Reliable Rate-Optimized Video Multicasting Services over LTE/LTE-A

Andrea Tassi*, Chadi Khirallah†, Dejan Vukobratović‡, Francesco Chiti*, John Thompson†, Romano Fantacci*
*Department of Electronics and Telecommunications, University of Florence, Florence, Italy
†School of Engineering, The University of Edinburgh, Edinburgh, UK
‡Dept. of Power, Electronics and Communication Engineering, University of Novi Sad, Serbia

Abstract—In this paper, we propose a novel advanced multi-rate design for evolved Multimedia Multicast/Broadcast Service (eMBMS) in fourth generation (4G) Long-Term Evolution (LTE)/LTE-Advanced (LTE-A) networks. The proposed design provides: i) reliability, based on random network coded (RNC) transmission, and ii) efficiency, obtained by optimized rate allocation across multi-rate RNC streams. The paper provides an in-depth description of the system realization and demonstrates the feasibility of the proposed eMBMS design using both analytical and simulation results. The system performance is compared with popular multi-rate multicast approaches in a realistic simulated LTE/LTE-A environment.

I. INTRODUCTION

Video content delivery over fourth generation (4G) networks, such as Long-Term Evolution (LTE)/LTE-Advanced (LTE-A), is estimated to grow 18-fold between 2011-2016 [1]. This growth is mainly attributed to the surge in demand for bandwidth-intensive mobile multimedia services by LTE-based smartphones, tablets and laptops. 3GPP standards provide a bandwidth efficient broadcast and multicast solution that is suitable for the simultaneous delivery of multimedia content to large groups of users; namely, the Multimedia Broadcast/Multicast Service (MBMS) [2].

Current MBMS services use the Unacknowledged Mode at the radio link control (RLC) layer for data transmission due to the lack of ability to process acknowledgement messages from users (UEs). Thus, MBMS avoids the use of complex automatic repeat request (ARQ) schemes that can result in significant packet delays caused by the retransmission processes that are used to recover lost packets. Nonetheless, in order to ensure reliable data delivery in multicast networks, in the case of delay-sensitive applications, 3GPP standards have integrated application-layer forward error-correction (AL-FEC) solutions. In particular, Raptor codes [3] are included in the MBMS standard to provide protection against channel erasures.

Recently, the proliferation of packet-level network coding techniques has increased interest in application-layer random network coding (AL-RNC) as a viable AL-FEC solution. However, AL-FEC leads to a large amount of redundancy (delays) on the end-to-end links, compared to the short real-time multimedia message transmission time. They also require additional resources over the entire data path, including a reliable over-provisioned core network infrastructure. For these reasons, in [4], the integration of RNC solutions at the medium access control (MAC) layer of the LTE/LTE-A radio access network (RAN), is proposed to exploit the very small MAC round-trip delay, in addition to a number of MAC-RNC benefits specific to media streaming services [5].

The focus of the paper is to investigate a novel advanced multi-rate design for the MBMS service. The proposed approach is based on the base station (eNB) delivering multiple MAC-RNC protected multicast streams matched to the set of heterogeneous user classes in the cell. The core of the proposed solution is an optimized rate-allocation strategy applied to share the available bandwidth resources among different multi-rate RNC streams. The paper demonstrates the feasibility of the proposed RNC-based rate-optimized MBMS service through analytical and numerical simulation results. Its efficiency is demonstrated by evaluating its performance against the most popular multi-rate multicast approaches with realistic LTE/LTE-A parameter settings.

The paper is organized as follows. Section II provides the necessary background on the existing MBMS service in LTE/LTE-A and an overview of the MAC-RNC proposal for LTE/LTE-A. In Section III, we present the proposed multi-rate MBMS solution. Section IV provides both analytical and simulation results. Finally, conclusions to the paper are given in Section V.

II. BACKGROUND

A. Multimedia Multicasting Services in LTE/LTE-A

3GPP standards provided several designs for MBMS starting from Release 6, which allowed point-to-multipoint transmission schemes [6]. In Release 8 [7], the enhanced MBMS (eMBMS) design is introduced with two proposed transmission schemes. The single-cell MBMS mode (SC-eMBMS) overcomes previous shortcomings by allowing user feedback on channel conditions and dynamic selection of the suitable modulation and coding (MC) mode. The multi-cell, or so called single frequency network MBMS mode (SFN-eMBMS), provides a coordinated effort of several eNBs to cover the network with the same physical signal. In this mode, a fixed MC scheme is adapted to match the edge-of-cell user requirements. In this paper, we focus on SC-eMBMS in order to exploit the dynamic data rate adaptation to different users in the cell [6].

B. Related Work

Several papers [8], [9], [10] addressed the problem of rate adaptation in multicast communication systems. The proposed techniques follow the fixed-rate or the multi-rate approaches in the system design. For the fixed-rate approach, the eNB
broadcasts a multicast flow to all UEs using the same MC scheme (i.e., the same transmission rate). One example is the Fixed Rate Strategy (FR-S) [8], where the eNB selects a rate allocation that satisfies a given set of Quality of Service (QoS) constraints across all UEs, including the cell-edge UEs. Unfortunately, FR-S can match only a very basic set of QoS requirements. An evolution of FR-S is represented by the Least Channel Gain Strategy (LCG-S) [9]. In the LCG-S case, a multicast flow is transmitted with a data rate that matches the delivered QoS to the state of UEs in the worst propagation conditions (not necessarily the cell-edge UEs).

Multi-rate rate adaptation strategies address the shortcomings of the fixed-rate approaches by transmitting several data streams of different levels of QoS that match the heterogenous UEs channel conditions in the cell. An example of the multi-rate strategy is the Information Decomposition Techniques (IDT) [10]. The IDT-based approach splits each multicast flow into several sub-flows so that UEs in the worst channel conditions would be able to decode the lowest channel rate sub-flows (i.e., lowest QoS), while UEs in a better channel conditions can decode, beside the basic sub-stream, the higher data rate sub-streams, and eventually combine those sub-streams to yield high QoS results. In other words, in contrast to the fixed-rate rate-allocation techniques, UEs experiencing ideal propagation conditions are no longer penalized, in terms of achieved rates/QoS, by the UEs in the worst channel conditions in the cell. A similar multi-rate approach for video delivery over broadband cellular networks has recently been investigated in [11].

C. MAC-RNC for Reliable Multicasting in LTE/LTE-A

In [4], a RNC-based solution integrated into the LTE/LTE-A RAN that is able to exploit very small round-trip delays at LTE MAC layer is considered. MAC-RNC is designed to improve performance, reduce complexity, increase possibilities for video-content awareness and fully exploit the benefits of rateless/network coding across the LTE RAN.

Fig. 1 shows the downlink flow of IP packets in the proposed MAC-based RNC approach in the LTE protocol stack. The main difference between the standardized solution is the MAC-RNC sublayer integrated on top of the MAC layer. In short, instead of segmenting each RLC layer packet (RLC PDU) into MAC packets (MAC PDUs) and transmitting each MAC PDU independently over the physical layer (PHY), the MAC-RNC solution segments the incoming RLC PDU into \( K \) equal size source symbols and produces a stream of RNC-coded symbols. An appropriate number of encoded symbols are grouped into a MAC PDU to fit the PHY transport block (TB) size, which depends on the selection of the MC scheme at the PHY layer. From each correctly received PHY TB at the UE, the set of encoded packets is extracted and delivered to the progressive GE decoder at the MAC-RNC decoding sublayer. As soon as \( K \) linearly independent encoded packets are received from the stream of MAC PDUs, the UE MAC entity reports an ACK message to end the RLC PDU delivery. The MAC-RNC scheme details are available in [4].

III. RATE-OPTIMISED RNC-CODED e-MBMS SYSTEM

A. SC-RNC-eMBMS System Model

We consider a multi-cell LTE/LTE-A system model that uses the proposed SC-RNC-eMBMS scheme to deliver video content to \( M \) multicast groups (MG) of users in the cell. Starting at the application layer, the video content is encapsulated into IP data streams that are segmented/concatenated into RLC PDUs (source messages) to be encoded using MAC layer RNC in order to ensure reliable delivery against channel erasures. More specifically, the RLC PDU is segmented into \( K \) information symbols that are RNC encoded, in a rateless fashion, into \( N \) symbols of length \( L \). The MAC layer broadcasts MAC PDUs which can be understood as containers that hold a predefined number of RNC encoded symbols (see Table I). The MAC layer of the eNB uses the reported channel condition (SINR) of the worst-case user in each MG to choose the MC schemes that are suitable for transmission to the different MGs at the PHY layer\(^1\). This means that MGs in better channel conditions are assigned higher rate MC schemes in order to increase their PHY layer transmission rates; while an MG in poorer channel conditions need to receive PHY TBs processed using a low rate MC scheme to be able to decode those TBs successfully. We assume that only a single MAC PDU can be placed within a PHY TB, and one PHY TB can be transmitted to a MG during a given transmission time interval (TTI).

The PHY TB transmission model for each MG requires: i) modeling the reported channel quality indicator (CQI) values as seen by the UEs, and ii) estimating PHY TB error rates for different CQI values as seen from the MAC layer. We model the reported CQI values and their dynamics using Finite-State Markov Chain (FSMC) channel models [12] as described in

\(^1\)The eNB calculates the probability of RLC PDU packet loss for delivery to the worst-case user, for the different MC schemes.

![Fig. 1. eNB RAN protocols: MAC-RNC solution.](image-url)
detail in [4]. In short, based on the 3GPP defined macrocellular path-loss model [13], we calculate the average SINR at the UE location, which is passed to the FSCM model whose states are aligned with SINR intervals corresponding to different CQI values.

We introduce a sequence of \( N_{\text{CQI}} = 16 \) concentric circles defined by the set of radii \( r = \{r_0 = 0, r_1, \ldots, r_{N_{\text{CQI}} - 1}\} \), where \( r_i \) is the distance between the eNB and the external edge of the \( i \)-th ring \((r_{i-1}, r_i)\). The \( i \)-th ring defines an area within which a UE can receive PHY TBs with error probability less than a threshold value \( P_e \), given that the eNB transmits using the \( i \)-th MC scheme (that corresponds to CQI\( = i \)). The set \( r \) can be obtained from the steady-state probabilities of the CQI-state FSCM model described in [4]; however, the details are omitted for brevity. Note that if a MG uses the \( i \)-th MC scheme, the UEs within the circle of radius \( r_i \) will receive the stream of PHY TBs with reconstruction probability of at least \( P_e \). Thus, the selection of the number of MGs \( M \) and the corresponding MC schemes can be mapped to a partition of \( r \), as shown in Table II \((M = 4, \text{MC} = \{12, 10, 6, 4\})\). In this paper we assumed a partition scheme such that the area of the circle of radius \( d_i \) (where \( d_i \) is the distance between the eNB and the outermost ring belonging to the \( i \)-th MG) increases close to linearly.

The capacity of PHY TBs broadcasted over the \( k \)-th MG depends on the MC mode selected and the amount of frequency resources allocated to the \( k \)-th MG:

\[
C_k = n_k \cdot N_{\text{RBP},k} \cdot L \quad [\text{bits}]
\]

where \( n_k \) and \( N_{\text{RBP},k} \) represent the number of coded symbols per PHY TB (as reported in Table I) and the frequency allocation (expressed in terms of number of RBPs), respectively. For LTE/LTE-A system parameters, we adopt the source/coded symbol length (in bits) of \( L = 32 \) [4]. Thus a single PHY TB on the \( k \)-th MG holds exactly \( F = n_k \cdot N_{\text{RBP},k} \) coded symbols. It is important to note that all the UEs belonging to the same MG will receive PHY TBs allocated over the same set of resource block paris (RBPs). Hence, in the rest of this paper, we use \( N_{\text{RBP},k} \) to refer to the frequency allocation to the \( k \)-th MG.

Finally, we assume that a set of UEs are placed in the macrocell area of radius \( R_c \) using a Poisson point process of average density \( \lambda \). The hit probability that a UE is placed at the \( i \)-th ring defined by \( r_{i-1} \leq r \leq r_i \) is \( p_{hi} = (r_i^2 - r_{i-1}^2)/(R_c^2) \). The average number of UEs per ring is \( n_i = \lambda \cdot p_{hi} \).

### B. Combinatorial Error Model for Multicast Communications

Let \( V_i \) (for \( i = 1, \ldots, \tilde{u}_k \)) be a random variable representing the number of PHY TB transmissions performed by the eNB to ensure the correct reception of a RLC PDU by the \( i \)-th UE in the \( k \)-th MG. In other words, \( V_i \) is the number of transmission attempts to ensure the correct reception of \( K \) linearly independent coded symbols by the \( i \)-th UE. Let \( \tilde{u}_k \) be the probability of reception of a PHY TB at the \( i \)-th UE. The cumulative density function (CDF) of \( U \) can be expressed as (see Eq. (12) of [15])

\[
Pr\{V_i \leq j\} = \sum_{n=1}^{\tilde{u}_k} u_i(a,j) w(a n_k N_{\text{RBP},k})
\]

where (see Eqs. (12) and (7) of [15])

\[
u_i(n, c) = \begin{cases} c \cdot 1 - p_i \cdot c - n & \text{if } K \leq n \leq c \\ 0 & \text{otherwise} \end{cases}
\]

and,

\[
w(c) = \begin{cases} \prod_{b=0}^{K-1} [1 - \frac{1}{q^a-b}] & \text{if } c \geq K \\ 0 & \text{otherwise} \end{cases}
\]

For \( j \cdot n_k \cdot N_{\text{RBP},k} < K \) we have that \( f_i(j) = 0 \).

From (2) and (3) the expected value of the random variable \( U \) can be expressed as [16]:

\[
E[U] = \zeta(\tilde{u}_k, N_{\text{RBP},k}) = \sum_{n=0}^{\infty} \left( \Pr\{U \leq n\} - \Pr\{U \leq n-1\} \right) = \sum_{n=0}^{\infty} \left( \prod_{r=1}^{\tilde{u}_k} \prod_{r=1}^{\tilde{u}_k} f_r(n) - \prod_{r=1}^{\tilde{u}_k} f_r(n-1) \right).
\]

The expression (6) is validated using numerical simulations for a network scenario composed of one MG that contains 10 UEs. Each UE receives PHY TBs with a delivery probability of \( p_e = 0.89 \). The RLC PDU length \( K = 100 \), the PHY TB capacity (in terms of coded symbols) is equal to 10, 50 or 100 and, the finite field size is \( q \in [2^2, 2^8] \) interval. The results are averaged over \( 10^4 \) simulation runs.

Fig. 2 shows the expected value \( U \) as a function of the generation length \((K)\) for different PHY TB sizes (TBS). From
Fig. 2, we observe that with the increase in the PHY TBS, the value of $U$ becomes less sensitive to the increase in the length of the RLC PDU. Fig. 2 confirms the match between the proposed analytical model and the simulation results.

C. Rate-Allocation Optimisation

In the proposed SC-RNC-eMBMS solution, the set of $M$ MGs share the fixed frequency resource allocation of $\hat{N}_{RBP}$ RBPs reserved for eMBMS. In this section, we formulate the rate-allocation optimization that targets optimal frequency resource sharing among MGs with the goal of matching the sustainable data rates requested by the set of heterogeneous UEs. The presented Rate Adaptation Model (RAM) fits to the system model where the optimal rate allocation is provided by the eNB, for example, for maximum user satisfaction during the transmission of layered scalable video source. We assume that the provided allocation (i.e., the dimensions in terms of RBPs of PHY TBs directed to each MG) cannot be changed during the video broadcasting process.

From Eq. (6) the $k$-th MG will receive an information stream characterized by a bit rate defined as follows:

$$b_k = b_k(\bar{u}_k, N_{RBP,k}) = \frac{L \cdot K}{TTT \cdot \xi_k(\bar{u}_k, N_{RBP,k})} \text{ [bps].}$$ (7)

In this specific context the RAM optimisation model can be defined as follows ($N^+ = N \setminus \{0\}$):

$$\text{minimize } \sum_{k=1}^{M} \delta_k \left(1 - \frac{\xi_k(\bar{u}_k, N_{RBP,k})}{\xi_k(\bar{u}_k, N_{RBP,k})} \right)$$

s.t.

$$b_k(\bar{u}_k, N_{RBP,k}) \geq v_k - \delta_k, \quad k = 1, \ldots, M$$

$$\delta_k \leq v_k, \quad k = 1, \ldots, M$$

$$\sum_{k=1}^{M} N_{RBP,k} \leq \hat{N}_{RBP}$$

$$N_{RBP,k} \in N^+, \quad k = 1, \ldots, M$$

where the $\xi_k$ parameter represents the lowest CQI index reported by the UEs belonging to the $k$-th MG normalized to the maximum index that can be reported by an UE (i.e., 15). Parameter $v_k$ (for $k = 1, \ldots, M$) represents the minimum sustained bit rate provided by the QoS profile characterizing the communication flow directed to the $k$-th MG. Finally, a feasible solution to the RAM optimization problem should be able to match all the minimum sustained bit rates characterizing each communication flow. However, it may happen that the number of RBPs required by some MGs is too large so that the constraint (11) cannot be met. For this reason obvious QoS constraints such as:

$$b_k(\bar{u}_k, N_{RBP,k}) \geq v_k, \quad k = 1, \ldots, M$$

have to be relaxed and restated as expressed by the constraints (9). Overall, the optimization objective is that violations to the QoS constraints are minimized (i.e., Eq. (8) minimizes the weighted sum of the overall system violations to the QoS). The RAM optimization problem has been solved by resorting to suitable derivative free methods [17], [18].

IV. ANALYTICAL AND SIMULATION RESULTS

The performance of the allocation strategy defined by the RAM optimization model (the RAM-S strategy) has been compared to two widely used rate allocation strategies [19]:

- The Fixed Rate Strategy (FR-S) - the eNB transmits only one downlink flow with a bitrate that matches the QoS requirements of the UE in the outermost cell-ring. The minimum sustained bitrates ($v_k$) assigned to cell-edge UEs are given in the second column of Table III;
- The Least Channel Gain Strategy (LCG-S) - the eNB broadcasts only a traffic flow with a bitrate that matches the $v_k$ value (where $h$ is the non-empty MG characterized by the lowest CQI index). The third column of Table III gives the considered minimum sustained bitrates ($v_k$).

The parameters $v_k$ (for $k = 1, \ldots, M = 4$) of the RAM-S are provided in the third column of Table III. Finally, in the following, we consider network scenarios characterized by an eNB inter-site distance (ISD) of 1000m.

Let $b_k$ be the average downlink bitrate characterizing the downlink flow directed to the $k$-th MG (defined by Eq. (7)). The average cell rate $\bar{b}$ can be defined as follows:

$$\bar{b} = \sum_{k=1}^{M} b_k \text{ [bps].}$$ (14)

Hence, $\bar{b}$ represents the actual bitrate that the UEs belonging to the innermost MG will be able to get by combining each multicast flow broadcasted by the eNB (see Sec. III-A).

Moreover, let $o_k = b_k/v_k$ be the gain index characterizing the $k$-th MG. If $o_k \geq 1$ then the QoS constraint of the data flow is guaranteed. The average cell gain index is defined as:

$$\bar{o} = \frac{\sum_{k=1}^{M} b_k}{\sum_{k=1}^{M} v_k} \frac{CQI_k}{o_k} .$$ (15)

It represents an aggregated measure of the capability of a given rate allocation strategy to match the required QoS constraints.

Fig. 3 shows the average cell rate as function of $K$ for the LCG-S, FR-S and the proposed RAM-S rate allocation strategies for $q = 2^4$. From Fig. 3 we note that the performance of the three strategies is not sensitive to the increase in the PHY TB size $F$ (see Sec. III-B). The same observations are obtained from Fig. 4 which demonstrates the average cell rate as a function of the field size $q$, assuming $K = 1125$ RNC symbols. Finally, Fig. 3 and Fig. 4 show a significant performance gain of almost 6-fold for the proposed RAM-S method compared to the LCG-S and FR-S schemes.

<table>
<thead>
<tr>
<th>Table III</th>
<th>Minimum sustained bitrate matched by the FA-S and LCG-S/RAM-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>725</td>
<td>1.4</td>
</tr>
<tr>
<td>625</td>
<td>1.5</td>
</tr>
<tr>
<td>1125</td>
<td>1.6</td>
</tr>
<tr>
<td>1200</td>
<td>1.6</td>
</tr>
<tr>
<td>1625</td>
<td>1.6</td>
</tr>
<tr>
<td>2125</td>
<td>1.6</td>
</tr>
</tbody>
</table>
Fig. 3. The average cell rate as function of the $K$ for different rate allocation strategies and considering $\lambda = 15$ UEs in the cell.

Fig. 4. The average cell rate as function of the finite field size ($q$) for different rate allocation strategies and considering $\lambda = 15$ UEs in the cell.

Fig. 5. The average cell rate as function of the Poisson $\lambda$ value, results are provided for different rate allocation strategies.

Fig. 6. The average cell gain index as function of the Poisson $\lambda$ value, results are provided for different rate allocation strategies.

V. CONCLUSIONS

In this paper we address the challenge of optimal rate allocation across multi-rate users in LTE/LTE-A networks, by using a novel combination of the MAC-layer RNC and the SC-MBMS approach. Our optimization and analytical model is confirmed by the simulation results which demonstrate significant throughput gains of 4-6 fold in average cell rate compared to the currently proposed multi-rate allocation schemes for LTE/LTE-A systems.

REFERENCES


