D2D Multihop Energy-Efficient Routing and OFDMA Resource Allocation in 5G Networks

Safwan Alwan*, Ilhem Fajjari† and Nadjib Aitsaadi‡
*University Paris-Est, LiSSi EA 3956, UPEC, F94400, Vitry-sur-Seine, France
†Orange Labs, F92320, Chatillon, France
‡University Paris-Est, LIGM-CNRS UMR 8049, ESIEE Paris, F93160, Noisy-le-Grand, France
Emails: safwan.alwan@univ-paris-est.fr, ilhem.fajjari@orange.com, nadjib.aitsaadi@esiee.fr

***Abstract*—In face of the rapidly increasing cellular traffic, 5G will employ offloading techniques to relieve the cellular infrastructure. The idea is to carry the traffic locally by User-Equipment (UE) relays or by other coexisting radio access technologies such as WiFi, Bluetooth, etc. To this end, in this paper, we propose an offloading scheme using multi-hop LTE-D2D communications and assisted by the operator. LTE-D2D UEs cooperate to carry intra-cell unicast/multicast traffic from sources to destinations by exploiting i) sidelink interfaces and ii) multi-hop paths. To increase the lifetime of the offloading system and to reduce the impact of the relaying process on the battery-limited UEs, we propose our energy-aware approach, named `JRRA-EE`, to solve jointly the routing and the OFDMA resource block allocation. We formulate our problem as a 0-1 Integer Linear Programming (ILP) model which is elaborated to take into consideration the realistic LTE-D2D capabilities and constraints. To gauge the effectiveness of our proposal, we implement the whole 3GPP LTE-D2D protocol stack in the NS-3 network simulator to simulate our approach. Based on extensive simulations, the obtained performances of `JRRA-EE` are better compared to other one-sided optimal strategies, including an energy non-aware variant, in terms of i) the network lifetime, ii) the packet loss and iii) the service interruption rate.**

***Index Terms*—LTE-D2D, Routing, OFDMA resource block allocation, Energy-aware offloading, Optimization.**

I. Introduction

In recent years, the Device-to-Device (D2D) communication paradigm has received much attention of the academic and industrial communities. Relying on the physical proximity of user terminals, D2D offers low-energy cost and short-distance communications. Furthermore, D2D allows reusing the existing of classical cellular hardware as well as the same frequency resources leading to improve the overall network and spectral efficiency. Apart from these advantages, D2D communication also enables many new applications such as proximity-based safety and commercial services, cooperative content sharing and relaying.

In this paper, we tackle both the routing and OFDMA resource block allocation for an energy-efficient offloading mechanism within the LTE networks for multicast and/or unicast flow-oriented application. Note that, from a formulation point of view, a unicast flow is a special case of a multicast (i.e., one destination). Specifically, within a single LTE eNodeB cell, our system utilizes the sub-network of LTE-D2D-enabled User Equipments (UEs) to route flows that originates from and terminates in the same macro-cell. As a matter of fact, while the data plane offloading rests with the UEs themselves, the whole operation is controlled by the base-station (the eNodeB/eNB). Many crowded-platform scenarios fit in the above description. Examples include content-sharing applications in stadium, train stations and airports. In fact, in such scenarios, we have a high density of quasi-stationary UEs during the event or the waiting period. As a consequence, the need to relieve the macro-cell and micro/femto-cells is really vital.

The offloading operation entails both routing and OFDMA Resource Block (RB) allocation of the D2D communications over the SideLinks (SLs) interfaces of UEs . Note that the SL makes use of the same hardware transceiver and frequency spectrum employed by the UpLink (UL) interface. Consequently, any UE cannot simultaneously communicate in both interfaces UL and SL. Moreover, the SL communication is half-duplex [1]. In other words, the UE cannot simultaneously send and receive on SL.

Our objective is to solve jointly the two sub-problems where the routing decision considers the i) available OFDMA RBs, ii) interferences, and iii) dynamic state of network. However, a practical LTE-D2D-based offloading should not be agnostic to the fact that UEs are usually battery-limited devices. Therefore, to increase utility and lifetime of the offloading sub-network, we formulate an energy-aware joint scheme, for the OFDMA RB allocation and the routing, as a 0-1 Integer Linear Problem (ILP). It is worth noting that our formulation includes realistic constraints related to the LTE-D2D such as the half-duplex of SLs, the contiguity of RB allocations, the total power consumption due to the i) the baseband processing, ii) the RF transmission/reception of the D2D sidelink interface. Solving ILP problems, in general, has been proven to be NP-hard and the optimal solution is unlikely to be found in polynomial time [2].

To cope with the potentially-exponential complexity, we propose a new scheme named Joint Routing and Resource Allocation Energy Efficient (`JRRA-EE`). Our proposal is a centralized strategy hosted in the eNodeB. This means that the D2D communications are under the supervision of the telecommunication operator. Indeed, only the data plane is offloaded and the control plane is still under the control of the operator. `JRRA-EE` is a two-stage scheme. First, a pre-routing stage is performed in order to reduce the space of solutions. Then, the ILP is solved making use of the Branch-and-Cut approach while considering only the reduced space of candidates. The LTE-D2D offloading mechanism aims to help the eNodeB under heavy traffic conditions. Note that our proposed offloading mechanism is presented as a complementary method to deliver multicast and/or unicast flows. In other words, the purpose of LTE-D2D offloading is to complement, not to compete with, the LTE macro/femto-cells conventional delivery methods. If the LTE-D2D path does not exist, the communication will be ensured over the traditional macro/femto-cells. To gauge the effectiveness of `JRRA-EE`, we implemented the whole LTE-D2D SL protocol stack in UEs in the NS-3 network simulator. Using extensive simulations, we compared our proposal `JRRA-EE` to other one-sided optimal strategies and a non-energy aware variant. One-sided optimal strategy is one which is optimal only in one sense either in resource block allocation or in routing. Based on simulation results, we establish that our proposal `JRRA-EE` outperforms the other strategies in terms of i) the network lifetime, ii) the packet loss and iii) the service interruption

ISBN 978-3-903176-08-9

rate.

The remainder of the paper is organized as follows. Section II will summarize the related strategies addressing energy-aware multihop D2D communications. In Section III, we will formulate the problem. Then, in Section IV, the proposal will be detailed. The simulation environment and the performance evaluation will be presented in Section V. Finally, Section VI concludes the paper.

II. Related Work

In this section, we summarize the most relevant related strategies found in the literature dealing with energy-aware routing the context of D2D communications. Note that D2D is employed as an umbrella term for technologies that include, amongst others, LTE-D2D and WiFi Direct.

In [3], the authors show that LTE-D2D cooperative relays save significant amounts of energy when compared to conventional Base Station (BS) to UE communications. In addition, the authors present a cooperative relaying scheme to improve the UE's battery life. The idea consists in maximizing of use of UEs with high battery levels to deliver the traffic of UEs with low energy. Numerical simulations show that the approach reduces the outage probability of the cellular cooperating UEs.

In [4], the authors present a scheme to deliver BS-to-UE video content delivery via a cooperative D2D multihop routing. The proposed scheme employs a generic framework to avoid disruption caused by the depletion of D2D UE's energy budget. Seeking to optimize the budget utility, the algorithm described jointly schedules the routes and traffic workloads depending on the energy efficiency of each D2D wireless link.

In [5], the authors design an energy efficient routing protocol in Wi-Fi Direct cluster-based networks. The designed protocol adopts ideas from `LEACH` and `HEED` protocols which are well-known in wireless sensor networks. Through numerical simulations, the authors demonstrate that the scheme considerably saves network's energy as compared to the conventional peer-to-peer mode of Wi-Fi Direct.

In [6], the authors propose a heuristic algorithm energy-efficient multi-hop routing algorithm for UE-UE unicast traffic. Both channel reusing and power allocation are jointly addressed to obtain a satisfactory solution. The simulations show that significantly improvements in energy-efficiency of the multi-hop D2D communication system.

In relation to our paper, [3] considers only UE-to-BS traffic where high-battery UEs help low-battery ones to relay their traffic to the BS. On the other hand, our work focuses on offloading UE-UE multicast and/or unicast traffic to alleviate the base-station. Similarly, [4] also tackles the BS-to-UEs multicast video traffic where UEs employ a distributed multi-path routing and caching technique. In comparison, despite being energy-budget aware like [4], our work focuses on central algorithms and flow-centric applications where the employed on-demand cluster formation and caching, in [4], cannot be used. In the same manner, the protocol in [5] cannot be adapted into LTE-D2D to serve our purpose, since the traffic model in WSN is multiple-sources-one-sink and the clustering technique is useless in our case. The closest work to ours is [6] despite its focus on unicasting UE-UE traffic. However, the authors assume generic assumptions about wireless technology where the medium is abstracted as whole channels not in terms of resource blocks. They also consider an analytical power consumption model for each D2D link while no energy-budget limitation is considered. On the other hand, we consider a model of multicasting UE-to-UEs traffic where the unicast model can be treated as a special case. Furthermore, we incorporate LTE-D2D specificities with an empirical power consumption model to cater for the efficient utilization of the allocated energy budget for the cooperative relaying process.

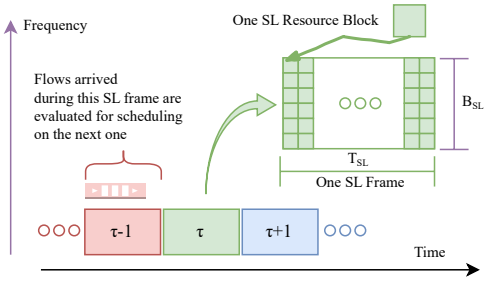


Fig. 1. Sidelink frame structure and scheduling.

III. Network and Problem Formulation

In this section, we will first provide a detailed model of our LTE-D2D system. Then, we will formulate our energy-aware routing and resource block allocation problem in LTE-D2D offloading networks.

A. System model

We consider N LTE-D2D enabled UEs who are located inside the LTE-A eNB's macro-cell. UEs are assumed to be quasi-stationary (e.g., located in the stadium) and are willing to offload the data plane of only **intra**-cell D2D traffic when it is expedient. The control plane is deployed in the eNB. The latter handles the offloading operations, over the D2D subnetwork, by continuously allocating OFDMA resource blocks during each sidelink (i.e., SL) frame. As depicted in Fig. 1, these operations are triggered each instant T where:

$$T = \tau \times T_{SL} \qquad \forall \tau \in \mathbb{N}$$

where T_{SL} and τ represent the duration of a SL frame and the frame index respectively. It is worth noting that the SL frame corresponds to the scheduling time unit in SL, which spans multiple one-millisecond time slots (i.e., multiple TTIs). Besides, it is characterized by a B_{SL} which corresponds to the total bandwidth of the SL communication, and is composed of Ω contiguous OFDMA resource blocks.

We model the D2D network as a symmetric directed graph $\mathbb{G} = (\mathcal{V}, \mathcal{E})$. Each node $v_i \in \mathcal{V}$ corresponds to one UE. An edge (i.e., sidelink) $e_{ij} \in \mathcal{E}$ between two nodes v_i and v_j exists if and only if the Signal to Noise Ratio (SNR), γ_{ij}, is greater than a predefined threshold, γ_{TOPO}. Formally,

$$\gamma_{ij} = \frac{g_{ij} P_{t,i}}{P_\sigma} \geq \gamma_{\text{TOPO}} \tag{1}$$

where i) $P_{t,i}$ is the power emitted by v_i, ii) P_σ corresponds to the thermal noise power, and iii) g_{ij} is the channel gain between the pair v_i and v_j which depends on the used channel model.

During each SL frame, the system's flows set, denoted by $\mathcal{F}$, is the union of two subsets $\mathcal{F}_\text{S}$ and $\mathcal{F}_\text{C}$ defined as follows:

- $\mathcal{F}_\text{S}$: set of *scheduled flows*. It corresponds to the on-going flows which are circulating on the LTE-D2D system. Hence, RBs need to be allocated for them in order to maintain their offloading operation.
- $\mathcal{F}_\text{C}$: set of *candidate flows*. It encompasses the flows which are waiting to be admitted during the next SL frame. Candidate flows are dynamically selected among those residing in the waiting queue, $\mathcal{F}_\text{W}$, based on the current availability of idle nodes $\mathcal{V}_\text{D}$.

Each multicast flow $f^k \in \mathcal{F}$ is characterized by a source node $s^k \in \mathcal{V}$, a destination group $\mathcal{D}^k \subseteq \mathcal{V}$ and a constant bit rate, R^k. Note that the unicast flow is a special case of the multicast flow in which $|\mathcal{D}^k| = 1$ (i.e., one destination). In this paper, we address only the Constant Bit-Rate (CBR)

TABLE I
NOTATION - SYSTEM MODEL

Symbol(s)	Meaning
v_n, e_{ij}, f^k	The n$^{\text{th}}$ node, the link from the i$^{\text{th}}$ to the j$^{\text{th}}$ node and the k$^{\text{th}}$ flow
$s^k, \mathcal{D}^k, \mathbb{T}^k$	The source, the destinations and the routing tree of f^k
R^k, D^k	The bit rate and the requested number of RBs of f^k
δ_x^y	Kronecker delta function which equals to 1 only when $x = y$ and 0 otherwise.
$\mathbb{1}_x^Y$	Set Y's indicator function which equals to 1 only when $x \in Y$ and 0 otherwise.
$\mathcal{O}(v_n), \mathcal{T}(v_n)$	Sets of outgoing (originating) links from v_n and terminating (incoming) links in v_n respectively
$x_{ij}^{h,k}$	Essential 0-1 decision variable that indicates whether the link e_{ij} is used to offload the flow f^k at the hop (tree level) number h
t_n^k	Auxiliary 0-1 variable indicating whether v_n acts as a (re-)transmitter for f^k. Note that at the source node, $t_{s^k}^k$ also indicates whether the flow is admitted or not.
H_n	Essential 0-1 decision variable that indicates the node v_n is scheduled to transmit during SL frames whose $p = \tau \bmod 2 = H_n$
$y_{u,n}$	Essential 0-1 decision variable that indicates that the RB pattern is allocated to the node v_n.
R_{ij}	Auxiliary 0-1 variable that indicates if the link e_{ij} is active.
R_n^ω	Auxiliary 0-1 variable indicates if the RB ω is allocated to v_n.
$R_n^{\omega,p}$	Auxiliary 0-1 variable indicates if the RB ω is allocated to v_n transmitting in the half duplex set p.
$R_{ij}^{\omega,p}$	Auxiliary 0-1 variable indicates if the RB ω is allocated to v_i transmitting to v_j in the half duplex set p.
$\phi_{n,ij}^{\omega,p}$	Auxiliary 0-1 variable indicates that v_n transmitting in the half duplex set p on the RB ω is interfering with the (active) link e_{ij}.

flows. Once admitted, f^k is carried throughout a routing tree $\mathbb{T}^k$ delivering its packets from s^k to $\mathcal{D}^k$. Note that if f^k is unicast flow, $\mathbb{T}^k$ is reduced to one branch (i.e., path). $\mathbb{T}^k$ is characterized by $h_{\max}$ levels which corresponds to the maximum number of hops from the root to the leaves.

Each UE is handled in an exclusive manner. This means that a given node can relay at most one flow at a time. Consequently, the routing trees (i.e., multicast) and/or branches (i.e., unicast) are mutually disjoint for concurrent flows.

B. Problem formulation

We address, in this paper, the energy-aware joint routing and OFDMA resource block allocation problem in LTE-D2D. The objective is to compute, for a given multicast flow $f^k \in \mathcal{F}$, the optimal routing tree while i) limiting the interferences between forwarding UEs, ii) minimizing the number of hops, iii) minimizing the communication energy consumption.

Conceptually, this problem can be decomposed into two sub-problems: i) routing and ii) resource allocation. However, following a cross-layer design, we propose to coordinate the resolution of the two sub-problems. In doing so, we aim to maximize the QoS and the efficiency of the system respectively from the point view of end-users and telecommunication operator. By such a joint treatment, enhanced results are obtained since the routing solution takes also in consideration, the induced wireless interferences in OFDMA RBs and the energy consumed in the transmission/reception operations.

In light of our adopted formalism shown in TABLE I, the routing problem can be formulated as following. Let $x_{ij}^{h,k}$ indicates whether the corresponding link e_{ij} is selected or not to be a part of a routing tree $\mathbb{T}^k$ for the flow f^k at the tree level (hop) h for $h = 0, 1, \ldots, h_{\max}$. To ensure a consistent tree structure, we introduce the following constraint which stipulates that a node v_n has at most one parent:

$$\sum_{\substack{e_{ij} \in \mathcal{T}(v_n) \\ 0 \le h \le h_{\max}, f^k \in \mathcal{F}}} \sum x_{ij}^{h,k} \le 1 \qquad \forall v_n \in \mathcal{V} \tag{2}$$

Note that $\mathcal{T}(v_n)$ corresponds to the set of incoming links to v_n. It is straightforward to see that this constraint ensures that a link cannot appear in more than one flow at a time.

Besides, we must ensure that only outgoing links from source are allowed at a tree's root (i.e., at $h = 0$). Formally,

$$x_{ij}^{h,k} \le \delta_0^h \cdot \delta_{v_i}^{s^k} + (1 - \delta_0^h)(1 - \delta_{v_i}^{s^k} - \delta_{v_j}^{s^k}) \quad \begin{matrix} \forall e_{ij} \in \mathcal{E} \\ \forall 0 \le h \le h_{\max} \\ \forall f^k \in \mathcal{F} \end{matrix} \tag{3}$$

We recall that δ_y^x corresponds to the Kronecker delta function which equals to 1 only when $x = y$ and 0 otherwise.

Also, we must guarantee that an outgoing link in the tree from a node v_n is possible at the level h if and only if an incoming link exists at the level $h - 1$. Formally,

$$x_{nm}^{h,k} \le \sum_{e_{ij} \in \mathcal{T}(v_n)} x_{ij}^{h-1,k} \qquad \begin{matrix} \forall e_{nm} \in \mathcal{E} \\ \forall 1 \le h \le h_{\max} \\ \forall f^k \in \mathcal{F} \end{matrix} \tag{4}$$

It is worth noting that constraints (2)–(4), also imply that a tree is a non-circular graph.

Besides, to prevent the addition of a needless branch stopping at a non-destination node, we require that only destination nodes are possible as leaves in a tree. Formally:

$$\sum_{\substack{e_{ij} \in \mathcal{T}(v_n) \\ 0 \le h \le h_{\max}}} \sum x_{ij}^{h,k} - \sum_{\substack{e_{ij} \in \mathcal{O}(v_n) \\ 0 \le h \le h_{\max}}} \sum x_{ij}^{h,k} \le \mathbb{1}_{v_n}^{\mathcal{D}^k} \qquad \begin{matrix} \forall v_n \in \mathcal{V} \\ \forall f^k \in \mathcal{F} \end{matrix} \tag{5}$$

where $\mathbb{1}_x^Y$ corresponds to a set Y's indicator function, which equals to 1 only when $x \in Y$ and 0 otherwise, and $\mathcal{O}(v_n), \mathcal{T}(v_n)$ are the sets of outgoing links from v_n and terminating links in v_n respectively.

To ensure that the tree is formed only when it provides a complete delivery to all destinations, we add the following constraints:

$$\sum_{\substack{e_{ij} \in \mathcal{T}(v_n) \\ 0 \le h \le h_{\max}}} \sum x_{ij}^{h,k} \ge \mathbb{1}_{v_n}^{\mathcal{D}^k} \cdot t_{s^k}^k \qquad \begin{matrix} \forall v_n \in \mathcal{V} \\ \forall f^k \in \mathcal{F} \end{matrix} \tag{6}$$

$$t_n^k \ge \sum_{0 \le h \le h_{\max}} x_{nm}^{h,k} \qquad \begin{matrix} \forall e_{nm} \in \mathcal{E} \\ \forall f^k \in \mathcal{F} \end{matrix} \tag{7}$$

$$t_n^k \le \sum_{\substack{e_{ij} \in \mathcal{O}(v_n) \\ 0 \le h \le h_{\max}}} \sum x_{ij}^{h,k} \qquad \begin{matrix} \forall v_n \in \mathcal{V} \\ \forall f^k \in \mathcal{F} \end{matrix} \tag{8}$$

where t_n^k is a 0-1 auxiliary variable which is fixed by constraints (7) and (8) to indicate whether v_n acts as a relay node (i.e., a non-leaf node for flow f^k).

In addition, to ensure that the concurrently-admitted flows have non overlapping relay nodes, a node v_n is required to transmit at most one flow. Formally,

$$\sum_{f^k \in \mathcal{F}} t_n^k \le 1 \qquad \forall v_n \in \mathcal{V} \tag{9}$$

However, a node v_n may act, at once, as a source and a destination for two distinct flows. This case is not permitted. Formally,

$$\sum_{f^k \in \mathcal{F}} \left(\delta_{v_n}^{s^k} + \mathbb{1}_{v_n}^{\mathcal{D}^k} \right) \cdot t_{s^k}^k \le 1 \qquad \forall v_n \in \mathcal{V} \tag{10}$$

Fig. 2 illustrates an example of a routing tree generated according to the above constraints.

In line with the LTE-D2D standard [1], UEs are characterized by *half-duplex* D2D transmission in the side-link

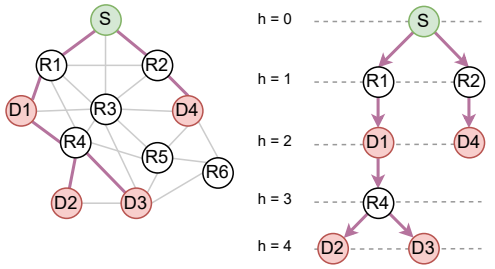


Fig. 2. Example of a constructed routing tree.

interface. Therefore, during a given SL, a node can act as a transmitter or a receiver but not both simultaneously. To cope with this hardware limitation, while reducing total end-to-end delay, we propose to schedule active links in an alternating fashion. To do so, we divide the set of active nodes $\mathcal{V}_\text{G} \subseteq \mathcal{V}$ into two *half-duplex sets*: $\mathcal{V}_\text{G}^0$ and $\mathcal{V}_\text{G}^1$. During a SL frame, the eNB scheduler addresses the RB allocations of one given set $\mathcal{V}_\text{G}^p$ depending on the parity p of the frame index τ (i.e., $p = \tau \mod 2$). Nodes belonging to the second set act as receivers. During the next frame, half-duplex sets switch their roles. As a result of this strategy, nodes in routing trees are scheduled according to the parity of their hop index in the routing tree (i.e., tree level). In other words, a parent node must belong to a different half-duplex set than its children. The node half-duplex allocation decision is embodied by the following constraint using the notations defined in TABLE I:

$$\sum_{\substack{f^k \in \mathcal{F} \\ 0 \le h \le h_\text{max}}} \sum x_{ij}^{h,k} \le H_i + H_j \le 2 - \sum_{\substack{f^k \in \mathcal{F} \\ 0 \le h \le h_\text{max}}} \sum x_{ij}^{h,k} \quad \forall e_{ij} \in \mathcal{E} \tag{11}$$

We assign a bandwidth B_SL, composed of Ω contiguous RBs, to the SL operation. Note that only contiguous RB allocations are feasible within this bandwidth because the SL has the same communication proprieties as the UL [7]. To do that, we enumerate all these allocations in the SL using a matrix $\mathcal{Z}_{\Omega \times U} = [z_{\omega,u}]$ in which columns represent the whole set contiguous patterns. The number of columns is given as $U = \frac{1}{2}\Omega(\Omega+1)$. Fo instance, all contiguous allocations for $\Omega = 4$ RBs are listed as columns in the following matrix:

$$\mathcal{Z}_{4\times 10} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

For example: the seventh column represents two RBs are allocated, namely the 3rd and 4th ones.

Furthermore, we model the RB allocation decision, for an active node v_n, as a set of respective 0-1 variables $y_{u,n} \forall u \in U$ which indicates the selected allocated pattern (i.e., column) u of the matrix $\mathcal{Z}$. This decision is constrained by:

$$\sum_{u=1}^{U} y_{u,n} \le \sum_{f^k \in \mathcal{F}} t_n^k \qquad \forall v_n \in \mathcal{V} \tag{12}$$

stipulating that only one allocation is possible for a node when it is acting as a (re-)transmitter for a flow.

To continue the formulation, additional auxiliary variables, whose definition are in TABLE I, are derived from the decision variables H_n and $y_{u,n}$ as detailed hereafter.

To indicate whether a RB ω is used by v_n, we introduce the 0-1 variable R_n^ω whose value is deduced from the respective variables $y_{u,n}$ by the following constraint:

$$R_n^\omega \triangleq \sum_{u=1}^{U} y_{u,n} z_{\omega,u} \qquad \begin{matrix} \forall v_n \in \mathcal{V} \\ \forall 1 \le \omega \le \Omega \end{matrix} \tag{13}$$

Furthermore, additional variables $R_n^{\omega,0}$ and $R_n^{\omega,1}$ are defined to indicate whether the RB ω is used by v_n in $\mathcal{V}_\text{G}^0$ or $\mathcal{V}_\text{G}^1$, i.e. half-duplex set of frames, respectively. Formally,

$$R_n^{\omega,0} \triangleq R_n^\omega - R_n^{\omega,1} \qquad \begin{matrix} \forall v_n \in \mathcal{V} \\ \forall 1 \le \omega \le \Omega \end{matrix} \tag{14}$$

$$R_n^{\omega,1} \triangleq H_n \cdot R_n^\omega \qquad \begin{matrix} \forall v_n \in \mathcal{V} \\ \forall 1 \le \omega \le \Omega \end{matrix} \tag{15}$$

B_n is the number of RB used by v_n and it is equal to:

$$B_n \triangleq \sum_{\omega=1}^{\Omega} R_n^\omega \qquad \forall v_n \in \mathcal{V} \tag{16}$$

An additional set of link-level auxiliary 0-1 variables are introduced as follows:

$$R_{ij} \triangleq \sum_{\substack{0 \le h \le h_\text{max} \\ f^k \in \mathcal{F}}} \sum x_{ij}^{h,k} \qquad \forall e_{ij} \in \mathcal{E} \tag{17}$$

$$R_{ij}^{\omega,p} \triangleq R_i^{\omega,p} \cdot R_{ij} \qquad \begin{matrix} \forall e_{ij} \in \mathcal{E} \\ \forall 1 \le \omega \le \Omega \\ \forall p \in \{0,1\} \end{matrix} \tag{18}$$

$$\phi_{n,ij}^{\omega,p} \triangleq R_n^{\omega,p} \cdot R_{ij}^{\omega,p} \qquad \begin{matrix} \forall v_n \in \mathcal{V}, \forall e_{ij} \in \mathcal{E} \\ \forall 1 \le \omega \le \Omega \\ \forall p \in \{0,1\} \end{matrix} \tag{19}$$

where R_{ij} indicates if e_{ij} is used for some flow. $R_{ij}^{\omega,p}$ indicates if the RB ω is used for the scheduled link e_{ij} during the p^th half duplex set. $\phi_{n,ij}^{\omega,p}$ is an interference indicator between node v_n and link e_{ij} on the RB ω.

To adhere to a linear formulation, a further step is needed to linearize Constraints (15), (18) and (19), which contain product terms. We make use of standard technique by introducing for each term $x \cdot y$ an additional auxiliary 0-1 variable λ_{xy} add three more linear constraints as follows:

$$(\lambda_{xy} \le x) \wedge (\lambda_{xy} \le y) \wedge (\lambda_{xy} \ge x + y - 1) \tag{20}$$

To increase the RB reutilization and reduce power consumption, we require that UEs cannot allocate RBs more than the flows' requests. Formally,

$$B_n \le \Omega + (D^k - \Omega) \cdot t_n^k \qquad \begin{matrix} \forall v_n \in \mathcal{V} \\ \forall f^k \in \mathcal{F} \end{matrix} \tag{21}$$

Note that the relation between flow bit-rate R^k and the respective demand for RBs D^k is defined by [7] as:

$$R^k = \frac{\text{TBS}\left(\text{MCS}, D^k\right)}{C} \qquad \text{[Mbps]} \tag{22}$$

where TBS is the MAC transport block size function in bits as defined in [7] considering a baseline *modulation and coding scheme* (MCS) for the SL. C is a constant equal to 1000.

In our model, we adopt a fixed power density scheme for the D2D emission power. In this scheme, the total emission power $S_{\text{tx},n}$ of a node is proportional to the number of allocated RBs B_n. Formally,

$$S_{\text{tx},n} = \Psi_{t,n} \cdot B_n \qquad \text{[mW]} \tag{23}$$

Furthermore, we assume a common emission power density for the D2D nodes (i.e., $\Psi_{t,n} = \Psi_t$, $\forall v_n \in \mathcal{V}$).

Following the same *per-RB* treatment and assuming *flat block-fading* channel model, the overall Signal-to-Interference-plus-Noise Ratio (SINR) on the link e_{ij} is equal to:

$$\gamma_{ij} = \frac{g_{ij}\Psi_{t,i}}{\sum\limits_{v_n \in \mathcal{V}} g_{nj}\Psi_{t,n} + \Psi_\sigma} \tag{24}$$

where Ψ_σ and $\Psi_{t,n}$ represent the spectral densities (per RB) of the thermal noise and the transmission from v_n, and g_{ij} is the channel gain between the node pair (v_i, v_j).

In face of the reutilization of RBs, system performance is limited by interference caused by nodes transmitting using the same RB. To optimize the performances by minimizing interferences, SINR must be upper-bounded by a common threshold γ. To formulate this constraint on RB allocations, we translate this limit (i.e., SINR $\le \gamma$) into the inequality

TABLE II
UE POWER CONSUMPTION MODEL PARAMETERS.

Parameter	Value	Parameter	Value
$P_{\text{tx}}^{\text{const}}$	883.52 mW	a_1^{rx}	24.8 mW
$P_{\text{rx}}^{\text{const}}$	878.1 mW	a_2^{rx}	7.86 mW
s_1^{tx}	0.2 dBm	a^{R}	8.16 mW
s_2^{tx}	11.4 dBm	b^{R}	0.97 mW/Mbps
s_1^{rx}	52.5 dBm	b_1^{tx}	0.78 mW/dBm
a_1^{tx}	23.6 mW	b_2^{tx}	17 mW/dBm
a_2^{tx}	45.4 mW	b_1^{rx}	0.04 mW/dBm
		b_2^{rx}	0.11 mW/dBm

$\mathcal{N}+\mathcal{I} \leq P_r/\gamma$ where P_r is the received power. Consequently,

$$\Psi_\sigma R_{ij}^{\omega,p} + \sum_{n\neq i} g_{nj}\Psi_t \cdot \phi_{n,ij}^{\omega,p} \leq \frac{g_{ij}\Psi_t}{\gamma} R_{ij}^{\omega,p} \quad \begin{matrix}\forall e_{ij}\in\mathcal{E} \\ \forall 1\leq\omega\leq\Omega \\ \forall p\in\{0,1\}\end{matrix} \quad (25)$$

where constraint (25) ensures that the SINR is below the threshold γ considering RB allocations and active nodes interfering in the same half-duplex set $\mathcal{V}_{\text{G}}^p$. The auxiliary 0-1 variable $R_{ij}^{\omega,p}$ indicates that the link e_{ij} is scheduled to transmit together with half-duplex set $\mathcal{V}_{\text{G}}^p$ on the RB ω.

To evaluate the effect of energy consumption, similar to [8], we make use of the following empirical model defined in [9], to calculate the total communication consumed power due to the D2D operations at both ends. At the transmitting end, the total consumed power is given by:

$$P_{\text{tx}}^{\text{D2D}} = P_{\text{tx}}^{\text{const}} + P_{\text{tx}}^{\text{RF}}(S_{\text{tx}}) \quad (26)$$

$$P_{\text{tx}}^{\text{RF}}(S_{\text{tx}}) = \begin{cases} b_1^{\text{tx}} \cdot S_{\text{tx}} + a_1^{\text{tx}} & \text{if } S_{\text{tx}} \leq s_1^{\text{tx}} \\ b_2^{\text{tx}} \cdot S_{\text{tx}} + a_2^{\text{tx}} & \text{if } s_1^{\text{tx}} < S_{\text{tx}} \leq s_2^{\text{tx}} \end{cases}$$

where $P_{\text{tx}}^{\text{D2D}}$ includes the constant term $P_{\text{tx}}^{\text{const}}$ related to the baseband circuit consumption when the D2D transmitter is active. $P_{\text{tx}}^{\text{RF}}$ is the total RF block consumption in terms of the power emitted S_{tx} from the antenna in dBm. Hence, the power consumed by a node v_n, due to the transmission of the flow f^k, can be estimated as:

$$\Pi_{\text{tx},n}^k = P_{\text{tx}}^{\text{const}} + P_{\text{tx}}^{\text{RF}}\left(S_{\text{tx,n}}^k\right) \quad [\text{mW}] \quad (27)$$

$$S_{\text{tx,n}}^k = \text{dBm}\left(\Psi_{t,n} D^k\right) \quad (28)$$

Similarly, at the receiving end of an active link, the total consumed power is equal to:

$$P_{\text{rx}}^{\text{D2D}} = P_{\text{rx}}^{\text{const}} + P_{\text{rx}}^{\text{RF}}(S_{\text{rx}}) + P_{\text{rx}}^{\text{BB}}(R) \quad (29)$$

$$P_{\text{rx}}^{\text{RF}}(S_{\text{rx}}) = \begin{cases} -b_1^{\text{rx}} \cdot S_{\text{rx}} + a_1^{\text{rx}} & \text{if } S_{\text{rx}} \leq -s_1^{\text{rx}} \\ -b_2^{\text{rx}} \cdot S_{\text{rx}} + a_2^{\text{rx}} & \text{if } S_{\text{rx}} > -s_1^{\text{rx}} \end{cases}$$

$$P_{\text{rx}}^{\text{BB}}(R) = b^{\text{R}} \cdot R + a^{\text{R}}$$

where $P_{\text{rx}}^{\text{D2D}}$ includes the constant term $P_{\text{rx}}^{\text{const}}$, related to the receiving circuit being active, and $P_{\text{rx}}^{\text{RF}}$, which gives the total RF block consumption in terms of the power received S_{rx} at the antenna in dBm. The additional term $P_{\text{rx}}^{\text{BB}}$ gives the rate-dependent power consumption in the base-band block of the device. Therefore, the power consumption at the receiver of an active link e_{ij}, due to the reception of the flow f^k, is estimated by:

$$\Pi_{\text{rx},ij}^k = P_{\text{rx}}^{\text{const}} + P_{\text{rx}}^{\text{RF}}\left(S_{\text{rx},ij}^k\right) + P_{\text{rx}}^{\text{BB}}\left(R^k\right) \quad [\text{mW}] \quad (30)$$

$$S_{\text{rx},ij}^k = \text{dBm}\left(g_{ij} \cdot \Psi_{t,i} D^k\right) \quad (31)$$

where R^k is the respective flow bit-rate defined in equation (22).

TABLE II illustrates the parameters values of the above power consumption model.

We model the impact of node participation in routing on its residual energy by proposing a ranking method that takes into consideration the current distribution of residual energy in the system. For each non-dead node v_n, we assign a fractional rank $\Lambda_n \in (0, 1]$ as following:

$$\Lambda_n(\tau) = \frac{1}{1 + \left\lfloor \frac{E_n(\tau) - E_{\min}(\tau)}{\sigma_E(\tau)} \right\rfloor} \quad (32)$$

where $E_n(\tau)$, $E_{\min}(\tau)$ and $\sigma_E(\tau)$ respectively represents i) node's residual energy, ii) minimum residual energy in the network, and iii) standard deviation of residual energy distribution at the beginning of the SL frame τ. Note that high fractional rank means high impact on the node's residual energy.

The time evolution of the residual energy is estimated at the eNodeB as detailed hereafter:

$$E_n(\tau) = E_n(\tau - 1) - P^{\text{D2D}} \cdot T_{\text{SL}} \quad (33)$$

$$P^{\text{D2D}} = \begin{cases} P_{\text{tx}}^{\text{D2D}} \text{ if } v_n \text{ was transmitting in frame } \tau - 1 \\ P_{\text{rx}}^{\text{D2D}} \text{ if } v_n \text{ was receiving in frame } \tau - 1 \\ 0 \quad \text{if } v_n \text{ was idle in frame } \tau - 1 \end{cases}$$

assuming that each node has initial energy budget $E_n(0)$.

Above, we have defined all the variables and constraints (2) – (19) addressing i) routing, ii) OFDMA RB allocation and iii) energy consumption. Now, we can complete the formulation by defining the objective function:

$$\max_{x_{ij}^{h,k}, H_n, \ldots} \quad \frac{1}{\aleph_{\text{B}}} \sum_{v_n\in\mathcal{V}} B_n + \frac{1}{\aleph_{\text{A}}} \sum_{f^k\in\mathcal{F}} t_{s^k}^k - \frac{1}{\aleph_{\text{R}}} \sum_{\substack{v_n\in\mathcal{V} \\ f^k\in\mathcal{F}}} \sum \Lambda_n t_n^k$$
$$- \frac{1}{\aleph_{\text{tx}}} \sum_{\substack{v_n\in\mathcal{V} \\ f^k\in\mathcal{F}}} \sum \Pi_{\text{tx},n}^k t_n^k - \frac{1}{\aleph_{\text{rx}}} \sum_{e_{ij}\in\mathcal{E}} \sum_{0\leq h\leq h_{\max}, f^k\in\mathcal{F}} \sum \Pi_{\text{rx},ij}^k x_{ij}^{h,k} \quad (34)$$

where the normalizing factors defined by:

$$\aleph_{\text{B}} \triangleq \Omega \cdot |\mathcal{V}|, \aleph_{\text{A}} \triangleq |\mathcal{F}_{\text{C}}|, \aleph_{\text{R}} \triangleq \sum_{v_n\in\mathcal{V}} \Lambda_n,$$
$$\aleph_{\text{tx}} \triangleq \sum_{\substack{v_n\in\mathcal{V} \\ f^k\in\mathcal{F}}} \sum \Pi_{\text{tx},n}^k, \aleph_{\text{rx}} \triangleq \sum_{e_{ij}\in\mathcal{E}, f^k\in\mathcal{F}} \sum \Pi_{\text{rx},ij}^k \quad (35)$$

The terms in the objective function in (34), represent *respectively* a normalized equal-weight multi-objective formulation of eNodeB goal to achieve the following objectives: i) increasing the number of RB allocated for each flow, ii) increasing the number of admitted flows in the system, iii) lowering the routing impact on nodes' residual energy, iv) lowering the power consumption in the relaying process at the transmitting side, v) lowering the power consumption in the relaying process at the receiving side.

To quantify the ILP model's size complexity, we cite its *column-size*, i.e., number of variables, and its *row-size*, i.e., number of constraints. As for our model, an asymptotic analysis shows that, in terms of $\mathbb{G}$, $\mathcal{F}$, Ω and $h_{\max}$, the ILP model has column-size of $\mathcal{O}\left(|\mathcal{V}|\,\Omega^2 + |\mathcal{V}|\,|\mathcal{E}|\,\Omega + |\mathcal{V}|\,|\mathcal{F}| + |\mathcal{E}|\,|\mathcal{F}|\,h_{\max}\right)$ and a row-size of $\mathcal{O}\left(|\mathcal{V}|\,|\mathcal{E}|\,\Omega + |\mathcal{V}|\,|\mathcal{F}| + |\mathcal{E}|\,|\mathcal{F}|\,h_{\max}\right)$.

IV. PROPOSAL: JRRA−EE

To solve our energy-aware joint routing and OFDMA resource block allocation problem, formulated in the above section as an ILP model, we propose a two-stage heuristic algorithm, based on the branch-and-cut method, named Joint Routing and Resource Allocation Energy Efficient (JRRA−EE). It is worth noting that our scheme is centralized and is handled by the eNodeB.

JRRA−EE adopts an online bulk strategy by considering for the SL frame τ resolution, all active scheduled flows and the waiting flows up to the previous SL frame. However, instead of considering all the waiting flows $\mathcal{F}_{\text{W}}$ for admittance, it proceeds by an initial stage of pre-routing to filter the waiting

Algorithm 1 JRRA−EE pseudo-code

for each SL frame τ **do**
  for each $f^k \in \mathcal{F}_{\text{A}}$ **do** ▷ Arriving Flows
    $\mathcal{F}_{\text{W}} \leftarrow \mathcal{F}_{\text{W}} \cup \{ f^k \}$
  end for
  for each $f^k \in \mathcal{F}_{\text{FIN}}$ **do** ▷ Finished Flows
    $\mathcal{V}_{\text{D}} \leftarrow \mathcal{V}_{\text{D}} \cup \text{NodesOF}\left(\mathbb{T}^k\right)$
  end for
  Execute Algorithm 2 ▷ Pre-routing
  Construct the ILP model as in formula (34)
  Solve the ILP model using branch-and-cut
  for each $f^k \in \mathcal{F}_{\text{C}}$ **do**
    if $t^k_{s_k} = 1$ **then** ▷ Flow is admitted
      Configure $\mathbb{T}^k$ according to $x^{h,k}_{ij}$
    end if
  end for
  $p \leftarrow \tau \mod 2$
  for each $v_n \in \mathcal{V}^p_{\text{G}}$ **do**
    Allocate RBs for v_n according to $y_{u,n}$
  end for
end for

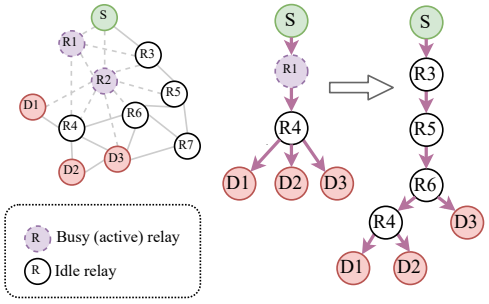


Fig. 3. Pre-routing tree formation and its deviation.

flows down to a set of candidate flows $\mathcal{F}_{\text{C}}$. The rational behind this initial stage is to reduce the size complexity of the ILP model by reducing the number of considered flows $\mathcal{F}$ and also by setting the model parameter $h_{\max}$ to a reasonable value. It is worth noting that high values for $h_{\max}$ implies more possible routing trees to discover while low values few routing trees and hence few admitted flow into the system. The pseudo-code of JRRA−EE illustrated in Algorithm 1.

The pre-routing stage proceeds as follows. For each waiting flow f^k, the eNodeB checks if it is possible to construct a routing tree from the source node to all destinations using *breadth-first-traversal* and considering only the currently idle nodes. We recall that each node cannot handle more than one flow. Note that such tree construction stops once all destinations are reached. If such *pre-routing* tree $\widetilde{\mathbb{T}}^k$ exists then the flow f^k is added to the set of candidate flows $\mathcal{F}_{\text{C}}$. In the other case, the flow is kept waiting for upcoming opportunities in subsequent frames. Thanks to the breadth-first-traversal, pre-routing trees are *well-balanced* as they tend to be short one-to-many routing trees. However, due to the dynamic state (e.g., end of current flows, low battery, etc.) of nodes, pre-routing trees also tend to deviate from this preferred condition as illustrated in Fig. 3. The pseudo-code of the pre-routing trees construction is illustrated in Algorithm 2.

Taking advantage of the dynamic nature of pre-routing trees construction stage, the latter goes one step further to set the parameter value $h_{\max}$ of the current ILP model based on the reported trees heights and those of the routing trees of active

Algorithm 2 Pre-routing of routing trees pseudo-code

Inputs: $\mathcal{V}_{\text{D}}, \mathbb{T}^k \, \forall f^k \in \mathcal{F}_{\text{S}}$, $\mathcal{F}_{\text{W}}$
Outputs: $\mathcal{F}_{\text{C}}, h_{\max}$

$\mathcal{F}_{\text{C}} \leftarrow \emptyset, h \leftarrow 0, \widetilde{h} \leftarrow 0$
for each $f^k \in \mathcal{F}_{\text{S}}$ **do** ▷ Trees of active flows
  if HeightOF $\left(\mathbb{T}^k\right) > h$ **then**
    $h \leftarrow$ HeightOF $\left(\mathbb{T}^k\right)$
  end if
end for
for each $f^k \in \mathcal{F}_{\text{W}}$ **do**
  if $\{v_{s^k}\} \cup \mathcal{D}^k \nsubseteq \mathcal{V}_{\text{D}}$ **then go to** 30
  end if
  $Q \leftarrow \emptyset$ ▷ New empty queue
  push v_{s^k} into Q
  $S \leftarrow \{v_{s^k}\}$
  LevelOF$(v_{s^k}) \leftarrow 0$
  while $Q \neq \emptyset \wedge \mathcal{D}^k \nsubseteq S$ **do** ▷ Breadth-first traversal
    $v_i \leftarrow$ Q.pop()
    if LevelOF$(v_i) > \widetilde{h}$ **then**
      $\widetilde{h} \leftarrow$ LevelOF(v_i)
    end if
    for each $e_{ij} \in \mathcal{O}(v_i)$ **do**
      if $v_j \notin S \wedge v_j \in \mathcal{V}_{\text{D}}$ **then**
        push v_j into Q
        $S \leftarrow S \cup \{ v_j \}$
        LevelOF$(v_j) \leftarrow$ LevelOF$(v_i) + 1$
      end if
    end for
  end while
  if $\mathcal{D}^k \subseteq S$ **then** ▷ Add f^k to candidates
    $\mathcal{F}_{\text{C}} \leftarrow \mathcal{F}_{\text{C}} \cup \{f^k\}$
  end if
end for
$h_{\max} \leftarrow \max\left\{h, \beta \cdot \widetilde{h}\right\}$ ▷ Update $h_{\max}$

scheduled flows as follows:

$$h_{\max} = \max\left\{ \max_{f^k \in \mathcal{F}_{\text{S}}} \mathfrak{H}\left(\mathbb{T}^k\right), \beta \max_{f^k \in \mathcal{F}_{\text{C}}} \mathfrak{H}\left(\widetilde{\mathbb{T}}^k\right) \right\} \quad (36)$$

where i) $\mathfrak{H}(\cdot)$ denotes the height-of-tree operator and ii) $\beta \geq 1$ is a "tradeoff-margin" factor to allow for longer routing trees to be explored and more flows to be admitted into the system when solving the current ILP model. After this initial stage, we drastically reduce the size of the solutions, hence the eNodeB can solve the resulted ILP model using the branch-and-cut method and the convergence time is tiny.

V. Performance Evaluation

In this section, we report the performance of our proposal JRRA−EE by performing a series of detailed simulations. We start by describing the network simulation environment setup. Afterwards, we define the performance metrics to evaluate our strategy. Finally, we analyze the results and discuss the effectiveness of our proposal JRRA−EE based on multiple-run simulations which invoke confidence-interval analysis with a confidence level of 95%.

A. Network simulation environment

We make use of NS-3 network simulator based on C++ language and widely used by the network research community. NS-3 supports a variety of conventional 3GPP LTE simulation scenarios through the module NS-3/LTE [10]. To realize the LTE-D2D standard, we integrate new features to NS-3 in order to support LTE-D2D protocol stack (side-link interface). In this context, we developed the necessary LTE-D2D procedures for the layers: PHY, MAC and PDCP/RLC. We also

TABLE III
SIMULATION PARAMETERS

Parameter	Value
Cell Radius Rcell	1 km
UL/SL Frequency f_{UL}	1930 MHz
UL/SL (Reference) Bandwidth B_{UL}	5 MHz (25 LTE RBs)
SL RBs Used Actually Ω	14 LTE RBs
SL frame (LTE-D2D SC-Period)	40 subframes (40 ms)
Data Part in SL frame	32 subframes
UE SL Power Transmit Density Ψ_t	-4 dBm/RB
Noise Spectral Density Ψ_n	-121.45 dBm/RB
LTE MCS Index used in SL	9 (QPSK)
UE Density λ_{UE}	$\{10, 15, 20, 25, 30, 35, 40\}$ per km^2
UE-UE SNR Threshold γ_{TOPO}	10 dB
Scheduling SINR Threshold γ	6 dB
UE Initial Energy Budget $E_n(0)$	3.856 Joules
Flow Simulation Period	10 seconds
Flow Arrival Process	Poisson Process
Flow Arrival Rates λ_{FL}	$\{10, 20\}$ flows/second
Flow Duration Random Variable	Exponential
Flow Duration Mean λ_{DUR}	1 second
Flow Bit Rate Classes	$\{25, 50, 75, 100, 125, 150, 175, 200\}$ kbps
Node-Flow Interest Probability ρ	0.1
$h_{\max}$ update factor β	1.5

implemented the signaling necessary to: i) configure the SL parameters, ii) establish SL Radio Bearers (SLRBs), and iii) exchange SL reports and grants.

In line with our formulation in Section III, we deploy one LTE macro-cell with radius Rcell $= 1$ km. The deployed UEs follow a Poisson Point Process distribution with a density λ_{UE} nodes per km^2 for values from $\{10, 15, 20, 25, 30, 35, 40\}$. The LTE macro-cell is configured to work with an UL/SL frequency of 1930 MHz (i.e., band 1) and a bandwidth of 5 MHz (i.e., 25 RBs). However, we assign only $\Omega = 14$ RBs for the actual SL bandwidth of D2D offloading operation. All UEs transmit on SL with a common power density of $\Psi_t = -4$dBm/RB which is equivalent to a maximum of 10 dBm over the whole 5 MHz. To model the SL path-loss (i.e., link gains g_{ij}), we make use of WINNER II B2-LOS channel model [11]. The SL frame duration is fixed to 40 milliseconds which corresponds to 40 LTE subframes. Note that only 32 subframes are actually used for data transmission while the initial 8 ones are used for SL control information. The eNodeB builds the D2D network topology making using of SNR reports and estimations (i.e., CQI metric). A communication link exists between two nodes if and only if the respective SNR is greater than a threshold $\gamma_{\text{TOPO}} = 10$ dB.

Simulated flows are generated following a Poisson process with an arrival rate equals to $\lambda_{\text{FL}} \in \{10, 20\}$ flows per second. Flow bit-rates are randomly selected from predefined Constant Bit Rate (CBR) classes. Flow duration distribution is simulated to follow an exponential random variable with a mean duration of $\lambda_{\text{DUR}} = 1$ second. Flows sources are selected according a random uniform distribution. As for destinations, they are selected for a given source assuming a *node-flow interest probability* of $\rho = 0.1$. In other words, once a flow source is selected, other nodes are evaluated for being interested in receiving the flow using Bernoulli trials with a success probability equals to ρ. TABLE III summarizes the main parameters used in simulations.

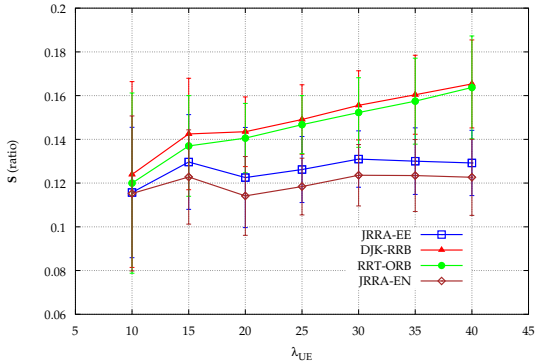
B. Performance metrics

As described in TABLE IV, we consider various metrics to evaluate purposes in our experiments. These metrics are grouped with respect to the following interests: i) I1: overall utility of offloading system, ii) I2: end-users' quality of service per flow, and iii) I3: energy consumption.

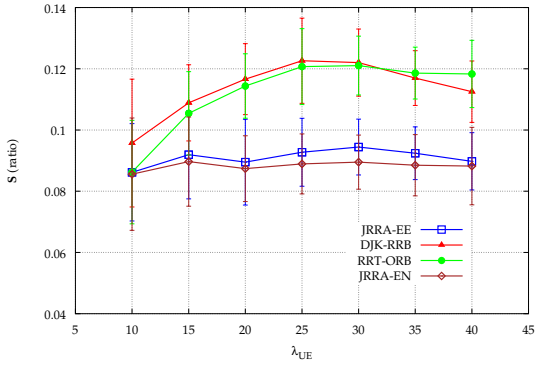
Unfortunately, the related strategies described in Section II cannot be compared with our proposal. The main reason

TABLE IV
PERFORMANCE METRICS

Metric	Definition	Interest
$\mathbb{S}$	ratio of the flows offloaded by the D2D subnetwork	I1
$\mathbb{I}$	ratio of interrupted flows (to the admitted ones) due to topology disruption by death of relays	I1,I3
$\mathbb{L}$	average of flow's packet loss in each simulation run	I2
$\mathbb{E}_n$	average network life time as n connected components	I3
$\mathbb{H}$	average height (hops) of trees in each simulation run	I2



(a) $\lambda_{\text{FL}} = 10$



(b) $\lambda_{\text{FL}} = 20$

Fig. 4. $\mathbb{S}$ w.r.t node density λ_{UE}.

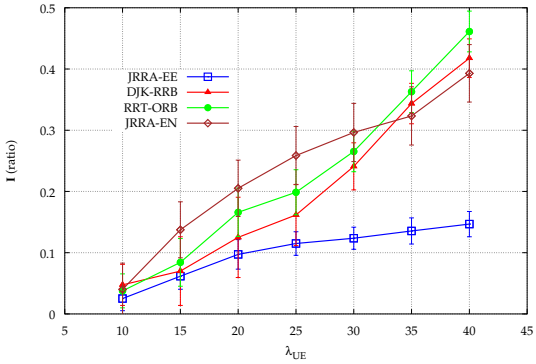
behind this is that the full optimization model considered in our paper (Section III) is not addressed at all in the related papers. Hence, we propose to compare JRRA-EE with the following variants:

1) DJK-RRB: is a pure path strategy that aims to find the optimal routing trees using the one-to-many version Dijkstra algorithm and then, allocates RB randomly.
2) RRT-ORB: is a pure resource block oriented strategy that finds the routing trees randomly using random walk on the topology graph, and allocates RB optimally.
3) JRRA-EN: is an energy non-aware variant of the original JRRA-EE. It relies, hence, on a modified objective function as described below:

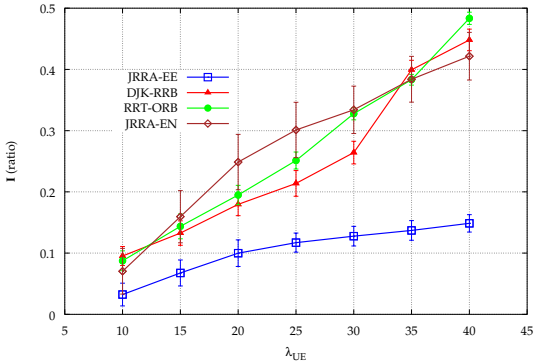
$$\max_{x_{ij}^{h,k}, H_n, \cdots} \quad \frac{1}{\aleph_{\text{B}}} \sum_{v_n \in \mathcal{V}} B_n + \frac{1}{\aleph_{\text{A}}} \sum_{f^k \in \mathcal{F}} t_{s^k}^k - \frac{1}{\aleph_{\text{N}}} \sum_{\substack{v_n \in \mathcal{V} \\ f^k \in \mathcal{F}}} \sum t_n^k$$

$$\text{with} \qquad \aleph_{\text{B}} \triangleq \Omega \cdot |\mathcal{V}|, \aleph_{\text{A}} \triangleq |\mathcal{F}_{\text{C}}|, \aleph_{\text{N}} \triangleq |\mathcal{V}| \qquad (37)$$

where the new normalized term $-\frac{1}{\aleph_{\text{N}}}[\cdots]$ represents the eNodeB's attempt to minimize the number of involved nodes in the offloading route.

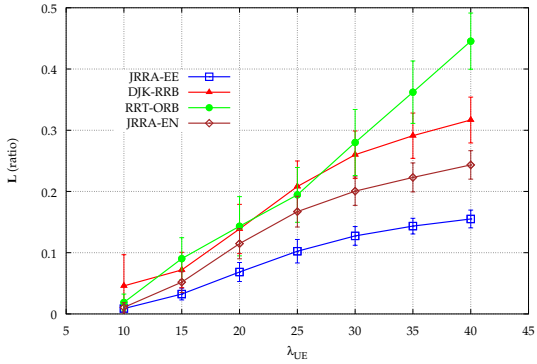


(a) $\lambda_{\text{FL}} = 10$

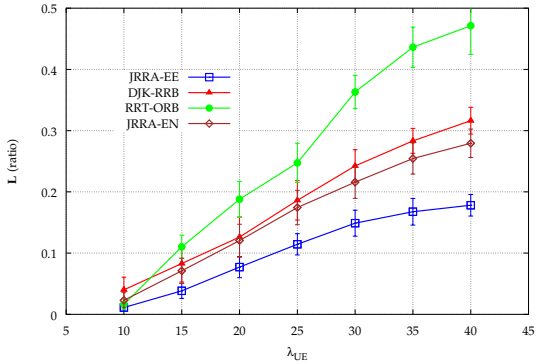


(b) $\lambda_{\text{FL}} = 20$

Fig. 5. $\mathbb{I}$ w.r.t node density λ_{UE}.



(a) $\lambda_{\text{FL}} = 10$



(b) $\lambda_{\text{FL}} = 20$

Fig. 6. $\mathbb{L}$ w.r.t node density λ_{UE}.

C. Simulation results

First, we evaluate the offloading capability of our approach compared with the related strategies. To do so, we measure the ratio of the flows offloaded by the D2D subnetwork. Fig. 4 illustrates $\mathbb{S}$ with respect to the density of UEs and under two traffic conditions $\lambda_{\text{FL}} = 10$ and $\lambda_{\text{FL}} = 20$ flows per second. We note that DJK−RRB generally outperforms the other strategies. This is expected since DJK−RRB routes flows over the fewest possible nodes (i.e., the smallest possible trees). Hence, it allows for more flows to be admitted. Taking DJK−RRB as a baseline, we note that our proposal JRRA−EE accepts, in average, around $\mathbb{S} = 12\%$ of the flows with $\lambda_{\text{FL}} = 10$ which is $\Delta\mathbb{S} = 4\%$ less than DJK−RRB as illustrated in Fig. 4a. On the other hand, Fig. 4b depicts $\mathbb{S}$'s variation under a higher traffic load. We notice that, for $\lambda_{\text{FL}} = 20$, the performance of JRRA−EE drops to around $\mathbb{S} = 9\%$. However, the most advantageous DJK−RRB also drops making the performance gap of JRRA−EE within $\Delta\mathbb{S} = 3\%$. It is straightforward to see that, even though JRRA−EE is outperformed by DJK−RRB and RRT−ORB, our proposal performs better than its energy non-aware variant JRRA−EN in terms of offloading capacity.

Fig. 5 depicts the ratio of interrupted flows according to the UEs' density for respectively $\lambda_{\text{FL}} = 10$ and $\lambda_{\text{FL}} = 20$ scenarios. From flows perspective, the admission rate (i.e., as illustrated in Fig. 4) alone is not sufficient and we have to ensure that the path is valid until the reception of all the packets. It is worth pointing out that our proposal achieves the lowest service interruption probability compared with the related strategies. Fig. 5a and Fig. 5b clearly demonstrate that JRRA−EE resists well to the traffic increase. In fact, it is able to maintain the service interruption rate $\mathbb{I}$ below 15% under both traffic conditions $\lambda_{\text{FL}} = 10$ and $\lambda_{\text{FL}} = 20$ flow per seconds. On the other hand, DJK−RRB which is able to maximize the offloading rate, struggles to resist to such an increase and may crash more than 40%.

Besides, to quantify the QoS in terms of packet error rate at the IP level, Fig. 6 illustrates the average flow's packet loss ($\mathbb{L}$) according to the UEs' density for respectively $\lambda_{\text{FL}} = 10$ and $\lambda_{\text{FL}} = 20$ scenarios. We recall that the packet loss does not come only from transmission error due to noise interference but may also be caused by the service interruption. Indeed, a flow may be disrupted because a relaying node has exhausted all its energy budget and consequently declared itself as dead. In Fig. 6, it is straightforward to see that JRRA−EE and JRRA−EN both outperform DJK−RRB and RRT−ORB thanks to their capability to take into consideration interference in OFDMA RB blocks allocation. However, RRT−ORB performs badly in general which may seems paradoxical. The rational behind this is RRT−ORB handles RBs allocation once the routes are randomly selected leaving few possibilities to allocate sufficient RBs to flows. It is straightforward to see that such a behaviour will lead to higher transmission delays. As a consequence, longer transmission delays paired with energy-agnostic node selection for routing is resulting in high packet loss due to the service disruption.
Being energy-aware makes JRRA−EE more robust against the packet loss. In fact, the latter succeeds to maintain $\mathbb{L}$ below 0.15 and 0.18 for both traffic conditions $\lambda_{\text{FL}} = 10$ and $\lambda_{\text{FL}} = 20$ flow per seconds respectively as depicted in Fig. 6a and Fig. 6b.

To highlight the energy efficiency of our proposal JRRA−EE, we make use of the metric $\mathbb{E}_n$ which measures the average lifetime of the D2D offloading system as a n connected components (i.e., evolution of network connectivity). Disruptions caused by nodes' energy shortage lead to the topology disconnection which, in its turn, degrades the overall utility of the D2D offloading system. In this regard, Fig. 7 highlights how JRRA−EE succeeds to keep the D2D topology

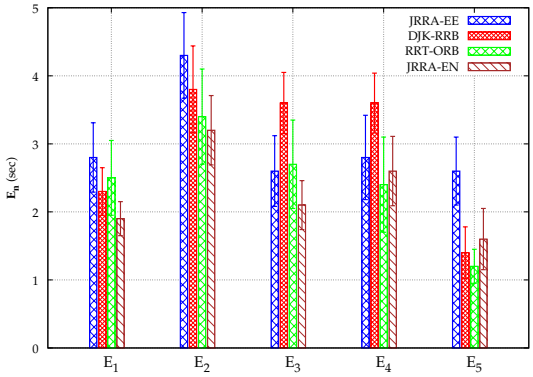
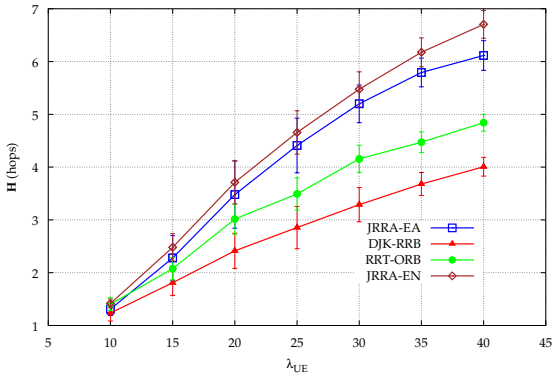
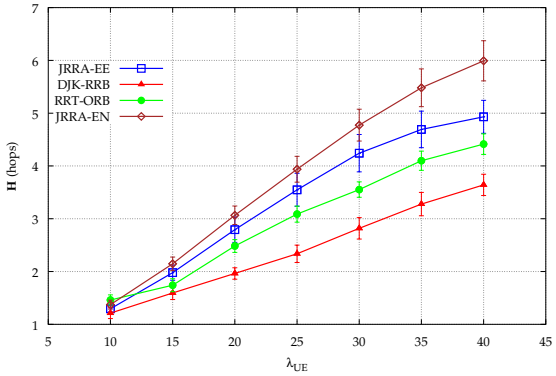


Fig. 7. Network connectivity lifetime $\mathbb{E}_n$ for $\lambda_{\text{FL}} = 20$.



(a) $\lambda_{\text{FL}} = 10$



(b) $\lambda_{\text{FL}} = 20$

Fig. 8. $\mathbb{H}$ w.r.t node density λ_{UE}

as few components as possible for longer times compared to the related schemes. It is worth noting that RRT−ORB and DJK−RRB achieve good performance in terms of network lifetime. The obtained results corroborate the previous ones depicted in Fig. 5 and Fig. 6. Indeed, the aforementioned strategies handle less traffic over the D2D sub-network due to their high service interruption level and packet loss rate. Consequently, nodes lifetime will be longer.

Fig. 8 illustrates the performance in terms of $\mathbb{H}$ metric which reflects the number of hops in the routing trees. This metric gives indication on the QoS presented to flows in terms of latencies where shorter is better. Specifically, the end-to-end and the average packet delays are in proportion to the product $\mathbb{H} \times T_{\text{SL}}$. Fig. 8a and Fig. 8b point out that the average number of hops increases almost linearly in accordance with the density of nodes λ_{UE}. As expected, DJK−RRB always yields the shortest number of hops by virtue of its strategy. We note that, in general, DJK−RRB and RRT−ORB lead to shorter paths and lower latencies compared with JRRA−EE. We recall, as illustrated above, that both DJK−RRB and RRT−ORB deteriorate service interruption and packet error rate metrics. In return, our proposal JRRA−EE outperforms its energy non-aware variant JRRA−EN.

In summary, network simulations show that JRRA−EE outperforms the variants in terms of network lifetime, packet loss and service interruption at the expense of small performance gaps with respect to latency and offloading capacity. Furthermore, JRRA−EE always outperforms its energy non-aware variant JRRA−EN.

VI. Conclusion

In this paper, we studied the LTE-D2D-based multihop offloading scheme for the intra-cell UE-to-UEs multicast/unicast flows. We proposed an energy-efficient offloading scheme to jointly solve the problem of multicast/unicast routing and the OFDMA resource block allocation. To increase the utility of the offloading system, the proposed scheme took into account the battery-limitation by defining an energy-budget for each cooperating UE. We also considered LTE-D2D-specific constraints: half-duplex operation and the contiguous resource block allocations. We formulated the problem as an ILP and we proposed a novel heuristic, named JRRA−EE, composed of a two-stage algorithm to solve it. We validated our proposal using the NS-3 simulator after implementing the whole LTE-D2D protocol stack. Through extensive simulations, we have shown that our proposed strategy JRRA−EE outperformed the related strategies. Performance gains manifested themselves as i) increased lifetime of the offloading network, ii) low packet loss and iii) low service interruption rates at the expense of small performance gaps in latency and offloading capacity.

References

[1] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (Release 14)," 3rd Generation Partnership Project (3GPP), TS 36.101, Sep. 2017. [Online]. Available: http://www.3gpp.org/DynaReport/36101.htm

[2] G. Nemhauser and L. Wolsey, "Computational complexity," in *Integer and Combinatorial Optimization*. John Wiley & Sons, Inc., 1988, pp. 114–145. [Online]. Available: http://dx.doi.org/10.1002/9781118627372.ch5

[3] T. Ta, J. S. Baras, and C. Zhu, "Improving smartphone battery life utilizing device-to-device cooperative relays underlaying lte networks," in *2014 IEEE International Conference on Communications (ICC)*, June 2014, pp. 5263–5268.

[4] B. Liu, Y. Cao, W. Wang, and T. Jiang, "Energy budget aware device-to-device cooperation for mobile videos," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–7.

[5] A. Laha, X. Cao, W. Shen, X. Tian, and Y. Cheng, "An energy efficient routing protocol for device-to-device based multihop smartphone networks," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 5448–5453.

[6] Z. Jingyi, L. Xi, and X. Quansheng, "Multi-hop routing for energy-efficiency enhancement in relay-assisted device-to-device communication," *The Journal of China Universities of Posts and Telecommunications*, vol. 22, no. 2, pp. 1–51, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S100588851560632X

[7] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 14)," 3rd Generation Partnership Project (3GPP), TS 36.213, Sep. 2017. [Online]. Available: http://www.3gpp.org/DynaReport/36213.htm

[8] M. Hoeyhtyae, A. Maemmelae, U. Celentano, and J. Roening, "Power-efficiency in social-aware d2d communications," in *European Wireless 2016; 22th European Wireless Conference*, May 2016, pp. 1–6.

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

[10] N. Baldo, M. Miozzo, M. Requena-Esteso, and J. Nin-Guerrero, "An open source product-oriented LTE network simulator based on ns-3," in *Proceedings of the 14th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*. ACM, 2011, pp. 293–298.

[11] Y. d. J. Bultitude and T. Rautiainen, "IST-4-027756 WINNER II D1. 1.2 V1. 2 WINNER II Channel Models," 2007.