A joint multicast/D2D learning-based approach to LTE traffic offloading

Filippo Rebecchi, Lorenzo Valerio, Raffaele Bruno, Vania Conan, Marcelo Dias de Amorim, Andrea Passarella

A Joint Multicast/D2D Learning-Based Approach to LTE Traffic Offloading

Filippo Rebecchi[a,b,*], Lorenzo Valerio[c], Raffaele Bruno[c], Vania Conan[b], Marcelo Dias de Amorim[a], Andrea Passarella[c]

[a] *Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, 4 place Jussieu 75005 Paris, France*
[b] *Thales Communications & Security, 4 av. des Louvresses, 92230 Gennevilliers, France*
[c] *IIT-CNR – Via Giuseppe Moruzzi, 1, 56124, Pisa, Italy*

Abstract

Multicast is the obvious choice for disseminating popular data on cellular networks. In spite of having better spectral efficiency than unicast, its performance is bounded by the user with the worst channel in the cell. To overcome this limitation, we propose to combine multicast with device-to-device (D2D) communications over an orthogonal channel. Such a strategy improves the efficiency of the dissemination process while saving resources at the base station. It is quite challenging, however, to decide which users should be served through multicast transmissions and which ones should receive the content via D2D communications. The progress of content dissemination through D2D communications depends on how users meet while on the move. The optimal decision for each content depends both on the status of the LTE channel (when the multicast transmission is executed) and on the evolution of the mobility process of the nodes from there on. We propose a learning solution based on a multi-armed bandit algorithm that dynamically selects the best allocation of users between multicast and D2D to guarantee the timely delivery of data. Numerical evaluations are performed to compare our proposal with the state-of-the-art scheme and an optimal but unfeasible strategy. We confirm that a proper mix of multicast and D2D helps operators save resources at the base station and that the learning algorithm can autonomously find a near-optimal configuration in a reasonable time.

Keywords: Cellular multicast, mobile data offloading, device-to-device communications, reinforcement learning, multi-armed bandit.

1. Introduction

Long Term Evolution (LTE) and LTE-A offer significant higher rates than their preceding technologies – up to 100 Mbps for LTE, and 500 Mbps for LTE-A [11]. Despite these numbers, delivering large amounts of data over cellular networks remains a challenge, with the radio access being the bottleneck. The situation will also be exacerbated by the predicted spectacular increase in mobile data demand for the coming years. It is foreseen that the demand for mobile data will grow at an exponential rate (57% compound annual growth rate between 2014 and 2019), while capacity will not match this pace (increasing only by a factor of 2 in the same time frame) [10]. Therefore, it is of paramount importance to find alternative solutions to reduce the burden on the cellular network.

It is possible to save considerable amount of cellular resources when one needs to distribute the same piece of data to a community of interested users grouped in a limited geographical area (i.e., when data requests are spatially and temporally correlated). In such situations, two possible approaches may address effectively the needs of cellular operators: *cellular multicast*[1] and *mobile data offloading*.

Lately, field trials for video service during crowded sport events like the super-bowl or soccer matches have tested the effectiveness of cellular multicast [1, 14]. By exploiting the broadcast nature of the wireless channel, multicast benefits from a single unidirectional link, shared among several user equipments (UEs) inside the same radio cell. This permits, in principle, a more efficient use of network resources with respect to the case where each UE is reached through dedicated unicast transmissions. However, despite its attractive features, cellular multicast presents intrinsic and still unresolved issues that limit its exploitation: (i) the rate of adaptation to the worst channel user and (ii) the lack of reliability. The reasons behind these inefficiencies are investigated in detail in Section 2.

On the other hand, re-routing part of the traffic to other

*Corresponding author
Email addresses: filippo.rebecchi@lip6.fr (Filippo Rebecchi), lorenzo.valerio@iit.cnr.it (Lorenzo Valerio), raffaele.bruno@iit.cnr.it (Raffaele Bruno), vania.conan@thalesgroup.com (Vania Conan), marcelo.amorim@lip6.fr (Marcelo Dias de Amorim), a.passarella@iit.cnr.it (Andrea Passarella)

[1] A more precise terminology would be "multicast/broadcast", because only a subset of nodes is concerned by the content (multicast), and the shared nature of the wireless medium (broadcast) is exploited to transmit data. For the sake of readability, in the following we will only employ the term "multicast".

types of network represents a very promising alternative to reduce the burden on the cellular network infrastructure [7, 13, 22]. Data offloading is considered as one of the key enabling technologies for 5G cellular network architectures [2, 24]. In this paper, we consider one type of data offloading based on opportunistic networks leveraging direct device-to-device (D2D) communications between mobile devices. While opportunistic networks offer additional capacity that can be leveraged to reduce congestion on the cellular network, timely delivery of content is an issue, due to the variability of human mobility and the resulting stochastic nature of forwarding events. When data must be delivered within a mandatory deadline, we need offloading solutions that reduce as much as possible the traffic carried by the cellular network, meeting at the same time the deadline.

In this paper, we *explore the combination of opportunistic offloading and multicasting*. Well-positioned UEs can participate in mitigating the inefficiencies of cellular multicast (where the UE with the worst radio conditions inherently limits the efficiency) by handing over content to nodes with bad cellular channel through opportunistic transmissions. Despite the benefits of this hybrid distribution strategy are evident, in its design, we faced several challenges specific to the opportunistic and wireless domains: (i) performance of the opportunistic delivery hinges on the mobility pattern of users, (ii) opportunistic networks can only guarantee a probabilistic assurance of data reception, and (iii) understanding which fraction of UEs to reach through multicast and D2D transmissions is vital to offer a minimal QoS while guaranteeing resource savings.

Since a truly optimal solution is not conceivable without precise knowledge of future contact patterns, we attack the problem from a more practical point of view. We apply a Reinforcement Learning (RL) approach to decide which fraction of UEs should be reached through a multicast transmission and which should be served using opportunistic communications. A central controller installed at the eNB (evolved Node B) decides, for each packet of a content item to be disseminated, which fraction of users to reach with a multicast transmission. Each decision results in a certain use of the cellular network resources, which generates a reward associated to that choice. This reward is then used to guide (probabilistically) the future choices of the controller. Due to the many similarities in the formulation, we adopt the well-known *multi-armed bandit* technique to implement this algorithm [26].

To fully understand the performance of this joint multicast/D2D approach, it is necessary to evaluate the amount of radio resources consumed at the base station. This motivates us to introduce a finer model of radio resource consumption with respect to those used in the literature. While this is well understood in the literature on physical aspects of cellular communications, existing proposals for opportunistic offloading do not consider heterogeneous channel conditions, assuming that delivering a given amount of data (i.e., a fixed size packet) to different users has always the same cost for the operator [17, 13]. Such an assumption does not hold in reality, as resource consumption varies according to the channel condition experienced by each user. In other words, transmitting the same piece of content to users with different channel conditions does lead to uneven costs at the eNB. To the best of our knowledge, we are the first to evaluate this aspect in the context of data offloading.

As a summary, the main contributions of this paper are:

- **Joint offloading strategy.** Our strategy employs direct D2D opportunistic communications to assist the cellular distribution via multicast.
- **RL-based multicast selection.** The multicast emission is driven by a RL algorithm. Exploiting the knowledge of past rounds, the algorithm tunes the set of nodes to be reached via multicast, allowing substantial savings at the cellular base stations.
- **Fine-grained resource consumption analysis.** We evaluate resource consumption employing the smallest radio resource unit that can be assigned to users for data transmission. This analysis shows that existing macroscopic techniques fail to capture actual system behaviors.
- **Performance evaluation.** The RL strategy permits saving consistent amount of radio resources at the eNB (up to 90% for different mobility models and short delay-tolerances). Even in the worst case, the RL approach approximates an unfeasible strategy that picks the best fixed fraction of multicast users after exhaustive search.

The remainder of the paper is organized as follows. We first present the motivation of our work in Section 2. Two different multi-armed bandit approaches are described along the joint distribution strategy in Section 3. We evaluate the proposed system using a realistic packet-level simulator in Section 4. We postpone the discussion on the related work to Section 5 so that the reader has enough material to capture our original contribution. We finally conclude the paper and identify topics for future research in Section 6.

2. Background and motivation

Opportunistic networks are self-organizing mobile networks where the existence of simultaneous end-to-end paths between nodes is not taken for granted, while disconnections and network partitions are the rule [20]. Opportunistic networks support multi-hop communication by temporarily storing messages at intermediate nodes, until the network reconfigures and better relays (towards the final destinations) become available.

LTE downlink transmission is based on OFDMA frames made of different frequency sub-carriers having a spacing

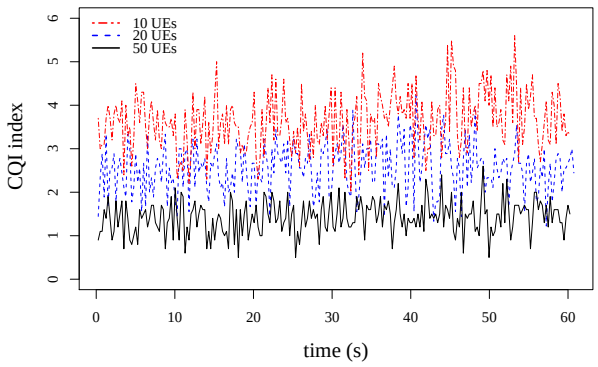


Figure 1: Minimum CQI for different multicast group sizes. 100 runs, confidence intervals are tight and not shown in figure.

Table 1: CQI / MCS Table for LTE [3].

CQI index	Modulation schema	code rate x 1024	Spectral Efficiency [bit/s/Hz]
0		out of range	
1	QPSK	78	0.1523
2	QPSK	120	0.2344
3	QPSK	193	0.3770
4	QPSK	308	0.6016
5	QPSK	449	0.8770
6	QPSK	602	1.1758
7	16-QAM	378	1.4766
8	16-QAM	490	1.9141
9	16-QAM	616	2.4063
10	64-QAM	466	2.7305
11	64-QAM	567	3.3223
12	64-QAM	666	3.9023
13	64-QAM	772	4.5234
14	64-QAM	873	5.1152
15	64-QAM	948	5.5547

of 15 kHz. OFDMA frames are further divided in the time and frequency domain to form the Resource Blocks (RBs), which are the smallest radio resource unit that can be allocated by the packet scheduler. The eNB (cellular base station for LTE) supports different modulation and coding schemes (MCS) to adapt transmission to the variable channel characteristics of users. The MCS determines how much data is transmitted over each RB. Channel adaptation is driven by Channel Quality Indicator (CQI) feedbacks from the UEs. The reported CQI is a number between 0 (worst) and 15 (best) as listed in Table 1. The CQI indicates the most efficient MCS giving a Block Error Rate (BLER) of 10% or less. In unicast transmission, the eNB selects the MCS and the resources to allocate to each UEs, based on this feedback regarding the channel state. A higher value of CQI allows the eNB to select an MCS such that it can transmit more information inside each RB. Thus, the number of RBs necessary to transmit a given amount of useful bits (or, equivalently, the amount of information transmitted per RB) is a typical measure of cost and thus efficiency of LTE transmissions.

Apart from unicast transmissions, new LTE releases propose also an optimized broadcast/multicast service through eMBMS (*enhanced Multimedia Broadcast Multicast Service*), a point-to-multipoint specification to transmit control/data information from the cellular base station (eNB) to a group of user entities (UEs) [16]. UEs interested in receiving multicast transmissions can subscribe – granted they are authorized – to the eMBMS service. After the announcement of all the existing services, subscribed UEs can join one or more multicast group(s) of interest, following the procedure detailed in [4]. All the users belonging to the same multicast group receive the same transmission. Channel heterogeneity (time varying and user-dependent) reduces the effectiveness of multicast because the eNB uses a single MCS for the entire multicast group. The selected MCS should be robust enough to ensure the successful reception and decoding of the data-frame for each UE in the multicast group. Thus, the worst channel among all the receivers dictates performance. It follows that an increase in the number of users in the multicast group boosts the probability that at least one user experiences bad channel conditions, degrading the overall throughput [8].

To exemplify the influence of poor quality users, we simulate a 500×500 m^2 single LTE cell with an increasing number of randomly located receivers using the ns-3 simulator [19]. Fig. 1 presents the minimum average channel quality in terms of CQI, reported at the eNB by users. In this configuration, users are static, and their location is uniformly distributed inside the eNB coverage area. From Fig. 1, we highlight two aspects. The first one is that, when users are uniformly placed, there is a high chance of having at least one user experiencing a poor channel quality (e.g., even with only 10 users in the cell, we still have the worst CQIs not greater than about 4). The second point is that this behavior worsens as more users are added in the area. The result is that augmenting the number of multicast receivers clearly affects the attainable cell throughput. Table 1 shows also that an UE with the best CQI could theoretically receive 37 times the throughput of a user with the lowest index.

This greatly motivates us to investigate methods to cope with the inefficiencies of multicast. We exploit the presence of alternative direct connectivity options available at UEs to relieve the cellular infrastructure load, while reducing the influence of users experiencing poor radio conditions.

3. A reinforcement learning approach for data dissemination

3.1. Content dissemination strategy and problem formulation

We address the dissemination of popular content to a set of N mobile UEs inside a single LTE cell. Each UE is a multi-homed device that embeds both a LTE interface and a short-range technology that allows D2D communications. In simulation, we consider IEEE 802.11g, however, the future integration of D2D capabilities within the LTE standard could be employed as well [12][2]. We want to transmit content with a guaranteed maximum *reception delay* D, at the smallest cost for the cellular infrastructure, i.e., using the minimum number of RBs. To this end, we exploit the possibilities offered by D2D connectivity and store-and-carry forwarding. In the following, we assume standard epidemic dissemination as far as opportunistic communications are concerned. Specifically, instead of addressing all interested UEs with a single multicast transmission – following the discussion in Section 2, this will likely result in a high cost in terms of used RBs – we address only a subset of the UEs (those in better channel quality), and exploit opportunistic D2D communications to reach the others. The challenging issue is that opportunistic dissemination is, by definition, unreliable, as it depends on many factors outside of the control of the cellular infrastructure (e.g., movement pattern of nodes, variable density of opportunistic neighbors, or interference on the D2D channel). Using only opportunistic communications it is thus impossible, in general, to guarantee the maximum reception delays to all the nodes. To achieve guaranteed delivery, we consider an acknowledgment mechanism, and *panic zone* retransmissions similarly to the proposition in [30]. Accordingly, all UEs that receive the required content send an acknowledgment to the central controller situated at the eNB through LTE. When the reception delay reaches its maximum value D, the central controller instructs the eNB to push all the missing data to UEs that have not been served yet using unicast transmissions. Of course, unicast transmissions represent the last opportunity to assure data reception[3]. In this scheme, the cost of disseminating content to interested UEs comes from i) the cost of the initial multicast transmission, and ii) the cost of the unicast transmissions in the panic zone.

Fig. 2 offers a representative example of the proposed dissemination strategy. To avoid the penalty due to the presence of UEs experiencing severe channel conditions, the eNB emits at a modulation that leaves them in outage. This is equivalent to restrict the access to the multicast group only to those UEs in "relatively" good channel conditions. In the opportunistic dissemination phase, *outaged* UEs benefit from nearby nodes, fetching data through out-of-band D2D transmissions. This cooperative strategy is expected to be more efficient in terms of cellular resource consumption than multicast alone, given that the cellular rate increases and the D2D links typically exploit a much larger bandwidth than cellular communications. Finally, panic injections assure data reception to all users.

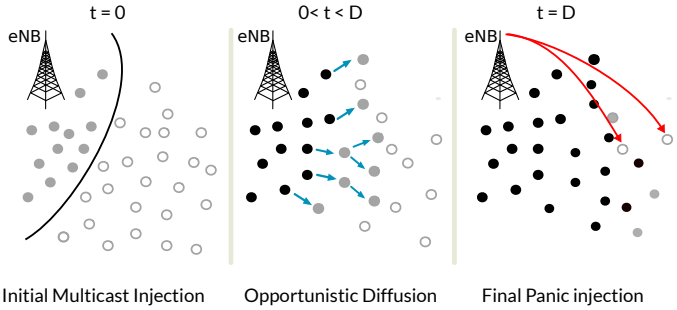


Figure 2: UEs can decode data with a maximum modulation schema depending on their channel quality. The eNB may decide to multicast at higher rate. UEs unable to decode data are reached through out-of-band D2D links and final panic retransmissions.

It is clear that such a scheme admits an optimal operating point. Reducing the set of UEs reached via the initial multicast transmission results in a lower cost for the multicast transmission. However, this may be paid with an additional cost for unicast transmissions in the panic zone, if the remaining UEs are not reached quickly enough through opportunistic dissemination. The challenge to identify this optimal operating point is that the cost of each possible configuration depends on future mobility of nodes, which is unknown at the time the multicast transmission needs to be configured.

More precisely, the problem we address is the following: *how to select the initial set of (seed) users to be reached using multicast transmissions with the objective of minimizing the total number of physical resource blocks (RBs) needed for content dissemination.*

In our scheme, we employ a single parameter I_0 to address this problem. I_0 is the fraction of UEs that should be reached via the initial multicast transmission. Assuming that, when the multicast transmission is configured, UEs are ranked by decreasing value of CQI, this means that the eNB reaches at least the best I_0 UEs in terms of channel quality. Optimally configuring I_0 is not trivial, because, while the cost of the multicast transmission is deterministic at the time when it is configured, the cost of the needed unicast transmissions in the panic zone is a stochastic variable, which depends on the pattern of mobility of UEs in the next D seconds.

We model this problem as a multi-armed bandit problem

[2]D2D communications will be soon possible within the LTE technology (and spectrum). However, in the proposed LTE-D2D extension, the eNB decides, as part of its scheduling process, which devices should communicate directly, not exploiting future direct communication opportunities enabled by nodes mobility.

[3]The need for a guaranteed reception method is a common issue for multicast due to its shared nature. UEs have no assurance of reception because the radio channel could suddenly degrade during data reception (e.g., due to fast fading or mobility). For this reason, mechanisms similar to panic retransmission are considered in several works in the literature [27, 21].

and we solve it through a Reinforcement Learning (RL) approach. As explained in detail in the rest of the section, our scheme is able to learn autonomously the best value of I_0 by observing the effect of different configurations on the resulting cost of disseminating a given content. Assuming that multiple content items need to be disseminated over time to a given set of UEs, our scheme actually learns the best *probability distribution* over the possible values of I_0, that results in the minimum cost in terms of RBs for a given stochastic mobility pattern of UEs. Without prior knowledge on the mobility patterns, and given that mobility is stochastic, learning the best distribution of I_0 is the only practical choice for a learning framework.

3.2. Multi-armed bandit background

Let us now briefly introduce the general formulation of a multi-armed bandit problem (bandit for short). In the simplest case, there is a set of K unknown probability distributions $\langle F_{D_1}, ..., F_{D_k} \rangle$ with associated expected values $\langle \mu_1, ..., \mu_k \rangle$ and variances $\langle \sigma_1^2, ..., \sigma_k^2 \rangle$.

For the sake of illustration, let us assume that F_{D_i} describes the distribution of the outcomes of the i^{th} arm on a slot machine (the bandit); the player is viewed as a gambler whose goal is to collect as much money as possible by pulling these arms over many turns. Initially, the distributions F_{D_i} are completely unknown to the player. At each turn, $t = 1, 2, ...$, the player selects an arm, with index $j(t)$, and obtains a reward $r(t) \sim D_{j(t)}$. Since the player does not know in advance the distribution F_{D_i}, it has to explicitly test the i^{th} action with a trial-and-error search. Therefore, the player has two conflicting objectives: on the one hand, finding out which distribution has the highest expected value (or explore the distribution space); on the other hand, gaining as much rewards as possible while playing (or exploit its knowledge). Reinforcement Learning algorithms specify a probabilistic strategy by which the player should choose an arm $j(t)$ at each turn. Clearly, the effectiveness of the solution depends on how the gambler handles the exploration/exploitation dilemma when testing the different arms iteratively. Exploitation maximizes its reward at present time; at the same time, exploration may lead to a greater total reward in the future.

3.3. Learning algorithm

The general multi-armed bandit formulation can be specialized as follows. First of all, in our problem each arm of the bandit corresponds to a different I_0 threshold (recall: the fraction of UEs that receive a packet via the multicast emission). Thus, K is the number of different thresholds chosen for multicast emission. It follows that F_{D_i} is the distribution of the amount of RBs employed during the entire dissemination process when $I_0 = i$ is used as threshold. More precisely, $D_i = m_i + X_i$, where m_i represents the amount of RBs necessary for a multicast transmission at the MCS needed to reach only the best I_0 ranked UEs, and X_i is the random variable that models the RBs used for the unicast transmissions during the panic zone. Note that, while m_i is fixed and known in advance, X_i depends on many factors, including the set of seeds that are activated, the network topology and node mobility, as well as the dissemination strategy. In our case, each turn corresponds to the dissemination of a content composed of a multitude of packets that are transmitted independently. After the deadline, the reward for each threshold is updated based on the outcome of the dissemination of each packet composing the content. Assuming that $I_0 = i$ was used for the n^{th} multicast transmission, the obtained reward is computed as:

$$\mu_i(n) = \frac{1}{m_i + x_i(n)}, \tag{1}$$

where $x_i(n)$ is the number of RBs that are used for the unicast transmissions in the n^{th} panic zone. Note that the higher the number of used RBs and the lower the reward. To dynamically estimate the *average* reward $\overline{\mu_i}(n)$ for each value of I_0 we use a classical exponential moving average with rate α:

$$\overline{\mu_i}(n) = \alpha \overline{\mu_i}(n-1) + (1-\alpha)\mu_i(n). \tag{2}$$

Now, we must define the policy to select at time $n+1$ the next I_0 value given the knowledge of the average rewards estimated at time n. Different learning methods have been proposed in the literature for the armed bandit problems.

The simplest one is the *ϵ-greedy* algorithm that selects with probability $(1-\epsilon)$ the value of I_0 with the maximum accumulated reward (greedy action), while it selects with probability ϵ one of the remaining I_0 values at random (with uniform probability) independently of the reward estimates (exploration action). More formally, let $\pi_{i(n)}$ be the probability to set $I_0 = i$ for the transmission of the n^{th} packet, and $i^*(n) = \text{argmax}_i\, \overline{\mu_i}(n-1)$. Then, in the ϵ-greedy algorithm it holds that $\pi_{i^*(n)} = 1 - \epsilon$.

Another class of learning algorithms is known as *pursuit* methods, in which the π probabilities are selected to strengthen the last greedy selection. Specifically, let $i^*(n)$ be the greedy value of I_0 defined above. Then, just prior to selecting the CQI for the transmission of the n^{th} packet, the greedy probability is reinforced as follows

$$\pi_{i^*(n)}(n) = \pi_{i^*(n)}(n-1) + \beta[\pi_{MAX} - \pi_{i^*(n)}(n-1)], \tag{3}$$

while all the non-greedy probabilities are updated as follows

$$\pi_{i(n)}(n) = \pi_{i(n)}(n-1) + \beta[\pi_{MIN} - \pi_{i(n)}(n-1)], i \neq i^*. \tag{4}$$

Here π_{MAX}, π_{MIN} are respectively the upper and the lower bound that the probability $\pi_{i(n)}(n)$ can take $\forall i, n$. In equations 3 and 4 the greedy choice is increased, but never more than π_{MAX}, and each non-greedy choice is reduced, but no less than π_{MIN}. This guarantees that the pursuit method is able to cope with the possible non-stationarity of the problem we are considering, i.e. the distribution of rewards can change over time due to the underlying

mobility. Compared to the pursuit method, the ϵ-greedy strategy presents a threshold effect by which the choice that has the maximum accumulated reward immediately gets the highest probability. In the pursuit method, the likelihood of the same option gradually increases by a factor β proportional to the distance to the maximum bound π_{MAX} (similar remarks hold for the non greedy choices). So in pursuit the evolution of the distribution over the possible choices is less drastic and more gradual.

3.4. Wrap-up of the dissemination strategy

As a summary, the key principles behind the joint multicast/D2D approach are: i) at initial time, the eNB sends data to the best I_0 CQI-ranked UEs through a single multicast emission. A RL algorithm is employed to learn the experimental distribution ($\pi_{i^*}(n)$) for the I_0 parameter; ii) the UEs that have received the data through the multicast emission start disseminating it in a D2D (epidemic) fashion; iii) before the maximum *deadline* D, we define a time interval, a *panic zone* where all the nodes that have not yet retrieved the content (either with the initial multicast emission or in D2D fashion) receive it through unicast cellular retransmissions. The proposed scheme allows all UEs to receive data by the deadline (as long as the panic zone is sufficiently large). It adapts to different *deadlines* – the larger ones allowing for more D2D dissemination. Its performance relies on the RL algorithm that permits the cellular base-station to learn by experience the best transmission rate for each multicast emission.

4. Performance Evaluation

In this section, we analyze the offloading performance of the proposed RL algorithms while distributing popular content under different UEs mobility and density configurations.

4.1. ns-3 implementation

We implemented the multi-armed bandit algorithm using ns-3, a packet-level network simulator which implements the full LTE and Wi-Fi stacks, allowing for very realistic simulations [6, 5]. Since ns-3 does not natively support cellular multicast, we modified it by implementing an additional module that interacts with the packet scheduler to emulate single-cell multicast. The multicast module decides, upon each transmission, the fraction of UEs to be reached directly, i.e. the I_0 parameter, based on the multi-armed bandit algorithm presented in Section 3. It receives the CQI from the standard LTE modules, and sets the MCS of the multicast transmission to reach the intended UEs. We fix the bandwidth allocated for the multicast service at 5 MHz. 3GPP standard recommends not to reserve more than 60% of RBs to multicast [16], so a 5 MHz value could represent respectively 50% or 25% of RBs in a typical 10 or 20 MHz deployment. The other simulation parameters for the LTE cell are listed in Table 2.

Table 2: ns-3 simulation parameters.

Parameter	**Value**
Cellular layout	Isolated cell, 1-sector
LTE downlink bandwidth	5 MHz (25 RBs)
Frequency band	1865 MHz (Band 3)
CQI scheme	Full Bandwidth
eNb TX-power	51 dBm
Pathloss	Cost 231
Fast fading	Extended Pedestrian A (EPA) model
Cell dimension	200×200 m^2
eNb position	RWP (100, 100), SMOOTH A (150,50), SMOOTH B (50,150)
eNb antenna height	30 m
UE antenna height	1.5 m
Multicast group size N	10, 25, 50 UEs
Max reception delay D	30, 60, 90 s

Additionally, we implemented DTN store-carry-forward routing mechanism at UEs to support D2D opportunistic communications. This is an implementation of the conventional epidemic forwarding mechanism [28]. Regardless of its reception method, an unexpired packet can be forwarded on the Wi-Fi interface upon meeting with neighbors. Neighbor discovery is implemented through a beaconing protocol. UEs periodically (every 250 ms) broadcast beacon messages containing their identifier and the list of buffered packets. Upon beacon reception, UEs update their vicinity information and can transmit packets opportunistically.

Implementation assumptions: In simulation we made the following simplifications:

- HARQ-level retransmissions and RLC-level feedback are disabled in multicast. This is a reasonable assumption: otherwise the eNB should merge the *ack/nack* messages received from all the UEs, and decide which is the best retransmission strategy. Instead, we guarantee the maximum content delivery time D with *panic zone* retransmissions.

- The PUCCH channel is employed to acknowledge data reception towards the eNB. Panic zone retransmissions are then triggered looking at the list of received acknowledgments.

- The RL algorithm proposed in Section 3.3 acts as a packet scheduler. It employs a cross-layer design at the eNB. By exploiting signaling from physical layer (i.e., the amount of RBs consumed and the CQI for each UEs are used to evaluate the reward), the algorithm decides the MCS of each multicast transmission.

We are aware that our simulation-based evaluation has some limits. First, we consider a simplified version of the eMBMS standard. The proposed approach requires deeper integration with the eNB scheduler. Moreover, we leave out the discussion on incentives that are vital to convince users to agree to spend their battery and storage resources to relay data to someone else. This is an orthogonal prob-

lem addressed in the opportunistic networking literature through appropriate mechanisms (see, for example, [32]).

4.2. Experimental setup

In evaluation, we consider a single eNB, a remote server that provides the content, and several LTE multicast capable UEs implementing also the DTN routing mechanism. To assess the capability of the RL strategies to save resources at the eNB in a high load scenario, we consider all the UEs connected to the same eNB as requiring the content. Simulations are conducted in pedestrian scenarios such as a shopping mall or a crowded touristic landmark. Possible contents of interest include location based broadcasting with advertisement, geo-relevant data, alerts, public utility information and over-the-air software updates.

Each content of interest is formed by a UDP constant bit-rate downlink flow, with packet size of 2048 bytes and a total size of 8 MB (in total, 4000 packets). Each packet is distributed independently using the RL algorithm described in Section 3. The traffic has a loose QoS guarantee on a per-content basis (meaning that individual packets can be delayed, but the entire content must reach the user within the given deadline D).

We implement the mobility of UEs according to two models: Random Way Point (RWP) and SMOOTH [18]. First, we evaluate the RWP model, where UEs move over the simulation area with speed falling between 1 and 2.5 m/s (pedestrian speed).[4] In this scenario, the eNB is placed in the center of the cell. Next, using SMOOTH, we focus on scenarios where the average distribution of the UEs is uneven across the cell. Apart from providing very realistic mobility, SMOOTH makes it possible to define landmark locations of different weights, where nodes head with variable probabilities. User density depends on the location, and we have a strong correlation between user positions (namely, around popular landmarks where contacts are frequent). Fig. 3, shows a synthetic map of waypoints generated with SMOOTH employing the parameters listed in Table 3. Taking advantage of this trace, we target two separate scenarios: one where the eNB is close to the most popular landmark (150,50), and the other, where the eNB is placed near the landmark with the lower weight (50,150). We test our RL algorithms in these two extreme scenarios with different correlations of movements.

All the simulation results are averages over 10 independent runs of 1 hour duration. Unless otherwise stated, confidence intervals are not shown, as they are very tight (usually in the order of 5% of the average value). We assess the performance for different values of N (the number of users inside the cell) and D (the maximum reception delay), so as to evaluate performance under different loads.

[4]While the realism of the RWP model is questionable in general, it has been demonstrated that it realistically reproduces movement patterns of groups of users moving in a confined physical area [9]. Therefore, we consider it appropriate for simulation in our target scenario.

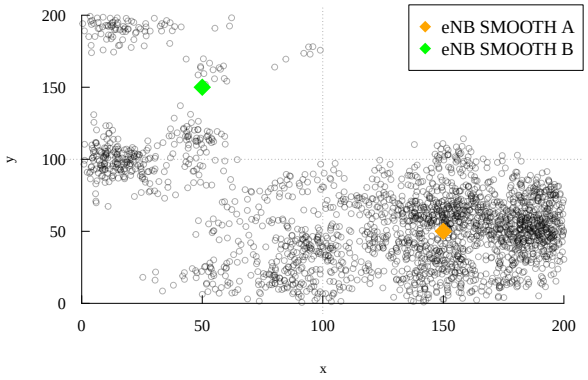


Figure 3: Synthetic map of waypoints generated for the SMOOTH-A and SMOOTH-B scenarios using the parameters listed in Table 3. eNB location for SMOOTH-A is set at (150, 50) and eNB location for SMOOTH-B is set at (50,150)

Table 3: SMOOTH-A and SMOOTH-B scenario parameters.

Parameter	**Value**
waypoints	1000
range R	25 m
clusters	5
alpha	4.0
beta	1.5
min pause	1 (*sec*)
max pause	60 (*sec*)

4.3. Reference strategies

We compare our proposal with four different strategies for content delivery. The main performance indexes we consider are (i) the number of RBs used by the eNB to deliver the content by the stated deadline, and (ii) the offloading ratio, i.e. the fraction of UEs that are served via D2D opportunistic communications with respect to the case where only multicast is used. Additionally, we are also concerned about the convergence time of the bandit algorithm. The considered strategies are the following ones:

- *Multicast-only* is the basic strategy, where UEs have no other means than the cellular network to receive data. In this case, the eNB always schedules multicast transmissions to reach all UEs. Unicast transmissions in the panic zone are used to guarantee the delivery of lost packets.
- *Fixed-best* maintains a static allocation of multicast users (I_0 is fixed during all the simulation duration). Since the best value of I_0 is unknown, we ran extensive simulations to find experimentally the I_0 value that minimizes the total number of RBs used for disseminating all contents. This strategy is clearly unfeasible in practice, but is used as a benchmark for the rest of the strategies, that are based on learning methods.
- *ϵ-greedy* implements the epsilon greedy version of the multi-armed bandit algorithm described in Section 3.

UEs	D	Pursuit	Epsilon	Fixed–Best
10	30s	69	65	55
	60s	66	64	51
	90s	59	61	66
25	30s	52	11	69
	60s	75	74	71
	90s	73	77	83
50	30s	25	–34	55
	60s	30	–19	72
	90s	76	80	85

(a) Aggregate over 1h trace

UEs	D	Pursuit	Epsilon	Fixed–Best
10	30s	73	63	63
	60s	73	70	54
	90s	75	73	73
25	30s	69	12	70
	60s	80	80	71
	90s	83	84	83
50	30s	54	–28	55
	60s	55	–31	72
	90s	88	87	89

(b) Instantaneous after 1h of learning

Figure 4: Cellular offloading ratio for the RWP scenario. Savings are referred to the multicast-only scenario (%): (a) considers the aggregate data savings over the entire 1 hour simulation; (b) considers only the final saving levels (after 1h of learning).

It selects the greedy value of I_0 following Eq. 2. In our implementation, we selected $\epsilon = 0.05$ and $\alpha = 0.5$. We motivate this choice as a trade-off between different requirements. We need to maintain the exploration phase active in order to counter the possible non-stationarity of the underlying process. However, transmitting with a wrong MCS (i.e., giving too much weight to the exploration phase) can lead to significant efficiency loss (due to the panic re-transmissions).

- *Pursuit* selects the I_0 transmission probability following Eqs. 3 and 4. In this case, the transmission probability pursues the greedy action by adapting the likelihood of emission to the temporal evolution of the system. For fairness, we employed the same update value of the ϵ-greedy strategy for α. Moreover, we fixed $\beta = 0.3$, $\pi_{MIN} = 0.01$, $\pi_{MIN} = 0.95$. We will explain better these choices later on.

Finally, both RL strategies are initialized very conservatively, with 95% of multicast emissions targeting all the present UEs. Specifically, considering N UEs in the system, we have that $\Pi_{i=N}(0) = 0.95$, and $\Pi_{i\neq N}(0) = 0.05/(N-1)$.

4.4. Fixed allocation vs. learning algorithms

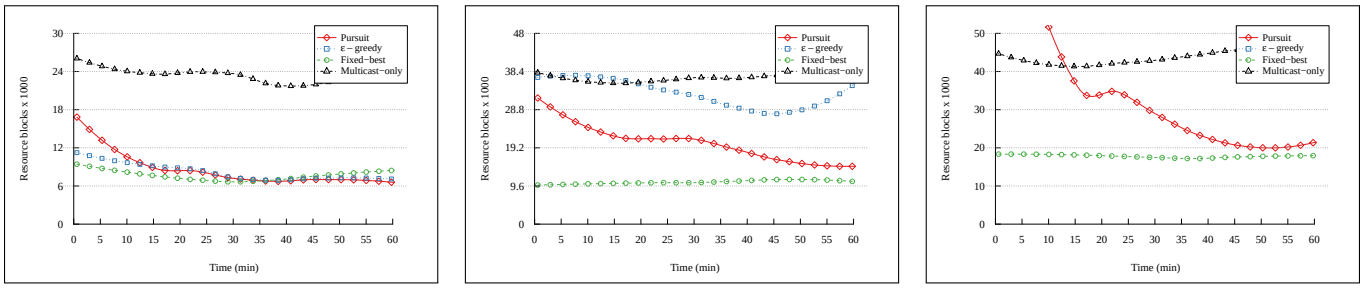
Fig. 4 provides a summary of the simulation results in terms of RB savings (aggregate over 1 hour of simulation, and at the end of 1h of learning) for the two considered algorithms, related to the basic *Multicast-only* approach and the RWP scenario. In general, the RL solution to the joint multicast-D2D problem approaches and even surpasses *Fixed-best* in more than one occasion. For instance, the pursuit method allows saving up to 88% of RBs for the 90s scenario. This result confirms that the right synergy in the utilization of multicast and D2D resources allows for significant resource savings at the eNB. Even with shorter deadlines, the pursuit method performs consistently, saving at least 54% of RBs after the training phase. The ϵ-greedy method instead, while behaving well in most cases, suffers from short deadlines and many UEs.

RL strategies can autonomously find the trade-off between multicast and D2D transmissions in a reasonable time (always less than 1 hour) - without extensively search all the entire parameter space (as *Fixed-best* does). The behavior of the bandit methods (*pursuit* and *ϵ-greedy*) and *Fixed-best* differs significantly. Our evaluation is based on the RBs usage at the eNB (Fig. 5), and on the reception method (Fig. 6). We fix the deadline, varying the number of multicast UEs in the cell from 10 to 50. Intuitively, increasing the number of UEs demanding for the same content should require more infrastructure resources. On the other hand, the number of contact opportunities increases as well, offering more possibilities to offload the network. We analyze the detailed evolution of our strategy considering only the shorter deadline (30 s) that represents the worst-case scenario. A similar analysis can be done with the longer deadlines (plotted in Appendix A).

Unlike many other works in the offloading literature, that focus on savings in terms of messages, the use of the ns-3 simulator allows us to evaluate precisely the amount of radio resources consumed at the eNB. As a qualitative example, Fig. 5 depicts the learning process compared to fixed allocations for the tightest deadline considered (30 *sec*). *Pursuit* and *ϵ-greedy* strategies need time to learn the most appropriate distribution for I_0. Once trained, their performance is often on par or even better than the best fixed-value strategy represented by *Fixed-best*, where the value of I_0 is fixed and pre-computed in advance. In this latter strategy, performance is stable over all the dissemination periods, but this figure is the outcome of an extensive trial and error simulation phase. Another advantage of the RL techniques is that, even when trained, they continue to explore the solution space, being able to cope with the possible non-stationarity of the contact process that rules the opportunistic diffusion. Conversely, *Fixed-best* is locked to a static value of the parameter I_0 and insensitive to mobility pattern variations.

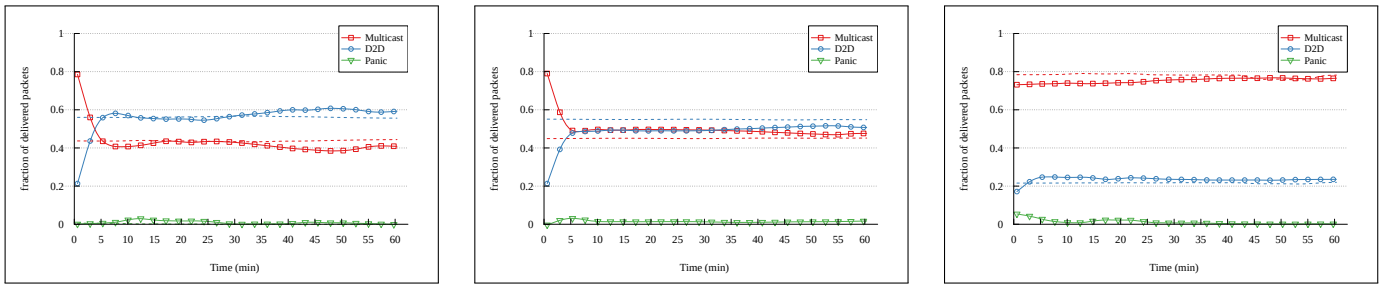
4.5. Comparison between ϵ-greedy and pursuit method

The ϵ-greedy method, owing to the *hard* selection of the greedy value of I_0, does not fit well scenarios where the underlying opportunistic diffusion process has significant variability. In these cases, *pursuit* is a better match. On the other hand, if performance of the opportunistic diffusion



(a) 10 UEs, *30 s* (b) 25 UEs, 30 *s* (c) 50 UEs, 30 *s*

Figure 5: RWP scenario – RBs usage for *Multicast-only* (black), ϵ-*greedy* (blue), *Fixed-best* (green), and *pursuit method* (red). Content is divided into 4000 packets of 2048 bytes. Plots are averaged over 10 runs, 95 % confidence intervals are not plotted but are knit.



(a) 10 UEs, 30 *s* (b) 25 UEs, 30 *s* (c) 50 UEs, 30 *s*

Figure 6: RWP scenario – Pursuit method, reception method. Dashed lines are the objective ratio for *Fixed-best*. Content is divided into 4000 packets of 2048 bytes. Plots are averaged over 10 runs, 95 % confidence intervals are not plotted but are knit.

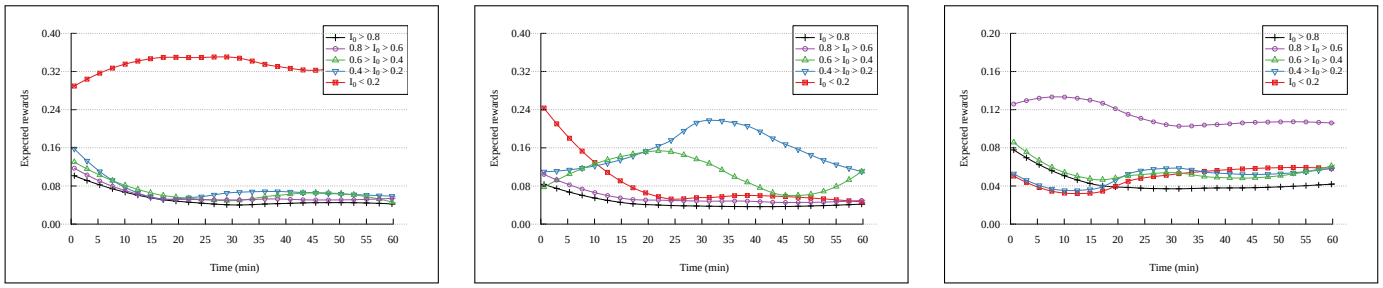
process is stable – i.e., in case of larger deadlines – ϵ-*greedy* allows for quicker convergence times (e.g., Fig. 5(a)).

In many cases, even when ϵ-*greedy* fails to converge, *pursuit* approaches the behavior of the *fixed-best* strategy. This depends on the fact that ϵ-*greedy* always selects the value of I_0 that maximizes the expected rewards. Instead, *pursuit* has an indirect selection method that better adapts to the temporal evolution of the system. The added complexity is however beneficial, as it translates into improved performance (Fig. 5(b) and Fig. 5(c)). The reinforcement given by Eq. 3, allows smoothing out the inherent variations in epidemic diffusion that prevent the proper prediction in the ϵ-*greedy* method. The effect appears when the number of targeted UEs increases while keeping a tight deadline (i.e., 30 *s* in our evaluations). In those scenarios, the variability in performance of the opportunistic diffusion prevents the ϵ-*greedy* method to learn properly the best distribution for selecting I_0. An example of this effect is illustrated in Fig. 5(c). In that case, the *pursuit method* succeeds in matching the *fixed-best* strategy. The ϵ-*greedy* method instead diverges nearly instantaneously, failing to learn an appropriate policy.

4.6. Detailed evaluation of the pursuit method

Considering the same deadline, increasing the number of UEs has the effect of reducing the share of D2D transmissions. We plot in Fig. 6 the fraction of packets partitioned by their reception method for the pursuit method. While a larger number of UEs should multiply the contact opportunities, many of them are not adequately exploited because UEs can transmit only to one neighbor at a time. The result is that fixing the deadline, the share of UEs addressed through D2D transmissions is upper bounded. Note, however, that for a larger number of UEs, even though the fraction of UEs addressed through D2D transmission is limited (e.g., 20% in the case of 50 UEs), the resulting advantage in terms of RB saving is much higher (around 55% for that case, from results in Fig. 4). In this case, the learning algorithm understands that it is better to serve the 20% of the UEs that are experiencing bad cellular quality through D2D. Serving them through multicast is expected to result in a too high cost in terms of RBs, due to the need of reducing too much the MCS. On the other hand, decreasing too much the share of UEs served by the multicast transmission brings the opposite effect, with a considerable amount of RBs spent for unicast transmissions in the panic zone.

Focusing on the convergence time, we note that there is a discrepancy between Fig. 5 and Fig. 6. Looking at Fig. 6, it seems that the convergence time is less than 10 minutes. Instead, the real convergence time, in terms of RBs, happens much later in time, and depends on the fine-tuning of I_0. The anomaly is justified by the fact that even a minimal amount of unicast retransmissions in the panic zone (such as those that may happen before finding



(a) 10 UEs, 30 s (b) 25 UEs, 30 s (c) 50 UEs, 30 s

Figure 7: RWP scenario – Pursuit method, average reward values for I_0. Content is divided into 4000 packets of 2048 bytes. Plots are averaged over 10 runs, 95 % confidence intervals are not plotted but are knit.

the best value for I_0) consumes much more resources than a single multicast emission.

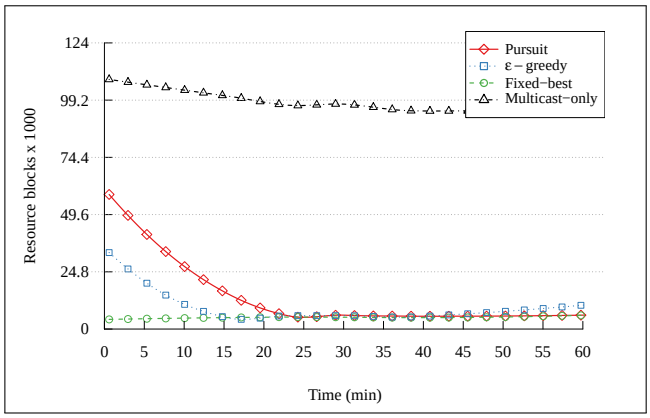
In two scenarios out of three (namely for 10 and 50 UEs), there is a set of values for I_0 that performs clearly better than the others do. In the 25 UEs scenario instead, the distribution of I_0 is more spread out. There is a clear tendency to prefer higher values of I_0 though, in the range between 0.2 and 0.6. Considering the detailed mechanisms of the *pursuit* method described in Sec. 3.3, Fig. 7 compares the average reward μ_i. We quantized the values of I_0 to form five levels. The best I_0 value is the one that is not affected too much by the loss in spectral efficiency due to the reduced multicast rate, but at the same time can guarantee a low penalty due to unicast panic re-injections. In the 10 UEs scenario, emitting with a multicast rate that targets one or two users ($I_0 \leq 0.2$) is sufficient to achieve high efficiency. On the other hand, we note that increasing the multicast group size, the best distribution of I_0 shifts towards higher values. Intuitively, the penalty due to panic re-injections is extremely severe in these cases and the pursuit algorithm tends to allocate more seeders in the opportunistic domain. Finally, the emission probability follows the pattern of the rewards with the exception that the greedy probability is always reinforced until the value π_{MAX} and the non-greedy probabilities are reduced until π_{MIN}.

4.7. Correlated mobility patterns

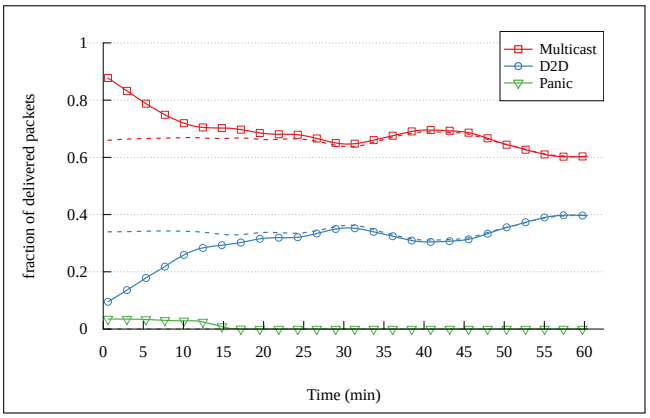
In order to better understand the performance of our RL approach, we evaluated the two SMOOTH scenarios introduced in Section 4.2. First, we ran the *SMOOTH-A* scenario with reception delay of 30 s, then, we switched to the *SMOOTH-B* scenario, maintaining the simulation parameters unchanged. We are interested in how our system reacts under different UEs densities in various locations and realistic mobility patterns. Figs. 8 and 9 propose the simulation results respectively for the SMOOTH-A and the SMOOTH-B scenarios with 25 UEs. As an initial consideration, if we focus only on the baseline *Multicast-only* strategy, we notice that the resource consumption is around three times higher than in the RWP case. The explanation for this observation lies in the fact that in both SMOOTH scenarios the eNB is skewed from the center of the mobility area (as it was in RWP). It follows that edge-located UEs are farther away from the eNB than in the RWP scenario (from Fig. 3 and Pythagoras' theorem, the additional distance between a UE and the eNB can be up to 70 m), worsening the overall transmission efficiency and increasing the amount of resources consumed at the eNB. This motivates even more the needs for efficient offloading strategies, since slight deterioration in channel quality brings large spikes in resource usage. Instead, since in the SMOOTH-A and SMOOTH-B scenarios the eNB location is symmetrical with respect to the center, we cannot perceive appreciable differences between the baseline RB consumption in these two scenarios.

In the *SMOOTH-A* scenario both RL algorithms converge towards an optimal level of resource usage very quickly (as shown in Fig. 8). Offloading savings due to D2D transmissions top a stunning 92% against the classic multicast-only mechanism. Indeed, the density of UEs around the eNB permits, most of the time, a resource-efficient multicast emission (i.e., with very high MCS) that covers at the same time a large part of requests. This is illustrated in Fig. 8(b), where we can appreciate that the larger fraction of packets are received through multicast. UEs heading towards distant waypoints are likely reached with D2D transmissions (representing around 40% of packet receptions). We notice that, in this scenario, ϵ-greedy benefits from its simple update strategy, showing a faster convergence time than pursuit.

Fig. 9 depicts the simulation results for the SMOOTH-B scenario. Evaluation outcomes differ sensibly from the prior scenario. This is a very demanding use case, with the majority of the UEs located at the edge of the cell. Consequently, most of the interactions between UEs take place further away from the eNB. Considering the reception methods, interestingly, D2D transmissions have a larger share. This is in line with our expectations: the multicast emission does not cover as many users as before, due to their distance from the eNB. The larger share of D2D transmissions is a direct consequence of this, along with the increased resource requirements for data dissemination due to a less efficient multicast emission. Also,

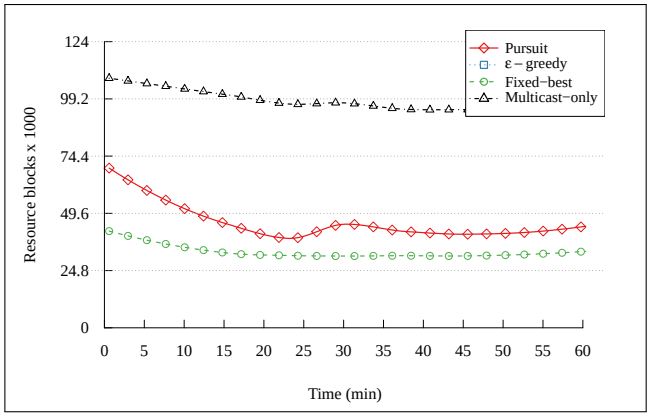


(a) Resource blocks

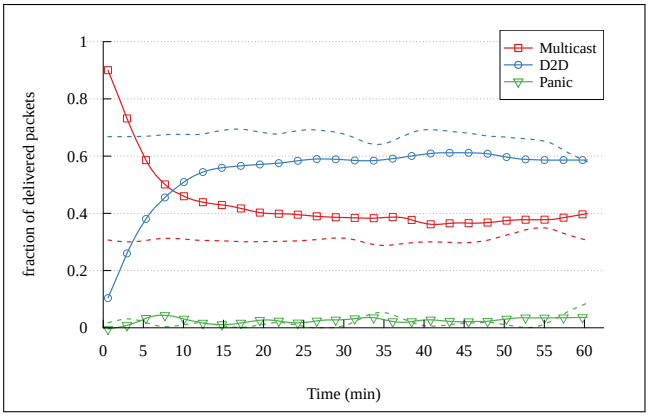


(b) Reception method

Figure 8: *SMOOTH-A* scenario, offloading performance for correlated mobility patterns with most of the waypoints near the eNB position.



(a) Resource blocks



(b) Reception method

Figure 9: *SMOOTH-B* scenario, offloading performance for correlated mobility patterns with most of the waypoints far away from the eNB position.

there are some fluctuations in the reception methods of the *Fixed-best* strategy (dashed lines in Fig. 9(b)), due to the non-stationarity of the underlying opportunistic diffusion process. The *pursuit* strategy works well, allowing discovering the optimal operating point and reaching a satisfactory trade-off between multicast and D2D transmissions. Despite the increased amount of panic zone retransmissions, they are kept to a reasonable level by *pursuit* – recall that panic zone retransmissions are unicast, thus they demand many resources. Similarly to the case with RWP and 50 UEs, the ϵ-*greedy* strategy is not plotted in Fig. 9. Indeed, it fails learning the appropriate policy. The choice that ϵ-*greedy* considers as the best one fluctuates a lot and does not stabilize. Therefore, the ϵ-*greedy*'s simple selection strategy fails to fill the gap with the *fixed-best* allocation. This forces many transmissions to the panic zone, resulting into an increased resource consumption. This argues in favor of the *pursuit* strategy, which at the price of a possibly longer convergence time, constantly converges, scoring better offloading performance.

Finally, by comparing Figure 8 and 9 we may appreciate that in SMOOTH-B, as expected, the amount of resources employed is higher for *fixed best* and *pursuit* (as the density of nodes is higher far away from the eNB). In any case, the use of D2D transmissions coupled with multicast still brings a very significant advantage over the *multicast-only* strategy.

5. Related work

Mobile data offloading. D2D communications have been the target of intensive studies as a method to relieve the pressure on the cellular infrastructure. Typically only unicast transmissions are considered. For instance, Han et al. identified the opportunity to save infrastructure data exploiting the social ties between users, proposing a subset selection mechanism based on contact history [13]. Similarly, Li et al. analytically formulated the problem of traffic offloading of multiple contents in a mobile environment. Under the assumption of Poisson contact, the optimal subset selection problem is solved under multiple constraints [17]. Barbera et al. analyzed contacts between

end-nodes in order to select a subset of socially important VIP users, which are turned into data forwarders [7]. The above-mentioned works consider only the optimal selection of the initial seeder nodes and not the control of the dissemination process through injections. Whitbeck et al. proposed an injection algorithm that follows target objective functions [30]. Some of the authors of the present work proposed a simple re-injection based scheme that takes into account the evolution of the opportunistic dissemination [22]. Finally, a reinforcement learning algorithm to identify which users are good data carriers was proposed in [29]. In all these works the principal metric is the amount of data (or messages) saved on the infrastructure link. While this is an influential driver for evaluation, it does not fully represent the real amount of saved resources at the base station. In addition, to the best of our knowledge, the interplay between multicast and D2D communications has not been addressed in this body of work.

D2D-aided multicast. Bhatia et al., proposed the use of D2D communications to improve performance of multicast in 3G cellular networks [8]. A multihop ad hoc network is modeled analytically. A near-optimal discovery algorithm selects the best data forwarder for receivers with poor channel quality. The authors in [31] devised an algorithm to figure out the optimal number of relays inside the cluster. The paper focuses on in-band D2D communications, such that considered in [2]. Similarly, in [25], only the cluster head receives the content and is in charge of D2D retransmission inside its cluster. No hints are given on how clusters are created and discovered, and how these techniques can be applied to LTE networks to reduce the cellular resource usage. Huo et al., proposed a cooperative multicast scheduling for 802.16 networks. A two phase schema is proposed, and all successful recipients of multicast participate in data retransmission using in-band D2D links [15]. To the best of our knowledge, the only work in this area focusing on the data offloading problem with guaranteed delivery is [23], which however does not offer any mechanism to devise the correct allocation of users between multicast and D2D transmissions.

6. Conclusion

In this work, we have presented a hybrid distribution strategy, jointly leveraging LTE multicast and opportunistic D2D communications to distribute popular content with guaranteed delays. Multicast is an advantageous option to distribute popular data to users co-located inside a cell. However, it does not offer any reception guarantee, and the user with the worst channel quality inside the multicast group dictates overall performance. We proposed a framework that exploits D2D capabilities at UEs to counter the inefficiencies of cellular multicast.

The proper balance of multicast and D2D transmissions is achieved using a multi-armed bandit learning strategy. We proposed and evaluated two different algorithms under variable multicast group size, reception deadlines and mobility patterns. Simulation results prove that D2D communications permit to configure multicast transmission in a more efficient way, saving more than up 90% of the radio resources at the base station. We have also shown that the tested learning algorithms are able to obtain performance comparable (and in several cases even superior) to the best possible strategy that uses a fixed split between multicast and D2D communications, which can only be identified *after* exhaustive search, and is thus practically unfeasible. On the other hand, the proposed learning algorithms are able to dynamically learn the best balance between multicast and D2D transmissions, and, to do so, need a reasonable learning time.

Acknowledgment

This work is partially supported by the European Commission in the framework of the FP7 Mobile Opportunistic Traffic Offloading (MOTO) project under grant agreement number 317959.

References

[1] TIM and Huawei: LTE Broadcast is coming, the new generation mobile TV, with the first live test in Italy. `http://www.telecomitalia.com/tit/en/archivio/media/note-stampa/market/2015/TIM-Huawei-LTE-broadcast-mobile-TV.html`.

[2] 3GPP. TSG SA: Feasibility study for proximity services (ProSe) (release 12), 2012.

[3] 3GPP. TS 36.213 V11.2.0 Rel.11: Evolved universal terrestrial radio access (e-utra); physical layer procedures, 2013.

[4] G. Araniti, M. Condoluci, and A. Molinaro. Resource management of multicast services over LTE. *Convergence of Broadband, Broadcast, and Cellular Network Technologies*, page 77, 2014.

[5] D. Azzarelli, E. Pierattelli, A. Marchetto, L. D'Orazio, F. Rebecchi, A. Passarella, R. Bruno, and G. Mainetto. Description and development of moto simulation tool environment release b, 2015. `http://cordis.europa.eu/docs/projects/cnect/9/317959/080/deliverables/001-D512v2Ares2015964509.pdf`.

[6] N. Baldo. The ns-3 lte module by the lena project.

[7] M. V. Barbera, A. C. Viana, M. D. de Amorim, and J. Stefa. Data offloading in social mobile networks through {VIP} delegation. *Ad Hoc Networks*, 19(0):92 – 110, 2014.

[8] R. Bhatia, L. Li, L. Haiyun, and R. Ramjee. ICAM: integrated cellular and ad hoc multicast. *IEEE Transactions on Mobile Computing*, 5(8):1004–1015, Aug 2006.

[9] C. Boldrini and A. Passarella. HCMM: Modelling spatial and temporal properties of human mobility driven by users social relationships. *Computer Communications*, 33(9):1056–1074, 2010.

[10] Cisco. Cisco visual networking index: Global mobile data traffic forecast update (2014 – 2019), 2015.

[11] E. Dahlman, S. Parkvall, and J. Skold. *4G: LTE/LTE-advanced for mobile broadband*. Academic Press, 2013.

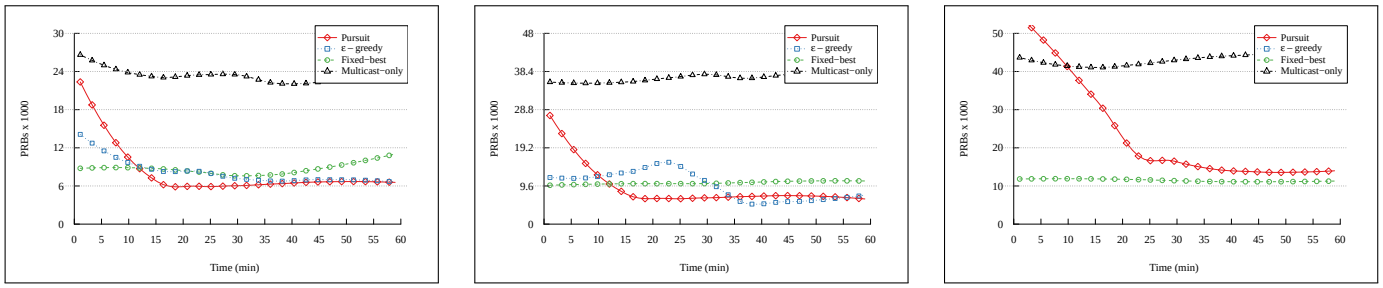
[12] K. Doppler, M. Rinne, C. Wijting, C. Ribeiro, and K. Hugl. Device-to-device communication as an underlay to lte-advanced networks. *IEEE Communication Magazine*, 47(12):42–49, Dec 2009.

[13] B. Han, P. Hui, V. S. A. Kumar, M. V. Marathe, J. Shao, and A. Srinivasan. Mobile data offloading through opportunistic communications and social participation. *IEEE Transactions on Mobile Computing*, 11(5):821–834, May 2012.

[14] Z. Honig. Verizon demonstrating LTE Multicast during Super Bowl XLVIII (hands-on video). http://www.engadget.com/2014/01/29/verizon-lte-multicast/.
[15] F. Hou, L. Cai, P.-H. Ho, X. Shen, and J. Zhang. A cooperative multicast scheduling scheme for multimedia services in ieee 802.16 networks. *IEEE Transactions on Wireless Communications*, 8(3):1508–1519, Mar. 2009.
[16] D. Lecompte and F. Gabin. Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: overview and rel-11 enhancements. *IEEE Communication Magazine*, 50(11):68–74, Nov. 2012.
[17] Y. Li, M. Qian, D. Jin, P. Hui, Z. Wang, and S. Chen. Multiple mobile data offloading through disruption tolerant networks. *IEEE Transactions on Mobile Computing*, 13(7):1579–1596, 2014.
[18] A. Munjal, T. Camp, and W. C. Navidi. Smooth: a simple way to model human mobility. In *ACM MSWiM*, pages 351–360, 2011.
[19] NS-3. Network simulator. http://www.nsnam.org.
[20] L. Pelusi, A. Passarella, and M. Conti. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. *IEEE Communications Magazine*, 44(11):134–141, 2006.
[21] M. Rahman, H. Cheng-Hsin, A. Hasib, and M. Hefeeda. Hybrid multicast-unicast streaming over mobile networks. In *IFIP Networking*, pages 1–9, Trondheim, Norway, June 2014.
[22] F. Rebecchi, M. D. de Amorim, and V. Conan. DROiD: Adapting to individual mobility pays off in mobile data offloading. In *IFIP Networking*, Trondheim, Norway, June 2014.
[23] F. Rebecchi, M. Dias de Amorim, and V. Conan. Flooding data in a cell: Is cellular multicast better than device-to-device communications? In *ACM CHANTS*, pages 19–24, Maui, HI, Sept. 2014.
[24] F. Rebecchi, M. Dias de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti. Data offloading techniques in cellular networks: A survey. *IEEE Communications Surveys & Tutorials*, 17(2):580–603, Secondquarter 2015.
[25] S. Spinella, G. Araniti, A. Iera, and A. Molinaro. Integration of ad-hoc networks with infrastructured systems for multicast services provisioning. In *IEEE ICUMT*, pages 1–6, St. Petersburg, Oct. 2009.
[26] R. S. Sutton and A. G. Barto. *Introduction to reinforcement learning*. MIT Press, 1998.
[27] G. Tan, S. Ma, D. Jiang, Y. Li, and L. Zhang. Towards optimum hybrid arq with rateless codes for real-time wireless multicast. In *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, pages 1953–1957. IEEE, 2012.
[28] A. Vahdat, D. Becker, et al. Epidemic routing for partially connected ad hoc networks. Technical report, Technical Report CS-200006, Duke University, 2000.
[29] L. Valerio, R. Bruno, and A. Passarella. Adaptive data offloading in opportunistic networks through an actor-critic learning method. In *ACM CHANTS*, pages 31–36, Maui, HI, 2014.
[30] J. Whitbeck, Y. Lopez, J. Leguay, V. Conan, and M. D. de Amorim. Push-and-track: Saving infrastructure bandwidth through opportunistic forwarding. *Pervasive and Mobile Computing*, 8(5):682–697, Oct. 2012.
[31] B. Zhou, H.Hu, S.Huang, and H.Chen. Intracluster device-to-device relay algorithm with optimal resource utilization. *IEEE Transactions on Vehicular Technology*, 62(5):2315–2326, June 2013.
[32] X. Zhuo, W. Gao, G. Cao, and Y. Dai. Win-Coupon: An incentive framework for 3G traffic offloading. In *IEEE International Conference on Network Protocols (ICNP)*, Vancouver, Canada, Oct. 2011.

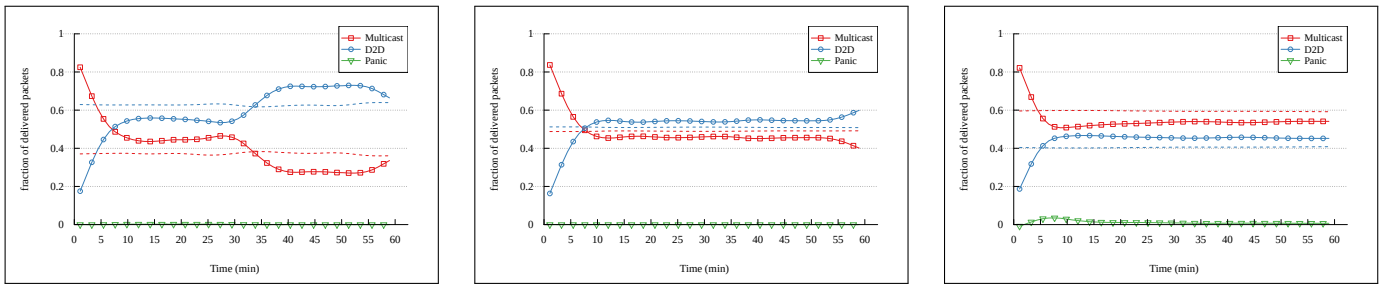
Appendix A. Additional figures

We include in the following page the figures for the simulations in the RWP scenario with deadline 60 and 90 seconds.



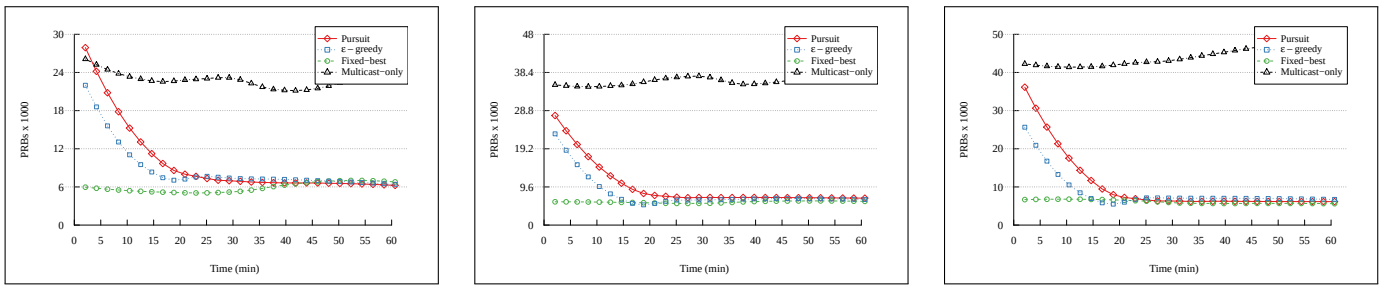
(a) 10 UEs, 60 s (b) 25 UEs, 60 s (c) 50 UEs, 60 s

Figure A.10: RBs usage for *Multicast-only* (black), ϵ*-greedy* (blue), *Fixed-best* (green), and *pursuit method* (red). Plots are averaged over 10 runs, 95 % confidence intervals are not plotted but are knit.



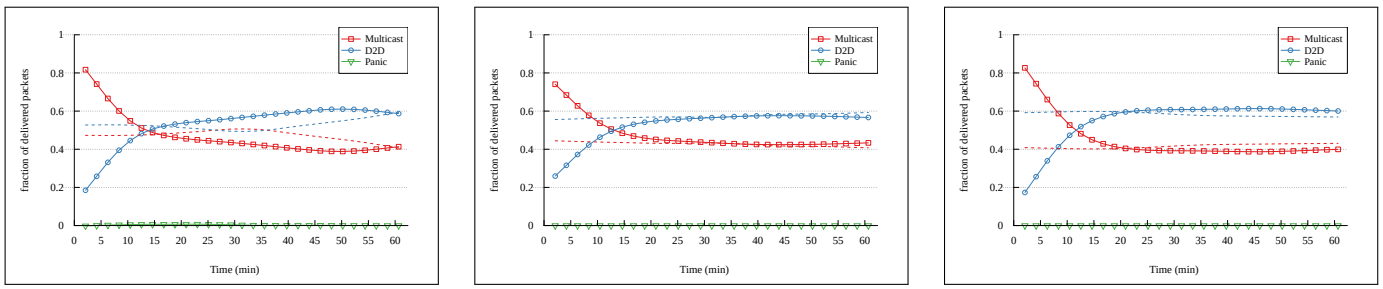
(a) 10 UEs, 60 s (b) 25 UEs, 60 s (c) 50 UEs, 60 s

Figure A.11: Pursuit method, reception method. Dashed lines are the objective ratio for *Fixed-best*. Plots are averaged over 10 runs, 95 % confidence intervals are not plotted but are knit.



(a) 10 UEs, 90 s (b) 25 UEs, 90 s (c) 50 UEs, 90 s

Figure A.12: RBs usage for *Multicast-only* (black), ϵ*-greedy* (blue), *Fixed-best* (green), and *pursuit method* (red). Plots are averaged over 10 runs, 95 % confidence intervals are not plotted but are knit.



(a) 10 UEs, 90 s (b) 25 UEs, 90 s (c) 50 UEs, 90 s

Figure A.13: Pursuit method, reception method. Dashed lines are the objective ratio for *Fixed-best*. Plots are averaged over 10 runs, 95 % confidence intervals are not plotted but are knit.