vukobratović

Abstract—The explosive growth of content-on-the-move, such as video streaming to mobile devices, has propelled research on multimedia broadcast and multicast schemes. Multi-rate transmission strategies have been proposed as a means of delivering layered services to users experiencing different downlink channel conditions. In this paper, we consider Point-to-Multipoint layered service delivery across a generic cellular system and improve it by applying different random linear network coding approaches. We derive packet error probability expressions and use them as performance metrics in the formulation of resource allocation frameworks. The aim of these frameworks is both the optimization of the transmission scheme and the minimization of the number of broadcast packets on each downlink channel, while offering service guarantees to a predetermined fraction of users. As a case of study, our proposed frameworks are then adapted to the LTE-A standard and the eMBMS technology. We focus on the delivery of a video service based on the H.264/SVC standard and demonstrate the advantages of layered network coding over multi-rate transmission. Furthermore, we establish that the choice of both the network coding technique and resource allocation method play a critical role on the network footprint, and the quality of each received video layer.

Index Terms—Network coding, multicast communication, multimedia communication, mobile communication, resource allocation, LTE-A, eMBMS, H.264/SVC.

I. INTRODUCTION

Multimedia multicast services will soon become a challenging issue to network service providers due to the increasing volume of multimedia traffic. Video content delivery represented 53% of the global mobile Internet traffic in 2013 and is expected to rise to 67% by 2018 [1]. Considering the recent developments in fourth generation (4G) communication networks, a notable fraction of multimedia services is anticipated to be delivered over cellular networks. As the number of users increases, Point-to-Multipoint (PtM) transmission of a multimedia service is the natural choice over multiple and independent Point-to-Point (PtP) sessions. For this reason, 4G cellular networks have native support for broadcasting and multicasting services [2]. Recent work proposes to exploit this attractive inherent feature of 4G networks for broadcasting next generation Digital Television (DTV) services [3]. Furthermore, service multicasting over 4G infrastructures could

A. Tassi and I. Chatzigeorgiou are with the School of Computing and Communications, Lancaster University, Lancaster, UK (email: {a.tassi, i.chatzigeorgiou}@lancaster.ac.uk).

D. Vukobratović is with the Department of Power, Electronics and Communication Engineering, University of Novi Sad, Serbia (email: dejanv@uns.ac.rs).

also be used to deliver extra content in event locations, such as instant replays in sport venues [4].

When a multicast service is transmitted by means of a single PtM data stream, the transmitting node sends the same data stream to all users. Given that users most likely experience heterogeneous propagation conditions, the transmission rate cannot be optimized for each user. Multirate Transmission (MrT) strategies overcome this issue by allowing users to recover different versions of the same PtM service [5]. This paper focuses on MrT strategies that are suitable for *layered services* [6]. A layered service consists of a *base layer* and multiple *enhancement layers*. The base layer allows each user to achieve a basic service quality, which is improved by using information conveyed by the enhancement layers. The ℓ-th enhancement layer can be used to improve the service quality of a user only if both the base and the first $\ell - 1$ enhancement layers have been successfully received by that user. In that context, a MrT strategy adapts the rate of each service layer by taking into account the heterogeneous propagation conditions between the transmitting node and the users.

The main goal of the considered family of MrT strategies is the maximization of the service level experienced by each user [7]. Most proposals divide users into multiple subgroups based on the user propagation conditions; each subgroup will eventually recover a different number of enhancement layers, in addition to the base layer. For example, [8], [9] propose MrT strategies which achieve the aforementioned goal by maximizing the sum of service layers recovered by each user. However, little attention has been paid to the definition of MrT strategies which can ensure that specific subsets of layers will be recovered by predetermined fractions of users.

Our work relies on the MrT principle but proposes resource allocation frameworks that differ from those in the literature in terms of the achieved goal. In particular, we have restated the MrT resource allocation problem from the point of view of the network service provider; we have chosen as the optimization goal the minimization of the total amount of required radio resources to deliver a PtM layered service. Furthermore, owing to the idea of a service-level agreement between the service provider and the cell users, the constraint sets of the proposed optimization frameworks ensure that at least a predetermined fraction of users shall recover a given number of service layers with a target probability. A key point in the proposed MrT frameworks is that reliability of PtM communications is improved by means of the Random Linear Network Coding (RLNC) principle [10]. In particular, the resource allocation goal is fulfilled by jointly optimizing both the transmission parameters and the employed RLNC scheme.

A. Related Works and Paper Contributions

In our system model, each service layer forming a PtM service is delivered over multiple orthogonal broadcast erasure subchannels. Even though Automatic Repeat-reQuest (ARQ) [11] and Hybrid ARQ error control protocols [12] are suitable for broadcast erasure channels, the required amount of user feedback becomes intractable as the number of users grows. In order to mitigate this issue, reliability of multicast communications can be improved by means of Application Level-Forward Error Correction (AL-FEC) techniques, for example schemes based on Luby transform or low-density parity-check codes [2]. Unfortunately, as noted by E. Magli *et al.* [13], this family of codes is designed to be applied over long source messages and, consequently, it introduces delay which is often undesirable in multimedia communications. In order to tackle this problem, several works propose the adoption of RLNC over one-hop broadcast networks [14]–[16]. A key point of RLNC implementations is that short source messages are preferred in order to reduce the decoding complexity and subsequently reduce the communication delay. Furthermore, various RLNC schemes for smartphones and low-end devices have been recently proposed, demonstrating that RLNC strategies are also affordable from the computational complexity point of view [17], [18]. For these reasons, our work adopts the RLNC principle to address the reliability issues of PtM layered service transmissions.

Since each layer of a service has a different importance level, Unequal Error Protection (UEP) can be used to link the level of importance that a service layer has to the required level of protection. The UEP concept has been frequently applied to FEC schemes, see for example Reed-Solomon or low-density parity-check codes [19], [20], but was later adapted for RLNC codes [21]. This paper deals with two different UEP RLNC schemes [21]: the *Non-Overlapping Window* (NOW-RLNC) and the *Expanding Window* RLNC (EW-RLNC). Coded packets associated with a service layer ℓ are generated from source packets of *layer ℓ only* in the case of NOW-RLNC or from source packets of *the first ℓ layers* in the case of EW-RLNC.

Various resource allocation strategies have been proposed to support the transmission of network-coded multimedia flows over unreliable networks [22]–[25]. In particular, [22] considers a system model where several *single-layer* multimedia flows are broadcast to users forming a wireless mesh network. Each user linearly combines those incoming flows that can be decoded by other neighbouring users. Similarly to [22], the system model presented in [23] is also concerned with a mesh network disseminating multimedia flows. However [23] considers layered multimedia streams whose reliability is improved by optimizing a distributed UEP RLNC implementation. In that case, each node realizes the UEP principle such that flows with high importance are more likely to be involved in linear combination operations. Differently to [22], [23], a two-hop content delivery network is studied in [24]. The source node applies network coding to combine packets that form a layered multimedia service. The coded packets are then stored into several intermediate nodes. Subsequently,

a single destination node retrieves the coded packets by connecting to the intermediate nodes via independent PtP sessions. According to the proposed UEP RLNC strategy in [24], which is valid for binary finite fields only, network-coded packets related to low-importance layers may depend on high-importance layers. Contrary to [22]–[24], [25] refers to a cellular network model, where the source node is in charge of generating and transmitting network coded packets to a *single user*. The user acknowledges successfully received packets to the source node. If the acknowledged message is not received, either the same or a new coded packet is transmitted. The core idea of [25] is that of optimizing the encoding process to minimize the total number of transmissions in a single PtP multimedia session.

In contrast to [22]–[24], our work refers to a typical cellular network topology, where the network coding operations are performed by the source node. Furthermore, this paper aims to jointly optimize the network coding process and the transmission parameters. In this way, we can view the RLNC implementation as a component which is fully integrated into the link adaptation framework of our communication system. Our proposal differs from [25] both in terms of the considered RLNC strategies and the nature of the delivered data streams. More specifically, [25] does not consider layered video services and, hence, does not investigate UEP RLNC strategies. Furthermore, the fact that the proposed scheme in [25] has not been integrated into a more generic link adaptation framework hinders its extensibility to the case of PtM services.

Our analysis refers to a generic cellular network model, in a purely standard-independent fashion. However, in order to demonstrate the practical value of the proposed resource allocation frameworks, we present a case study, which refers to the 3GPP Long Term Evolution-Advanced (LTE-A) standard. The proposed implementation shows how our resource allocation frameworks can be adopted for the delivery of multimedia multicast services over the existing and, by following the same implementation guidelines, how can be also extended to next-generation networks.

LTE-A integrates the evolved Multimedia Broadcast Multicast Service (eMBMS) framework, which enables it to handle multicast and broadcast services [26]. In the proposed implementation, we refer to multimedia multicast services that make use of the widely used H.264 video encoding standard and its scalable extension, known as Scalable Video Coding (H.264/SVC), which is gaining popularity [27]. In line with our considered layered message structure, the H.264/SVC encoder transforms a raw video stream into a layered service, such that *enhancement layers* improve the resolution of a *base video layer* of a stream [28]. In order to make the considered network-coded service delivery suitable for multicasting over an LTE-A network, we have adopted the proposal of integrating a RLNC encoder into the LTE-A protocol stack, as described in [29]. In its original version, the proposed integration refers to a system model according to which a *PtP data stream* is transmitted by a base station to a *single user*, either directly or via a relay node. The system design proposed in [29] was later enhanced in [30]

in order to broadcast H.264/SVC video streams as eMBMS flows. Concerning the optimization frameworks that will be presented, this work builds on and extends the idea presented in [30]. In particular, [30] provides a resource allocation model minimizing the total number of transmission attempts needed to broadcast a H.264/SVC video stream. Even though we aim at fulfilling the same objective, this paper significantly differs to [30] in terms of the considered radio resource model. We refer to a generic system model where coded packets are transmitted over a set of orthogonal subchannels. Unlike [30], we develop resource allocation frameworks which *allow* coded packets associated with different video layers to be mixed within the same subchannel to enhance user performance, both in the case of NOW- and EW-RLNC. For any of the proposed resource allocation models, we provide efficient heuristic strategies capable of finding a good quality resource allocation solution in a finite number of steps.

With regards to the coding schemes that we will refer to, unlike [29] and [30], this work focuses on NOW- and EW-RLNC schemes suitable for layered service transmissions. In addition, the authors of [29], [30] did not optimize the bit length of source packets used to represent the transmitted layered service; the source packet bit length is given a priori. This paper proposes a model for optimizing the source packet bit length to fit the transmission constraints of the communication standard in use. Since the bit length of source packets is constrained to be smaller than or equal to a maximum target value, the number of source packets representing a layered service can be upper-bounded. Hence, this work can represent the same layered service with a smaller number of source packets, compared to what proposed in [30]. We remark that the number of source packets has a significant impact on the computation complexity of the RLNC decoding phase [10].

The remaining part of the paper is organized as follows. In Section II, we present the considered standard-independent system model and derive the necessary theoretical foundations needed to assess the performance of NOW- and EW-RLNC. Section III builds upon the aforementioned system model the proposed resource allocation models suitable for optimizing layered multicast communications. Section IV shows, as a case study, how the proposed modelling and resource allocation frameworks can be implemented in a practical communication system, such as LTE-A. Analytical results are discussed in Section V, whereas Section VI summarizes the main findings of the paper.

II. System Parameters and Performance Analysis

We consider an one-hop wireless communication system composed of one source node and U users. Each transmitted data stream is delivered to users through C orthogonal broadcast erasure subchannels. In our system model we have that all the data streams are conveyed to the users according to the RLNC principle. As a consequence, each subchannel delivers streams of network-coded packets (henceforth referred to as *coded packets* for brevity) that may be associated with one or more data streams. Furthermore, we impose that the maximum length of a stream, in terms of the number of coded packets

TABLE I
Commonly used notation.

$\hat{B}_c$	Maximum number of coded packets that can be transmitted over subchannel c
m_c	Modulation and Coding Scheme (MCS) adopted by subchannel c
$p_u(m)$	Packet Error Rate (PER) of user u when MCS with index m is adopted
$\hat{p}$	The reception of a coded packet is acceptable if $p_u(m) \leq \hat{p}$
p_u	Defined in (3) and approximated by (28)
$M(u)$	The greatest value of m for which $p_u(m) \leq \hat{p}$
H	Source/coded packet bit length
$\mathbf{x}$	Layered source message that consists of K equal-length source packets
$\mathbf{w}_\ell$	Set of k_ℓ source packets composing the ℓ-th layer
$\mathbf{W}_\ell$	Set of K_ℓ source packets belonging to the first ℓ service layers
$n^{(\ell,c)}$	Number of coded packets related to layer ℓ layer and transmitted over subchannel c
$\mathbf{n}_u$	Vector $\{n_{1,u}, \ldots, n_{L,u}\}$, where $n_{\ell,u}$ is defined by (2)
$\mathbf{N}_u$	Vector $\{N_{1,u}, \ldots, N_{L,u}\}$, where $N_{\ell,u}$ is defined by (8)
$\mathrm{P}_{1:\ell}^{\mathrm{NOW}}(\mathbf{n}_u)$	Probability that user u will recover the first ℓ service layers, in the case of the NOW-RLNC
$\mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u)$	Probability of user u recovering the ℓ-th window, in the case of the EW-RLNC
$\mathrm{P}_{1:\ell}^{\mathrm{MrT}}$	Probability of user u recovering the ℓ-th window, when the MrT strategy is in use

that can be transmitted over the c-th subchannel during a given time interval, for $c = 1, \ldots, C$, is fixed and equal to $\hat{B}_c$. In particular, we assume that indexes $c = 1, 2, \ldots, C$ are assigned to subchannels so that the relation $\hat{B}_1 \leq \hat{B}_2 \leq \ldots \leq \hat{B}_C$ holds. For clarity, Table I summarizes the symbols commonly used in the paper.

Each element of a coded packet stream is delivered by means of a specific Modulation and Coding Scheme (MCS), which is identified by nonnegative integer m. We denote by $p_u(m)$ the Packet Error Rate (PER) that a user u experiences when m is the index of the adopted MCS. If m' and m'' are indexes identifying two different MCSs and $m' \leq m''$, then the MCS described by m'' either uses a higher modulation order or reduced error-correcting capability than the MCS represented by m'. Naturally, for the same channel conditions, it follows that $p_u(m') \leq p_u(m'')$ also holds. In general, we regard reception of a coded packet as being *acceptable* if $p_u(m)$ is equal to or smaller than a predetermined threshold $\hat{p}$. To this end, if user u can choose from a range of MCSs, we define $M(u)$ as the greatest value of m for which $p_u(m) \leq \hat{p}$, that is

$$M(u) \doteq \{\, m \mid p_u(m) \leq \hat{p} \ \wedge \ p_u(m+1) > \hat{p} \,\}. \quad (1)$$

In the system model presented in this paper, we also impose that coded packets transmitted through the c-th subchannel shall use the same MCS, characterized by index m_c. As will become evident in the rest of the paper, the determination of the optimal MCS for each subchannel, $m_1, \ldots, m_C$, is part of the proposed resource allocation strategies.

Let $\mathbf{x} = \{x_1, \ldots, x_K\}$ be a layered source message that consists of K equal-length source packets, classified into L service layers. For simplicity and without loss of generality, we assume that packets in the source message are arranged in order of decreasing importance. In other words, the first service layer appears at the beginning of the source message and is followed by progressively less important layers, until the last and least important L-th service layer. If the ℓ-th layer consists of k_ℓ data packets, we observe that $K = \sum_{\ell=1}^{L} k_\ell$. Throughout this paper, we define the Quality-of-Service (QoS) level experienced by a user as the number of *consecutive*

source message layers that can be recovered, starting from the first layer.

In the remainder of this section, we present the layered RLNC strategies under consideration. In addition we provide accurate models to evaluate the probability that a source message transmitted by means of NOW-RLNC and EW-RLNC is correctly received by a user. Theoretical results discussed in the rest of this section are general and apply to any cellular system model, where: (i) data flows can be delivered by using different MCSs, and (ii) each source message layer is broadcast through independent communication subchannels.

A. Performance of Non-Overlapping Window RLNC

We first consider the case where the source node uses the RLNC principle on each individual layer of the source message. Let us define K_ℓ as $K_\ell = \sum_{i=1}^{\ell} k_i$. The source node will linearly combine the k_ℓ data packets composing the ℓ-th layer $\mathbf{w}_\ell = \{x_i\}_{i=K_{\ell-1}+1}^{K_\ell}$ and will generate a stream of $n_\ell \geq k_\ell$ coded packets $\mathbf{y} = \{y_j\}_{j=1}^{n_\ell}$, where $y_j = \sum_{i=K_{\ell-1}+1}^{K_\ell} g_{j,i} \cdot x_i$. Coefficient $g_{j,i}$ is uniformly selected at random over a finite field $\mathbb{F}_q$ of size q. We refer to this encoding strategy as NOW-RLNC throughout this paper.

A stream of coded packets associated with a service layer can be broadcast to the U users over a single subchannel or multiple subchannels. Let $n^{(\ell,c)}$ signify the number of coded packets that are related to the ℓ-th layer and are transmitted over the c-th subchannel. We expect that some or all of these $n^{(\ell,c)}$ coded packets will be received by user u, if the predetermined PER requirement is met, i.e. $M(u) \geq m_c$. Otherwise, user u will not recover any of the $n^{(\ell,c)}$ coded packets. We can express the maximum number of coded packets associated with the ℓ-th layer that user u can collect from the C subchannels as

$$n_{\ell,u} = \sum_{c=1}^{C} n^{(\ell,c)} I\left(M(u) \geq m_c\right) \tag{2}$$

where $I(\cdot)$ is an indicator function where $I(\cdot) = 1$ if its input argument is true, otherwise $I(\cdot) = 0$.

To simplify our analysis, we introduce p_u as the maximum PER value experienced by user u across all subchannels that offer acceptable reception and convey at least one coded packet (namely, $\sum_{\ell=1}^{L} n^{(\ell,c)} > 0$), that is

$$p_u \doteq \max_{c=1,\ldots,C} \left(p_u(m_c) \,\middle|\, M(u) \geq m_c \wedge \sum_{\ell=1}^{L} n^{(\ell,c)} > 0 \right). \tag{3}$$

Based on [31], we can infer that if $n_{\ell,u}$ coded packets are transmitted over those subchannels such that $M(u) \geq m_c$, user u will recover the ℓ-th layer with probability

$$P_\ell^{\mathrm{NOW}}(n_{\ell,u}) = \sum_{r=k_\ell}^{n_{\ell,u}} P_R(n_{\ell,u}, r)\, P_{D,\ell}(r) \tag{4}$$

where

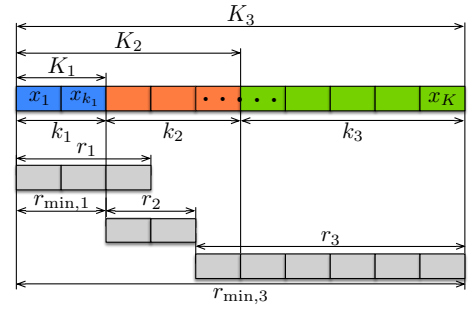$$P_R(n_{\ell,u}, r) = \binom{n_{\ell,u}}{r} p_u^{n_{\ell,u}-r} (1-p_u)^r \tag{5}$$

Fig. 1. Expanding Window source message model, a possible combination of coded packets that have been received and graphic interpretation of $r_{\min,1}$ and $r_{\min,3}$.

represents the probability that r out $n_{\ell,u}$ coded packets are successfully received by user u, when the PER is given by (3). In addition, the term

$$P_{D,\ell}(r) = \prod_{i=0}^{k_\ell-1} \left[1 - \frac{1}{q^{r-i}}\right] \tag{6}$$

is the probability that k_ℓ out of r received coded packets are linearly independent, i.e., $P_{D,\ell}(r)$ is the probability that the source packets forming $\mathbf{w}_\ell$ can be recovered [32]. The joint probability that user u will recover the first ℓ service layers, i.e. $1, 2, \ldots, \ell$, is simply the product of the ℓ individual probabilities, which can be written as

$$P_{1:\ell}^{\mathrm{NOW}}(\mathbf{n}_u) = \prod_{i=1}^{\ell} P_i^{\mathrm{NOW}}(n_{i,u}) \tag{7}$$

where $\mathbf{n}_u = \{n_{1,u}, \ldots, n_{L,u}\}$.

B. Performance of Expanding Window RLNC

We will now shift our focus onto a different RLNC approach known as the expanding window principle, which was presented in [21]. To this end, we consider the same layered source message $\mathbf{x}$ as before, and define the ℓ-th window $\mathbf{W}_\ell$ as the set of source packets belonging to the first ℓ service layers. As depicted in Fig. 1, a window spanning over the first ℓ layers can be expressed as $\mathbf{W}_\ell = \{\mathbf{w}_i\}_{i=1}^{\ell}$ or, equivalently, $\mathbf{W}_\ell = \{x_j\}_{j=1}^{K_\ell}$. In the case of EW-RLNC, the source node (i) linearly combines data packets belonging to the same window, (ii) repeats this process for all windows, and (iii) broadcasts each stream of coded packets associated with each window over one or more subchannels.

In a similar fashion to the NOW case, we define $N^{(\ell,c)}$ as the number of coded packets that are associated with the ℓ-th window and are transmitted over the c-th subchannel. The maximum number of coded packets related to the ℓ-th window that user u can collect through the C subchannels is

$$N_{\ell,u} = \sum_{c=1}^{C} N^{(\ell,c)} I\left(M(u) \geq m_c\right). \tag{8}$$

Using (8), we can obtain vector $\mathbf{N}_u = \{N_{1,u}, \ldots, N_{L,u}\}$, which describes the maximum number of transmitted coded packets, related to each window, that can be collected by user u. The objective of this section is to derive a closed-form

expression for $\mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u)$, which denotes the probability of user u recovering the ℓ-th window and thus retrieving the first ℓ layers. To do that, we shall first consider vector $\mathbf{r} = \{r_1, \ldots, r_L\}$, which describes the number of successfully received coded packets associated with each window, and study the requirements for which $\mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u) > 0$.

For a given set of received coded packets $\mathbf{r}$, we define the minimum number of coded packets associated with the ℓ-th expanding window, denoted as $r_{\min,\ell}$, which shall be successfully received such that the probability of recovering $\mathbf{W}_\ell$, by considering just coded packets associated with the first ℓ expanding windows, is non-zero. Clearly, for $\ell = 1$, we have that $r_{\min,1} = K_1$. Indeed, as per the properties of random network coding, the first window ($\ell = 1$) is likely to be decoded ($\mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u) > 0$) only if: (i) the number of received coded packets pertaining to the first window is at least equal to the number of source packets comprising that window ($r_1 \geq K_1$), or (ii) the probability of recovering a larger window is greater than zero.

Consider Fig. 1, which provides a graphical interpretation of $r_{\min,1}$ and $r_{\min,3}$. In the reported example, given that $r_{\min,1} = K_1$, we note that $r_{\min,1} + r_2$ is less than K_2. Hence, the set of source packets $\mathbf{W}_2$ cannot be recovered because the number of linearly independent coded packets associated with the first two windows cannot be equal to K_2. However, in this case, the value of r_3 is such that $r_{\min,1} + r_2 + r_3$ is equal to K_3. This means that the probability of having K_3 linearly independent coded packets and recovering $\mathbf{W}_3$ is greater than zero. We also note that, in the considered example, the value of r_3 is the smallest one such that $\mathrm{P}_{1:3}^{\mathrm{EW}}(\mathbf{N}_u) > 0$ holds. Hence, $r_{\min,3} = r_{\min,1} + r_2 + r_3$. In general, the remaining values of $r_{\min,\ell}$, for $\ell = 2, \ldots, L$, can be computed using the following recursion:

$$r_{\min,\ell} = K_\ell - K_{\ell-1} + \max(r_{\min,\ell-1} - r_{\ell-1}, \ 0) \qquad (9)$$

which asserts that the probability of decoding the first ℓ layers is non-zero if the number of received coded packets related to the ℓ-th window is at least equal to the size difference between windows ℓ and $\ell-1$, complemented by a possible packet deficit carried over from window $\ell-1$.

Having derived an expression for $r_{\min,\ell}$, for $N_{\ell,u} > 0$, the probability of user u recovering the first ℓ layers, $\mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u)$, can be written as the probability $\mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u, \mathbf{r})$ of successfully receiving $\mathbf{r} = \{r_t\}_{t=1}^\ell$ coded packets *and* recovering the ℓ-th window, summed over all valid values of $\mathbf{r}$. In other words, we can write

$$\mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u) = \sum_{r_1=0}^{N_{1,u}} \cdots \sum_{r_{\ell-1}=0}^{N_{\ell-1,u}} \sum_{r_\ell=r_{\min,\ell}}^{N_{\ell,u}} \mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u, \mathbf{r}). \qquad (10)$$

Let

$$\mathrm{P}_{\mathrm{R}}(\mathbf{N}_u, \mathbf{r}) = \prod_{i=1}^\ell \binom{N_{i,u}}{r_i} p_u^{N_{i,u}-r_i} (1-p_u)^{r_i} \qquad (11)$$

be the probability of receiving r_i out of $N_{i,u}$ coded packets, where the PER is given by (3), for any $i = 1, \ldots, \ell$. Of course, in this case, the term $n^{(\ell,c)}$ in (3) is replaced with $N^{(\ell,c)}$. The relation

$$\mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u, \mathbf{r}) = \mathrm{P}_{\mathrm{R}}(\mathbf{N}_u, \mathbf{r}) \, \mathrm{P}_{\mathrm{D},1:\ell}(\mathbf{r}) \qquad (12)$$
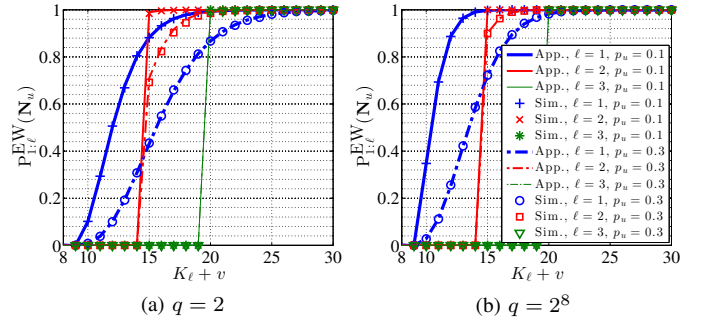

(a) $q = 2$ (b) $q = 2^8$

Fig. 2. Performance comparison between the approximated and simulated version of $\mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u)$, refer to the legend of Fig. 2b for both figures.

considers all possible combinations of receiving $\mathbf{r}$ coded packets out of $\mathbf{N}_u$ packets, multiplied by the probability $\mathrm{P}_{\mathrm{D},1:\ell}(\mathbf{r})$ of successfully decoding the source message $\mathbf{W}_\ell$. Similarly to (6), $\mathrm{P}_{\mathrm{D},1:\ell}(\mathbf{r})$ represents the probability of having K_ℓ linearly independent coded packets out of the $\sum_{i=1}^\ell r_i$ received ones.

Owing to the lack of an accurate expression for $\mathrm{P}_{\mathrm{D},1:\ell}(\mathbf{r})$, we approximated it by using (6). Let $\mathbf{r}' = \{r_t'\}_{t=1}^L$ be a vector of ℓ elements, where $r_t' = 0$ if $t \neq \ell$, otherwise $r_\ell' = \sum_{j=1}^\ell r_j$. It is straightforward to note that the relation $\mathrm{P}_{\mathrm{D},1:\ell}(\mathbf{r}) \leq \mathrm{P}_{\mathrm{D},1:\ell}(\mathbf{r}')$ holds. In addition, from (6), we understand that $\mathrm{P}_{\mathrm{D},1:\ell}(\mathbf{r}')$ is equal to $\mathrm{P}_{\mathrm{D},\ell}(r_\ell')$. For these reasons, we decide to approximate $\mathrm{P}_{\mathrm{D},1:\ell}(\mathbf{r})$ as follows:

$$\mathrm{P}_{\mathrm{D},1:\ell}(\mathbf{r}) \simeq \mathrm{P}_{\mathrm{D},\ell}\left(\sum_{j=1}^\ell r_j\right) = \prod_{i=0}^{K_\ell-1} \left[1 - \frac{1}{q^{(\sum_{j=1}^\ell r_j)-i}}\right]. \quad (13)$$

In order to inspect the quality of the considered approximation, we compared probabilities obtained by using (10) with those obtained by computer simulations, for different values of $p_u = 0.1$ or 0.3 and finite field sizes $q = 2$ or $q = 2^8$. In particular, Fig. 2 compares both the approximated and the simulated value of $\mathrm{P}_{1:\ell}^{\mathrm{EW}}(\mathbf{N}_u)$, where $N_{\ell,u} = K_\ell + v$, for $\ell = 1, \ldots, 3$ and $v \geq 0$. We consider $K_1 = 5$, $K_2 = 10$ and $K_3 = 15$. Note that the maximum performance gap between the approximated and the simulated results occurs for $p_u = 0.3$ and it is smaller than 0.017 for $q = 2$, and 0.004 for $q = 2^8$. The performance gap between approximated and simulated results becomes negligible for an increasing value of q.

III. PROPOSED MULTI-CHANNEL RESOURCE ALLOCATION MODELS AND HEURISTIC STRATEGIES

In this section, we propose strategies that can be used to allocate coded packets over the set of communication subchannels. All the proposed optimization models jointly optimize the MCSs associated with each subchannel and the number of coded packet transmissions. The objective of the proposed models aim at minimizing the total number of coded packet transmissions needed to deliver service layers. This minimization is constrained by the fact that (at least) a predetermined fraction of users shall be able to recover a given set of service layers with (at least) a target probability. For each proposed optimization model, efficient heuristic strategies are provided.
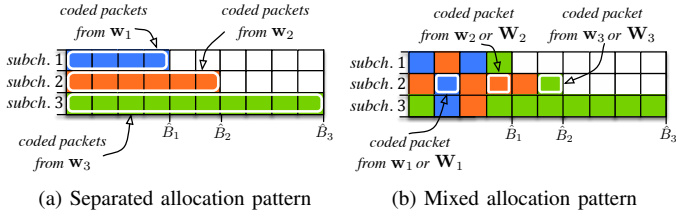
(a) Separated allocation pattern   (b) Mixed allocation pattern

Fig. 3. Considered radio allocation patterns, for $C = 3$ subchannels ("subch.") and $L = 3$ layers.

Before going into the details of the proposed resource allocation models, we consider the following allocation patterns:

- Separated Allocation (SA) pattern (Fig. 3a), where a stream of coded packets associated with a service layer or expanding window shall be mapped on one subchannel only. This means that coded packets belonging to different layers or windows cannot be mixed within the same subchannel.
- Mixed Allocation (MA) pattern (Fig. 3b), where coded packets belonging to different service layers or windows can be delivered through the same subchannel.

In this section we refer to the generic system model described in Section II. Hence, the resource allocation frameworks that will be presented are also generic and standard-independent.

A. Non-Overlapping Window Resource Allocation Strategies

Consider a system where the source node delivers the layered service by means of the NOW-RLNC principle. From (7), we define the indication variable $\lambda_{u,\ell}$ as follows:

$$\lambda_{u,\ell} = I\left(\mathrm{P}_{1:\ell}^{\mathrm{NOW}}(\mathbf{n}_u) \geq \hat{P}\right). \tag{14}$$

In other words, $\lambda_{u,\ell} = 1$, if u can recover the first ℓ layers with a probability value that is equal to or greater than a target value $\hat{P}$, otherwise $\lambda_{u,\ell} = 0$. Equivalently, we can say that if $\lambda_{u,\ell} = 1$, u achieves the QoS level ℓ with at least a probability of $\hat{P}$.

The resource allocation model that we propose for the case of NOW-RLNC employing SA (NOW-SA) can be formulated as follows:

$$\text{(NOW-SA)} \quad \min_{\substack{m_1,\ldots,m_C \\ n^{(1,c)},\ldots,n^{(L,c)}}} \sum_{\ell=1}^{L}\sum_{c=1}^{C} n^{(\ell,c)} \tag{15}$$

$$\text{subject to} \quad \sum_{u=1}^{U}\lambda_{u,\ell} \geq U\,\hat{t}_\ell \qquad \ell = 1,\ldots,L \tag{16}$$

$$m_{c-1} < m_c \qquad c = 2,\ldots,L \tag{17}$$

$$0 \leq \sum_{\ell=1}^{L} n^{(\ell,c)} \leq \hat{B}_c \quad c = 1,\ldots,C \tag{18}$$

$$n^{(\ell,c)} = 0 \qquad\qquad \text{for } \ell \neq c \tag{19}$$

where the objective function (15) represents the overall number of coded packet transmissions needed to deliver all the L service layers. Furthermore, constraint (16) ensures that the fraction of users that can recover the first ℓ service layers is equal to or greater than a desired value $\hat{t}_\ell$. In order to let

---

Procedure 1 Subchannel MCSs optimization.

```
1: c ← C
2: v ← m_MAX and
3: while c ≥ 1 do
4:     repeat
5:         m_c ← v
6:         v ← v − 1
7:     until |U^(m_c)| ≥ U · t̂_c or v < m_min
8:     c ← c − 1
9: end while
```

the model exploit user heterogeneity, constraint (17) avoids the situation in which two subchannels are transmitted using the same MCS. Constraint (18) ensures that the number of coded packets delivered by any subchannel does not exceed $\hat{B}_c$. Constraint (19) avoids that coded packets associated with different service layers are mixed within the same subchannel. Hence, in this case, C has to be equal to or greater than L.

Considering the case of a MA pattern, the service delivery based on the NOW-RLNC approach can be optimized by means of a new optimization model which we shall refer to as NOW-MA. This new optimization model has the same definition of the NOW-SA but, in this case, we remove constraint (19). In this way, coded packets associated with different service layers can be delivered by means of the same subchannel and L can be different from C.

Unfortunately, both the NOW-SA and NOW-MA are hard integer optimization problems because of constraints (16) and (18) that introduce strong coupling relations among delivered service layers. To this end, we propose a couple of two-step heuristic strategies suitable for deriving, in a finite number of iterations, good quality solutions for both aforementioned problems. In particular, the idea underlying each heuristic approach is that of separating the optimization of MCS (associated with each subchannel) from the number of coded packets (related to each service layer) to be delivered.

Considering the SA pattern, the first step of the proposed heuristic strategy aims at optimizing variables m_c, for $c = 1,\ldots,C$. In this case, the value of C has to be equal to L because of the nature of the considered allocation pattern. Furthermore, without loss of generality, we assume that the coded packet stream associated with layer ℓ is delivered by means of the ℓ-th subchannel[1]. Let $\mathcal{U}^{(m_c)}$ be a set of users such that $u \in \mathcal{U}^{(m_c)}$ if $M(u) \geq m_c$. The first step of the heuristic aims at selecting the value of m_c such that the cardinality of $\mathcal{U}^{(m_c)}$, denoted as $|\mathcal{U}^{(m_c)}|$, is equal to or greater than $U \cdot \hat{t}_c$. In particular, this heuristic step, reported in Procedure 1, can be summarized as follows:

(i) Starting from the maximum MCS index m_{MAX} and $c = C$, we select the greatest MCS index such that the number of users in $\mathcal{U}^{(m_c)}$ is equal to or greater than $U \cdot \hat{t}_c$.

(ii) Then, the index c is decreased and the previous step is repeated by considering the MCS index range which goes from $m_c - 1$ to the minimum MCS index m_{min}.

(iii) The procedure iterates while $\ell \geq 1$.

The second step of the heuristic strategy aims at optimizing the variables $n^{(\ell,\ell)}$ (for $\ell = 1,\ldots,L$). In particular, let

---

[1]To this end, in the case of the SA pattern we reference both subchannels and service layers with the same index ℓ.

Procedure 2 Coded packet allocation for a NOW-RLNC service delivery using the MA pattern.

```
1:  c ← 1
2:  n̄^(ℓ,c) ← 0 for any ℓ = 1, ..., L and c = 1, ..., C
3:  n̄ = {n̄^(ℓ)}^L_{ℓ=1}, where n̄^(ℓ) ← 0 for any ℓ = 1, ..., L
4:  for l ← 1, ..., L do
5:      while P^NOW_{1:ℓ}(n̄) < P̂ and c ≤ C do
6:          n̄^(ℓ,c) ← n̄^(ℓ,c) + 1
7:          n̄^(ℓ) ← Σ^C_{t=1} n̄^(ℓ,t) for any ℓ = 1, ..., L
8:          if Σ^L_{t=1} n̄^(t,c) = B̂_c then
9:              c ← c + 1
10:         end if
11:     end while
12:     if P^NOW_{1:ℓ}(n̄) < P̂ and c > C then
13:         no solution can be found.
14:     end if
15: end for
```

$\tilde{n}^{(\ell)}$ be the value of $n^{(\ell,\ell)}$ provided by the heuristic, where $\tilde{\mathbf{n}} = \{\tilde{n}^{(t)}\}^L_{t=1}$. That optimization is summarized as follows:

(i) For any value of $\ell = 1, \ldots L$, $\tilde{n}^{(\ell)}$ is set equal to k_ℓ while $\tilde{n}^{(t)}$, for $t = \ell + 1, \ldots, L$, is set to zero. Then the value of $\tilde{n}^{(\ell)}$ is progressively increased until $P^{NOW}_{1:\ell}(\tilde{\mathbf{n}}) \geq \hat{P}$ does not hold and $\tilde{n}^{(\ell)} \leq \hat{B}_\ell$.

(ii) The procedure iterates while $\ell \leq L$.

It is straightforward to note that the aforementioned heuristic step requires a number of iterations which is equal to or less than $\sum^L_{t=1} \left(\hat{B}_t - k_t + 1 \right)$.

Moving on to the MA pattern, to simplify our analysis, we impose that the number of subchannels has to be equal to the number of service layers, hence, $L = C$. However, the heuristic strategy we propose *does not impose* that all the subchannels have to be used to deliver coded packets. This means that some subchannels could remain unassigned at the end of the allocation process. Concerning the first step of the heuristic strategy, we refer to the same procedure proposed for the SA pattern. For the second heuristic step, in this case, we refer to Procedure 2, which behaves as follows:

(i) We define $\overline{n}^{(\ell,c)}$, for $\ell = 1, \ldots, L$ and $c = 1, \ldots, C$, as the value of $n^{(\ell,c)}$ provided by the heuristic step. At the end of each iteration of the for-loop (lines 4-15), a set of values $\overline{n}^{(\ell,1)}, \overline{n}^{(\ell,2)}, \ldots, \overline{n}^{(\ell,C)}$ are derived, for every service layer. In particular, within the iteration associated with layer ℓ, the value of $\overline{n}^{(\ell,c)}$ is incremented (lines 5-11) as long as the probability of recovering the first ℓ layers is smaller than $\hat{P}$ and $\sum^L_{t=1} \overline{n}^{(t,c)} \leq \hat{B}_c$. If the c-th subchannel cannot hold more packets, the procedure switches to the next subchannel (lines 8-10).

(ii) If the overall number of packets that can be conveyed by all the subchannels is not enough to deliver the coded packet stream associated with the first ℓ layers, the procedure cannot provide a valid allocation (lines 12-14).

It is straightforward to note that Procedure 2 requires at most $\sum^C_{t=1} \hat{B}_t$ iterations.

Consider the second heuristic step of both SA and MA cases; both procedures generate *the same* optimized number of coded packets associated to each service layer. The only difference between the two allocation patterns is that, in the second case, coded packets associated to the same service layer may be transmitted over multiple subchannels.

B. Expanding Window Resource Allocation Strategy

Similar to the NOW-RLNC case, we propose an optimization model suitable for the EW-based service delivery. Due to space limitations, we just focus on the MA allocation pattern.

Before giving the definition of the proposed EW-MA allocation model, it is worth recalling that, from the definition of the EW principle (see Section II-B), we know that user u can recover the first ℓ service layers if the ℓ-th window is recovered, or any window t, for $t = \ell + 1, \ldots, L$, is recovered. Hence, we understand that user u will recover the first ℓ service layers at least with probability $\hat{P}$ if any of the windows $\ell, \ell + 1, \ldots, L$ are recovered (at least) with probability $\hat{P}$. For brevity, from (10), we define the following indicator variable[2]

$$\mu_{u,\ell} = I \left(\bigvee^L_{t=\ell} \left\{ P^{EW}_{1:t}(\mathbf{N}_u) \geq \hat{P} \right\} \right). \tag{20}$$

In other words, $\mu_{u,\ell}$ is equal to one, if u achieve a QoS level equal to or greater than ℓ with at least a probability of $\hat{P}$.

The resource allocation model we propose, called EW-MA, can be expressed as follows:

$$\text{(EW-MA)} \quad \min_{\substack{m_1,\ldots,m_C \\ N^{(1,c)},\ldots,N^{(L,c)}}} \sum^L_{\ell=1} \sum^C_{c=1} N^{(\ell,c)} \tag{21}$$

$$\text{subject to} \quad \sum^U_{u=1} \mu_{u,\ell} \geq U \, \hat{t}_\ell \qquad \ell = 1, \ldots, L \tag{22}$$

$$m_{c-1} < m_c \qquad c = 2, \ldots, L \tag{23}$$

$$0 \leq \sum^L_{\ell=1} N^{(\ell,c)} \leq \hat{B}_c \quad c = 1, \ldots, C \tag{24}$$

Also in this case, the objective function (21) expresses the overall number of coded packet transmissions. Furthermore, constraint (22) imposes that the first ℓ service layers are recovered at least with probability $\hat{P}$ by a fraction of users which shall not be smaller than $\hat{t}_\ell$. Similarly to the NOW-SA and NOW-MA models, constraints (23) and (24) allow the model to exploit the heterogeneity of users.

Unfortunately, the EW-MA model is also a complex integer optimization problem, whose complexity is caused by the coupling constraints among optimization variables given by (22) and (24). To this end, once again, we resort to a two-step heuristic strategy to find a good quality solution of EW-MA, in a finite number of steps.

Once more, for the first step, we refer to the same procedure adopted for the NOW-based allocation models. Let us define $\overline{N}^{(\ell,c)}$ as the value of $N^{(\ell,c)}$ provided by the heuristic step and $\overline{\mathbf{N}} = \{\overline{N}^{(\ell)}\}^L_{\ell=1}$, where $\overline{N}^{(\ell)} = \sum^C_{c=1} \overline{N}^{(\ell,c)}$. Starting from Procedure 2, the second heuristic step has been defined as follows:

(i) For $\ell = 1$ and $c = 1$, $\overline{N}^{(\ell,c)}$ is set to one, while $\overline{N}^{(\ell',c')} = 0$, for $\ell' = \ell, \ldots, L$ and $c' = c, \ldots, C$. The value of $\overline{N}^{(\ell,c)}$ is gradually increased until $P^{EW}_{1:\ell}(\overline{\mathbf{N}}) \geq \hat{P}$ does not hold and $\sum^C_{t=1} \overline{N}^{(\ell,t)} \leq \hat{B}_t$.

---

[2]In this paper, we refer to the logic expression $s_1 \vee s_2 \vee \cdots \vee s_T$ as $\bigvee^T_{t=\ell} s_t$, where $s_1, \ldots, s_T$ are logic statements.

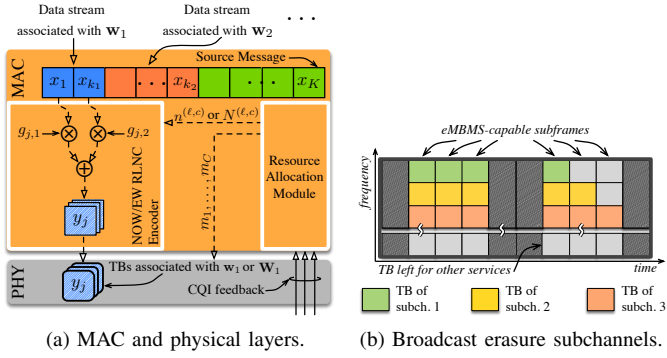(a) MAC and physical layers.

(b) Broadcast erasure subchannels.

Fig. 4. A part of the considered LTE-A protocol stack and model of $C = 3$ broadcast erasure subchannels that span one LTE-A radio frame.

If the subchannel c cannot hold more coded packets, coded packets will be gradually allocated on the next subchannel and the index c is set equal to $c + 1$.

(ii) The value of the index ℓ is increased and the previous steps are repeated. The procedure iterates while $\ell \leq L$ and $c \leq C$.

Finally, likewise to Procedure 2, the aforementioned heuristic step iterates for at most $\sum_{t=1}^{C} \hat{B}_t$ times.

IV. H.264/SVC SERVICE DELIVERY OVER LTE-ADVANCED eMBMS NETWORKS

In order to give an overview of a possible practical implementation of the proposed standard-independent modelling and resource allocation strategies, we refer to the LTE-A standard. Since the first release of LTE-A, PtM communications are managed by means of the eMBMS framework [2].

In the remaining part of the paper, we concentrate on a particular way of delivering PtM services, known as Single Cell-eMBMS (SC-eMBMS) transmission mode [26]. More precisely, we consider a network scenario formed by a base station, henceforth referred to as *target base station*, which delivers a layered video service to a set of users forming a Multicast Group (MG), hereafter called *target MG*. We also assume that all the multicast users are associated to the target base station. In addition, the target base station is surrounded by several interfering base stations, which impair service transmissions to the target MG.

A. Network-coded Video Transmission over eMBMS Networks

In our network scenario, the PtM multimedia service multicast by the target base station is a H.264/SVC video stream formed by L different layers. In particular, the first layer, called *base layer*, provides a basic reconstruction quality, which is gradually improved by the remaining $L - 1$ layers, called *enhancement layers*. In agreement with the layered message structure presented in Section II, the level of the user QoS improves as the number of *consecutive layers* (starting from the base layer) that can be successfully recovered increases.

We assume that each video layer is provided as an independent input of the LTE-A stack. More specifically, the data stream of each layer passes through the Packet Data Conversion Protocol and Radio Link Control layers then, it is forwarded to the Media Access Control (MAC) layer. Since

each video layer has to be delivered by means of the NOW- or EW- RLNC approaches (see Section II-A and II-B), we refer to a modified MAC layer, similar to that proposed in [29], which is in charge of all the network coding operations.

The layered video service produced by a H.264/SVC encoder can be modeled as a stream of Group of Pictures (GoPs) [27]. Each GoP is characterized by fixed number of frames and has a fixed time duration d_{GoP}. In particular, the value of d_{GoP} can be easily obtained by dividing the number of frames of a GoP with the video frame rate. Since the next GoP should be recovered (with a certain QoS) at least by the end of the currently reproduced one, the transmission time of each GoP shall not exceed d_{GoP}.

Given that the decoding process of a H.264/SVC service is performed on a GoP-by-GoP basis, a GoP in our system model represents a layered source message to be delivered according to the network coding principle. We recall from Section II-A that k_ℓ is the number of source packets forming the ℓ-th layer $\mathbf{w}_\ell$ of the source message. Consider Fig. 4a, the MAC layer segments the data stream, forwarded by the higher protocol layers and associated with the ℓ-th video layer of a GoP, into k_ℓ source packets with the same bit length H. Let ν_ℓ be the bitrate associated with the ℓ-th video layer observed at the MAC layer. The term k_ℓ can be defined as $k_\ell = \lceil (\nu_\ell \cdot d_{\mathrm{GoP}}) / H \rceil$.

For each GoP, the MAC layer produces streams of coded packets, each of which has the same size of a source packet. In particular, in the case of the NOW-RLNC, the MAC layer produces one stream of coded packets per video layer. On the other hand, in the case of the EW-RLNC case, one stream per window is generated. We assume that the selection process of coding coefficients is initialized by random number generator (RNG) seeds that are delivered to the multicast user as part of LTE-A signalling data. Delivered RNG seeds are used to re-generate coding coefficients [29].

Each coded packet is forwarded to the physical layer and mapped onto one Transport Block (TB). A TB is a frequency-time structure characterized by a fixed transmission time duration $d_{\mathrm{TTI}} = 1\,\mathrm{ms}$ equal to one Transmission Time Interval (TTI). Each TB may consist of one or more "resource block pairs", which are frequency-time resource units that span a bandwidth of $180\,\mathrm{kHz}$ and have the same transmission time duration of a TB. In other words, the TB bandwidth is an integer multiple of $180\,\mathrm{kHz}$. Furthermore, a TB is transmitted with a certain MCS [26].

We remark that the actual number of bits (referred to as *bit capacity*, in this paper) that a resource block pair can hold depends on the MCS in use. Assuming that a TB can hold just one coded packet, both the number of resource block pairs per TB and the source/coded packet size H have to be selected in order to fit, as tightly as possible, the bit capacity of a TB. To this end, let $\mathrm{N}_{\mathrm{B},m}$ and $\mathrm{N}_{\mathrm{C},m}$ be the number of resource block pairs forming a TB and the bit capacity of a resource block pair, for the m-th MCS, respectively. In this paper, H and $\mathrm{N}_{\mathrm{B},m}$ values have been obtained by solving the following

min-max problem:

$$\min_{N_{B,\min},\ldots,N_{B,MAX},H} \max \quad N_{B,m}N_{C,m} - H \tag{25}$$

$$\text{subject to} \quad N_{B,m}N_{C,m} \geq H \quad m = m_{\min},\ldots,m_{MAX} \tag{26}$$

$$N_{B,m} \leq \hat{N}_B \quad m = m_{\min},\ldots,m_{MAX} \tag{27}$$

where the objective function (25) minimizes the maximum unused bit capacity per TB, for all the possible MCSs. Constraint (26) ensures that the TB bit capacity is at least equal to H, for any MCSs. In addition, constraint (27) imposes that the number of resource block pairs per TB does not exceed a maximum value equal to $\hat{N}_B$. Note that (25)-(27) is an integer optimization problem but it has a modest complexity and can be solved by means of a basic branch-and-bound strategy [33].

B. MAC Layer Augmented Resource Allocation Capabilities

Even though the eMBMS framework enables LTE-A to manage PtM service transmission, the standard delegates the definition and implementation of all the resource allocation operations to the manufactures. However, the standard imposes that the MAC layer is in charge of all the scheduling and resource allocation tasks [34]. For these reasons, we assume that the considered network coding-capable MAC layer is also in charge of allocating resources according to the resource allocation strategies presented in Section III. To this end, we update the subchannel definition given in Section II.

Consider Fig. 4b, which shows the structure of one LTE-A radio frame. One frame is composed of 10 subframes, each subframe has a transmission time duration equal to 1 TTI. At most 6 out of 10 subframes of a radio frame can be used to deliver eMBMS traffic [26], while the remaining subframes are dedicated to PtP traffic. Consider subchannel c, we remark that the maximum number $\hat{B}_c$ of coded packets that can be transmitted over it, during a given time interval, is fixed. Since a TB can hold just one coded packet, we define the subchannel c, as shown in Fig. 4b, as a group of $\hat{B}_c$ TBs, transmitted over eMBMS-capable subframes. In particular, we impose that just one TB per-subchannel can be delivered during a TTI.

For simplicity, in the considered LTE-A scenario, we assumed that $\hat{B}_c = \hat{B}$ (for $c = 1,\ldots,C$), and that the considered fraction of eMBMS-capable subframes per radio frame is 0.6, i.e., 6 out of 10 subframes. Due to the fact that each GoP shall be delivered before transmission of the next GoP begins, the value of $\hat{B}_c$ shall not be greater than $\hat{d}_{GoP} = \lfloor 0.6 \cdot (d_{GoP}/d_{TTI}) \rfloor$ TTIs.

In LTE-A systems, the reception of TB, which adopts a given MCS, is acceptable as long as the TB error rate experienced by a user u is equal to or smaller than 0.1 [26]. The standard allows users to provide Channel Quality Indicator (CQI) feedback to the base station about their propagation conditions. In particular, the CQI feedback provided by a user u indicates the greatest MCS index $m \in [1,15]$ (see Table 7.2.3-1 [35]) such that the TB error probability of u is equal to or smaller than 0.1 [26]. To this end, we set $\hat{p} = 0.1$ in (1). Obviously, the actual PER experienced by each user of the target MG is unknown to the target base station. However, as reported in Fig. 4a, the LTE-A standard imposes that CQI

feedback are directly forwarded to the MAC layer. Hence, it is reasonable to assume that the proposed resource allocation strategies can easily access the CQI information. Owing to the lack of knowledge of the user PER, the target base station approximates the user PER as $p_u(m_c) = \hat{p}$ if $M(u) \geq m_c$, otherwise $p_u(m_c) = 1$. As a consequence, the definition of p_u, provided by (3), is updated as follows:

$$p_u \cong \max_{c=1,\ldots,C} \left(\hat{p} \mid M(u) \geq m_c \wedge \sum_{\ell=1}^{L} n^{(\ell,c)} > 0 \right) \tag{28}$$

where $M(u)$ is equal to the MCS index reported in the CQI feedback provided by user u. For the sake of clarity, we note that the approximation of p_u, given in (28), is considered only by the target base station during the resource allocation operations. On the other hand, all the analytical results and performance assessment, presented in the following sections, will refer to the user PER expression provided in (3).

Consider again Fig. 4a, all the resource allocation operations can be ideally modelled as a functional block of the MAC layer. In the case of the proposed resource allocation strategies, the resource allocation module provides the optimized $n^{(\ell,c)}$ or $N^{(\ell,c)}$, for $\ell = 1,\ldots,L$ and $c = 1,\ldots,C$, to the network coding encoder. In addition, the optimized MCS values $m_1,\ldots,m_C$, associated to each subchannel, are forwarded to the physical layer, which is in charge of transmitting each TB.

Even though this section considered the LTE-A standard, we point out what follows: (i) The generic modelling of Sections II and III can be easily adapted to any OFDMA-based system able to manage PtM communications and hence also future LTE-A releases, (ii) The considered RLNC schemes and the proposed resource allocation strategies should be plugged into the protocol stack layer in charge of allocating radio resources and, (iii) Our practical implementation proposal can be easily adapted to any kind of multimedia layered service.

V. ANALYTICAL RESULTS

We investigate the performance of the proposed resource allocation strategies by considering an LTE-A network formed by a 19 macro-base stations. In particular, we assume that the cell controlled by the target base station (hereafter called *target cell*) is surrounded by 18 interfering macro-base stations, organized in two concentric rings. Each base station manages three hexagonal sectors per cell. Concerning the physical layer and transmission parameters, we referred to the 3GPP's benchmark simulation scenario, called *Case 1* scenario [36], where base stations are characterized by an inter-site-distance of $500\,\text{m}$. Furthermore, we assumed that users forming the target MG are placed outdoors. Hence, all the physical layer parameters have been set by following the guidelines provided in Tables A.2.1.1-2 and A.2.1.1.2-3 of [36]. The first part of Table II summarizes all the remaining system parameters we considered.

In order to provide an effective user QoS assessment, we considered a user distribution characterized by a high heterogeneity from the point of view of the experienced propagation conditions. This means that each user is characterized by a different Signal to Interference plus Noise Ratio (SINR) and

TABLE II
MAIN SIMULATION PARAMETERS.

	Parameter	Value
Physical Layer	Inter-Site-Distance	500 m
	System Bandwidth	20 MHz
	Transmission Scheme	SISO
	Duplexing Mode	FDD
	Carrier Frequency	2 GHz
	Transmission Power	46 dBm per sector
	Base Station and User Antenna Gains	see Table A.2.1.1-2 [36]
	Pathloss and Penetration Loss	see Table A.2.1.1.2-3 [36]
	Channel Model	ITU-T PedA [37]
MAC Layer	C	3 subchannels
	$\hat{\mathrm{N}}_B$	6 resource block pairs
	$\mathrm{N}_{C,m}$, for $m = 4,\dots,15$	see Table 10.1 [26]
	$\mathrm{N}_{B,m}$, for $m = 4,\dots,15$	6, 4, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1
	H	32 KB
	$\hat{P}$	0.99
	$\hat{t}_\ell$, for $\ell = 1,\dots L$	0.99, 0.8, 0.6
	$\hat{B}_c$, for $c = 1,\dots,C$	$K + \lceil K/2 \rceil$ TB transmissions
	q	$[2, 2^8]$

TABLE III
H.264/SVC VIDEO STREAMS CONSIDERED.

Stream	No. Frames per GoP	Frame Rate (fps)	$\nu_1, \dots, \nu_L$ (Mbps)	$\rho_1, \dots \rho_L$ (dB)
News CIF $L = 3$ $\hat{d}_{GoP} = 320$	16	30	2.45, 2.45, 7.35	31.6, 37.4, 43.7
Blue Planet $L = 3$ $\hat{d}_{GoP} = 320$	16	24	1.96, 2.94, 19.60	36.6, 44.5, 51.2

hence, a different PER. In particular, we refer to a target MG of $U = 80$ users that are placed along the radial line representing the symmetry axis of one sector of the target cell. The first user is placed at a distance of $90\,\mathrm{m}$ from the target base station and the distance between two consecutive users is $2\,\mathrm{m}$.

In this performance investigation, we refer to two different video streams encoded using the H.264/SVC Coarse Grain Scalability (CGS) principle. Each layer of a CGS stream successively increases the fidelity of any video frame. In order to do so, H.264/SVC CGS adopts those forms of spatial scalability such that the combination of one or more consecutive layers gives the same spatial frame resolution [38]. Both video streams belong to the video trace database, provided as a companion of [28], and developed for network performance evaluation purposes. The first stream is the *News CIF* (352×288) video sequence [39] composed by $L = 3$ layers, with GoPs of size 16 frames and video frame rate of 30 frame-per-second (fps). The second stream is the *Blue Planet* (1920×1088) video sequence [40] that consists of $L = 3$ layers, GoPs of size 16 frames and video frame rate of 24 fps.

It is worth noting that the bit rate of the video stream obtained by combining all the layers of Blue Planet is 2.8-times greater than that of News CIF video stream. In addition to the main characteristics of the considered video streams, Table III gives the maximum bitrate $\nu_1 \dots, \nu_L$ per-video layer, for each stream. Furthermore, as a performance metric of the video fidelity, Table III provides also the average Peak Signal-to-Noise Ratio (PSNR) ρ_ℓ achieved after successfully recovering the first ℓ video layers, for $\ell = 1,\dots, L$ [28].

In order to inspect the impact of different resource allocation models on the target MG, we developed a system level MATLAB simulator. In particular, we refer to the simulation framework proposed in [29]. Hence, given the physical layer parameters of Table II and Eq. (1) of [29], we evaluated the average SINR value associated to each user in the target MG. Consider (3), in order to assess the user performance, we need the PER value $p_u(m_c)$ associated to the user u and MCS m_c. Unlike [29], we relied on the LTE-A downlink link level simulator presented in [37] to obtain the value of $p_u(m_c)$, as a function of the average user SINR. In particular, for any average SINR value, $p_u(m_c)$ is set equal to the PER value

obtained from the LTE-A downlink link level simulator and averaged over 10^4 simulation runs. Since we are concerned with stationary and low-mobility users, link level simulations have been performed by considering the ITU-T PedA channel model [37]. Hence, by using (1), it is straightforward to emulate the CQI feedback that users provide to the target base station (see Section IV-B).

We remark that the MCS index advertised by CQI feedback may span the interval $[1, 15]$. Since the bitrates ensured by MCSs 1-3 are too small[3] compared to bitrates of the considered video streams, users providing CQI feedback with MCS indexes less than 4 are excluded from the optimization process. For this reason, we set $m_{\min}$ equal to 4, while m_{MAX} is kept equal to 15.

Each video layer of a video stream is delivered by the target base station over $C = 3$ subchannels, as described in Sections II and III. As noted in Section IV, the number $\mathrm{N}_{B,m}$ of resource block pairs forming a TB depends on the MCS index m used to transmit it. Assuming that each TB cannot consists of more than $\hat{\mathrm{N}}_B = 6$, the solution to problem (25)-(27) is reported in Table II. We remark also that the source/coded packet bit size H is part of the aforementioned solution.

Consider the remaining MAC layer simulation parameters of Table II, they are related to the resource allocation strategies (see Section III). In particular, we assumed that consecutive video layers, starting from the base layer, shall be recovered with at least a probability of $\hat{P} = 0.99$. Furthermore, we imposed that at least 99% and 60% of the users forming the target MG shall experience the basic or the maximum QoS, respectively. For simplicity we assume that any subchannel consists of the same number of TBs. Having in mind that the transmission time duration of any layer of a GoP shall not be greater than $\hat{d}_{\mathrm{GoP}}$ we set $\hat{B}_c$ equal to $K + \lceil K/2 \rceil$, as a case of study.

A. Performance Metrics and Benchmark

Performance has been evaluated in terms of the total number of TB transmissions τ needed to deliver all video layers of a GoP. In the remaining part of the paper, we will refer to τ as the *resource footprint*. From the expressions of the objective functions (15) and (21), τ can be defined as follows:

$$\tau = \begin{cases} \sum_{\ell=1}^{L} \sum_{c=1}^{C} n^{(\ell,c)}, & \text{for NOW-RLNC} \\ \sum_{\ell=1}^{L} \sum_{c=1}^{C} N^{(\ell,c)}, & \text{for EW-RLNC} \end{cases} \tag{29}$$

---

[3]For a TB formed by one resource block pair, MCS index $m = 3$ ensures a bitrate smaller than 26.7 kbps, at net of all the signalling information.

where the values of $n^{(\ell,c)}$ and $N^{(\ell,c)}$ have been optimized by the resource allocation strategies presented in Section III. From (7) or (10) we also evaluated user performance in terms of the probability that a user u recovers the first ℓ video layers. Furthermore, we considered, as a third performance metric, the maximum PSNR that user u can achieve, defined as:

$$\rho(u) = \begin{cases} \max\limits_{\ell=1,\ldots,L} \left\{ \rho_\ell \ P_{1:\ell}^{\text{NOW}}(\mathbf{n}_u) \right\}, & \text{for NOW-RLNC} \\ \max\limits_{\ell=1,\ldots,L} \left\{ \rho_\ell \ P_{1:\ell}^{\text{EW}}(\mathbf{N}_u) \right\}, & \text{for EW-RLNC.} \end{cases} \quad (30)$$

Since the users of the target MG are regularly placed on the symmetry axis of the cell-sector, the value of $\rho(u)$ can be equivalently expressed in terms of the distance between the user u and the target base station. In a similar way, parameter $\hat{t}_\ell$ can be interpreted as the minimum distance, from the centre of the target cell, where a user shall recover the first ℓ video layers with a probability of at least $\hat{P}$.

We provide performance comparisons among solutions of NOW-SA, NOW-MA and EW-MA, obtained by the proposed heuristic strategies and by directly solving the aforementioned problems using a genetic strategy (we refer to this kind of solutions as *direct solutions*) [41]. Even though, the direct solution can be considered as a good approximation of the optimal solution of the proposed problems, it is worth noting that a genetic strategy cannot be considered a viable alternative to solve the proposed optimization models in a practical scenario because of its computational complexity [42].

Both the direct and the heuristic solutions of the proposed resource allocation strategies have been compared with a MrT transmission strategy that relies on a standard LTE-A protocol stack. In other words, we referred to a protocol stack which does not adopt RLNC-based service multicasting and does not rely on any AL-FEC strategy. For the implementation of the considered MrT strategy, we refer to the resource allocation strategy proposed in [8], [9] which aims at maximizing the sum of the video quality experienced by each user. In particular, this goal is achieved by optimizing the MCS index m_ℓ used to deliver the TB stream holding data associated with the ℓ-th video layer, for $\ell = 1, \ldots, L$.

It is worth noting that both [8] and [9] implicitly refer to a concept that is similar to the SA pattern. Specifically, data streams associated to different video layers are independently transmitted to the target MG. Assume that the ℓ-th video layer is delivered with the MCS with index m_ℓ. We understand that, in the case that the target base station relies on the standard LTE-A protocol stack, the uncoded transmission of TBs associated to video stream ℓ is equivalent to the transmission of all the k_ℓ TBs defining the ℓ-th layer $\mathbf{w}_\ell$ of a GoP. In order to make fair comparisons, we referred here to the same values of $N_{B,m}$ reported in Table II. For these reasons, the probability $P_{1:\ell}^{\text{MrT}}$ that user u recovers the first ℓ layers can be expressed as $P_{1:\ell}^{\text{MrT}} = P_{1:\ell-1}^{\text{MrT}} \cdot [1 - p_u(m_\ell)]^{k_\ell}$ where, $P_{1:1}^{\text{MrT}} = [1 - p_u(m_1)]^{k_1}$. In this case, the maximum PSNR that u can achieve is $\rho(u) = \max\limits_{\ell=1,\ldots,L} \left\{ \rho_\ell \cdot P_{1:\ell}^{\text{MrT}} \right\}$. Hence, we
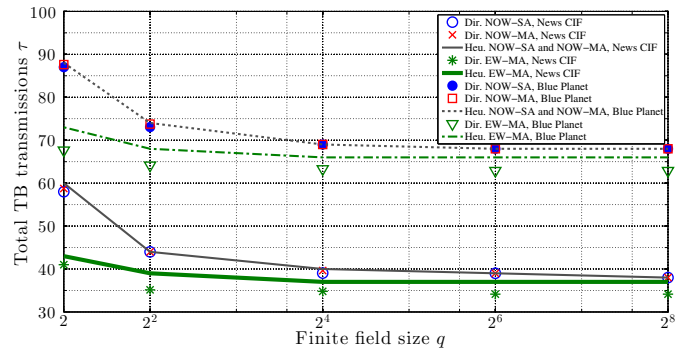

Fig. 5. Overall number of TB transmissions associated with all the proposed resource allocation frameworks.

expressed the considered MrT strategy as follows:

$$(\text{MrT}) \quad \max_{m_1,\ldots,m_L} \sum_{u=1}^{U} \rho(u) \quad (31)$$

$$\text{subject to} \quad m_{\ell-1} < m_\ell \qquad \ell = 2, \ldots, L. \quad (32)$$

As well as in the case of the proposed resource allocation strategies, the exact value of $p_u(m_\ell)$ is unknown at the target base station side. Hence, during the resource allocation based on MrT, the PER expression is approximated as $p_u(m_\ell) \cong \hat{p}$ if $M(u) \geq m_\ell$, otherwise $p_u(m_\ell) \cong 1$.

B. Assessment of the Heuristic Solutions

Let us start our performance investigation from Fig. 5, it compares the number of TB transmissions, represented by τ, which are associated with the direct ("Dir.") and heuristic ("Heu.") solutions, of all the proposed resource allocation strategies, as a function of the finite field size q over which all the RLNC-related operations are performed. The figure shows results for both News CIF and Blue Planet streams. Due to the fact that τ represents the value of the objective functions of the proposed optimization models, it allows us to inspect the performance gap between each proposed heuristic strategy and the corresponding direct solution. We remark that the number of TB transmissions associated with a direct solution is unlikely to be greater than that associated with a heuristic solution [41]. As clearly shown in Fig. 5, the performance gap between the heuristic and the direct solutions is negligible. In particular, the gap is at most equal to 2, 1 and 5 TBs for the NOW-SA, NOW-MA and EW-MA models, respectively. For this reason, in the rest of this section, we refer only to the heuristic solutions of the proposed resource allocation models.

We also observe in Fig. 5 that the value of τ, of any resource allocation model, decreases as the the value of q increases. We understand that, for an increasing value of q, the probability of receiving coded packets that are linearly dependent with the previous ones decreases. As a consequence, the resource footprint of each allocation strategy decreases, as the finite field size increases. However, for small finite field sizes, there is a remarkable gap between any solution based on a NOW-RLNC strategy and the direct/heuristic EW-MA solution. In particular, for $q = 2$, the gap between the heuristic solution of EW-MA and, either NOW-SA or NOW-MA, is equal to 17 TBs.
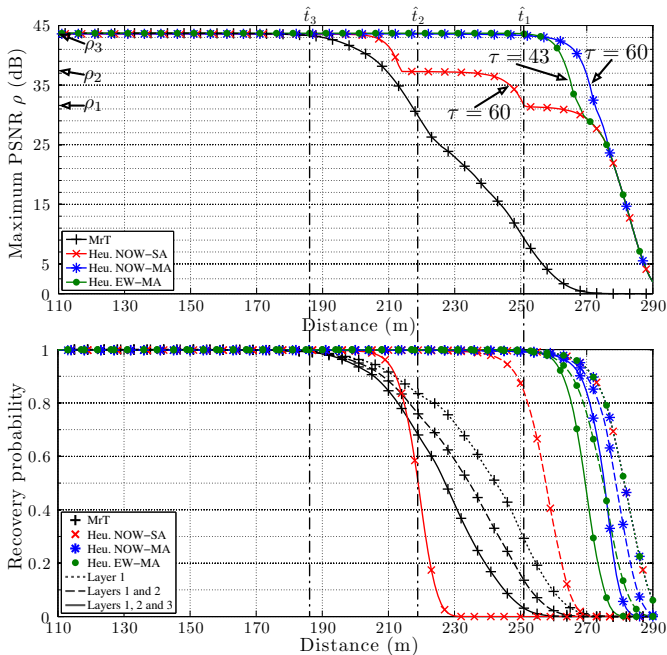
Fig. 6. Maximum PSNR and probability of recovering a given set of video layers associated with stream News CIF, for $q = 2$.

Fig. 7. Maximum PSNR and probability of recovering a given set of video layers associated with stream Blue Planet, for $q = 2$.

C. Performance Evaluation of the Proposed Frameworks

Focusing on a finite field with $q = 2$, Fig. 6 shows both the maximum PSNR ρ, and the probability of recovering the first ℓ video layers as a function of the distance from the centre of the cell. We recall that the MrT strategy does not rely on any AL-FEC or RLNC-based strategy. In addition, due to the fact that the MrT aims at maximizing the sum of the video quality achieved by all the users, its performance in terms of coverage diverges from both that of the proposed strategies and the target performace. In particular, we note that the base video layer can be received at least with a probability of 0.99 up to a distance of $188\,\mathrm{m}$ from the centre of the cell. The MrT performance confirms the idea underlying the proposed optimization strategies; that is defining allocation models where the constraint set ensures that a target coverage and objective function minimizes the amount of resources needed to deliver the multicast service.

We see also in Fig. 6 that, even though all the proposed allocation models meet the coverage constraints, strategies based on the MA pattern provide better coverage than that associated with the SA pattern. In particular, due to the fact that MA pattern can exploit the user heterogeneity better than the SA one, both NOW-MA and EW-MA can successfully deliver all the video layers up to a distance of $252\,\mathrm{m}$. On the other hand, the NOW-SA model ensures the maximum service quality only up to $203\,\mathrm{m}$. Furthermore, from Fig. 6, we understand that the τ value of the heuristic EW-MA strategy is $\sim 28\%$ smaller than that of the NOW-MA and NOW-SA heuristic solutions. In particular, we can argue that both the NOW-MA and EW-MA strategies achieve almost the same coverage performance but the second one requires a smaller resource footprint. Finally, as expected (see Section III), both the heuristic NOW-SA and NOW-MA models are characterised by the same values of τ.

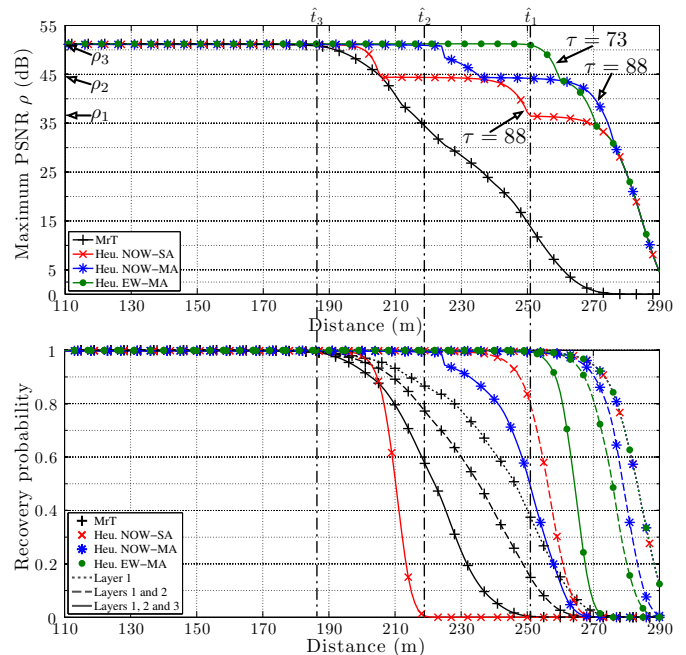Fig. 7 compares the same performance metrics considered in

Fig. 6 (for $q = 2$), associated with the stream Blue Planet. We remark, the overall bitrate of stream Blue Planet is greater than that of the stream News CIF. Also in this case, we note that all the proposed resource allocation solutions meet the target service constraints. As shown by Fig. 5, the τ value associated with the heuristic EW-MA strategy is $\sim 17\%$ smaller than that of the heuristic NOW-SA/NOW-MA solution. Furthermore, the EW-MA strategy provides a resource allocation solution such that all the video layers can be successfully recovered up to a distance of $252\,\mathrm{m}$, which is $28\,\mathrm{m}$ greater by than that ensured by NOW-MA. In accordance with stream News CIF, we observe that the heuristic NOW-SA provides allocation solutions such that all the video layers can be recovered up to a distance that is $27\,\mathrm{m}$ ($55\,\mathrm{m}$) smaller, respectively, than that associated with the heuristic NOW-MA (EW-MA). Finally, also in this case, the performance of MrT diverges from the performance of the proposed strategies.

Fig. 8 shows the value of ρ associated with the streams News CIF and Blue Planet, as a function of distance from the centre of the cell, for $q = 2^8$. For the sake of comparison, we also report the performance of MrT even if it does not depend on the value of q. We recall from Fig. 5 that the performance gap, in terms of the value of τ, between the heuristic NOW-SA/NOW-MA and EW-MA solutions is small (2 TBs). As expected, the heuristic NOW-MA solution provides a service coverage that overlaps with that given by the heuristic EW-MA, in the case of both video streams. We can thus conclude that NOW-MA and EW-MA strategies perform similarly both in terms of resource footprint and service coverage, for large value of q. Furthermore, even though the NOW-SA approach is characterized by the same resource footprint of NOW-MA, the achieved service coverage still diverges from that of NOW-MA and EW-MA. Once more, this performance gap is caused by the fact that the NOW-SA
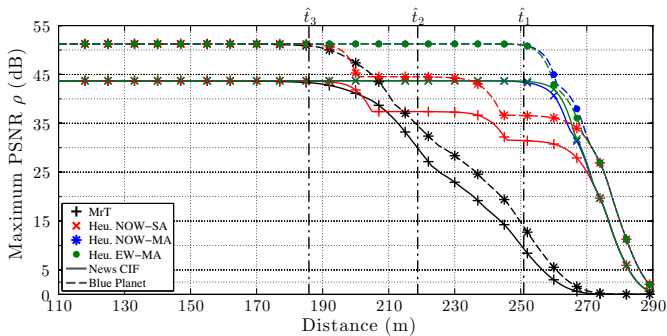
Fig. 8. Maximum PSNR associated with video streams News CIF and Blue Planet, for $q = 2^8$.

approach cannot exploit user heterogeneity. Finally, we remark that, also in this case, all the proposed allocation models meet the required coverage constraints.

We demonstrated that the proposed resource allocation frameworks fulfil the desired goals set in Section I, namely (i) to ensure the desired QoS levels to at least a target fraction of users, and (ii) to minimize the required number of TB transmissions. In particular, we established that each proposed framework ensures a service coverage, which not only meets the target performance but also outperforms the service coverage provided by the considered MrT strategy. A fact that should be kept in mind is that the increased service coverage factor depends on the user propagation conditions and the bitrate of each service layer. However, if the transmitted multicast services have similar bitrates, the increased service coverage can be directly translated into an enlargement of the inter-site-distance or a better placement of the base stations.

VI. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this paper, we studied a general system model in which a source node uses point-to-multipoint (PtM) transmission to multicast a layered message to a group of users. The number of consecutive layers recovered by a user determines the QoS level of that user. In order to improve communication reliability, we considered Random Linear Network Coding (RLNC) and we investigated two different implementations, which are suitable for layered source messages: the Non-Overlapping Window (NOW-RLNC) and the Expanding Window (EW-RLNC) schemes. We derived accurate closed-form expressions for the probability of recovering a predetermined set of consecutive message layers for both NOW-RLNC and EW-RLNC and we used these expressions to assess the user QoS. To maintain the generality of the system model and facilitate its extendibility to 4G and next-generation standards, we assumed that a layered source message can be transmitted over multiple orthogonal communication subchannels.

Based on this assumption, we developed resource allocation frameworks which aim to minimize the overall number of coded packet transmissions. The proposed frameworks allocate coded packets of the same layer or the same expanding window either to a single or to multiple subchannels; we called the former pattern Separated Allocation (SA) while the latter pattern Mixed Allocation (MA). A key point in the formulation of the resource allocation problems is that the derived solutions ensure that predetermined fractions of users can achieve the desired QoS with at least a target probability. We explained that both SA and MA are computationally complex integer problems but we proposed heuristic strategies which are capable of obtaining good-quality solutions in a finite number of steps.

As a case study, we presented a possible integration of the RLNC-based schemes into the standard LTE-A Media Access Control (MAC) layer and the adaptation of the developed resource allocation frameworks to LTE-A systems. In addition, we described how the resulting modified MAC layer can be used to efficiently deliver a layered multimedia stream compliant with the H.264/SVC standard over an LTE-A network that operates in the Single-Cell eMBMS mode.

In order to investigate the performance of the proposed schemes, we referred to an LTE-A network scenario defined by 3GPP to benchmark urban cellular network deployments. Furthermore, we considered two video traces – one of low bitrate and the other of high bitrate – both of which are publicly available for network performance evaluation. The first part of our investigation compared heuristic solutions to solutions obtained by directly solving the optimization problems and established that our proposed heuristic strategies indeed produce good-quality solutions. In the second part of our analysis, we demonstrated that both NOW and EW schemes can offer the same quality of service, in terms of PSNR, as conventional multi-rate transmission (MrT) but over a much longer distance. For a 99% probability of recovering the base video layer, we showed that the proposed strategies can achieve a coverage that is greater than that of a conventional MrT strategy by a factor of at least 1.35. Furthermore, we unveiled that EW-MA can achieve similar coverage to that of NOW-SA and NOW-MA but at a notable resource advantage when binary network coding is used. More specifically, EW-MA can reduce packet transmissions by 28% and 17% for the case of the considered low and high bitrate streams, respectively. Nevertheless, we clarified that as the field size of network coding increases, the NOW and EW schemes perform similarly.

Future research directions involve the optimization of the sparsity of RLNC as well as the definition of different optimization objectives. In this paper, we employed the classic implementation of RLNC, where coding coefficients are randomly selected over a finite field. It is well known from the literature that the coding coefficient selection can be biased in order to increase the probability of selecting a zero coefficient. We understand that, as the sparsity of a coding vector increases, the RLNC decoding complexity decreases. However, the more zero coefficients a coding vector has, the higher the probability is that a user receives linearly dependent coded packets. Owing to the lack of a theoretical characterization of the tradeoff between sparsity and decoding complexity, we will strive to reinterpret both the NOW-RLNC and EW-RLNC approaches. The resulting theoretical characterization will allow us to jointly optimize transmission parameters and the sparsity of RLNC.

REFERENCES

[1] Cisco Visual Networking Index, "Global mobile data traffic forecast update, 2013-2018," 2014.

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

[3] G. Walker, J. Wang, C. Lo, X. Zhang, and G. Bao, "Relationship Between LTE Broadcast/eMBMS and Next Generation Broadcast Television," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 185–192, Jun. 2014.

[4] P. Rysavy, "Challenges and Considerations in Defining Spectrum Efficiency," *Proc. of the IEEE*, vol. 102, no. 3, pp. 386–392, Mar. 2014.

[5] W. tian Tan and A. Zakhor, "Video Multicast using Layered FEC and Scalable Compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 373–386, Mar. 2001.

[6] S. Lakshminarayana and A. Eryilmaz, "Multirate Multicasting With Intralayer Network Coding," *IEEE/ACM Trans. Netw.*, vol. 21, no. 4, pp. 1256–1269, Aug. 2013.

[7] R. Afolabi, A. Dadlani, and K. Kim, "Multicast Scheduling and Resource Allocation Algorithms for OFDMA-Based Systems: A Survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 240–254, First 2013.

[8] H. Zhang, Y. Zheng, M. Khojastepour, and S. Rangarajan, "Scalable Video Streaming over Fading Wireless Channels," in *Proc. of IEEE WCNC 2009*, Budapest, Hungary, HU, Apr. 2009.

[9] D. Munaretto, D. Jurca, and J. Widmer, "Scalable video broadcast in cellular networks: impact on QoS and network resources," in *Proc. of IEEE ISCC 2010*, Riccione, Italy, IT, 2010.

[10] M. Médard and A. Sprintson, *Network Coding: Fundamentals and Applications*, ser. Academic Press. Elsevier, 2012.

[11] B.-S. Kim, S. W. Kim, and R. Ekl, "OFDMA-Based Reliable Multicasting MAC Protocol for WLANs," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 3136–3145, Sep. 2008.

[12] J. Kim, H. Jin, D. K. Sung, and R. Schober, "Optimization of Wireless Multicast Systems Employing Hybrid-ARQ with Chase Combining," *IEEE Trans. Veh. Technol.*, vol. 59, no. 7, pp. 3342–3355, Sep. 2010.

[13] E. Magli, M. Wang, P. Frossard, and A. Markopoulou, "Network Coding Meets Multimedia: A Review," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1195–1212, 2013.

[14] M. Ghaderi, D. Towsley, and J. Kurose, "Reliability Gain of Network Coding in Lossy Wireless Networks," in *Proc. of IEEE INFOCOM 2008*, Phoenix, Arizona, US-AZ, Apr. 2008.

[15] A. Eryilmaz and L. Ying, "Scaling Laws with Network Coding," in *Network Coding: Fundamentals and Applications*, ser. Academic Press. Elsevier, 2012, pp. 235–265.

[16] N. Xie and S. Weber, "Network coding broadcast delay on erasure channels," in *Proc. of Information Theory and Applications Workshop 2013*, San Diego, California, US-CA, Feb. 2013.

[17] D. Ferreira, R. Costa, and J. Barros, "Real-Time Network Coding for Live Streaming in Hyper-Dense WiFi Spaces," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 4, pp. 773–781, Apr. 2014.

[18] M. V. Pedersen, D. E. Lucani, F. H. Fitzek, C. W. Sorensen, and A. S. Badr, "Network coding designs suited for the real world: What works, what doesn't, what's promising," in *Proc. of IEEE Information Theory Workshop 2013*, Sevilla, Spain, ES, Sep. 2013.

[19] Y. Huo, M. El-Hajjar, R. Maunder, and L. Hanzo, "Layered Wireless Video Relying on Minimum-Distortion Inter-Layer FEC Coding," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 697–710, Apr. 2014.

[20] H. Ha and C. Yim, "Layer-Weighted Unequal Error Protection for Scalable Video Coding Extension of H.264/AVC," *IEEE Trans. Consum. Electron.*, vol. 54, no. 2, pp. 736–744, May 2008.

[21] D. Vukobratović and V. Stanković, "Unequal Error Protection Random Linear Coding Strategies for Erasure Channels," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1243–1252, May 2012.

[22] H. Seferoglu and A. Markopoulou, "Video-Aware Opportunistic Network Coding over Wireless Networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 713–728, Jun. 2009.

[23] N. Thomos, J. Chakareski, and P. Frossard, "Prioritized Distributed Video Delivery With Randomized Network Coding," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 776–787, Aug. 2011.

[24] K. Nguyen, T. Nguyen, and S.-C. Cheung, "Video Streaming with Network Coding," *J. Signal Process. Syst.*, vol. 59, no. 3, pp. 319–333, Jun. 2010.

[25] D. Nguyen, T. Nguyen, and X. Yang, "Joint Network Coding and Scheduling for Media Streaming Over Multiuser Wireless Networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1086–1098, Mar. 2011.

[26] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons, 2011.

[27] ITU-T H.264, "Advanced Video Coding for Generic Audiovisual Services," Tech. Rep., Nov. 2007.

[28] P. Seeling and M. Reisslein, "Video Transport Evaluation With H.264 Video Traces," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 1142–1165, Fourth 2012.

[29] C. Khirallah, D. Vukobratović, and J. Thompson, "Performance Analysis and Energy Efficiency of Random Network Coding in LTE-Advanced," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4275–4285, 2012.

[30] A. Tassi, C. Khirallah, D. Vukobratovic, F. Chiti, J. Thompson, and R. Fantacci, "Resource Allocation Strategies for Network-Coded Video Broadcasting Services over LTE-Advanced," *IEEE Trans. Veh. Technol.*, 2014, in press.

[31] F. Chiti, R. Fantacci, F. Schoen, and A. Tassi, "Optimized Random Network Coding for Reliable Multicast Communications," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1624–1627, 2013.

[32] O. Trullols-Cruces, J. Barcelo-Ordinas, and M. Fiore, "Exact Decoding Probability Under Random Linear Network Coding," *IEEE Commun. Lett.*, vol. 15, no. 1, pp. 67–69, Jan. 2011.

[33] P. Belotti, "COUENNE: Convex Over and Under ENvelopes for Nonlinear Estimation." [Online]. Available: http://www.coin-or.org/Couenne/

[34] 3GPP TS 36.321 v12.2.1, "Medium Access Control (MAC) protocol specification," 2014.

[35] 3GPP TS 36.213 v12.2.0, "Physical layer procedures," 2014.

[36] 3GPP TR 36.814 v9.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects," 2010.

[37] S. Schwarz, J. Ikuno, M. Simko, M. Taranetz, Q. Wang, and M. Rupp, "Pushing the limits of LTE: A survey on research enhancing the standard," *IEEE Access*, vol. 1, pp. 51–62, 2013.

[38] R. Gupta, A. Pulipaka, P. Seeling, L. Karam, and M. Reisslein, "H.264 Coarse Grain Scalable (CGS) and Medium Grain Scalable (MGS) Encoded Video: A Trace Based Traffic and Quality Evaluation," *IEEE Trans. Broadcast.*, vol. 58, no. 3, pp. 428–439, Sep. 2012.

[39] "YUV Video Sequences." [Online]. Available: http://trace.eas.asu.edu/yuv/

[40] "Video Traces 2 for Blue Planet." [Online]. Available: http://trace.eas.asu.edu/videotraces2/cgs/1080/BluePlanet_G16B3_DQP8_4EL_48_40_32_24_16/

[41] K. Deep, K. Pratap Singh, M. Kansal, and C. Mohan, "A real coded genetic algorithm for solving integer and mixed integer optimization problems," *Applied Mathematics and Computation*, vol. 212, no. 2, pp. 505–518, 2009.

[42] D. Goldberg, *Genetic Algorithms*. Pearson Publishing, 2013.