A Network-coded Cooperation Protocol for Efficient Massive Content Distribution

Israel Leyva-Mayorga*, Roberto Torre†, Sreekrishna Pandi†, Giang T. Nguyen†, Vicent Pla*,
Jorge Martinez-Bauset*, and Frank H.P. Fitzek†
*ITACA Research Institute, Universitat Politècnica de València, 46022, Valencia, Spain
Email: {isleyma, vpla, jmartinez}@upv.es
†Deutsche Telekom Chair of Communication Networks
Technische Universität Dresden, 01062 Dresden, Germany
Email: {firstname}.{lastname}@tu-dresden.de

***Abstract*—Massive content delivery in cellular networks is in the spotlight of the research community as data traffic is increasing at an incredibly fast pace. The existing LTE-A implementation for content broadcast presents several issues such as indoor coverage, along with low energy and spectral efficiency. Therefore, novel systems that provide efficient massive content delivery and reduced energy consumption are needed. In this paper we present a massive content distribution protocol that combines the benefits of cooperative mobile clouds (CMCs) with Random Linear Network Coding (RLNC) through multicast WiFi links. Our main goal is to offload data traffic from the LTE-A link and to reduce the energy consumption at the cooperating UEs. We solve the problem of excessive signaling that oftentimes arises in cooperative approaches by eliminating feedback messages within the CMCs. Instead, we provide a simple but accurate analytic model to correctly configure the number of coded transmissions to be performed within the CMCs. Results show that energy savings of more than 37 percent can be achieved with our protocol when compared to direct content download from the cellular base station. Furthermore, bandwidth utilization at the LTE-A link is sharply reduced.**

***Index Terms*—Cooperative networks; content distribution; LTE-A; Random Linear Network Coding (RLNC); traffic offload.**

I. Introduction

Wireless data traffic is increasing dramatically. For instance, the amount of traffic transmitted in 2016 grew 63 percent when compared to 2015. By 2021, a data traffic of 49 exabytes per month by 2021 is expected [1]. This represents an increase of around 700 percent and more than three quarters of this traffic will be caused by mobile video.

Nowadays, the user equipments (UEs) that request access to a given content (e.g., video streaming) through LTE-A are connected via a unicast link from the cellular base station (eNB), regardless of the number of UEs that request the exact same content. Therefore a large number of replicated unicast sessions are created if multiple users located within the same cell request the same content simultaneously. For example, such a scenario can occur with the passengers in the same train, with the audience in a music festival or in a football stadium, or by players in augmented reality mobile games.

The industry is aware that the current LTE-A system will not be able to handle the expected increase in data traffic in the coming years. Consequently, several systems have been deployed in order to provide multicast capabilities to LTE-A. The one that took advantage in the first years of this decade was the LTE-A evolved Multimedia Broadcast Multicast Service (eMBMS) [2]; a multicast implementation through LTE small cells. But, several issues were detected during its implementation, such as reduced transmission range, high energy consumption, and poor spectral efficiency [3]. Therefore, different content distribution mechanisms to reduce the amount of traffic requested directly from cellular networks must be designed.

Cooperative mobile clouds (CMCs) are a promising solution to the described content distribution scenario [4]. A CMC is a cooperative architecture in which a group of UEs share the available wireless resources opportunistically. For instance, UEs cooperate through a short-range technology, such as WiFi, to reduce the consumed resources in the LTE-A link.

Some content distribution systems have been proposed in the literature, but short-range unicast sessions are oftentimes used [5], [6]. Since the UEs within a CMC are closely located, the use of multicast short-range links for content distribution is possible and much more efficient than the use of independent unicast sessions. It is in multicast wireless networks where Network Coding (NC) schemes have proven to be highly valuable to ensure a high data rate and a low error rate [7].

Random Linear Network Coding (RLNC) is one of the most widely used NC schemes [8]. In RLNC, the transmitter combines the packets contained in its coding matrix to produce coded packets. For this, each packet is multiplied by a coefficient chosen randomly from a Galois-field of size q, GF(q). The systematic RLNC is a variant of RLNC in which the packets are first sent without coding; then coded packets are transmitted. It has been observed that systematic RLNC results in a higher probability of decoding the generation and reduces the decoding complexity at the UEs when compared

This research has been supported in part by European Union's H2020 research and innovation program under grant agreement H2020-MCSA-ITN-2016-SECRET 722424 and by the Ministry of Economy and Competitiveness of Spain under Grant TIN2013-47272-C2-1-R and Grant TEC2015-71932-REDT.

The work of I. Leyva-Mayorga was supported by grant 383936 CONACYT-Gobierno del Estado de México 2014.

to full-vector RLNC [9].

The combination of cooperative approaches such as CMCs with RLNC schemes has lead to the innovative communication paradigm of Network-Coded Cooperation (NCC) [10], [11]. NCC has the potential to provide increased performance in multicast applications [12]. In this paper, we propose an NCC protocol for the efficient content distribution in cellular networks. It comprises two phases, namely the cellular and CMC phases. In the cellular phase, the eNB segments the requested content in batches of size g packets; hereafter we refer to the batch size, g, as the generation size. These g packets are transmitted to a CMC through time multiplexed unicast links. Then, in the CMC phase, the UEs cooperate under the systematic RLNC scheme to distribute these packets through multicast WiFi links.

One of the main drawbacks of existing cooperative systems is the transmission of a large number of feedback messages within the CMCs, which are needed to keep track of the state of the UEs [13]. Hence, in this study we propose to eliminate the transmission of feedback in the CMCs and instead use a simple but accurate analytical model to calculate the number of coded transmissions needed to receive the whole generation reliably at each and every one of the UEs in the CMC. By doing so, our NCC protocol can be correctly configured.

However, two main challenges arise when modeling the multicast transmissions under an RLNC scheme. The first challenge is to model a multicast problem with multiple sources. That is, the content is distributed among the UEs in the CMC and the packets received at each node are not present at the remaining UEs. Single-source multicast scenarios under RLNC schemes have been studied in the literature and the formulation of the exact decoding probability is not trivial [14]. Concretely, exact formulations only exist for the case of one source and two destinations; lower bounds must be used for a higher number of destinations. In our model, we incorporate a lower bound to solve this problem, whose accuracy under the systematic RLNC scheme has been confirmed in [14].

The second challenge is to model the inclusion of packets received from both, the eNB and CMC neighbors in the coding matrix of the UEs. This approach enhances the throughput when compared to other policies like, for example, only include packets received directly from the eNB in the coding matrix [15]. Needless to say, finding an accurate expression for the linear independence of every coded packet transmission in our scenario is not straightforward. Therefore, we use an upper bound for the probability of linear independence of coded packets and evaluate its accuracy in Section IV.

We also use our analytical model to calculate the energy consumption at the UEs. Results show that energy savings of more than 37 percent when compared to single unicast content distribution can be achieved with our protocol in addition to a reduced use of LTE-A bandwidth.

The remainder of the paper is organized as follows. The state-of-the-art is presented in Section II. Then, we describe our NCC protocol in Section III and the analytical model in Section IV. We present the parameters that lead to the adequate configuration of the system, along with the achievable gains and overhead in Section V. Section VI concludes the paper.

TABLE I
COMPARISON BETWEEN RELATED SYSTEMS

		LTE-A	NC	Short-range	FEC
eMBMS	[2]	✓	✗	✗	Raptor Codes
MicroCast	[5]	✓	✓	Unicast	Network Coding
CoopStream	[6]	✓	✓	Unicast	RLNC
NCVCS	[18]	✗	✓	Multicast	Network Coding
NCC system	[11]	✓	✓	Multicast	RLNC

II. Related Work

As mentioned above, the eMBMS system has several drawbacks. Therefore, diverse solutions to the massive content distribution in LTE-A have been developed. For instance, the idea of organizing microcells in cloudlets was first described in [16]. Cooperative relaying was proved to increase network performance in [17] whereas the advantages of Network Coding were first shown in [7]. Moreover, the interplay between subgrouping in cloudlets and network coding was first proposed in [10].

Despite the clear advantage of short-range NC multicast in the cloud, most existing cooperative systems consider unicast short-range data transmissions. Some examples are the MicroCast [5], and CoopStream [6] systems, whose main focus is to offload data traffic from the eNB. Clearly, the performance of all these previous technologies might increase by using WiFi multicast in the short-range.

In [18] a new Network-coding-based video conference system for mobile devices in multicast network (NCVCS) was presented. NCVCS demonstrates the advantages of multicast over short-range unicast, but is still incomplete since it lacks the cellular communication backhaul.

The main motivation for this paper is the NCC system first proposed in [19]. The main focus of this NCC network was to offload the LTE-A network, but also, important throughput and energy gains were observed. Consequently, two demonstrators were built and presented at MWC 2017, and IEEE CCNC 2018/CES 2018 [11].

Regarding the analytical modeling multicast under RLNC, a thorough study on the decoding probability in a one-source multicast scenario with both, full-vector and systematic RLNC is conducted in [14]. Here the authors highlight the importance of the correlation between the packets received at each node, and conclude that the effect of this correlation is negligible for the systematic RLNC but not for full-vector RLNC. As it will be seen in Section IV, we deal with a similar but more complex problem because in our NCC protocol: a) content distribution within the CMCs is performed through multiple multicast sessions; b) the eNB distributes the data packets among the UEs; and c) coding is performed by combining the packets received from the eNB and from neighboring UEs.

III. Proposed NCC protocol

In this paper, we propose and evaluate the performance of an NCC protocol for efficient massive content distribution. In

our protocol, groups of UEs called CMCs are formed. For this, let n be the maximum number of UEs that are allowed in a CMC, hereafter referred to as the cloud size; n is signaled by the eNB as a configuration parameter. CMCs are groups of at most n UEs that: a) have LTE-A connection to the same eNB; b) request access to the exact same content; and c) are fully interconnected by a short-range technology, namely WiFi. It is is out of the scope of this paper to develop the rules and the protocol for the formation of CMCs. Instead, we focus on the content distribution once the CMCs have been formed.

Content distribution occurs in two phases, the cellular and CMC phases. In the cellular phase, the eNB segments the requested content in batches of g data packets; g is the generation size. These g packets are transmitted to a CMC through n time-multiplexed unicast sessions. Then, at the CMC phase, the UEs first multicast the packets received from the eNB without coding. Afterwards, the UEs multicast coded packets to recover the errors that may have occurred in the previous transmissions. The cellular and CMC phases are now described in detail.

Cellular phase: The eNB transmits the g data packets to the n UEs through n unicast sessions. In LTE-A, data transmission takes place in a slotted channel, whose minimum scheduling unit is one subframe, with duration $d_s = 1$ ms [20].

The n unicast sessions are time-multiplexed, so packets are transmitted to the UEs, one at a time, in a round-robin fashion. For this, each of the n UEs is assigned an index, in the set $N = \{i \in \mathbb{Z}_+ \mid i \leq n\}$, that defines the order in which they will receive the data packets from the eNB.

In this paper, we assume the CMCs are closely located to the eNB so that no wireless channel errors can occur during the cellular phase. This is a valid assumption as the considered data rate is relatively low (see Table III), LTE-A is provided with highly reliable data transmission mechanisms such as HARQ, and is set to modify the modulation and coding scheme (MCS) if the packet error rate (PER) is higher than 0.1 [21, Sec. 7.2.3]. As such, the cellular phase is comprised of g transmissions, distributed among the n UEs, which must cooperate to distribute these packets in the CMC phase.

CMC phase: The UEs are in charge of redistributing the g packets received from the eNB in the CMC. Since no feedback messages are transmitted, the eNB must inform the number of time slots allocated for the content distribution within the CMC to the UEs.

The index i assigned to each UE in the cellular phase is used to create a TDMA schedule. At each time slot, a UE performs a WiFi multicast packet transmission to the remaining $n-1$ UEs in the CMC. The transmitting UE changes at each time slot to uniformly distribute energy consumption among the CMC members. Please observe that the time slot duration at this phase is not necessarily the same as that of the LTE-A subframe, hence a higher or lower data rate can be used.

At the end of the cellular phase, g_i packets are present at the ith UE and these are not present in the remaining $n-1$ UEs. The systematic RLNC scheme is implemented in this phase, hence, g_i packets are transmitted without coding by the ith UE. Therefore, the first g packet transmissions within the CMC are not coded; hereafter we refer to these as the systematic transmissions. That is, each UE will forward the packets received at the cellular phase. Then, coded packet transmissions are performed in order to recover the errors that occurred during the g systematic transmissions.

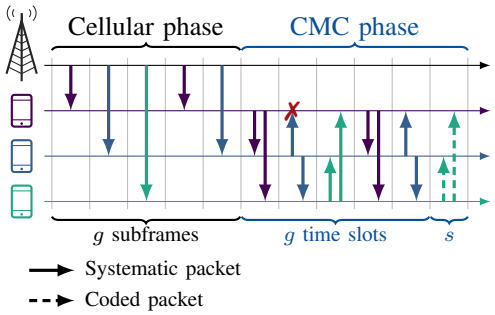

Fig. 1. Timing diagram for the proposed NCC protocol given $n = 3$, $g = 5$, and $s = 1$. The error that occurred in the second time slot is recovered at the first coded packet transmission.

Exactly g time slots are needed for the transmission of the g systematic packets. Therefore, the eNB only has to calculate the number of time slots allocated for the transmission of coded packets, s. When $g + s$ time slots have elapsed, the eNB continues with the transmission of the next generation if needed, hence a new cellular phase begins. Otherwise, data transmission is terminated.

The timing diagram of our NCC protocol for $n = 3$, $g = 5$, and $s = 1$ is illustrated in Fig. 1. In the diagram, an error occurs at the second time slot. This error is recovered at the first coded packet transmission because UEs include packets transmitted by neighboring UEs in their coding matrices.

IV. Analytical Model

In this section we provide a simple but accurate analytical model to obtain s^*, defined as the minimum number of coded transmissions performed in the CMC (i.e., from every $i \in N$), s, that are needed to achieve a desired reliability, τ. Once s^* has been obtained, the maximum throughput and the average energy consumption per UE can be easily calculated.

To find s^*, let S be the RV that defines the total number of coded packet transmissions needed to decode the generation at each of the n UEs. That is, S has a phase-type distribution that describes the probability that the coding matrix of each UE in the CMC is full rank, which occurs when the g degrees of freedom (DOFs) have been obtained by the n UEs. Hereafter we refer to S as the probability of successful content distribution. The domain of S is the number of time slots allocated for the transmission of coded packets, s.

At the end of the cellular phase, g source packets have been distributed among the n UEs in the CMC following a round-robin scheduling. We define g_i as the total number of data packets received by the ith UE in the cellular phase. Clearly, g_i is also the number of systematic transmissions performed

by the ith UE in the CMC phase, and is given as

$$g_i = \left\lceil \frac{g-(i-1)}{n} \right\rceil. \quad (1)$$

To proceed, we define the set of neighbors of the ith UE as $N_i = \{j \mid j \in N \setminus i\}$. Next, we define the stochastic process $X_{t_i}^{(i)}$ as the rank of the coding matrix of the ith UE at time index $t_i \in \mathbb{Z}_{\geq 0}$, whose support for any t_i is $x = \{0, 1, \ldots, g\}$. Time index t_i is defined as the number of coded packet transmissions towards the ith UE (i.e., from every $j \in N_i$).

Let $X_0^{(i)}$ be RV that defines the rank of the coding matrix of the ith UE at the end of the systematic transmissions and ϵ be the PER at the WiFi links (i.e., we assume the same ϵ for each pair of UEs, $\{i, j\}$). The pmf of $X_0^{(i)}$ is given as

$$p_{X_0}(x; i) = \Pr\left[X_0^{(i)} = x\right] = \binom{g-g_i}{x-g_i}(1-\epsilon)^{x-g_i}\,\epsilon^{g-x}. \quad (2)$$

Please note that only systematic transmissions have been performed up to this point, so $X_0^{(i)}$ is also the number of non-zero columns in the coding matrix of the ith UE at $t_i = 0$.

Now coded packet transmissions are performed at every $t_i \geq 1$. Let, $T^{(i)}$ be the RV that defines the number of coded transmissions from the $j \in N_i$ UEs needed for the coding matrix of the ith UE to be full rank. $T^{(i)}$ also has a phase-type distribution whose domain is the set of values for the time t_i. We calculate the CDF of $T^{(i)}$ as

$$F_T(t_i; i) = F_{X_{t_i}}(g; i) = \Pr\left[X_{t_i}^{(i)} = g\right]. \quad (3)$$

Clearly, $T^{(i)}$ depends on the PER, ϵ, and on the probability of linear independence of each of the t_ith coded packet transmissions, denoted as $\mathbb{P}(t_i)$. Nevertheless, the correlation between the packets received at each pair of UEs is needed in order to obtain the exact value for $\mathbb{P}(t_i)$. Therefore, we define the stochastic process $Z_{t_i}^{(i,j)}$ as the number DOFs that are missing from the coding matrices of both, the ith and jth UEs at t_i. The joint pmf of $X_0^{(i)}$ and $Z_0^{(i,j)}$ is given as

$$p_{X_0 Z_0}(x, z \mid i, j) = \epsilon^{g-x+z} \sum_u \left[\binom{g_j}{u} \binom{\gamma}{x-g_i-u} \times \binom{\gamma-x+g_i+u}{z} (1-\epsilon)^{\gamma+u-z} \right] \quad (4)$$

where $\gamma = g - g_i - g_j$ and u represents the number of DOFs in the coding matrix of the ith UE that were transmitted by the jth UE. The summation in (4) is performed in the set of possible values, $\{u \in \mathbb{Z}_{\geq 0} | \max\{0, x-\gamma-g_i+z\} \leq u \leq \min\{g_j, x-g_i\}\}$. The exact value of $\mathbb{P}(t_i)$ for a given x and z is defined as

$$\begin{aligned}\mathbb{P}(t_i \mid x, z) &= P\left[X_{t_i+1}^{(i)} = x+1 \mid X_{t_i}^{(i)} = x \cap Z_{t_i}^{(i,j)} = z\right] \\ &= 1 - q^{x+z-g}. \end{aligned} \quad (5)$$

That is, $\mathbb{P}(t_i)$ depends on x and z, but also on the selected Galois-field size, q, and generation size, g.

Clearly, different pairs of UEs, $\{i, j\}$, may have different joint distributions of $X_0^{(i)}$ and $Z_0^{(i,j)}$, as these depend on g_i, g_j, and γ. Furthermore, the joint pmf of $X_{t_i}^{(i)}$ and $Z_{t_i}^{(i,j)}$ is different for each t_i. That is, the joint pmf of $X_{t_i}^{(i)}$ and $Z_{t_i}^{(i,j)}$ must be calculated for every possible t_i and for each $\{i, j\}$ in order to calculate the exact $\mathbb{P}(t_i)$. This makes our problem intractable even for small values of n and t_i.

Instead, we approximate $\mathbb{P}(t_i)$ by assuming that every one of the missing DOFs in the decoding matrix of the ith receiving UE is present in the jth transmitting UE. That is, $\Pr\left[Z_{t_i}^{(i,j)} = 0\right] = 1$ for each t_i, i, and j, which gives

$$\mathbb{P}'(t_i) = \mathbb{P}(t_i \mid x, 0) = 1 - q^{x-g}; \quad (6)$$

this is clearly an upper bound for $\mathbb{P}(t_i)$ and allows us to use the pmf of $X_{t_i}^{(i)}$ alone instead of the joint pmf of $X_{t_i}^{(i)}$ and $Z_{t_i}^{(i,j)}$ to calculate $T^{(i)}$.

Naturally, (6) is exact for for $q \approx \infty$ and also for $n = 2$ since $g = g_1 + g_2$ in this latter case. The mean squared error (MSE) of the upper bound in (6) can be calculated as

$$\text{MSE}\left[\mathbb{P}'(t_i)\right] = \sum_{\forall x,z} p_{X_{t_i} Z_{t_i}}(x, z \mid i, j)\left(q^{x+z-g} - q^{x-g}\right)^2. \quad (7)$$

Table II shows the MSE for the first coded transmission in the CMC, MSE $[\mathbb{P}'(0)]$, for characteristic values of n, g, ϵ, and q. The first coded transmission for $n = 3$ and for $n = 100$ is performed by the second and the first UEs, respectively. Therefore, MSE $[\mathbb{P}'(0)]$ was obtained with $i = 1$ and $j = 2$ for $n = 3$, and with $i = 2$ and $j = 1$ for $n = 100$.

TABLE II
MSE BETWEEN THE APPROXIMATE AND EXACT PROBABILITY OF LINEAR INDEPENDENCE OF THE FIRST CODED PACKET TRANSMISSION.

	$n=3$		$n=100$	
	$g=10$	$g=100$	$g=10$	$g=100$
$\epsilon = 0.02$				
$q=2$	$6.95\cdot10^{-5}$	$3.09\cdot10^{-4}$	$1.77\cdot10^{-4}$	$5.58\cdot10^{-4}$
$q=2^8$	$1.64\cdot10^{-8}$	$5.38\cdot10^{-8}$	$4.12\cdot10^{-8}$	$8.20\cdot10^{-8}$
$\epsilon = 0.16$				
$q=2$	$2.54\cdot10^{-3}$	$1.33\cdot10^{-5}$	$4.63\cdot10^{-3}$	$5.92\cdot10^{-7}$
$q=2^8$	$4.86\cdot10^{-7}$	$1.53\cdot10^{-10}$	$7.69\cdot10^{-7}$	$1.44\cdot10^{-12}$

Clearly, (6) provides a highly accurate approximation, and the parameter that has the greater impact on accuracy is the field size, q. Concretely, a relatively high error is only obtained with $q = 2$ by setting: a) a short g and high PER; and b) a large g and low PER. As it will be seen in Section V, the error introduced by this approximation in the pmf of S is negligible. Therefore, (6) is used throughout the remainder of the paper.

Now we proceed to obtain the probability of successful distribution, S. For this, let $\mathbf{C}$ be a coding matrix of size $r \times c$ s.t. $r \in \mathbb{Z}_{\geq 0}$ and $\{c \in \mathbb{Z}_+ \mid c \leq g\}$, whose elements are selected uniformly at random from GF(q). The probability that matrix $\mathbf{C}$ is full rank, denoted as $F(r, c)$, is

$$F(r,c) = \begin{cases} 0 & \text{for } r < c, \\ \prod_{j=0}^{c-1}\left(1-q^{j-r}\right) & \text{otherwise.} \end{cases} \quad (8)$$

Then we use (8) to obtain the CDF of $T|X_0^{(i)}$ as

$$F_{T|X_0}\left(t_i \mid x; i\right) = \sum_{u=g-x}^{t_i} \binom{t_i}{u} \left(1-\epsilon\right)^u \, \epsilon^{t_i-u} \, F\left(u, g-x\right) \quad (9)$$

which allows us to calculate the marginal CDF of $T^{(i)}$,

$$\begin{aligned} F_T\left(t_i; i\right) &= \sum_{x=g_i}^{g} p_{X_0}\left(x; i\right) \, F_{T|X_0}\left(t_i \mid x; i\right) \\ &= \sum_{u=g}^{g+t_i} \left(1-\epsilon\right)^{u-g_i} \epsilon^{g+t_i-u} \\ &\quad \times \sum_{x=x_{\min}}^{g} \binom{t_i}{u-x} \binom{g-g_i}{x-g_i} F\left(u-x, g-x\right) \end{aligned} \quad (10)$$

where $x_{\min} = \max\{g_i, u-t_i\}$.

To obtain the distribution of S, we first define the number of coded transmissions towards the ith UE, t_i, as a function of the number of time slots allocated for the transmission of coded packets, s, as

$$t_i = f(s,i) = s + g_i - \left\lceil \frac{g+s-(i-1)}{n} \right\rceil. \quad (11)$$

That is, t_i transmissions will be performed by the UEs in N_i at time index s. Then, we use (10) to define the CDF of S as

$$F_S\left(s; n\right) \equiv \Pr\left[\bigcap_{i=1}^{n} X_{f(s,i)}^{(i)} = g\right]. \quad (12)$$

But obtaining $F_S\left(s; n\right)$ is complicated. On the other hand,

$$F'_S\left(s; n\right) = \prod_{i=1}^{n} \Pr\left[X_{f(s,i)}^{(i)} = g\right] = \prod_{i=1}^{n} F_T\left(f(s,i); i\right) \quad (13)$$

can be easily obtained and has been observed to be a tight lower bound for $F_S\left(s; n\right)$ under the systematic RLNC for a wide range of values of q and g [14]. Therefore, we use $F'_S\left(s; n\right)$ to calculate s^* for a given desired reliability, τ, as

$$s^* \equiv \min_s \left\{ s \mid F'_S\left(s; n\right) \geq \tau \right\}. \quad (14)$$

That is, τ is a threshold for S and its value is selected depending on the needs of the content distribution application.

From there, we calculate the throughput per UE, given in bits per second. If the same data rate, R, is used at the LTE-A and at the WiFi links, the throughput per UE is given as

$$R_{\text{ue}}(n) = \frac{\ell}{d_s} \frac{g}{2g+s^*} = \frac{R}{2+\frac{s^*}{g}}. \quad (15)$$

Please observe that (15) can be easily extended for the case of different data rates in the LTE-A and WiFi links.

Finally, we calculate the average energy consumption per UE, $\overline{E}_{\text{ue}}$. For this, let

$$\mathbb{E}\left[T^{(i)} \mid s^*\right] = \sum_{t=0}^{f(s^*,i)} t \; p_T(t; i) \quad (16)$$

TABLE III
PARAMETER SETTINGS

Parameter	Symbol	Settings
Generation size	g	100 packets
Field size	q	$\{2, 2^8\}$
Cloud size	n	$\{2, 3, \ldots, 100\}$ UEs
Desired reliability	τ	$1 - 10^{-5}$
Packet erasure rate (PER)	ϵ	$\{0.2, 0.4, 0.8, 0.16\}$
Subframe duration	d_s	1 ms
Packet length	ℓ	1470 bytes
Data rate at the LTE-A and WiFi links	R	11.76 Mbps
Power consumption for LTE-A reception	$P_{\text{cel,rx}}$	924.57 mW
Power consumption for WiFi transmission	$P_{\text{wifi,tx}}$	442.60 mW
Power consumption for WiFi reception	$P_{\text{wifi,rx}}$	442.60 mW

be the expected number of subframes that the ith UE is in reception mode and in which coded packets are transmitted. Then we calculate $\overline{E}_{\text{ue}}(n)$ as

$$\begin{aligned} \overline{E}_{\text{ue}}(n) = \frac{1}{n\, d_s} \Bigg[& g\, P_{\text{cel,rx}} + (g+s^*)\, P_{\text{wifi,tx}} \\ & + \left(n\, g + \sum_{i=1}^{n} \mathbb{E}\left[T^{(i)} \mid s^*\right] - g_i \right) P_{\text{wifi,rx}} \Bigg] \end{aligned} \quad (17)$$

where $P_{\text{cel,rx}}$, $P_{\text{wifi,rx}}$, and $P_{\text{wifi,tx}}$ define the power consumption during reception in the LTE-A link, and reception and transmission in the WiFi link, respectively.

V. RESULTS

In this section we first compare the results obtained by our model with those obtained by Monte Carlo simulations. Afterwards we present the optimal number of coded transmissions, s^*, as a function of the cloud size, n. Then we evaluate the energy savings that can be achieved with CMCs and, finally, we discuss the impact on the throughput of our NCC protocol.

The subframe duration is $d_s = 1$ ms and a typical packet length, $\ell = 1470$ bytes, is selected, which gives an LTE-A data rate of $R = 11.76$ Mbps. We assume this same data rate for the WiFi links. Power consumption parameters were obtained from the LTE-A and WiFi energy consumption models in [22] and [23], respectively. We assume the same energy per bit is consumed during transmission and reception over WiFi. Other parameter settings are listed in Table III.

We developed a C-based simulator to assess the accuracy of our model. The number of simulation runs is set to ensure the relative margin of error for each point in the pmf of successful content distribution, S, is less than 0.5 percent at a 95 percent confidence interval.

The accuracy of our model is assessed by means of the Jensen-Shannon divergence (JSD), which measures the increase in the Shannon's entropy when an approximated pmf is assumed to be the real pmf of a RV. Table IV shows the JSD between the pmf of S obtained by our model, $p'_S(s;n)$, and by simulations, $p_{S_{\text{sim}}}(s;n)$, calculated as

$$\begin{aligned} \text{JSD}\left[p'_S(s;n)\right] \equiv & H\left[\frac{p_{S_{\text{sim}}}(s;n) + p'_S(s;n)}{2}\right] \\ & - \frac{H\left[p_{S_{\text{sim}}}(s;n)\right] + H\left[p'_S(s;n)\right]}{2} \end{aligned} \quad (18)$$

TABLE IV
JSD BETWEEN THE PMFS OF SUCCESSFUL CONTENT DISTRIBUTION OBTAINED BY OUR MODEL AND BY SIMULATIONS.

	$n=3$		$n=100$	
	$g=10$	$g=100$	$g=10$	$g=100$
$\epsilon=0.02$				
$q=2$	$1.14\cdot10^{-5}$	$5.35\cdot10^{-5}$	$3.43\cdot10^{-2}$	$2.80\cdot10^{-3}$
$q=2^8$	$9.34\cdot10^{-6}$	$3.53\cdot10^{-6}$	$6.29\cdot10^{-5}$	$3.36\cdot10^{-6}$
$\epsilon=0.16$				
$q=2$	$1.14\cdot10^{-3}$	$6.38\cdot10^{-5}$	$7.02\cdot10^{-3}$	$9.45\cdot10^{-6}$
$q=2^8$	$8.25\cdot10^{-5}$	$1.73\cdot10^{-5}$	$3.35\cdot10^{-5}$	$1.73\cdot10^{-5}$

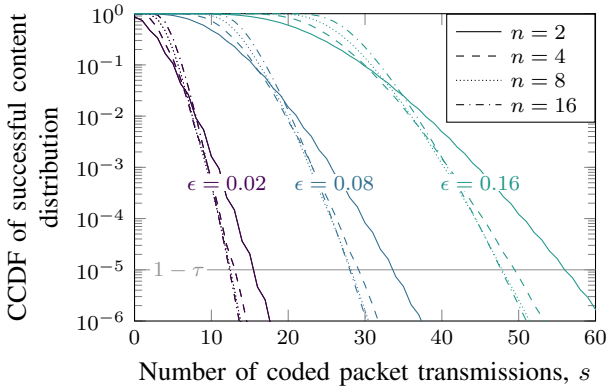

Fig. 2. Complementary CDF (CCDF) of successful content distribution, S, for $q=2^8$, $\epsilon=\{0.02,0.08,0.16\}$, and $n=\{2,4,8,16\}$ in logarithmic scale.

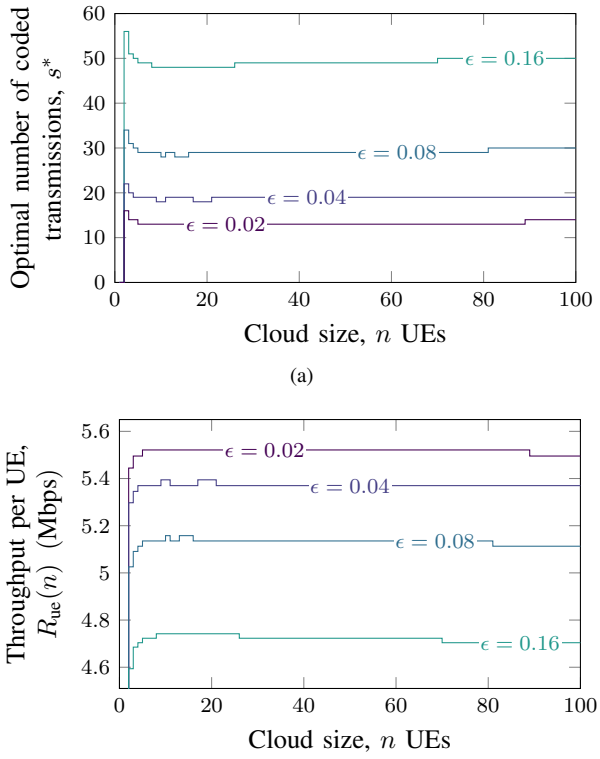

Fig. 3. (a) Optimal number of coded packet transmissions, s^*, and (b) throughput per UE given $\tau=1-10^{-5}$ and $q=2^8$.

where $H[\cdot]$ is the base-e Shannon's entropy. As such, the JSD is upper bounded by $\log 2$ and a JSD of zero indicates both pmfs are identical. Hence, $0\leq \text{JSD}[\cdot]\leq \log 2$.

As it can be seen in Table IV, the JSD is extremely low regardless of the cloud and generation sizes, and is only relatively high for the combinations of $q=2$ with a large n. As a reference, the JSD $[p'_S(s;n)]=2.48\cdot10^{-5}$ for $n=2$, $g=100$, and $\epsilon=0.16$, where our formulations are exact. One million simulations were performed for this case, hence, we consider all cases that lead to a comparable JSD to be exact. It is important to mention that the greatest errors between both CDFs are obtained when $F'_S(s;n)$ is lower than $F_{S_{\text{sim}}}(s;n)$. This is a consequence of lower bound (13), which has a greater impact on accuracy than upper bound (6).

We begin the analysis of our NCC protocol by comparing the complementary CDF (CCDF) of successful content distribution, $1-F'_S(s;n)$, for $n=\{2,4,8,16\}$ and $\epsilon=\{0.02,0.08,0.16\}$ in Fig. 2. As it can be seen, large cluster sizes reduce the number of transmissions if a high reliability (i.e., $\tau\geq 1-10^{-2}$) is needed. The reason for this is that the ratio of transmissions from the UEs in N_i to total transmissions in the CMC phase, t_i/s^*, increases with n. In other words, the frequency of the packets transmitted in the CMC towards each of the n UEs increases with n.

The effect of cloud size on performance can be clearly observed in Fig 3, where we show s^* and the throughput per UE as a function of n and ϵ for $q=2^8$. Specifically, the selection of $n=2$ results in the largest s^* and the lowest throughput. Conversely, the cloud sizes that maximize throughput are in the order of 20 UEs for each of the values of ϵ and the decrease in throughput for $n\geq 20$ is minimal. If we select $q=2$ instead of $q=2^8$, throughput is reduced 6 and 3 percent for $\epsilon=0.02$ and for $\epsilon=0.16$, respectively.

Please observe that the maximum achievable throughput per UE within the CMC is lower than that of a single unicast LTE-A session, $R=11.76$ Mbps. For example, $R_{\text{ue}}(n)\approx R/2.3$ for all n given $\epsilon=0.08$. This slight decrease in throughput is the main overhead of our NCC protocol, and, as described by (15), occurs because g packet transmissions are performed in the cellular phase, followed by g systematic and s^* coded transmissions in the CMC phase. In exchange, the amount of consumed wireless resources are sharply reduced when compared to distribution over parallel unicast sessions.

Now we showcase the main benefit of our NCC protocol: the sharp reduction in the energy consumption at the UEs. For this, we show an area plot of the average energy consumption per UE as a function of n in Fig 4 for $\epsilon=0.16$. Colors indicate the energy consumption at each interface (i.e., LTE-A reception, WiFi reception, and WiFi transmission). For example, the energy consumption for the direct transmission of the g packets in the LTE-A link is 92.45 mJ. On the other hand, the energy consumption per UE for $n=20$ is 57.96 mJ and is further reduced as n increases. Therefore, energy savings of more than 37 percent can be achieved with our NCC protocol, even with relatively small cloud sizes and a high PER.

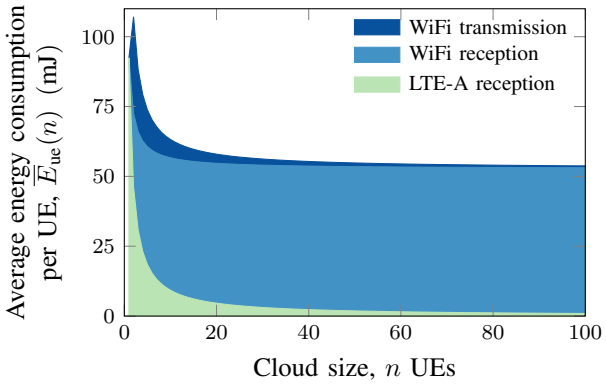

Fig. 4. Average energy consumption per UE given $\epsilon = 0.16$ and $q = 2^8$.

Fig. 4 also shows that the main contributing factor to the overall energy savings is that the number of packets transmitted from the eNB to each UE decreases as n increases. Conversely, the number of packets transmitted through WiFi to each UE increases with n, but the power consumption during reception in WiFi is much lower than that in LTE-A. Finally, the energy consumed for WiFi transmissions is the least contributing factor to overall energy consumption and becomes particularly small for large settings for n.

VI. Conclusion

In this paper we presented an NCC protocol and a simple but accurate analytical model that allows to fine-tune its parameters. Results show that important energy savings of more than 37 percent can be achieved with our protocol, even with a relatively low cloud sizes and high PERs. The main overhead of our protocol is the decrease in throughput when compared to data transmission through parallel unicast LTE-A links. But an eNB can only serve a limited number of unicast sessions in parallel at a high data rate (or at any data rate if the number of UEs is extremely large). Hence the difference in throughput in a practical implementation will be much lower than the one reported in this paper.

The presented model was used to configure our protocol. It includes a lower bound for the probability that multiple UEs decode the generation and an upper bound for the probability of linear independence of coded packets. The first bound is commonly used in the literature and its accuracy has been confirmed under certain conditions in a different study. We evaluated the error of the upper bound and observed that its impact is negligible when compared to the lower bound.

A relevant characteristic that is not captured by our model is that the PER between some pairs of UEs increases with the cloud size in practical implementations. As a result, important differences in the PER between pairs of UEs are expected if large clusters are formed. Hence, the achievable throughput will be limited by the maximum PER in the CMC. Building on this, we advise to set the exact cloud size that results in the maximum throughput (i.e., in the order of 20 UEs in our scenario). By doing so, small clusters will be formed and still important energy savings will be achieved.

References

[1] Cisco, "Cisco visual networking index: Forecast and methodology, 2016–2021," Cisco Technologies, Tech. Rep., Mar. 2017, white paper.

[2] V. Solutions, "LTE multimedia broadcast multicast services (MBMS)," Tech. Rep., 2015, white paper.

[3] EBU, "Delivery of broadcast content over LTE networks," Tech. Rep., July 2014.

[4] M. V. Pedersen and F. H. P. Fitzek, "Mobile clouds: The new content distribution platform," *Proc. IEEE*, vol. 100, no. Special Centennial Issue, pp. 1400–1403, 2012.

[5] L. Keller, A. Le, B. Cici, H. Seferoglu, C. Fragouli, and A. Markopoulou, "MicroCast: Cooperative video streaming on smartphones," in *Proc. Int. Conf. Mobile Syst., Appl., and Services (MobiSys)*, 2012, pp. 57–70.

[6] L. Aymen, B. Ye, and T. M. T. Nguyen, "Offloading performance evaluation for network coding-based cooperative mobile video streaming," in *Proc. Int. Conf. Network Future (NOF)*, 2016, pp. 1–5.

[7] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[8] T. Ho, M. Medard, J. Shi, M. Efiros, and D. R. Karger, "On randomized network coding," in *Proc. Annu. Allerton Conf. Communication, Control, and Computing*, 2003, pp. 11–20.

[9] A. L. Jones, I. Chatzigeorgiou, and A. Tassi, "Binary systematic network coding for progressive packet decoding," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 4499–4504.

[10] M. D. Renzo, M. Iezzi, and F. Graziosi, "On diversity order and coding gain of multisource multirelay cooperative wireless networks with binary network coding," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1138–1157, 2013.

[11] S. Pandi, R. T. Arranz, G. T. Nguyen, and F. H. P. Fitzek, "Massive video multicasting in cellular networks using network coded cooperative communication," in *Proc.15th IEEE Annu. Consumer Commun. Networking Conf. (CCNC)*, 2018.

[12] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.

[13] Z. Chang, S. Zhou, T. Ristaniemi, and Z. Niu, "Collaborative mobile clouds: An energy efficient paradigm for content sharing," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 186–192, 2018.

[14] E. Tsimbalo, A. Tassi, and R. J. Piechocki, "Reliability of multicast under random linear network coding," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2547–2559, 2018.

[15] M. Tömösközi, F. H. P. Fitzek, D. E. Lucani, M. V. Pedersen, P. Seeling, and P. Ekler, "On the packet delay characteristics for serially-connected links using random linear network coding with and without recoding," in *Proc. Eur. Wireless Conf.*, 2015, pp. 1–6.

[16] F. Fitzek, M. Katz, and Q. Zhang, "Cellular controlled short-range communication for cooperative P2P networking," *Wireless Pers. Commun.*, vol. 18, no. 1, pp. 141–155, 2009.

[17] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[18] L. Wang, Z. Yang, L. Xu, and Y. Yang, "NCVCS: Network-coding-based video conference system for mobile devices in multicast networks," *Ad Hoc Networks*, vol. 45, pp. 13–21, 2016.

[19] R. Torre, "Offloading traffic from cellular networks using network coding and cooperation," Master Thesis, TU Dresden, May 2017.

[20] 3GPP, *Physical channels and modulation*, TS 36.211 V14.2.0 36.211 V14.2.0, Apr. 2017.

[21] ——, *Physical layer procedures*, TS 26.213 V13.0.0 36.213 V13.0.0, May 2016.

[22] M. Lauridsen, L. Noël, T. B. Sørensen, and P. Mogensen, "An empirical lte smartphone power model with a view to energy efficiency evolution," *Intel® Technol. J.*, vol. 18, no. 1, pp. 172–193, 2014.

[23] L. Sun, H. Deng, R. K. Sheshadri, W. Zheng, and D. Koutsonikolas, "Experimental evaluation of WiFi active power/energy consumption models for smartphones," *IEEE Trans. Mobile Comput.*, vol. 16, no. 1, pp. 115–129, 2017.