Joint Mode Selection and Resource Allocation for Mobile Relay-Aided Device-to-Device Communication

Rui Tang[1], Jihong Zhao[1, 2], Hua Qu[1], Zhengcang Zhu[1] and Yanpeng Zhang[3]
[1] School of Electronics and Information Engineering, Xi'an Jiaotong University
Xi'an 710049, Shaanxi - P. R. China
[e-mail: allanxjtu@163.com]
[2] School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications
Xi'an 710061, Shaanxi - P. R. China
[3] School of Software Engineering, Xi'an Jiaotong University
Xi'an 710049, Shaanxi - P. R. China
*Corresponding author: Rui Tang

Abstract

Device-to-Device (D2D) communication underlaying cellular networks is a promising add-on component for future radio communication systems. It provides more access opportunities for local device pairs and enhances system throughput (ST), especially when mobile relays (MR) are further enabled to facilitate D2D links when the channel condition of their desired links is unfavorable. However, mutual interference is inevitable due to spectral reuse, and moreover, selecting a suitable transmission mode to benefit the correlated resource allocation (RA) is another difficult problem. We aim to optimize ST of the hybrid system via joint consideration of mode selection (MS) and RA, which includes admission control (AC), power control (PC), channel assignment (CA) and relay selection (RS). However, the original problem is generally NP-hard; therefore, we decompose it into two parts where *a hierarchical structure* exists: (i) PC is *mode-dependent*, but its optimality can be perfectly addressed for any given mode with additional AC design to achieve individual quality-of-service requirements. (ii) Based on that optimality, the joint design of MS, CA and RS can be viewed from the graph perspective and transferred into the *maximum weighted independent set* problem, which is then approximated by our greedy algorithm *in polynomial-time*. Thanks to the numerical results, we elucidate the efficacy of our mechanism and observe a resulting gain in MR-aided D2D communication.

Keywords: Device-to-Device communication; mobile relay; mode selection; resource allocation; system throughput; graph theory

1. Introduction

Device-to-Device (D2D)[1] communication is viewed as a promising add-on component in future radio communication systems [1-3]. It is a novel communication paradigm where an instance of user equipment (UE) transmits data directly to another UE in the vicinity without requiring a base station (BS) as a relay. D2D has the potential to enhance spectral efficiency and improves quality-of-service (QoS). Moreover, it offloads the relay burden from the BS, shortens transmission latency and enables emerging context-aware applications [4]. For these reasons, mainstream telecommunications organizations and companies, such as the 3rd Generation Partnership Project [4-5], the Nokia Research Center [1], the European Union's Research and Innovation Funding Programme [2] and IEEE 802.15 (Personal Area Networks) [3], are addressing the standardization of D2D communication over licensed bands.

Table 1. The abbreviations used most often in this paper

D2D: Device-to-Device	CL: cellular link	DL: D2D link
UE: user equipment	DT: D2D transmitter	DR: D2D receiver
MR: mobile relay	AF: amplify-and-forward	DF: decode-and-forward
ST: system throughput	EC: energy conservation	EE: energy efficiency
MS: mode selection	RA: resource allocation	PC: power control
AC: admission control	CA: channel assignment	RS: relay selection

Unfortunately, intra-cell interference is inevitable as a result of spectral sharing; therefore, mitigating severe cross-tier interference while exploring the proximate gain residing in adjacent communicating UEs is a key challenge when integrating D2D communication into cellular networks. Resource allocation (RA) has proven to be efficient for addressing the cross-tier interference issue and enhancing both system-wide and individual performance. In addition, apart from conventional RA techniques, e.g., admission control (AC), power control (PC) and channel assignment (CA), mode selection (MS) is introduced to each applying D2D link (DL), i.e., a D2D transmitter (DT) and its corresponding D2D receiver (DR), as a novel resource. Three transmission modes are available as explained below.

- *Mode 1: Cellular mode.* The DT transmits to the DR in a traditional way, with the data relayed through the BS on a dedicated channel orthogonal to all cellular links (CLs).
- *Mode 2: Overlaid D2D mode.* The DT transmits the data directly to the DR without being relayed via BS on a dedicated channel orthogonal to CLs.
- *Mode 3: Underlaid D2D mode.* The DT transmits the data directly to the DR without being relayed via BS while sharing the same channel with an existing CL.

MS design ushers in new diversity and requires careful cooperation with conventional RA techniques, and thus has motivated many studies in the recent literature. [6-11] focused on the scenario involving one CL and one DL. Specifically, [6-7] maximized system throughput (ST) of two such links with QoS protection only for the prioritized CL via joint MS and PC. Those results were extended [8-11] by introducing further fixed relays to facilitate the performance

---

[1] Because readers may find the frequent use of abbreviations in some sections and feel difficult to read, **Table 1** lists the abbreviations most commonly used in this paper.

of the CL. Considering *Mode 2* and *Mode 3* under fixed power settings, [8] found that *Mode 3* was preferable when the received strength of the desired signal at DR exceeded a threshold, but the study neglected the interference imposed from DL to CL. Compared to [8], PC was complemented in [9] for ST maximization under the same simplified interference model. The preceding drawback was resolved in [10], where only MS was considered using a fixed power setting. The QoS protection of the DL was first raised in [11], in which MS and PC were jointly designed to optimize both ST and energy conservation (EC) under a given data rate constraint, but the imposed interference from DL to CL was also ignored as in [8-9]. The investigation was expanded in [12-20] to a multiuser scenario that further explored the gain when the allocations of multidimensional resources were in cooperation. Specifically, [12-13] considered a code-division-multiple-access based system and optimized EC [12] and energy efficiency (EE) [13], respectively, by focusing on MS and PC, but as the problems were NP-hard, they derived the optimality via an exhaustive brute-force search and did not fully emphasize the feasibility check or AC. A lightly loaded system was considered in [14-15], but that study accounted for the selection only between *Mode 1* and *Mode 2*, i.e., each channel could be only assigned to at most one link. An attempt to maximize EC under individual QoS requirements by joint MS, CA and PC was made in [14], which proposed a heuristic algorithm to solve the originally NP-hard problem. In contrast, [15] modeled the joint MS and CA problem into a stationary stochastic process under time-varying channel conditions, which was optimally solved by an opportunistic sub-channel scheduling algorithm. Compared to [12-15], [16-20] dealt with all three modes. In particular, [16-17] came up with a two-step mechanism to determine the MS based on partial path-gain information in step one and to adjust the RA of each link individually using a specified priority order in step two. The joint MS, CA and PC problem for ST was decomposed in [18] into two parts: PC optimization as in [21], and the joint MS and CA, which was transformed into an integer linear programming (ILP) task. Again, this was generally NP-hard; consequently, two heuristic algorithms were devised to approximate the optimality according to system load level. The study by [19] can be viewed as an enhanced version of [16-17] and intended to maximize EC, while MS and RA of each link was performed progressively based on the superiority of channel conditions and the gap between the currently achieved data rate and the individual QoS requirement. Another study [20] also accounted for modulation and coding schemes compared to [16-19]. Their solution was made up of three components with the different variables optimized sequentially. Furthermore, the precoding, PC and CA of underlaid DLs were jointly studied in [22] from the game perspective in the uplink multiple inputs and multiple outputs (MIMO) cellular network. In contrast to the scheme in [22], the above variables were merged into one single variable, i.e., PORAV^{dm}, in [23] for a multi-cell MIMO cellular system, and the new variable maintained the group sparse structure, which could be exploited to derive a stationary solution using advanced information theory. But, the greedy optimizations of sum data rate or ST in [22-23] will cause the serious fairness problem due to the absence of QoS protection for the underlaid DLs.

Inspired by [6-23], it is beneficial to incorporate multidimensional resource variables into the optimization model, but any gain will decrease when the distance between DLs rises, as shown in [18]; thus, *Mode 1* and *Mode 2* become more preferable (as in [14-15]) even under high system load levels, which results not only in poor spectral efficiency but also in a small number of admitted links. As the number of UEs has increased exponentially during recent years [1-3], it may be wise to extend the range of D2D communication via mobile relays (MR) [24-25] to allow more admissions and spectral sharing opportunities for DLs. Consequently, the design of MR-aided DLs has inspired another line of research as a means of cooperative communication [26], and introduced two additional transmission modes.

- *Mode 4: Overlaid MR-aided mode.* The DT transmits to the DR with data relayed via another UE, i.e., an MR, on a dedicated channel orthogonal to all CLs.

- *Mode 5: Underlaid MR-aided mode.* The DT transmits to the DR with data relayed via an MR while sharing the same channel with an existing CL.

When MRs are enabled, the MS mechanisms proposed in [6-20] are no longer efficient, and new questions are raised, e.g., determining when MR-aided modes are preferable and which MR should be chosen. In addition, when *Mode 4* or *Mode 5* is selected, there are choices to be made about which channel and how much power should be assigned to the two-hop links described above under a given optimization target with additional individual QoS requirement. In [27], a distributed relay selection (RS) scheme was proposed to maximize EC together with PC and CA, but that study considered only one DL and only *Mode 3* and *Mode 5*. In [28], a RS scheme was proposed to reduce the averaged transmission time delays from a cross-layer perspective in a multiuser scenario, but the QoS protection was not fully stressed due to the absence of PC, and only two modes were involved, as in [27]. Another study [29] analyzed the outage probability of *Mode 5* in a two-way amplify-and-forward (AF) relay protocol [26] with fixed power settings. The achievable transmission capacity of DLs under a QoS requirement was minimized in [30] with respect to the power ratio between the CLs and DLs and a relay selection range based on stochastic geometry. However, only two modes were involved, just as in [27-28] with the decode-and-forward (DF) relay protocol [26]. In addition, the above performance was derived under an aloha-like multiple access protocol, which was not aligned with cellular system standardization, e.g., long term evolution (LTE) applies a channelized multiple access protocol for better interference coordination with a central controller (BS) for multiuser scheduling. The performance of two-hop D2D communications was evaluated in [31], which devised a heuristic MS along with a distributed PC to balance the optimization of ST and EC, but the design of both RS and CA was omitted.
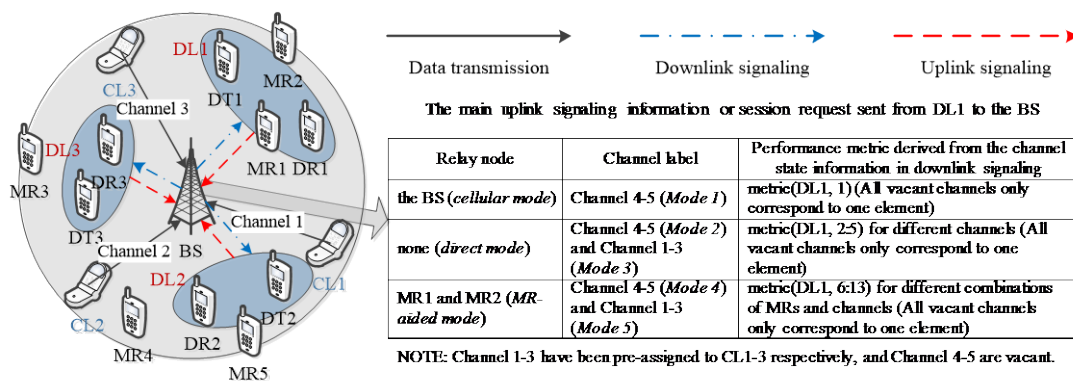

Fig. 1. An illustration of the scenario and content in this paper

Motivated by [6-23, 27-31], our goal is to present a comprehensive view for the joint optimization of MS and RA by including AC, PC, CA and RS in our hybrid cellular system when MR is further enabled to facilitate DLs (under either the AF or DF relay protocol) in case the channel condition of their desired link is unfavorable. Our contributions are threefold:

- In terms of scenario and content, "Mobile Relay-Aided Device-to-Device Communication" in the title indicates MRs *have the potential to* facilitate local pairs, but *cellular mode* and *direct mode* are also options. As shown in **Fig. 1**, in the uplink signaling

information or the session request sent from DL1 to the BS, all five modes are available compared to [6-23, 27-31], and each local pair is required to evaluate the performance metric for each mode, i.e., $metric(DL1,1{:}13)$ for DL1, which is determined by multiple factors including: (i) channel state information, e.g., the cross-tier interference between CLs and DLs and the channel condition between local pairs and their adjacent MRs, and (ii) the system settings, e.g., the relay protocol implemented at the MRs and the power budget for the DLs. The final decision about MS and RA is made at the BS based on the global requests with additional consideration for (iii) system load level, e.g., the number of vacant channels and that incurred by applying local pairs. As far as we know, we are the first to consider this problem and attempt to maximize the ST of our hybrid system by incorporating all available resource variables raised in [6-19, 21, 27-31] with additional QoS protection for all involved links missed in [6-13, 15, 22-23, 28-29, 31]. The problem holds great significance for future D2D based network architectures, e.g., for mobile cloud computing as proposed in [32], but the joint optimization of MS and RA is a mixed-integer non-linear programming problem, which is notoriously NP-hard.

- In terms of methodology and theoretical analysis, because PC variables and joint MS, CA and RS variables are separable, the original problem can be decomposed into two parts: (i) the PC part under specified MS, CA and RS conditions, and (ii) the joint optimization of MS, CA and RS. Moreover, the two sub-problems are *related* and maintain *a hierarchical structure*. Specifically, the PC part is *mode-dependent*, but its optimality can be perfectly characterized through additional AC design for any given mode, and based on the above optimality, the joint MS, CA and RS can be viewed from the graph perspective and proved to be equivalent to the *maximum weighted independent set* (MWIS) problem. Therefore, we adopt a heuristic algorithm to approximate the global optimality *in polynomial-time* and achieve a tradeoff between performance and complexity. In addition, the framework of our mechanism is also applicable to other targets by either (i) switching the weights of the graph model, e.g., to maximize the number of admitted DLs, or (ii) reformulating the PC model, e.g., to optimize EC or EE.

- In terms of performance comparison and evaluation, we compare our mechanism with (i) *forced cellular mode*, (ii) *forced direct mode*, (iii) *forced MR-aided mode*, and (iv) the case when MR is disabled, i.e., the only options are *cellular and direct mode*, as well as (v) the upper bound benchmark derived by solving the original problem through a professional optimization toolbox. We observe the variations of the performance metrics, i.e., ST and the number of admitted DLs, through changes in the system load level, the number of MRs, the maximum distance for DLs and the maximum transmit power of DLs, respectively, which help verify the efficacy of our mechanism and show the gain when the MRs are enabled and the optimizations of multidimensional variables are in cooperation.

The remainder of this paper is organized as follows: In Section 2, we briefly introduce the system model and performance evaluation in all five available transmission modes, followed by the formulation of our joint MS and RA problem. In Section 3, we decompose the original problem into two parts and address the optimality of each part in sequence. Section 4 discusses the extensive simulations we conducted to demonstrate the efficacy of our joint optimization mechanism. Finally, we draw the conclusions in Section 5.

2. Model Framework

2.1 System Model

We consider an isolated cell by assuming that the inter-cell interference is well-controlled via cell planning [33] and that only uplink channels are available to be reused by DLs. We denote the number of channels as N, where the first M ($M \leq N$) have been pre-assigned to CLs using scheduling that is out of the scope of this paper. Therefore, the system load level is defined as $\eta = M / N$, and there are $L = N - M$ remaining vacant channels. There are J DLs trying to get admitted into the system using one of the five available transmission modes as explained in Section 1. Moreover, there are K vacant MRs with the potential to provide relay service for DLs. Compared to Layer 3 fixed relays [34], MRs are both energy- and function-limited, so it is reasonable to assume that each MR can provide the relay service for at most one flow at a time, and channel reassignment is forbidden, i.e., the same channel must be used during the two-hop transmission. In addition, to avoid the negative impact of dense spectral sharing and various signaling measurements among multiple UEs [35-36], we add the restrictions that one channel can be reused by at most one DL and that one DL can reuse at most one channel. Finally, the entire set of channels and vacant channels can be defined as $H = \{1,...,N\}$ and V, respectively, so if $M < N$, $V = \{M+1,...,N\}$; otherwise, $V = \varnothing$. We also assume that the label of a CL is the same as that of the channel it is assigned to, so the set of CLs is $C = \{1,...,M\}$, while the set of DLs and MRs are $D = \{1,...,J\}$ and $R = \{1,...,K\}$, respectively. Finally, **Fig. 1** shows an illustration of our system model where three local pairs (DL1-3) are requesting for admission, and five MRs (MR1-5) are available to provide two-hop transmission for adjacent DLs, and three CLs (CL1-3) are already in service and pre-assigned to channels 1-3, respectively, and channels 4-5 are unused. The major symbols used in this paper are listed in **Table 2**.

Table 2. The major symbols used in this paper

Symbols	Definitions
$H = \{1,...,N\}$	set of entire channels, and N is the cardinality of H
$C = \{1,...,M\}$	set of CLs, and M, $M \leq N$ is the cardinality of C
V	set of vacant channels, and if $M < N$, $V = \{M+1,...,N\}$ and its cardinality is $L = N - M > 0$; otherwise, $V = \varnothing$ and its cardinality is 0
$D = \{1,...,J\}$	set of DLs, and J is the cardinality of D
$R = \{1,...,K\}$	set of MRs, and K is the cardinality of R
$A = \{1,...,L+J+K\}$	set of relay nodes, and $L + J + K$ is the cardinality of A
η	system load level of existing cellular network, and $0 \leq \eta = M / N \leq 1$
p_j^D, p_m^C, p_k^R	transmit power at the j-th DL, the m-th CL and the k-th MR, respectively
$p_{\max}^D, p_{\max}^C, p_{\max}^R$	maximum transmit power of DLs (in non-*cellular mode*), CLs (and DLs when they are in *cellular mode*) and MRs, respectively
$g_{i,j}^{XY}$	aggregate path gain from the transmitter of the j-th link in Y to the receiver of the i-th link in X, and in particular, if $j = B$ or $i = B$, it stands for the transmitter or the receiver at the BS
σ^2	addictive white Gaussian noise (AWGN) at each receiver
θ_j^D, θ_m^C	minimum SINR requirements of the j-th DL and the m-th CL, respectively

$x_{j,a,n}$	joint MS, CA and RS binary variables, and if the j-th DL communicates on the n-th channel via the a-th relay node, then $x_{j,a,n}=1$; otherwise, $x_{j,a,n}=0$.
$\gamma_{j,l,1}^{D\ (1)},\gamma_{j,l,2}^{D\ (1)}$	SINRs of the first and second hop of the j-th DL, respectively, when it is in *Mode 1* (*cellular mode*)
$R_{j,l}^{D\ (1)}$	data rate of the j-th DL when it is in *Mode 1* (*cellular mode*)
$\gamma_{j,l}^{D\ (2)}$	SINR of the j-th DL when it is in *Mode 2* (*overlaid direct mode*)
$R_{j,l}^{D\ (2)}$	data rate of the j-th DL when it is in *Mode 2* (*overlaid direct mode*)
$\gamma_{j,m}^{D\ (3)},\gamma_{m}^{C(3)}$	SINRs of the j-th DL and the m-th CL, respectively, when the j-th DL is in *mode 3* (*underlaid direct mode*) and reuses the same channel as the m-th CL
$R_{j,m}^{D\ (3)}$	sum data rate of the j-th DL and the m-th CL when the j-th DL is in *mode 3* (*underlaid direct mode*) and reuses the same channel as the m-th CL
$\gamma_{j,k,l}^{D\ (4)},\gamma_{k,j,l}^{D\ (4)}$	SINRs of the first and second hop of the j-th DL, respectively, when it is in *Mode 4* (*overlaid MR-aided mode*) relayed via the k-th MR on the l-th vacant channel
$R_{j,k,l}^{D\ (4-1)},R_{j,k,l}^{D\ (4-2)}$	data rate of the j-th DL when it is in *Mode 4* (*overlaid MR-aided mode*) relayed via the k-th MR on the l-th vacant channel under the AF and DF, respectively
$\gamma_{j,k,m}^{D\ (5F)},\gamma_{m}^{C(5F)}$	SINRs of the first hop of the j-th DL and the m-th CL, respectively, when the j-th DL is in *Mode 5* (*underlaid MR-aided mode*) relayed via the k-th MR on the m-th channel and the m-th CL occurs in the first hop of the j-th DL's transmission
$\gamma_{k,j,m}^{D\ (5F)}$	SINRs of the second hop of the j-th DL when it is in *Mode 5* (*underlaid MR-aided mode*) relayed via the k-th MR on the m-th channel and the m-th CL occurs in the first hop of the j-th DL's transmission
$R_{j,k,m}^{D\ (5-1)},R_{j,k,m}^{D\ (5-2)}$	sum data rate of the j-th DL and the m-th CL when the j-th DL is in *Mode 5* (*underlaid MR-aided mode*) relayed via the k-th MR on the m-th channel under the AF and DF, respectively, and the m-th CL occurs in the first hop of the j-th DL's transmission
$\gamma_{j,k,m}^{D\ (5S)}$	SINRs of the first hop of the j-th DL when it is in *Mode 5* (*underlaid MR-aided mode*) relayed via the k-th MR on the m-th channel and the m-th CL occurs in the second hop of the j-th DL's transmission
$\gamma_{k,j,m}^{D\ (5S)},\gamma_{m}^{C(5S)}$	SINRs of the second hop of the j-th DL and the m-th CL, respectively, when the j-th DL is in *Mode 5* (*underlaid MR-aided mode*) relayed via the k-th MR on the m-th channel and the m-th CL occurs in the second hop of the j-th DL's transmission
$R_{j,k,m}^{D\ (5-3)},R_{j,k,m}^{D\ (5-4)}$	sum data rate of the j-th DL and the m-th CL when the j-th DL is in *Mode 5* (*underlaid MR-aided mode*) relayed via the k-th MR on the m-th channel under the AF and DF, respectively, and the m-th CL occurs in the second hop of the j-th DL's transmission

We assume all the involved links experience independent block fading, and the overall path gain from the transmitter of the m-th CL to the BS is defined as $g_{B,m}^{CC}=G_{B,m}^{CC}\alpha_{B,m}^{CC}$, where $G_{B,m}^{CC}$ is the large-scale distance-dependent path gain, and includes antenna gains at both transmitter and receiver, and $\alpha_{B,m}^{CC}$ is the small-scale channel fading component. In a like manner, the overall path gain from the transmitter of the m-th CL to the j-th DR and to the k-th MR can be defined as $g_{j,m}^{DC}$ and $g_{k,m}^{RC}$, respectively. Moreover, the path gain from the j-th DT to its desired DR, to the BS and to the k-th MR, and those from the k-th MR to the j-th DR and to the BS, as well as that from the BS to the j-th DR, are denoted as $g_{j,j}^{DD},g_{B,j}^{CD},g_{k,j}^{RD},g_{j,k}^{DR},g_{B,k}^{CR},g_{j,B}^{DC}$, respectively.

We assume that the BS is able to obtain full channel state information, which can be done by decoding the sounding reference signal in LTE based cellular networks. Consequently, we are dedicated to designing centralized joint MS and RA mechanisms for better performance and to setting an upper bound benchmark for distributed ones.

2.2 Performance Evaluation of All Five Available Transmission Modes

In this section, we evaluate the contribution of a DL to ST in any transmission mode. To avoid loss of generality, we consider only the j-th DL ($j \in D$). A transmission frame is divided into two slots. If *Mode 2* and *Mode 3* are chosen, the direct link occupies both slots; otherwise, the DT transmits to the relay node (the BS or MR) in the first slot, while the corresponding relay node transmits to the DR in the second slot.

A. Mode 1: Cellular mode. The j-th DL transmits to the BS on the l-th vacant uplink channel ($l \in V$) in the first slot, and the BS decodes and forwards the data to the corresponding DR on a downlink channel in the second slot. We assume that when the j-th DL is in *cellular mode*, its channel will not be reused by other DLs. Therefore, the signal-to-interference-plus-noise ratio (SINR) of the first hop is then reduced to the signal-to-noise ratio (SNR), denoted as $\gamma_{j,l,1}^{D\,(1)}$ and defined in (1-a) below,

$$\gamma_{j,l,1}^{D\,(1)} = p_j^D g_{B,j}^{CD} / \sigma^2, \forall l \in V \quad (a); \quad \gamma_{j,l,2}^{D\,(1)} = p_j^B g_{j,B}^{DC} / \sigma^2, \forall l \in V \quad (b) \tag{1}$$

where p_j^D is the transmit power of the j-th DT, and σ^2 is the addictive white Gaussian noise (AWGN) at each receiver. Similarly, the SINR of the second hop can also be reduced to the SNR as $\gamma_{j,l,2}^{D\,(1)}$, defined in (1-b), due to the orthogonality in downlink cellular scheduling, where p_j^B is the transmit power of the BS. To guarantee a successful transmission, a SINR threshold, i.e., θ_j^D , should be satisfied for each hop and achieved by analyzing the QoS class identifier in the user bearer. Finally, if $\gamma_{j,l,1}^{D\,(1)} \geq \theta_j^D, \gamma_{j,l,2}^{D\,(1)} \geq \theta_j^D$, then the data rate of the j-th DL under *Mode 1* can be denoted as $R_{j,l}^{D,(1)}$ in (2); otherwise, $R_{j,l}^{D,(1)} = -\infty$, indicating infeasibility.

$$R_{j,l}^{D,(1)} = \min\{\log_2(1 + \gamma_{j,l,1}^{D\,(1)}), \log_2(1 + \gamma_{j,l,2}^{D\,(1)})\} / 2, \forall l \in V \tag{2}$$

In general, the BS is equipped with a much larger power budget than UEs, i.e., 46 dBm at the BS compared to 23 dBm at a UE in the LTE system. Thus we assume the downlink rate can always satisfy $\log_2(1 + \gamma_{j,l,2}^{D\,(1)}) \geq \log_2(1 + \gamma_{j,l,1}^{D\,(1)})$, and $R_{j,l}^{D,(1)}$ can be simplified into (3) as in [12-20].

$$R_{j,l}^{D,(1)} = \log_2(1 + \gamma_{j,l,1}^{D\,(1)}) / 2, \forall l \in V \tag{3}$$

B. Mode 2: Overlaid direct mode. The j-th DT transmits to its corresponding DR using both slots on the l-th vacant channel ($l \in V$). The SNR of a single-hop transmission can be denoted as $\gamma_{j,l}^{D\,(2)}$ in (4). If $\gamma_{j,l}^{D\,(2)} \geq \theta_j^D$, then the data rate under *Mode 2* can be defined in (5) as $R_{j,l}^{D,(2)}$; otherwise, $R_{j,l}^{D\,(2)} = -\infty$, indicating infeasibility.

$$\gamma_{j,l}^{D\,(2)} = p_j^D g_{j,j}^{DD} / \sigma^2, \forall l \in V \tag{4}$$

$$R_{j,l}^{D\,(2)} = \log_2(1 + \gamma_{j,l}^{D\,(2)}), \forall l \in V \tag{5}$$

C. Mode 3: Underlaid direct mode. The j-th DT transmits to its corresponding DR using both slots by reusing the same channel with the m-th CL ($m \in C$). In this way, mutual interference is imposed between the two links, and their SINRs can be defined in (6) as $\gamma_{j,m}^{D\,(3)}$ and $\gamma_m^{C\,(3)}$, respectively, where p_m^c is the transmit power of the m-th CL. Apart from θ_j^D , θ_m^c is introduced

to guarantee the QoS requirement for the prioritized CL, and if $\gamma_{j,m}^{D}{}^{(3)} \geq \theta_j^D, \gamma_m^{C(3)} \geq \theta_m^C$, the data
rate under *Mode 3* is defined in (7) as $R_{j,m}^{D}{}^{(3)}$; otherwise, $R_{j,m}^{D}{}^{(3)} = -\infty$, indicating infeasibility.

$$\gamma_{j,m}^{D}{}^{(3)} = p_j^D g_{j,j}^{DD} / (p_m^C g_{j,m}^{DC} + \sigma^2), \gamma_m^{C(3)} = p_m^C g_{B,m}^{CC} / (p_j^D g_{B,j}^{CD} + \sigma^2), \forall m \in C \tag{6}$$

$$R_{j,m}^{D}{}^{(3)} = \log_2(1 + \gamma_{j,m}^{D}{}^{(3)}) + \log_2(1 + \gamma_m^{C(3)}), \forall m \in C \tag{7}$$

D. Mode 4: Overlaid MR-aided mode. The *j*-th DT transmits to the *k*-th MR ($k \in R$) in the first
slot, while the *k*-th MR relays the data to the *j*-th DR in the second hop, and both hops occupy
the same *l*-th vacant channel ($l \in V$). Thus the SNR of each hop can be defined in (8) as $\gamma_{j,k,l}^{D}{}^{(4)}$
and $\gamma_{k,j,l}^{D}{}^{(4)}$, respectively,

$$\gamma_{j,k,l}^{D}{}^{(4)} = p_j^D g_{k,j}^{RD} / \sigma^2, \forall k \in R, l \in V \quad (a); \quad \gamma_{k,j,l}^{D}{}^{(4)} = p_k^R g_{j,k}^{DR} / \sigma^2, \forall k \in R, l \in V \quad (b) \tag{8}$$

where p_k^R is the transmit power of the *k*-th MR in the second hop. Then, the data rate of the
two-hop transmission can be classified into one of the following two cases based on the relay
protocol implemented at the MRs:

(i) *Case 1: under the AF relay protocol.* If $\gamma_{j,k,l}^{D}{}^{(4)} \geq \theta_j^D, \gamma_{j,k,l}^{D}{}^{(4)}\gamma_{k,j,l}^{D}{}^{(4)} / (\gamma_{j,k,l}^{D}{}^{(4)} + \gamma_{k,j,l}^{D}{}^{(4)} + 1) \geq \theta_j^D$,
the data rate under *Mode 4* can be defined in (9) as $R_{j,k,l}^{D}{}^{(4-1)}$ [26]; otherwise, $R_{j,k,l}^{D}{}^{(4-1)} = -\infty$,
indicating infeasibility.

$$R_{j,k,l}^{D}{}^{(4-1)} = \log_2(1 + \gamma_{j,k,l}^{D}{}^{(4)}\gamma_{k,j,l}^{D}{}^{(4)} / (\gamma_{j,k,l}^{D}{}^{(4)} + \gamma_{k,j,l}^{D}{}^{(4)} + 1)) / 2, \forall k \in R, l \in V \tag{9}$$

(ii) *Case 2: under* the *DF relay protocol.* If $\gamma_{j,k,l}^{D}{}^{(4)} \geq \theta_j^D, \gamma_{k,j,l}^{D}{}^{(4)} \geq \theta_j^D$, the data rate under *Mode
4* can be defined in (10) as $R_{j,k,l}^{D}{}^{(4-2)}$ [26]; otherwise, $R_{j,k,l}^{D}{}^{(4-2)} = -\infty$, indicating infeasibility.

$$R_{j,k,l}^{D}{}^{(4-2)} = \min\{\log_2(1 + \gamma_{j,k,l}^{D}{}^{(4)}), \log_2(1 + \gamma_{k,j,l}^{D}{}^{(4)})\} / 2, \forall k \in R, l \in V \tag{10}$$

E. Mode 5: Underlaid MR-aided mode. The *j*-th DT transmits to the *k*-th MR ($k \in R$) in the first
slot, while the *k*-th MR relays the data to the *j*-th DR in the second hop, and both hops occupy
the same channel with the *m*-th CL ($m \in C$). To create more opportunity for successful
transmission, we assume that the CL occurs in only one of the two slots, thus the performance
of this mode could be divided into one of the following four cases depending on the hop
occupied by the *m*-th CL and the relay protocol implemented at the MRs.

When the *m*-th CL occupies the first hop, the SINR of the first hop in the *j*-th DL and the
m-th CL can be defined in (11) as $\gamma_{j,k,m}^{D}{}^{(5F)}$ and $\gamma_m^{C(5F)}$, respectively, while the SNR of the second
hop in the *j*-th DL is similar to that in (8-b) and defined in (12) as $\gamma_{k,j,m}^{D}{}^{(5F)}$,

$$\gamma_{j,k,m}^{D}{}^{(5F)} = p_j^D g_{k,j}^{RD} / (p_m^C g_{k,m}^{RC} + \sigma^2), \gamma_m^{C(5F)} = p_m^C g_{B,m}^{CC} / (p_j^D g_{B,j}^{CD} + \sigma^2), \forall k \in R, m \in C \tag{11}$$

$$\gamma_{k,j,m}^{D}{}^{(5F)} = p_k^R g_{j,k}^{DR} / \sigma^2, \forall k \in R, m \in C \tag{12}$$

(i) *Case 1: under AF relay protocol.* If $\gamma_{j,k,m}^{D}{}^{(5F)} \geq \theta_j^D, \gamma_{j,k,m}^{D}{}^{(5F)}\gamma_{k,j,m}^{D}{}^{(5F)} / (\gamma_{j,k,m}^{D}{}^{(5F)} + \gamma_{k,j,m}^{D}{}^{(5F)} + 1) \geq \theta_j^D$
and $\gamma_m^{C(5F)} \geq \theta_m^C$, the data rate under *Mode 5* is defined in (13) as $R_{j,k,m}^{D}{}^{(5-1)}$;
otherwise, $R_{j,k,m}^{D}{}^{(5-1)} = -\infty$, indicating infeasibility.

$$R_{j,k,m}^{D}{}^{(5-1)} = [\log_2(1 + \gamma_{j,k,m}^{D}{}^{(5F)}\gamma_{k,j,m}^{D}{}^{(5F)} / (\gamma_{j,k,m}^{D}{}^{(5F)} + \gamma_{k,j,m}^{D}{}^{(5F)} + 1)) + \log_2(1 + \gamma_m^{C(5F)})] / 2, \forall k \in R, m \in C \tag{13}$$

(ii) *Case 2: under the DF relay protocol.* If $\gamma_{j,k,m}^{D}{}^{(5F)} \geq \theta_j^D, \gamma_{k,j,m}^{D}{}^{(5F)} \geq \theta_j^D, \gamma_m^{C(5F)} \geq \theta_m^C$, the data rate
under *Mode 5* is defined in (14) as $R_{j,k,m}^{D}{}^{(5-2)}$; otherwise, $R_{j,k,m}^{D}{}^{(5-2)} = -\infty$, indicating infeasibility.

$$R_{j,k,m}^{D}{}^{(5-2)} = [\min\{\log_2(1 + \gamma_{j,k,m}^{D}{}^{(5F)}), \log_2(1 + \gamma_{k,j,m}^{D}{}^{(5F)})\} + \log_2(1 + \gamma_m^{C(5F)})] / 2, \forall k \in R, m \in C \tag{14}$$

When the m-th CL occupies the second hop, the SINR of the second hop in the j-th DL and the m-th CL can be defined in (15) as $\gamma_{k,j,m}^{D}{}^{(5S)}$ and $\gamma_{m}^{C(5S)}$, respectively, while the SNR of the first hop in the j-th DL is similar to that in (8-a) and defined in (16) as $\gamma_{j,k,m}^{D}{}^{(5S)}$,

$$\gamma_{k,j,m}^{D}{}^{(5S)} = p_k^R g_{j,k}^{DR} / (p_m^C g_{j,m}^{DC} + \sigma^2), \gamma_m^{C(5S)} = p_m^C g_{B,m}^{CC} / (p_k^R g_{B,k}^{CR} + \sigma^2), \forall k \in R, m \in C \qquad (15)$$

$$\gamma_{j,k,m}^{D}{}^{(5S)} = p_j^D g_{k,j}^{RD} / \sigma^2, \forall k \in R, m \in C \qquad (16)$$

(iii)   *Case   3:   under   the   AF   relay   protocol.*   If $\gamma_{j,k,m}^{D}{}^{(5S)} \geq \theta_j^D, \gamma_{j,k,m}^{D}{}^{(5S)} \gamma_{k,j,m}^{D}{}^{(5S)} / (\gamma_{j,k,m}^{D}{}^{(5S)} + \gamma_{k,j,m}^{D}{}^{(5S)} + 1) \geq \theta_j^D$ and $\gamma_m^{C(5S)} \geq \theta_m^C$, the data rate under *Mode 5* is defined in (17) as $R_{j,k,m}^{D}{}^{(5-3)}$; otherwise, $R_{j,k,m}^{D}{}^{(5-3)} = -\infty$, indicating infeasibility.

$$R_{j,k,m}^{D}{}^{(5-3)} = [\log_2(1 + \gamma_{j,k,m}^{D}{}^{(5S)} \gamma_{k,j,m}^{D}{}^{(5S)} / (\gamma_{j,k,m}^{D}{}^{(5S)} + \gamma_{k,j,m}^{D}{}^{(5S)} + 1)) + \log_2(1 + \gamma_m^{C(5S)})] / 2, \forall k \in R, m \in C \quad (17)$$

(iv) *Case 4: under DF relay protocol.* If $\gamma_{j,k,m}^{D}{}^{(5S)} \geq \theta_j^D, \gamma_{k,j,m}^{D}{}^{(5S)} \geq \theta_j^D, \gamma_m^{C(5S)} \geq \theta_m^C$, the data rate under *Mode 5* is defined in (18) as $R_{j,k,m}^{D}{}^{(5-4)}$; otherwise, $R_{j,k,m}^{D}{}^{(5-4)} = -\infty$, indicating infeasibility.

$$R_{j,k,m}^{D}{}^{(5-4)} = [\min\{\log_2(1 + \gamma_{j,k,m}^{D}{}^{(5S)}), \log_2(1 + \gamma_{k,j,m}^{D}{}^{(5S)})\} + \log_2(1 + \gamma_m^{C(5S)})] / 2, \forall k \in R, m \in C \qquad (18)$$

2.3 Problem Formulation

To give a simplified form of the original problem, we first introduce binary variables $x_{j,a,n}$ with respect to the triple $(j,a,n) \in D \times A \times H$, where $A = \{1,...,L, L+1,...., L+J, L+J+1,..., L+J+K\}$ is used to denote the set of relay nodes, and two direct modes, i.e., *Mode 2* and *Mode 3*, can be viewed as two special cases when DR coincides with the relay node. We assume that if $a \in A_1 = \{1,...,L\}$, it represents the case when BS is the relay node, i.e., in *Mode 1*, and if $a \in A_2 = \{L+1,...,L+J\}$, it represents the case when the $(a-L)$-th DR is the relay node, i.e., in direct modes (*Mode 2* and *Mode 3*); otherwise, i.e., if $a \in A_3 = \{L+J+1,...,L+J+K\}$, it represents the case when the $(a-L-J)$-th MR is the relay node, i.e., in the MR-aided relay modes (*Mode 4* and *Mode 5*). Finally, if the j-th DL communicates on the n-th channel via the a-th relay node, then $x_{j,a,n} = 1$; otherwise, $x_{j,a,n} = 0$. Moreover, $R_{j,a,n}$ is introduced to denote the data rate related to the triple $(j,a,n) \in D \times A \times H$. In this way, the joint MS and RA for ST maximization can be formulated as *Problem 1 (P1)* in (19),

$$\underline{Problem\ 1:}\text{maximize}_{x_{j,a,n}, p_j^D, p_n^C, p_k^R} \sum_{j \in D} \sum_{a \in A} \sum_{n \in H} x_{j,a,n} R_{j,a,n}(p_j^D, p_n^C, p_k^R) \qquad (a)$$

$$\text{subject to}: x_{j,a,n} = \{0,1\} \qquad (b); \quad \sum_{a \in A} \sum_{n \in H} x_{j,a,n} \leq 1, \forall j \in D \qquad (c)$$

$$\sum_{j \in D} \sum_{n \in H} x_{j,a,n} \leq 1, \forall a \in A \qquad (d); \quad \sum_{a \in A} \sum_{j \in D} x_{j,a,n} \leq 1, \forall n \in H \qquad (e) \qquad (19)$$

$$0 \leq p_j^D \leq p_{\max}^C, \forall a \in A_1, 0 \leq p_j^D \leq p_{\max}^D, \forall a \in A - A_1, 0 \leq p_n^C \leq p_{\max}^C, 0 \leq p_k^R \leq p_{\max}^R \qquad (f)$$

where (19-a) conveys that we aim to maximize ST with respect to $x_{j,a,n}$ and p_j^D, p_n^C, p_k^R, and the designs of MS, CA and RS have been included in $x_{j,a,n}$. It's noteworthy that the individual QoS requirement has been implied in the value of $R_{j,a,n}$, and specifically, if feasibility is violated, $R_{j,a,n} = -\infty$ as described in Section 2.2. (19-b) shows the binary feature of $x_{j,a,n}$, and the constraints from (19-c) to (19-e) limit a DL to occupying at most one flow at a time, each relay node can serve at most one flow at a time, and dense spectral sharing is forbidden, respectively. Finally, (19-f) represents the local power budgets at the DTs, CLs and MRs. Apparently, *P1* requires mixed-integer non-linear programming, which is in general NP-hard, thus efficient mechanisms are preferable for achieving a tradeoff between performance and complexity.

3. Joint MS and RA Mechanism

Because the constraints in (19) are separable between $x_{j,a,n}$ and p_j^D, p_n^C, p_k^R, we can switch the order of the two maximizations in (19-a) and derive the compound optimality in sequence, i.e., $\underset{x_{j,a,n},p_j^D,p_n^C,p_k^R}{maximize} \sum_{j\in D} \sum_{a\in A} \sum_{n\in H} x_{j,a,n} R_{j,a,n} = \underset{x_{j,a,n}}{maximize} \sum_{j\in D} \sum_{a\in A} \sum_{n\in H} x_{j,a,n} \underset{p_j^D,p_n^C,p_k^R}{maximize} R_{j,a,n}$. Therefore, *P1* can be decomposed into two sub-problems: The one only pertinent to p_j^D, p_n^C, p_k^R as *Problem 2 (P2)* in (20), and the one only pertinent to $x_{j,a,n}$ as *Problem 3 (P3)* in (21). Specifically, *P2* is intended to achieve the optimal $R_{j,a,n}$ for any given triple $(j,a,n) \in D \times A \times H$, which is then used as a weight value to variable $x_{j,a,n}$ in *P3*, i.e., $R_{j,a,n}^*$ in (21-a). Thus, the above two sub-problems maintain *a hierarchical structure*, and *P3* optimizes $x_{j,a,n}$ based upon the optimality of *P2*. In what follows, we are going to elaborate on the design for each sub-problem sequentially as follows: (i) investigating how to derive all feasible $R_{j,a,n}^*$ by PC with AC, and (ii) using all $R_{j,a,n}^*$, investigating how to effectively solve the constrained ILP for the joint MS, CA and RS.

$$\underline{Problem\ 2}: maximize_{p_j^D, p_n^C, p_k^R} R_{j,a,n}(p_j^D, p_n^C, p_k^R) \qquad\qquad (a)$$

$$subject\ to : 0 \le p_j^D \le p_{max}^C, \forall a \in A_1, 0 \le p_j^D \le p_{max}^D, \forall a \in A - A_1, 0 \le p_n^C \le p_{max}^C, 0 \le p_k^R \le p_{max}^R \quad (b) \qquad (20)$$

$$\underline{Problem\ 3}: maximize_{x_{j,a,n}} \sum_{j\in D} \sum_{a\in A} \sum_{n\in H} x_{j,a,n} R_{j,a,n}^* \qquad\qquad (a)$$

$$subject\ to : x_{j,a,n} = \{0,1\} \quad (b); \quad \sum_{a\in A}\sum_{n\in H} x_{j,a,n} \le 1, \forall j \in D \quad (c) \qquad (21)$$

$$\sum_{j\in D}\sum_{n\in H} x_{j,a,n} \le 1, \forall a \in A \quad (d); \quad \sum_{a\in A}\sum_{j\in D} x_{j,a,n} \le 1, \forall n \in H \quad (e)$$

3.1 The Optimization of PC with Additional AC

Under any given triple $(j,a,n) \in D \times A \times H$, the optimality of *P2* is *mode-dependent* and highly coupled with both the spectral sharing condition and the selected relay node as well as with its adopted relay protocol. So, we address the optimality of PC and AC separately for each mode.

3.1.1 The Modes on Dedicated Channels (*Mode 1*, *Mode 2* and *Mode 4*)

Because mutual interference is canceled in this case, the optimal data rate is achieved when the DT and/or the MR emits at its maximum power on the condition that feasibility is guaranteed.

For the triple $(j,a,n) \in D \times A \times H$, if $a \in A_1 = \{1,...,L\}, n \in V$, this means that the *j*-th DL is in *Mode 1*, and $R_{j,a,n}^* = R_{j,n}^{D,(1)}(p_{max}^C)$ with (1-a) and (3) if $\gamma_{j,n,1}^{D,(1)}(p_{max}^C) = p_{max}^C g_{B,j}^{CD} / \sigma^2 \ge \theta_j^D$; $R_{j,a,n}^* = -\infty$ otherwise. Moreover, if $a = L + j \in A_2 = \{L+1,...,L+J\}, n \in V$, then the *j*-th DL is in *Mode 2*, and $R_{j,a,n}^* = R_{j,n}^{D,(2)}(p_{max}^D)$ with (4) and (5) if $\gamma_{j,n}^{D,(2)}(p_{max}^D) = p_{max}^D g_{j,j}^{DD} / \sigma^2 \ge \theta_j^D$; $R_{j,a,n}^* = -\infty$ otherwise. Finally, if $a = L + J + k \in A_3 = \{L+J+1,...,L+J+K\}, n \in V$, it means that the *j*-th DL is in *Mode 4* with data relayed via the *k*-th MR. When AF is applied, $R_{j,a,n}^* = R_{j,k,n}^{D,(4-1)}(p_{max}^D, p_{max}^R)$ with (8) and (9) if $\gamma_{j,k,n}^{D,(4)}(p_{max}^D) = p_{max}^D g_{k,j}^{RD} / \sigma^2 \ge \theta_j^D$ and $\gamma_{j,k,n}^{D,(4)}(p_{max}^D)\gamma_{k,j,n}^{D,(4)}(p_{max}^R) / (\gamma_{j,k,n}^{D,(4)}(p_{max}^D) + \gamma_{k,j,n}^{D,(4)}(p_{max}^R) + 1) \ge \theta_j^D$ where $\gamma_{k,j,n}^{D,(4)}(p_{max}^R) = p_{max}^R g_{j,k}^{DR} / \sigma^2$; $R_{j,a,n}^* = -\infty$ otherwise, and when DF is applied, $R_{j,a,n}^* = R_{j,k,n}^{D,(4-2)}(p_{max}^D, p_{max}^R)$ with (8) and (10) if $\gamma_{j,k,n}^{D,(4)}(p_{max}^D) \ge \theta_j^D$ and $\gamma_{k,j,n}^{D,(4)}(p_{max}^R) \ge \theta_j^D$; $R_{j,a,n}^* = -\infty$ otherwise.

3.1.2 The Modes on Reused Channels (*Mode 3* and *Mode 5*)

A. Mode 3: For the triple $(j,a,n) \in D \times A \times H$, if $a = L + j \in A_2, n \in C$, then the *j*-th DL chooses underlaid direct mode, and *P2* is specified into *Problem 4 (P4)*. We first check its feasibility,

i.e., whether the constraint set is non-empty or the SINR requirement in both (22-d) and (22-e) can be achieved within (22-b) and (22-c), which can be summarized in **Proposition 1**. Then, we proceed to derive its optimality upon above feasibility as concluded in **Proposition 2**.

$$\underline{Problem\ 4}: maximize_{p_n^C, p_j^D} R_{j,n}^{D\ (3)}(p_n^C, p_j^D) = \log_2(1 + \gamma_{j,n}^{D\ (3)}) + \log_2(1 + \gamma_n^{C(3)}) \tag{a}$$

$$subject\ to: 0 \le p_j^D \le p_{\max}^D \quad (b); \quad 0 \le p_n^C \le p_{\max}^C \quad (c) \tag{22}$$

$$\gamma_{j,n}^{D\ (3)} = p_j^D g_{j,j}^{DD} / (p_n^C g_{j,n}^{DC} + \sigma^2) \ge \theta_j^D \quad (d); \quad \gamma_n^{C(3)} = p_n^C g_{B,n}^{CC} / (p_j^D g_{B,j}^{CD} + \sigma^2) \ge \theta_n^C \quad (e)$$

Proposition 1. *P4* is feasible if $b_1, b_2 > 0, b_3 / b_2 \le \min\{b_1, p_{\max}^D\}$, where $b_1 = (p_{\max}^C g_{B,n}^{CC} / \theta_n^C - \sigma^2) / g_{B,j}^{CD}$, and $b_2 = 1 - \theta_j^D \theta_n^C g_{j,n}^{DC} g_{B,n}^{CD} / (g_{j,j}^{DD} g_{B,n}^{CC}), b_3 = \theta_j^D \theta_n^C g_{j,n}^{DC} \sigma^2 / (g_{B,n}^{CC} g_{j,j}^{DD}) + \theta_j^D \sigma^2 / g_{j,j}^{DD}$.

Proof. As one's transmit power is proportional to its SINR, we can derive the Pareto-optimal power for EC by forcing (22-d) and (22-e) to take equality as in (23). From (23-b), we have $p_n^C = \theta_n^C (p_j^D g_{B,j}^{CD} + \sigma^2) / g_{B,n}^{CC}$, and to make it achievable within $0 \le p_n^C \le p_{\max}^C$ for any given $p_j^D \ge 0$, $b_1 > 0$ and $p_j^D \le b_1$. By substituting p_n^c with $\theta_n^C (p_j^D g_{B,j}^{CD} + \sigma^2) / g_{B,n}^{CC}$ in (23-a), we have $b_2 p_j^D = b_3$ and to satisfy $0 \le p_j^D \le \min\{b_1, p_{\max}^D\}$, $b_2 > 0, p_j^D = b_3 / b_2 \le \min\{b_1, p_{\max}^D\}$, thus requires $b_3 / b_2 \le \min\{b_1, p_{\max}^D\}$. ∎

$$\gamma_{j,n}^{D\ (3)} = p_j^D g_{j,j}^{DD} / (p_n^C g_{j,n}^{DC} + \sigma^2) = \theta_j^D \quad (a); \qquad \gamma_n^{C(3)} = p_n^C g_{B,n}^{CC} / (p_j^D g_{B,j}^{CD} + \sigma^2) = \theta_n^C \quad (b) \tag{23}$$

Proposition 2. When feasibility is guaranteed as claimed in **Proposition 1**, the optimal power pair, i.e., (p_n^{C*}, p_j^{D*}), to *P4* lies within P, X, Q, Y, I, which is both feasible and reaps the largest $R_{j,n}^{D\ (3)}$ in (22-a), i.e., $(p_n^{C*}, p_j^{D*}) = argmax_{\{P, X, Q, Y, I\}} R_{j,n}^{D\ (3)}(p_n^C, p_j^D)$ with $P = (\theta_n^C (p_{\max}^D g_{B,j}^{CD} + \sigma^2) / g_{B,n}^{CC}, p_{\max}^D)$, $X = (p_{\max}^C, (p_{\max}^C g_{B,n}^{CC} / \theta_n^C - \sigma^2) / g_{B,j}^{CD})$, $Q = ((p_{\max}^D g_{j,j}^{DD} / \theta_j^D - \sigma^2) / g_{j,n}^{DC}, p_{\max}^D)$, $Y = (p_{\max}^C, \theta_j^D (p_{\max}^C g_{j,n}^{DC} + \sigma^2) / g_{j,j}^{DD})$ and $I = (p_{\max}^C, p_{\max}^D)$.

Proof. Based on [37], for any feasible power pair (p_n^C, p_j^D), $R_{j,n}^{D\ (3)}(\alpha p_n^C, \alpha p_j^D)$ is strictly increasing in $\alpha > 1$ when $(\alpha p_n^C, \alpha p_j^D)$ is still feasible with the help of (24), thus the optimal $R_{j,n}^{D\ (3)}(p_n^C, p_j^D)$ is obtained when at least one of the upper bound constraints in (22-b) and (22-c) is activated. Therefore, in what follows, we proceed to analyze the properties of $R_{j,n}^{D\ (3)}(p_{\max}^C, p_j^D)$ and $R_{j,n}^{D\ (3)}(p_n^C, p_{\max}^D)$. By the monotonicity of $\log_2(\cdot)$ function, $\max_{p_j^D} R_{j,n}^{D\ (3)}(p_{\max}^C, p_j^D)$ is reduced to $\max_{p_j^D} \psi \triangleq (1 + \gamma_{j,n}^{D\ (3)})(1 + \gamma_n^{C(3)})|_{(p_{\max}^C, p_j^D)}$, so we derive the first- and second-order derivatives of ψ in (25), if $g_{j,j}^{DD} \sigma^2 - g_{B,j}^{CD}(p_{\max}^C g_{j,n}^{DC} + \sigma^2) \ge 0$, $\partial \psi / \partial p_j^D > 0$ and ψ is strictly increasing in p_j^D; otherwise, $\partial^2 \psi / \partial (p_j^D)^2 > 0$, and ψ is thus convex in p_j^D. Due to the symmetry, $(1 + \gamma_{j,n}^{D\ (3)})(1 + \gamma_n^{C(3)})|_{(p_n^C, p_{\max}^D)}$ is also either strictly increasing or convex in p_n^C. In conclusion, (p_n^{C*}, p_j^{D*}) can lie only at corner points of the feasible zone when $p_n^C = p_{\max}^C$ and $p_j^D = p_{\max}^D$, and the structure of the optimality can be categorized into the three cases illustrated in **Fig. 2**, where the shadow region is the feasible zone constructed by (23) and $p_n^C = p_{\max}^C, p_j^D = p_{\max}^D$. Specifically, when $(p_j^D)_Y \ge p_{\max}^D$ as shown in **Fig. 2(a)**, where $(p_j^D)_Y$ is the p_j^D at point Y, the CL tends to impose a strong interference to the DL, i.e., a small $g_{B,j}^{CD} / g_{B,n}^{CC}$ and a large $g_{j,n}^{DC} / g_{j,j}^{DD}$, and (p_n^{C*}, p_j^{D*}) lies within P, Q. In contrast, when $(p_j^D)_X \ge p_{\max}^D \ge (p_j^D)_Y$ as shown **Fig. 2(b)**, where $(p_j^D)_X$ is the p_j^D at point X, the above mutual interference is weak, i.e., both $g_{B,j}^{CD} / g_{B,n}^{CC}$ and $g_{j,n}^{DC} / g_{j,j}^{DD}$ are small, and (p_n^{C*}, p_j^{D*}) lies within P, Y, I. Finally, when $p_{\max}^D \ge (p_j^D)_X$ as shown in **Fig. 2(c)**, the DL tends to impose a strong interference to the CL, i.e., a small $g_{j,n}^{DC} / g_{j,j}^{DD}$ and a large $g_{B,j}^{CD} / g_{B,n}^{CC}$, and (p_n^{C*}, p_j^{D*}) lies within X, Y. ∎

$$\gamma_{j,n}^{D\ (3)}(\alpha p_n^C,\alpha p_j^D) = \frac{\alpha p_j^D g_{j,j}^{DD}}{\alpha p_n^C g_{j,n}^{DC} + \sigma^2} = \frac{p_j^D g_{j,j}^{DD}}{p_n^C g_{j,n}^{DC} + \sigma^2/\alpha} > \gamma_{j,n}^{D\ (3)}(p_n^C,p_j^D)$$

$$\gamma_n^{C(3)}(\alpha p_n^C,\alpha p_j^D) = \frac{\alpha p_n^C g_{B,n}^{CC}}{\alpha p_j^D g_{B,j}^{CD} + \sigma^2} = \frac{p_n^C g_{B,n}^{CC}}{p_j^D g_{B,j}^{CD} + \sigma^2/\alpha} > \gamma_n^{C(3)}(p_n^C,p_j^D) \left.\right\} R_{j,n}^{D\ (3)}(\alpha p_n^C,\alpha p_j^D) > R_{j,n}^{D\ (3)}(p_n^C,p_j^D) \quad (24)$$

$$\frac{\partial \log_2(1+x)}{\partial x} = \log_2 e \frac{1}{1+x} > 0, \forall x > 0$$

$$\frac{\partial \psi}{\partial p_j^D} = \frac{g_{j,j}^{DD}(p_j^D g_{B,j}^{CD} + \sigma^2)^2 + g_{B,n}^{CC} p_{\max}^C [g_{j,j}^{DD}\sigma^2 - g_{B,j}^{CD}(p_{\max}^C g_{j,n}^{DC} + \sigma^2)]}{(p_{\max}^C g_{j,n}^{DC} + \sigma^2)(p_j^D g_{B,j}^{CD} + \sigma^2)^2} \quad (a)$$

$$(25)$$

$$\frac{\partial^2 \psi}{\partial (p_j^D)^2} = \frac{2g_{B,j}^{CD} g_{B,n}^{CC} p_{\max}^C [g_{B,j}^{CD}(p_{\max}^C g_{j,n}^{DC} + \sigma^2) - g_{j,j}^{DD}\sigma^2]}{(p_{\max}^C g_{j,n}^{DC} + \sigma^2)(p_j^D g_{B,j}^{CD} + \sigma^2)^3} \quad (b)$$

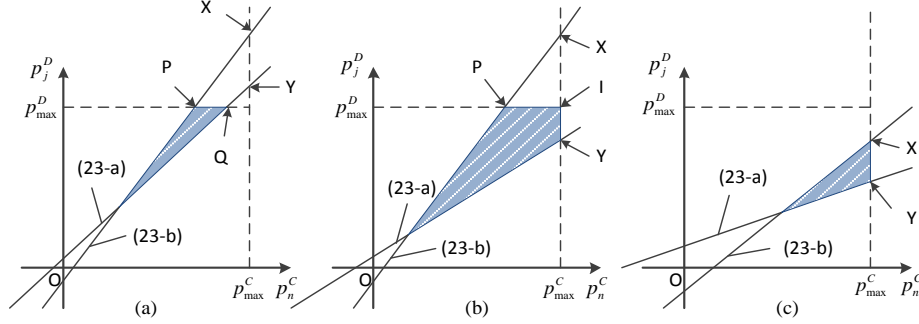**Fig. 2.** Structure of optimality in *P4* for *Mode 3*

B. Mode 5: For the triple $(j,a,n) \in D \times A \times H$, if $a = L + J + k \in A_3, n \in C$, then the *j*-th DL chooses the underlaid MR-aided mode with data relayed via the *k*-th MR. When the AF relay protocol is adopted at the MRs and the *n*-th CL occupies the first slot, *P2* is specified into *Problem 5 (P5)*, and the feasible condition and the optimality upon feasibility can be jointly summarized below, in **Proposition 3**.

$$\underline{Problem\ 5}: \underset{p_n^C,p_j^D,p_k^R}{maximize}\ 2R_{j,k,n}^{D\ (5-1)} = \log_2(1 + \gamma_{j,k,n}^{D\ (5F)} \gamma_{k,j,n}^{D\ (5F)} / (\gamma_{j,k,n}^{D\ (5F)} + \gamma_{k,j,n}^{D\ (5F)} + 1)) + \log_2(1 + \gamma_n^{C(5F)}) \quad (a)$$

$$subject\ to: 0 \le p_j^D \le p_{\max}^D, 0 \le p_n^C \le p_{\max}^C, 0 \le p_k^R \le p_{\max}^R \quad (b); \qquad \gamma_n^{C(5F)} = p_n^C g_{B,n}^{CC} / (p_j^D g_{B,j}^{CD} + \sigma^2) \ge \theta_n^C \quad (c)$$

$$\gamma_{j,k,n}^{D\ (5F)} = p_j^D g_{k,j}^{RD} / (p_m^C g_{k,n}^{RC} + \sigma^2) \ge \theta_j^D \quad (d); \quad \gamma_{j,k,n}^{D\ (5F)} \gamma_{k,j,n}^{D\ (5F)} / (\gamma_{j,k,n}^{D\ (5F)} + \gamma_{k,j,n}^{D\ (5F)} + 1) \ge \theta_j^D \quad (e)$$

$$(26)$$

Proposition 3: *P5* is feasible if $c_1, c_2 > 0, c_3/c_2 \le \min\{c_1, p_{\max}^D\}$ and $\beta \underline{\underline{\Delta}} \gamma_{k,j,m}^{D\ (5F)}(p_{\max}^R) > \theta_j^D$, where $c_1 = (p_{\max}^C g_{B,n}^{CC} / \theta_n^C - \sigma^2)/g_{B,j}^{CD}, c_2 = 1 - \alpha_{j,k}^D \theta_n^C g_{k,n}^{RC} g_{B,j}^{CD} / (g_{k,j}^{RD} g_{B,n}^{CC}), c_3 = \alpha_{j,k}^D \theta_n^C g_{k,n}^{RC} \sigma^2 / (g_{B,n}^{CC} g_{k,j}^{RD}) + \alpha_{j,k}^D \sigma^2 / g_{k,j}^{RD}$ with $\alpha_{j,k}^D = \theta_j^D(1+\beta)/(\beta - \theta_j^D)$. Upon feasibility, the optimal power pair to *P5* lies within P', X', Q', Y', I' with $p_k^R = p_{\max}^R$, which is both feasible and reaps the largest $R_{j,k,n}^{D\ (5-1)}(p_n^C, p_j^D, p_{\max}^R)$, i.e., (p_n^{C*}, p_j^{D*}) $= \underset{\{P',X',Q',Y',I'\}}{argmax}\ R_{j,k,n}^{D\ (5-1)}(p_n^C, p_j^D, p_{\max}^R)$, $P' = (\theta_n^C(p_{\max}^D g_{B,j}^{CD} + \sigma^2)/g_{B,n}^{CC}, p_{\max}^D)$, $X' = (p_{\max}^C, (p_{\max}^C g_{B,n}^{CC} / \theta_n^C - \sigma^2)/g_{B,j}^{CD})$, $Q' = ((p_{\max}^D g_{k,j}^{RD} / \alpha_{j,k}^D - \sigma^2)/g_{k,n}^{RC}, p_{\max}^D), Y' = (p_{\max}^C, \alpha_{j,k}^D(p_{\max}^C g_{k,n}^{RC} + \sigma^2)/g_{k,j}^{RD})$ and $I' = (p_{\max}^C, p_{\max}^D)$.

Proof. We note that the objective function $R_{j,k,n}^{D\ (5-1)}(p_n^C, p_j^D, p_k^R)$ is strictly increasing in p_k^R, thus $p_k^R = p_{\max}^R$ when optimal. If we define $\beta \underline{\underline{\Delta}} \gamma_{k,j,m}^{D\ (5F)}(p_{\max}^R) = p_{\max}^R g_{j,k}^{DR} / \sigma^2$, *P5* can be reduced to (27).

$$maximize_{p_n^C, p_j^D} 2R_{j,k,n}^{D}{}^{(5-1)}(p_n^C, p_j^D, p_k^R) = \log_2(1 + \beta\gamma_{j,k,n}^{D}{}^{(5F)} / (\gamma_{j,k,n}^{D}{}^{(5F)} + \beta + 1)) + \log_2(1 + \gamma_n^{C(5F)}) \quad (a)$$

$$subject\ to : 0 \le p_j^D \le p_{\max}^D, 0 \le p_n^C \le p_{\max}^C \qquad (b);\ \gamma_n^{C(5F)} = p_n^C g_{B,n}^{CC} / (p_j^D g_{B,j}^{CD} + \sigma^2) \ge \theta_n^C \quad (c) \quad (27)$$

$$\gamma_{j,k,n}^{D}{}^{(5F)} = p_j^D g_{k,j}^{RD} / (p_n^C g_{k,n}^{RC} + \sigma^2) \ge \theta_j^D(1 + \beta)/(\beta - \theta_j^D) \quad (d);\ \beta = p_{\max}^R g_{j,k}^{DR} / \sigma^2 > \theta_j^D \qquad (e)$$

Compared to *P4*, (27) has an additional constraint (27-e), but it is independent to p_n^C, p_j^D, thus **Proposition 1** can be readily extended here for the feasibility check only by switching $g_{j,j}^{DD}, g_{j,n}^{DC}, \theta_j^D$ into $g_{k,j}^{RD}, g_{k,n}^{RC}, \theta_j^D(1 + \beta)/(\beta - \theta_j^D)$, respectively, in b_1, b_2, b_3, and further limiting that (27-e) is maintained. Moreover, compared to *P4*, (27) has a different objective function, but it can be readily verified that (27-a) has the same properties as (22-a), elaborated in **Proposition 2** as follows: (i) For any feasible power pair (p_n^C, p_j^D), $R_{j,n}^{D}{}^{(5-1)}(\alpha p_n^C, \alpha p_j^D)$ is still strictly increasing in $\alpha > 1$ when $(\alpha p_n^C, \alpha p_j^D)$ is still feasible with the help of (24) and $\partial \log_2(1 + \beta x/(x + \beta + 1))/\partial x > 0$, and (ii) For $\xi \underline{\underline{\Delta}}(1 + \beta\gamma_{j,k,n}^{D}{}^{(5F)} / (\gamma_{j,k,n}^{D}{}^{(5F)} + \beta + 1))(1 + \gamma_n^{C(5F)})|_{(p_{\max}^C, p_j^D)}$, we can derive the first-order derivative as in (28), where $I_{k,n}^D = p_{\max}^C g_{k,n}^{RC} + \sigma^2, I_{j,n}^C = p_j^D g_{B,j}^{CD} + \sigma^2$, and if (29) is valid, $\partial \xi / \partial p_j^D \ge 0$ or $\partial \xi / \partial p_j^D \le 0$; otherwise, we can readily verify that $\partial^2 \xi / \partial (p_j^D)^2 > 0$, and we omit the math for brevity. In contrast to **Proposition 2**, the symmetry is violated, so we proceed to analyze the property of $\varpi \underline{\underline{\Delta}}(1 + \beta\gamma_{j,k,n}^{D}{}^{(5F)} / (\gamma_{j,k,n}^{D}{}^{(5F)} + \beta + 1))(1 + \gamma_n^{C(5F)})|_{(p_n^C, p_{\max}^D)}$, the first-order derivative of which is shown in (30), where $J_{k,n}^D = p_n^C g_{k,n}^{RC} + \sigma^2, J_{j,n}^C = p_{\max}^D g_{B,j}^{CD} + \sigma^2$. The minimum value of $\partial \xi / \partial p_j^D$ is obtained when $p_n^C = 0$, so if (31) is valid, $\partial \varpi / \partial p_n^C > 0$; otherwise, we can readily verify that $\partial^2 \varpi / \partial (p_n^C)^2 > 0$, and we omit the math for brevity. Therefore, both $\xi(p_j^D)$ and $\varpi(p_n^C)$ are either monotone or convex, thus the optimality to *P5* adopts the identical structure of that in *P4* only by switching the channel coefficients at the potential optimal points P', X', Q', Y', I'. Similar to **Fig. 2**, we first filter out the feasible points from P', X', Q', Y', I', and then find out the one with the maximum objective function in (27-a). The details are omitted for brevity.∎

$$\frac{\partial \xi}{\partial p_j^D} = \frac{\beta(\beta + 1)g_{k,j}^{RD}I_{j,n}^C(p_{\max}^C g_{B,n}^{CC} + I_{j,n}^C) - (\beta + 1)g_{B,j}^{CD}p_{\max}^C g_{B,n}^{CC}I_{k,n}^D(\gamma_{j,k,n}^{D}{}^{(5F)} + 1)(\gamma_{j,k,n}^{D}{}^{(5F)} + \beta + 1)}{(\gamma_{j,k,n}^{D}{}^{(5F)} + \beta + 1)^2 I_{k,n}^D I_{j,n}^{C}{}^2} \quad (28)$$

$$(g_{B,j}^{CD}I_{k,n}^D - g_{k,j}^{RD}\sigma^2)[(\beta + 1)g_{B,j}^{CD}I_{k,n}^D - g_{k,j}^{RD}(p_{\max}^C g_{B,n}^{CC} + \sigma^2)] \le 0 \quad (29)$$

$$\frac{\partial \varpi}{\partial p_n^C} = \frac{-\beta(\beta + 1)g_{k,n}^{RC}p_{\max}^D g_{k,j}^{RD}(p_n^C g_{B,n}^{CC} + J_{j,n}^C) + (\beta + 1)g_{B,n}^{CC}J_{k,n}^{D}{}^2(\gamma_{j,k,n}^{D}{}^{(5F)} + 1)(\gamma_{j,k,n}^{D}{}^{(5F)} + \beta + 1)}{(\gamma_{j,k,n}^{D}{}^{(5F)} + \beta + 1)^2 J_{j,n}^C J_{k,n}^{D}{}^2} \quad (30)$$

$$g_{B,n}^{CC}[p_{\max}^{D}{}^2 g_{k,j}^{RD2} + (\beta + 2)\sigma^2 p_{\max}^D g_{k,j}^{RD} + (\beta + 1)(\sigma^2)^2] \ge \beta g_{k,n}^{RC}p_{\max}^D g_{k,j}^{RD}J_{j,n}^C \quad (31)$$

Remark 1. When the *n*-th CL occupies the second slot, *P2* can be specified similarly to *P5*. Compared to (27), the channel coefficients between the two interfering links are changed, but the structure of the optimality stays the same, so conclusions on the feasibility check and the optimality when the *n*-th CL occupies the second slot have the same form as **Proposition 3** only by changing $g_{k,j}^{RD}, g_{B,j}^{CD}, g_{k,n}^{RC}, \beta, p_j^D, p_{\max}^D$ into $g_{j,k}^{DR}, g_{n,k}^{CR}, g_{j,n}^{DC}, \phi, p_k^R, p_{\max}^R$, respectively, in c_1, c_2, c_3 and P', X', Q', Y', I' after $\phi \underline{\underline{\Delta}} p_{\max}^D g_{k,j}^{RD} / \sigma^2 > \theta_j^D$. In sum, if we denote the optimal power pairs in the above two cases as $(p_{n,AF}^{CF*}, p_{j,AF}^{D*})$ and $(p_{n,AF}^{CS*}, p_{j,AF}^{R*})$, respectively, then the optimal data rate of *Mode 5* under the AF relay protocol is $R_{j,a,n}^* = \max\{R_{j,k,n}^{D}{}^{(5-1)}(p_{n,AF}^{CF*}, p_{j,AF}^{D*}, p_{\max}^R), R_{j,k,n}^{D}{}^{(5-3)}(p_{n,AF}^{CS*}, p_{\max}^D, p_{j,AF}^{R*})\}$.

On the other hand, when the DF relay protocol is adopted at the MRs and the *n*-th CL occupies the first slot, *P2* is specified into *Problem 6 (P6)*, and the feasible condition and the optimality upon feasibility can be jointly summarized below, in **Proposition 4**.

$\underline{Problem\ 6}:maximize_{p_n^C,p_j^D,p_k^R}\,2R_{j,k,n}^{D\ (5-2)}(p_n^C,p_j^D,p_k^R)=\log_2(1+\min\{\gamma_{j,k,n}^{D\ (5F)},\gamma_{k,j,n}^{D\ (5F)}\})+\log_2(1+\gamma_n^{C(5F)})\quad(a)$

$subject\ to:0\le p_j^D\le p_{\max}^D,0\le p_n^C\le p_{\max}^C,0\le p_k^R\le p_{\max}^R\ \ (b);\quad \gamma_n^{C(5F)}=p_n^C g_{B,n}^{CC}/(p_j^D g_{B,j}^{CD}+\sigma^2)\ge\theta_j^D\quad(c)\,(32)$

$\gamma_{j,k,n}^{D\ (5F)}=p_j^D g_{k,j}^{RD}/(p_n^C g_{k,n}^{RC}+\sigma^2)\ge\theta_j^D\qquad(d);\qquad\qquad \gamma_{k,j,n}^{D\ (5F)}=p_k^R g_{j,k}^{DR}/\sigma^2\ge\theta_j^D\quad(e)$

Proposition 4: *P6* is feasible if $d_1,d_2>0,d_3/d_2\le\min\{d_1,p_{\max}^D\}$ and $\beta=\gamma_{k,j,n}^{D\ (5F)}(p_{\max}^R)\ge\theta_j^D$, where

$d_1=(p_{\max}^C g_{B,n}^{CC}/\theta_n^C-\sigma^2)/g_{B,j}^{CD},d_2=1-\theta_j^D\theta_n^C g_{k,n}^{RC}g_{B,j}^{CD}/(g_{k,j}^{RD}g_{B,n}^{CC})$ and $d_3=\theta_j^D\theta_n^C g_{k,n}^{RC}\sigma^2/(g_{B,n}^{CC}g_{k,j}^{RD})+\theta_j^D\sigma^2/g_{k,j}^{RD}$.

Upon feasibility, the optimal power pair for *P6* lies within $P^{\cdot},X^{\cdot},Q^{\cdot},Y^{\cdot},S^{\cdot},Z^{\cdot},I^{\cdot}$ with $p_k^R=p_{\max}^R$, which is both feasible and reaps the largest $R_{j,k,n}^{D\ (5-2)}(p_n^C,p_j^D,p_{\max}^R)$ in (33-a), i.e., $(p_n^{C*},p_j^{D*})=$

$\underset{\{P^{\cdot},X^{\cdot},Q^{\cdot},Y^{\cdot},S^{\cdot},Z^{\cdot},I^{\cdot}\}}{argmax}\ R_{j,k,n}^{D\ (5-2)}(p_n^C,p_j^D,p_{\max}^R),P^{\cdot}=(\theta_n^C(p_{\max}^D g_{B,j}^{CD}+\sigma^2)/g_{B,n}^{CC},p_{\max}^D),X^{\cdot}=(p_{\max}^C,(p_{\max}^C g_{B,n}^{CC}/\theta_n^C-\sigma^2)/g_{B,j}^{CD})$,

$Q^{\cdot}=((p_{\max}^D g_{k,j}^{RD}/\theta_j^D-\sigma^2)/g_{k,n}^{RC},p_{\max}^D),Y^{\cdot}=(p_{\max}^C,\theta_j^D(p_{\max}^C g_{k,n}^{RC}+\sigma^2)/g_{k,j}^{RD}),S^{\cdot}=((p_{\max}^D g_{k,j}^{RD}/\beta-\sigma^2)/g_{k,n}^{RC},p_{\max}^D)$,

$Z^{\cdot}=(p_{\max}^C,\beta(p_{\max}^C g_{k,n}^{RC}+\sigma^2)/g_{k,j}^{RD})$ and $I^{\cdot}=(p_{\max}^C,p_{\max}^D)$.

Proof. Due to the characteristics of DF-based two-hop transmissions (i.e., the maximum data rate of the second hop is achieved when $p_k^R=p_{\max}^R$, and the performance of the two-hop transmission is determined by the hop with the worst data rate), *P6* can be reformulated below,

$maximize_{p_j^D,p_k^R}\,2R_{j,k,n}^{D\ (5-2)}(p_n^C,p_j^D,p_k^R)=\log_2(1+\gamma_{j,k,n}^{D\ (5F)})+\log_2(1+\gamma_n^{C(5F)})\qquad(a)$

$subject\ to:0\le p_j^D\le p_{\max}^D,0\le p_n^C\le p_{\max}^C\qquad(b);\qquad \gamma_n^{C(5F)}=p_n^C g_{B,n}^{CC}/(p_j^D g_{B,j}^{CD}+\sigma^2)\ge\theta_n^C\quad(c)\quad(33)$

$\beta\ge\gamma_{j,k,n}^{D\ (5F)}=p_j^D g_{k,j}^{RD}/(p_n^C g_{k,n}^{RC}+\sigma^2)\ge\theta_j^D\quad(d);\qquad \beta=\gamma_{k,j,n}^{D\ (5F)}(p_{\max}^R)=p_{\max}^R g_{j,k}^{DR}/\sigma^2\ge\theta_j^D\quad(e)$

Compared to *P4*, an additional upper bound constraint is imposed on SINR $\gamma_{j,k,n}^{D\ (5F)}$, i.e., $\beta\ge\gamma_{j,k,n}^{D\ (5F)}$ in (33-d), but it is always valid after $\beta\ge\theta_j^D$, so the conclusion of the feasibility check in *P4* can be directly extended here only by switching $g_{j,j}^{DD},g_{j,n}^{DC}$ into $g_{k,j}^{RD},g_{k,n}^{RC}$ respectively, on the condition that (33-e) is further satisfied. Moreover, the objective function in (33-a) is in the same form as (22-a) in *P4*, so similar to **Proposition 2**, both $(1+\gamma_{j,k,n}^{D\ (5F)})(1+\gamma_n^{C(5F)})|_{(p_{\max}^C,p_j^D,p_{\max}^R)}$ and $(1+\gamma_{j,k,n}^{D\ (5F)})(1+\gamma_n^{C(5F)})|_{(p_n^C,p_{\max}^D,p_{\max}^R)}$ are either monotonically increasing or convex in p_j^D and p_n^C, respectively. Finally, the optimal power pair can lie only at corner points of the feasible zone when $p_n^C=p_{\max}^C$ and $p_j^D=p_{\max}^D$, i.e., (p_n^{C*},p_j^{D*}) lies within the points $P^{\cdot},X^{\cdot},Q^{\cdot},Y^{\cdot},S^{\cdot},Z^{\cdot},I^{\cdot}$, which is both feasible and reaps the largest $R_{j,k,n}^{D\ (5-2)}(p_n^C,p_j^D,p_{\max}^R)$ in (33-a), i.e., $(p_n^{C*},p_j^{D*})=argmax_{\{P^{\cdot},X^{\cdot},Q^{\cdot},Y^{\cdot},S^{\cdot},Z^{\cdot},I^{\cdot}\}}R_{j,k,n}^{D\ (5-2)}(p_n^C,p_j^D,p_{\max}^R)$, $P^{\cdot}=(\theta_n^C(p_{\max}^D g_{n,j}^{CD}+\sigma^2)/g_{B,n}^{CC},p_{\max}^D)$, $X^{\cdot}=(p_{\max}^C,(p_{\max}^C g_{B,n}^{CC}/\theta_n^C-\sigma^2)/g_{B,j}^{CD}),Q^{\cdot}=((p_{\max}^D g_{k,j}^{RD}/\theta_j^D-\sigma^2)/g_{k,n}^{RC},p_{\max}^D),Y^{\cdot}=(p_{\max}^C,\theta_j^D(p_{\max}^C g_{k,n}^{RC}+\sigma^2)/g_{k,j}^{RD})$, $S^{\cdot}=((p_{\max}^D g_{k,j}^{RD}/\beta-\sigma^2)/g_{k,n}^{RC},p_{\max}^D),Z^{\cdot}=(p_{\max}^C,\beta(p_{\max}^C g_{k,n}^{RC}+\sigma^2)/g_{k,j}^{RD})$ and $I^{\cdot}=(p_{\max}^C,p_{\max}^D)$. Specifically, if $(p_n^C)_{P^{\cdot}}\ge(p_n^C)_{S^{\cdot}}$ where $(p_n^C)_{P^{\cdot}},(p_n^C)_{S^{\cdot}}$ are the p_n^C at points $P^{\cdot},S^{\cdot}$, constraint (33-e) is redundant, and the structure of (p_n^{C*},p_j^{D*}) is exactly the same as that in *P4*, as illustrated in **Fig. 3(a)** through **Fig. 3(c)** by switching P,X,Q,Y,I into $P^{\cdot},X^{\cdot},Q^{\cdot},Y^{\cdot},I^{\cdot}$, where the shadow region is the feasible zone constructed by (34) and $p_n^C=p_{\max}^C,p_j^D=p_{\max}^D$; otherwise, the structure of (p_n^{C*},p_j^{D*}) can be depicted as in **Fig. 3(d)** through **Fig. 3(f)**. In **Fig. 3(d)**, (p_n^{C*},p_j^{D*}) lies within $S^{\cdot},Q^{\cdot}$. In **Fig. 3(e)**, (p_n^{C*},p_j^{D*}) lies within $S^{\cdot},Y^{\cdot},I^{\cdot}$. In **Fig. 3(f)**, (p_n^{C*},p_j^{D*}) lies within $Y^{\cdot},Z^{\cdot}$.∎
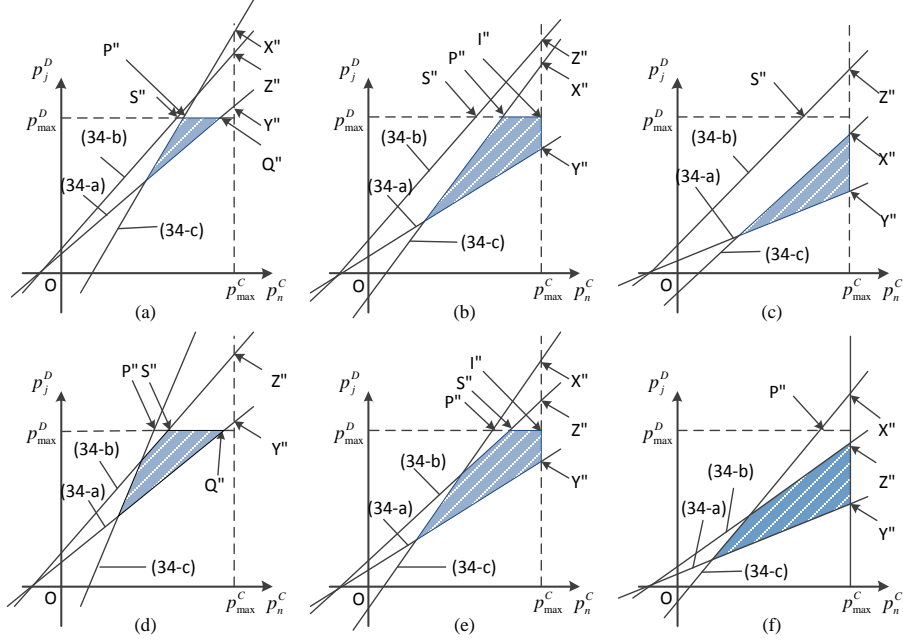
Fig. 3. Structure of optimality in *P6* for *Mode 5* when the DF relay protocol is adopted at MRs

$$p_j^D g_{k,j}^{RD} / (p_n^C g_{k,n}^{RC} + \sigma^2) = \theta_j^D \quad (a); \quad p_j^D g_{k,j}^{RD} / (p_n^C g_{k,n}^{RC} + \sigma^2) = \beta \quad (b); \quad p_n^C g_{B,n}^{CC} / (p_j^D g_{B,j}^{CD} + \sigma^2) = \theta_n^C \quad (c) \quad (34)$$

Remark 2. When the *n*-th CL occupies the second slot, *P2* can be specified similarly to *P6*. Compared to (33), the channel coefficients between the two interfering links are changed, but the structure of the optimality stays the same, so conclusions on the feasibility check and the optimality when the *n*-th CL occupies the second slot have the same form as **Proposition 4** only by switching $g_{k,j}^{RD}, g_{k,n}^{RC}, g_{B,j}^{CD}, \beta, p_j^D, p_{max}^D$ into $g_{j,k}^{DR}, g_{j,n}^{DC}, g_{B,k}^{CR}, \phi, p_k^R, p_{max}^R$, respectively, in d_1, d_2, d_3 and $P'', X'', Q'', Y'', S'', Z'', I''$ after $\phi \geq \theta_j^D$. In sum, if we denote the optimal power pairs in the above two cases as $(p_{n,DF}^{CF*}, p_{j,DF}^{D*})$ and $(p_{n,DF}^{CS*}, p_{j,DF}^{R*})$, respectively, then the optimal data rate of *Mode 5* under the DF relay protocol is $R_{j,a,n}^* = \max\{R_{j,k,m}^{D\ (5-2)}(p_{n,DF}^{CF*}, p_{j,DF}^{D*}, p_{max}^R), R_{j,k,m}^{D\ (5-4)}(p_{n,DF}^{CS*}, p_{max}^D, p_{j,DF}^{R*})\}$.

In summary, our original joint MS and RA problem, i.e., *P1* in (19), can be perfectly decomposed into two sub-problems, i.e., *P2* in (20) and *P3* in (21), due to the separability of PC variables and joint MS, CA and RS variables. In addition, the above two sub-problems are *highly related* and *maintain a hierarchical structure*, i.e., *P3* optimizes $x_{j,a,n}$ based on $R_{j,a,n}^*$ of *P2*, so we analyze *P2* in front of *P3*. From (20), we learn that the design of PC and AC in *P2* is *mode-dependent*, thus we address its optimality separately for each of the five available modes as described in Section 2.2. For the modes on dedicated channels, i.e., *Modes* 1, 2 and 4, the maximum data rate is achieved when all involved transmitters emit at the largest power as articulated in Section 3.1.1. While for the modes on reused channels, i.e., *Modes* 3 and 5, the optimal power pair for the maximum data rate lies within a finite set of points. In particular, when a DL is in *Mode 3*, *P2* is specified into *P4*, and the optimality of *P4* is concluded in **Proposition 2** upon the feasible conditions given in **Proposition 1**. When a DL is in *Mode 5* and the AF relay protocol is applied at its MR, *P2* is specified into *P5*, and the optimality of *P5* is summarized with the help of **Proposition 3** and **Remark 1**. Besides, when a DL is in *Mode 5* and the DF relay protocol is adopted at its MR, *P2* is specified into *P6*, and the optimality of *P6* is outlined with the help of **Proposition 4** and **Remark 2**. Based on the above optimality of

P2 under all five available modes, i.e., $R^*_{j,a,n}, \forall (j,a,n) \in D \times A \times H$, *P1* is reduced to *P3* in (21), which is the joint design of MS, CA and RS and will be studied in the following subsection.

3.2 Joint Design of MS, CA and RS

With all feasible $R^*_{j,a,n}$ in hand, *P3* is intended to maximize the sum data rate or the ST of all admitted DLs by joint MS, CA and RS under the restrictions that (i) each DL chooses at most one mode and one flow, i.e., constraint (21-c), (ii) each MR serves at most one flow at a time, i.e., constraint (21-d), and (iii) each channel is reused by at most one DL (no dense spectral sharing), i.e., constraint (21-e). In what follows, we view *P3* in (21) from the perspective of graph theory, and an undirected weighted graph $G = (V, E, W)$ is formulated as specified below:

- V is the set of vertices, which is made up of all feasible triples with the help of the PC and AC as articulated in Section 3.1, i.e., $V = \{v = (j,a,n) \mid R^*_{j,a,n} > 0, \forall j \in D, a \in A, n \in H\}$. And vertex $v = (j,a,n)$ means that the *j*-th DL can be accessed into the system with the help of the *a*-th relay node on the *n*-th channel as defined in Section 2.3. Besides, the above transmission specified by triple (j,a,n) maintains the feasibility and reaps a maximum data rate of $R^*_{j,a,n} > 0$ with the help of the optimizations in Section 3.1.

- E is the set of edges, and every two vertices will have an edge if the triples they represent overlap, i.e., $E = \{(v_i, v_j) \mid v_i, v_j \in V, v_i \cap v_j \neq \varnothing\}$, indicating a resource-sharing collision between the above two vertices in DLs, relays and channels as depicted in (21-c) through (21-e). And for two arbitrary vertices, i.e., $v_1 = (j_1, a_1, n_1) \in V, v_2 = (j_2, a_2, n_2) \in V$, if $j_1 = j_2$, then these two vertices share a same DL, and this conflicts with constraint (21-c). Similarly, if $a_1 = a_2$ or $n_1 = n_2$, then the above two vertices share a same MR or channel, and this violates constraint (21-d) or constraint (21-e), respectively. Therefore, a feasible solution to *P3* in (21) must correspond to a set of vertices where no edge exists.

- W is the set of weights, and each element is associated with a vertex in V. For vertex $v = (j,a,n) \in V$, the weight is denoted as $W(v)$ and $W(v) = R^*_v = R^*_{j,a,n} > 0$.

It is worth noting that dense spectral sharing is disabled as assumed in Section 2.1 or (21-e), so there is no co-channel interference among admitted DLs, whereas the co-channel cross-tier interference between DLs and CLs does exist and has been amply addressed in Section 3.1 to derive the weights of vertices in our graph model. Then, we give definitions of *independent set* (IS) and *maximum weighted independent set* (MWIS) below before our final conclusion on *P3*.

Definition 1. Given an undirected weighted graph model $G = (V, E, W)$, a subset of vertices, i.e., $T \subseteq V$, is an IS if there is no edge between any two nodes in T, i.e., $(v_1, v_2) \notin E, \forall v_1 \in T, v_2 \in T - \{v_1\}$, and if the sum weight is the maximum among all ISs, T is then called the MWIS in G.

Proposition 5. To derive the optimal solution to *P3* is equivalent to searching for the MWIS of our undirected weighted graph as formulated above.

Proof. *P3* is in respect to the binary variables $x_v = x_{j,a,n}$, and if $x_v = 1$, vertex $v = (j,a,n)$ is selected; otherwise, vertex $v = (j,a,n)$ is left unselected. In this sense, *P3* can be thought of as choosing a subset from V, e.g., $T \subseteq V$, and the triples represented by every two different vertices in T do not overlap, as interpreted from the constraints from (21-c) to (21-e); therefore, T falls into an IS of G by the definition of E. To maximize its sum weight as in (21-a), *P3* searches for the MWIS of our undirected weighted graph by **Definition 1**. ∎

Table 3. A greedy algorithm that approximates an answer to our MWIS problem (**Algorithm 1**)

Initiation. Construct our undirected weighted graph $G = (V, E, W)$ as specified above. Here, $x_v, v \in V$ is a set of binary variables introduced to show whether the triple represented by vertex $v \in V$ is selected into the final MWIS. Initially, $x_v = 0, v \in V$.

Perform iteratively until no vertices are left in the graph.
(i) Derive the metric of each vertex in the current graph, i.e., $metric(v) = W(v) / \sum_{u \in NB^+(v)} W(u), \forall v \in V$.
(ii) Find the vertex with the maximum metric, i.e., $v^* = argmax_{v \in V} metric(v)$, and update $x_{v^*} = 1$. If there are ties, break them arbitrarily.
(iii) Remove v^* and the vertices that have an edge in common with v^* from V, thus updating V to $V = V - NB^+(v^*)$.

Unfortunately, MWIS is a well-known NP-complete problem [38], so it is preferable to find a heuristic algorithm to approximate to the optimality with polynomial-time complexity. In this sense, we adopt the greedy scheme proposed in [38] to tackle the above issue, i.e., **Algorithm 1** as shown in **Table 3**, which outputs an IS with a sum weight of at least $\sum_{v \in V} [W(v)^2 / \sum_{u \in NB^+(v)} W(u)]$ where $NB^+(v_i) = \{v_i\} \cup \{v_k \in V \mid (v_i, v_k) \in E\}$ [38, Theorem 3.11]. In **Algorithm 1**, the vertices in the final MWIS are selected in an iterative manner. In each iteration, the vertex selection criteria is performed by evaluating a metric, i.e., $W(v) / \sum_{u \in NB^+(v)} W(u)$, as specified in *Step (i)*, where the numerator $W(v)$ indicates the weight or the contribution to the overall ST of vertex v, and the denominator $\sum_{u \in NB^+(v)} W(u)$ indicates the sum weight of its neighbor vertices or the potential loss imposed on the overall ST by the admission of vertex v. Therefore, to achieve a higher ST, it's preferable to admit the vertex with the largest metric, i.e., $v^* = argmax_{v \in V} metric(v)$, as specified in *Step (ii)*. Due to the resource-sharing constraint shown in (21-c) through (21-e), when a vertex is selected, neighboring vertices will have no chance of being selected and are thus eliminated from V as specified in *Step (iii)*. The iteration continues until no vertices are left in the graph, i.e., $V = \varnothing$.

There are three components that comprise **Algorithm 1**: (i) The vertices of G are constructed during initiation. The maximum number of vertices is $JL + JN + JKN$, where JL denotes the vertices for *Mode 1*, JN denotes the vertices for *Modes 2 and 3*, and JKN denotes the vertices for *Modes 4 and 5*. Moreover, with the help of Section 3.1, the optimality of the PC part occurs when either the DT or MR transmits at the maximum power as given in Section 3.1.1 for the modes on dedicated channels, i.e., *Modes* 1, 2 and 4, or when the transmit power pair of all involved links is within a finite set of points as shown in Section 3.1.2 for the modes on reused channels, i.e., *Modes* 3 and 5. Therefore, the complexity of the PC part can simply be denoted as $O(1)$, and the sum complexity of component (i) is thus $O((JL + JN + JKN) \cdot 1) = O(JL + JN + JKN)$. (ii) The neighbor table of G is constructed during initiation. The complexity of this part is related to the maximum number of nodes from graph theory and on the order of $O((JL + JN + JKN)^2)$. (iii) Iterative vertex selection is performed. The main computations at each iteration lie primarily in finding the vertex with the maximum weight, as in *Step (ii)*. In the q-th iteration, the number of vertices is at most $(L - q)(L + J + K - q)(N - q)$, thus the complexity of component (iii) is on the order of $O(\sum_{q=0:\Delta_1} (L - q)(L + J + K - q)(N - q)) < O(\sum_{q=0:\Delta_1} (\Delta_2 - q)^3) < O(\Delta_1 (2\Delta_2 - \Delta_1)^3)$, where $\Delta_1 = \min\{J, N\}$ and $\Delta_2 = \max\{L + J + K, N\}$. Consequently, our algorithm to *P3*, i.e., **Algorithm 1**, is solvable in polynomial-time and thus greatly outweighs the exhaustive brute-force search scheme, whose

complexity is $O(\binom{L+J+K}{J} J! \max\{J,N\}^3)$, where $\binom{L+J+K}{J} J!$ derives all possible matchings

from D to A, and $\max\{J,N\}^3$ derives the optimal matching from each matching in $D \times A$ to N by the *Hungary algorithm*.

Remark 3. It is noteworthy that our joint resource optimization mechanism can be extended to optimize other targets only by switching (i) the weights in the graph formulation or (ii) the optimization model of *P2*. Specifically, for any vertex $v = (j,a,n)$, (i) when $W(v)$ is switched into a binary value: If $W(v) = 1$, it means the *j*-th DL can get admitted into the system on the *n*-th channel with data relayed via the *a*-th relay node; otherwise, it indicates infeasibility. Our mechanism is then changed to maximize the number of admitted DLs; (ii) when we switch the target in *P2*, i.e., $R_{j,a,n}$ in (20-a), into $-(p_n^C + p_j^D + p_k^R)$ or $R_{j,a,n} / (p_n^C + p_j^D + p_k^R)$, and also change the meaning of weights in *P3* accordingly, our mechanism is then converted to optimize EC or EE. We leave out the details due to space limitations and simply focus on the optimization of ST.

4. Numerical Results

Consider an isolated circular cell with a radius of 300 m, where the BS is located at cell center, and all transmitters are uniformly and independently distributed in the cell area. In addition, each DR is uniformly distributed in the disk area of its corresponding DT with a maximum radius of r. We assume that the BS and each UE is equipped with a single and omni-directional antenna with a processing gain of 14/0 dBi for BS/UEs and that the path loss model is the same as [18]. The bandwidth of each channel is 180 kHz, and the power spectral density of AWGN is -174 dBm/Hz, with a noise figure of 5/9 dB for BS/UEs.

A detailed description of the algorithms compared in this part is given below: (i) *Alg1: Forced cellular mode*, which is only available on vacant channels. To maximize ST, we need only arrange the DLs in a descending order by $R_{j,n}^{(1)}(p_{\max}^C), \forall j \in D$ for any $n \in V$, and then admit the first L DLs. The complexity of the above process is $O(J)$. (ii) *Alg2: Forced D2D mode*. The problem for ST maximization can be reduced into a bipartite matching where $R_{j,n}^{(2)}(p_{\max}^D), \forall n \in V, R_{j,n}^{(3)}(p_n^{C*}, p_j^{D*}), \forall n \in C$ is the utility for the (*j-n*) DL-channel pair, and can be optimality solved by the *Hungary algorithm* with the complexity being $O(\max(J,N)^3)$. (iii) *Alg3: Forced MR-aided mode* under the DF relay protocol. (iv) *Alg4: Forced MR-aided mode* under the AF relay protocol. The problems in the above *forced MR-aided modes* can be formulated similarly to *P1* in (19) only by switching A into A_3, and thus can be effectively solved by our mechanism. (v) *Alg5*: When only *cellular mode* and *D2D mode* are enabled, the problem is still NP-hard, but it can be well approximated by the two algorithms in [18] based on the system load level. (vi) *Our1*: When all five transmission modes are optional and the *DF relay protocol* is adopted at the MRs, we apply **Algorithm 1** to solve *P3*. (vii) *Our2*: When all five transmission modes are optional and the *AF relay protocol* is adopted at the MRs, we apply **Algorithm 1** to solve *P3*. (viii) *Opt1/Opt2/Opt3/Opt4*: We solve the same problem as in *Alg3/Alg4/Our1/Our2* using *Mosek*, which is a commercial optimization toolbox for complex problems, including the ILP. Finally, we observe the variations in ST and the number of admitted DLs using the change in system parameters as shown from **Fig. 4** to **Fig. 7**. In **Fig. 4**, we change η from 0 to 1 using a step size of 0.1 by fixing $K = 100, r = 200m, p_{\max}^D = p_{\max}^R = 20mW$. In **Fig. 5**, we change K from 50 to 150 using a step size of 10 by fixing $\eta = 0.8, r = 200m$ and $p_{\max}^D = p_{\max}^R = 20mW$. In **Fig. 6**, we change r from 150 m to 250 m using a step size of 10 m by

fixing $\eta = 0.8, K = 100, p_{max}^D = p_{max}^R = 20mW$. In **Fig. 7**, we change p_{max}^D from 5 mW to 55 mW using a step size of 5 mW by fixing $\eta = 0.8, K = 100, r = 200m$. In addition, we set $N = J = 20, \theta_j^D = 16dB, \forall j$ and $\theta_n^C = 10dB, \forall n, p_{max}^C = 200mW$, and all results are averaged over 1000 samples.
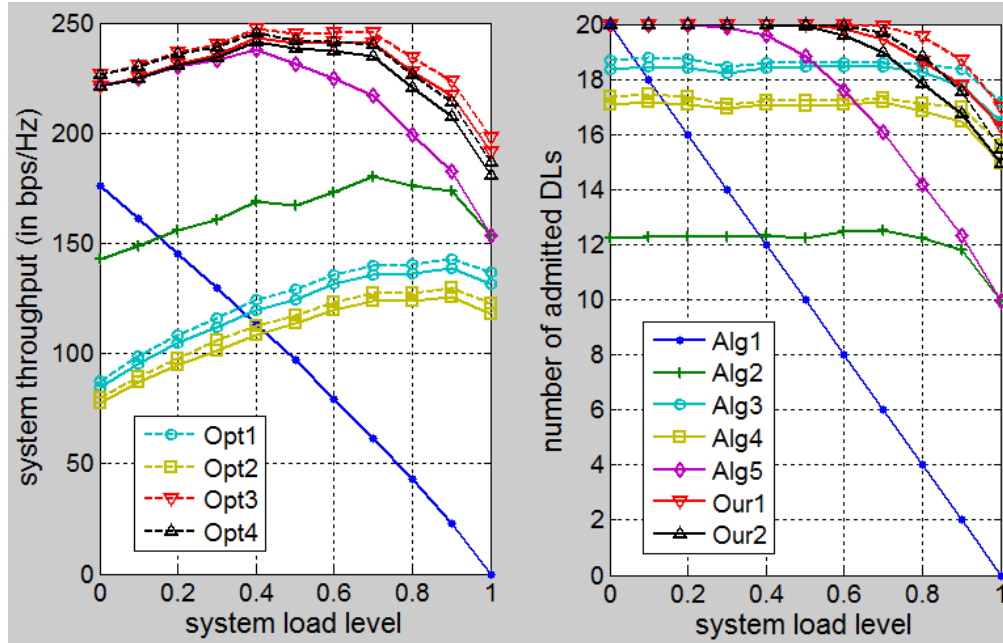

Fig. 4. Performance variation with change in system load level

As depicted in **Fig. 4**, as η rises, the number of vacant channels decreases; because DLs in *forced cellular mode* can use only dedicated channels, the number of admitted DLs and the ST in *Alg1* drops sharply. Specifically, when $\eta = 1$, no DLs can be admitted. As η rises, two changes are introduced that influence the ST for all algorithms except *Alg1*. First, spectral reuse is enabled, and the data rate of CLs will be accounted in the aggregate throughput as defined in (7), (13), (14), (17) and (18), thus enlarging ST. Second, with the decrease in L, some DLs in unfavorable channel conditions that were originally on dedicated channels are removed from the system, which lowers the ST. Finally, the ST of all algorithms except *Alg1* first increases and then decreases according to the degree of the above two influences. Compared to *Alg2*, *Alg3* and *Alg4* suffer an inferior ST due to the two-hop feature, but reap a higher number of admitted DLs with the help of the MRs. Compared to *Alg4*, *Alg3* reaps a higher ST under the DF relay protocol at the MRs by comparing (13), (17) and (14), (18), but may suffer from security issues because a UE can decode others' data. The same phenomenon can also be observed by comparing *Our1* and *Our2*. When $\eta \leq 0.3$, *Alg5* achieves almost the same ST as our two schemes because dedicated cellular and direct mode favor contributing a higher ST compared to the MR-aided relay mode. As η rises, spectral sharing is inevitable, and some DLs cannot get admitted under *Mode 1* due to channel shortages or under *Mode 3* due to severe mutual interference, but they can still be accessed with the help of the MRs, as in our two schemes, i.e., the gain of the MRs. Specifically, when $\eta = 1$, *Alg5* obtains the same performance as *Alg2* because *Mode 1* is forbidden, and *Our1* or *Our2* outwits *Alg5* by 24.70% or 17.86% in ST, respectively, and by 63.37% or 50.05% in the number of admitted DLs, respectively. Moreover, *Alg3* or *Alg4* harvests almost the same number of admitted DLs as

Our1 or *Our2*, but suffers a loss of 45.25% or 53.10% in ST, respectively, due to the two-hop feature. **Algorithm 1** is effective in solving the ILP in *P3*, while *Alg3/Alg4/Our1/Our2* suffer averaged losses of only 3.22/3.22/2.25/2.26% in ST, respectively, and 1.71/1.62/2.12/2.37% in the number of admitted DLs, respectively, compared to *Opt1/Opt2/Opt3/Opt4*.
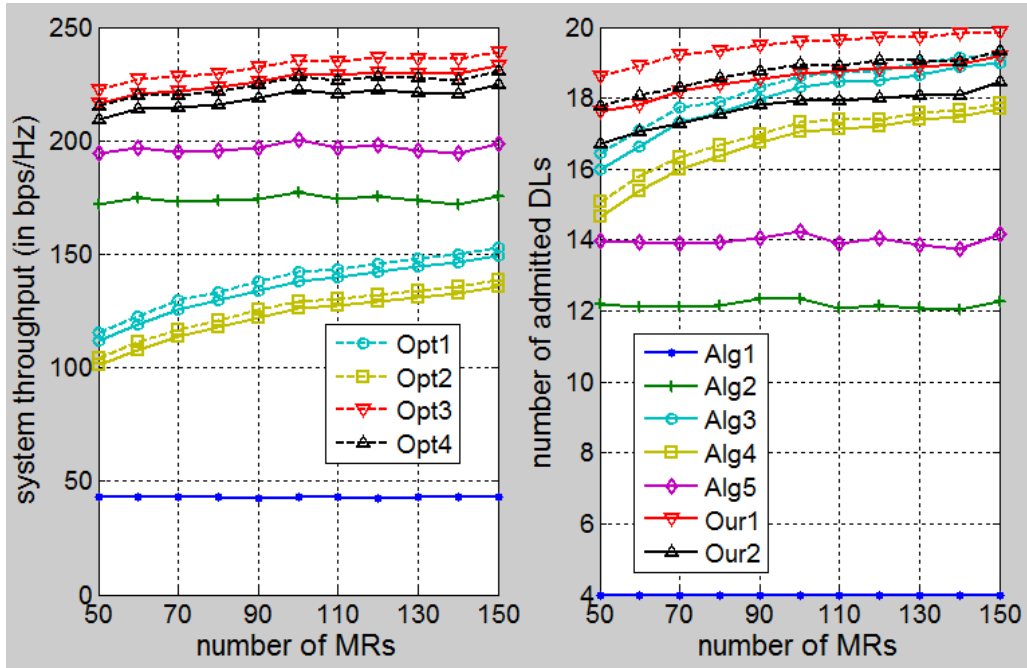

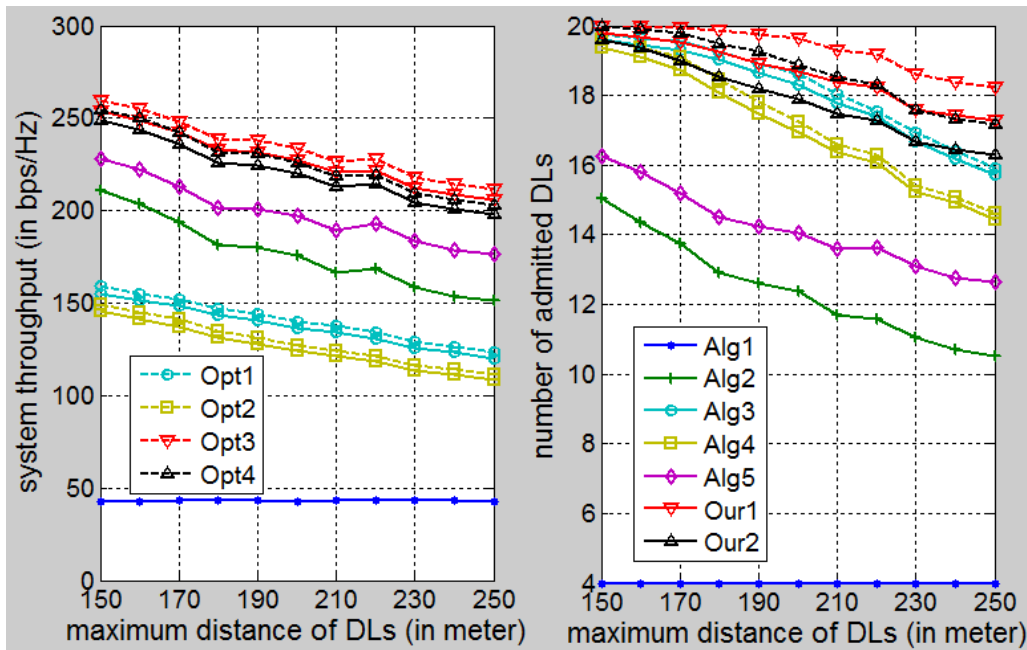**Fig. 5.** Performance variation with change in number of MRs

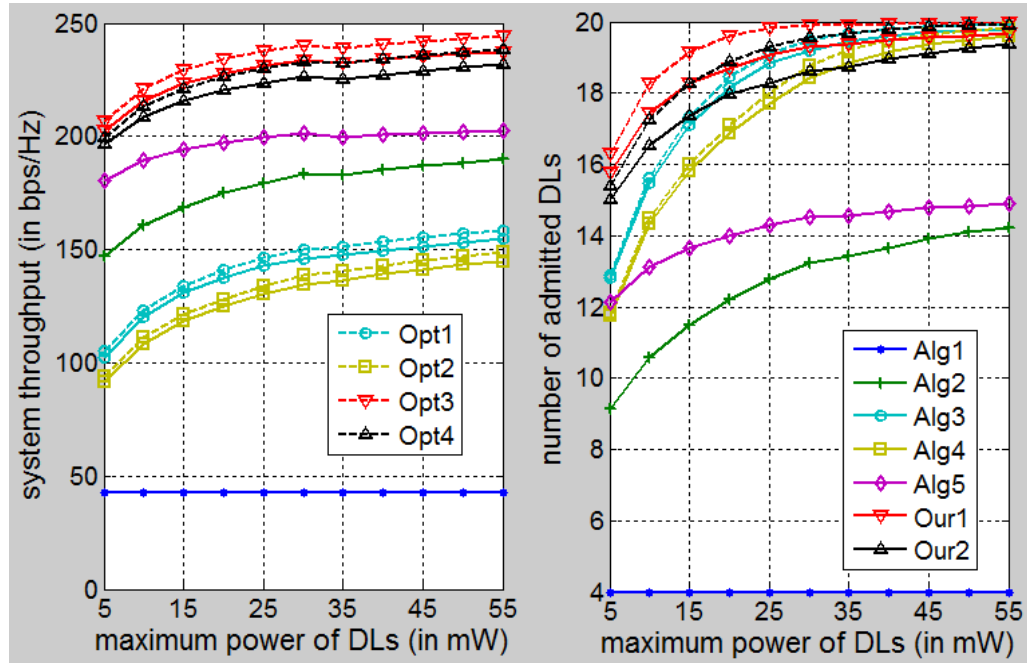**Fig. 6.** Performance variation with change in maximum distance of DLs

Fig. 7. Performance variation with change in maximum power of DLs

As shown in **Fig. 5**, the performances of *Alg1*, *Alg2* and *Alg5* are not affected by the change in the number of MRs because MRs are not involved in their calculations. As K increases, it becomes more likely that two distant UEs could be admitted with the relay of an MR, thus contributing to an enlarged ST value in algorithms where the MRs are involved. As delineated in **Fig. 6**, as the maximum distance of DLs increases, each link suffers a raised path loss, and may fail to achieve its SINR threshold especially in the case of spectral sharing, which results in a reduced number of admitted DLs and an inferior ST using all algorithms except *Alg1*. As illustrated in **Fig. 7**, as the maximum power budget of DLs increases, the system tends to maintain a larger feasible zone under *Mode 3* and *Mode 5*; thus all algorithms except *Alg1* harvest a rise in the number of admitted DLs, thereby enhancing ST. In sum, from **Fig. 5** to **Fig. 7**, the STs and the number of admitted DLs in *Our1* and *Our2* vary with different settings of system parameters, but they always outperform the rest by further exploring the benefits of MRs, especially under high system load levels. Moreover, they achieve polynomial-time complexity with only minor performance loss and are applicable when only MR-aided modes are optional, as in *Alg3* and *Alg4*. Specifically, *Our1* suffers averaged performance losses of only 2.71/2.55/2.62% in ST and of 4.70/3.86/3.08% in the number of admitted DLs with changes in $K, r, p_{max}^{D}(p_{max}^{R})$ when compared to *Opt3*, and verifies the effectiveness of our **Algorithm 1** in solving the original ILP in *P3*, which can also be observed via the comparison between *Alg3/Alg4/Our2* and *Opt1/Opt2/Opt4*. The details are omitted for brevity. Moreover, the performance gap between *Alg3* (*Alg4*) and *Opt1* (*Opt2*) is relatively smaller than that between *Our1* (*Our2*) and *Opt3* (*Opt4*) because the number of vertices in our graph model (searching space) is reduced when A is changed to A_3.

5. Conclusions

We have investigated the joint design of MS and RA for ST maximization in hybrid cellular networks with MR-aided D2D communication. Due to the high complexity of the problem, we decompose it into two parts and address each part sequentially. It is noteworthy that the above two parts are highly correlated, and the optimality of the PC part helps to construct the graph formulation in the second part. With the help of simulations, we illustrate that our scheme can approximate the optimality in polynomial-time, and observe the gain when MR is enabled to facilitate DLs and multiple resource variables are in elaborate cooperation, which provides more admission and spectral sharing opportunities, especially under high system load levels. In the future, we plan to develop distributed mechanisms and to consider scenarios when dense spectral reuse [35-36] and multi-hop transmission are enabled, which will play an important role in future user-oriented radio communication systems (5G) in improving individual QoS experiences as well as resource utilization, e.g., in the promising D2D-based heterogeneous network architecture for mobile cloud computing [32]. Moreover, we will include (massive) MIMO, i.e., the spatial-dimension resource, into the design of RA, which is another crucial technique addressed in future 5G radio communication systems [2-3, 22-23] and complements the D2D-based heterogeneous network architecture [32].

References

[1]   K. Doppler, M. Rinne, C. Wijting and et al, "Device-to-Device communication as an underlay to LTE-Advanced networks," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 42-49, Dec., 2009. Article (CrossRef Link)

[2]   A. Osseiran, F. Boccardi, V. Braun and et al, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26-35, May, 2014. Article (CrossRef Link)

[3]   J. Qiao, X. Shen, J. W. Mark and et al, "Enabling device-to-device communications in millimeter-wave 5G cellular networks," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 209-215, Jan., 2015. Article (CrossRef Link)

[4]   X. Lin, J. G. Andrews, A. Ghosh and R. Ratasuk, "An overview of 3GPP device-to-device proximity services," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 40-48, April, 2014. Article (CrossRef Link)

[5]   3GPP, "3rd Generation Partnership Project; Technical Specification Group RAN; Study on LTE device to device proximity services-Radio aspects (Release 12)," *TR 36. 843, V1.0.0*, Nov., 2013. Article (CrossRef Link)

[6]   K. Doppler, C.-H. Yu, C. B. Ribeiro and P. Jänis, "Mode selection for device-to-device communication underlaying an LTE-Advanced network," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1-6, April 18-21, 2010. Article (CrossRef Link)

[7]   C.-H. Yu, K. Doppler, C. B. Ribeiro and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2752-2763, Aug., 2011. Article (CrossRef Link)

[8]   Z. Liu, T. Peng, S. Xiang and W. Wang, "Mode selection for device-to-device (D2D) communication under LTE-Advanced networks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 5563-5567, June 10-15, 2012. Article (CrossRef Link)

[9]   S. Xiang, T. Peng, Z. Liu and W. Wang, "A distance-dependent mode selection algorithm in heterogeneous D2D and IMT-Advanced network," in *Proc. of IEEE Globecom Workshop*, pp. 416-420, Dec. 3-7, 2012. Article (CrossRef Link)

[10] Q. Wang, W. Wang, S. Jin and et al, "Mode selection for D2D communication underlaying a cellular network with shared relays," in *Proc. of 6th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1-6, Oct. 23-25, 2014. Article (CrossRef Link)

[11] S. Shalmashi, E. Björnson, S. B. Slimane and M. Debbah, "Closed-form optimality characterization of network-assisted device-to-device communications," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 508-513, April 6-9, 2014. Article (CrossRef Link)

[12] S. Hakola, T. Chen, J. Lehtomäki and T. Koskela, "Device-to-Device (D2D) communication in cellular network - Performance analysis of optimum and practical communication mode selection," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1-6, April 18-21, 2010. Article (CrossRef Link)

[13] M. Jung, K. Hwang and S. Choi, "Joint mode selection and power allocation scheme for power-efficient device-to-device (D2D) Communication," in *Proc. of IEEE 75th Vehicular Technology Conference Spring (VTC-S)*, pp. 1-5, May 6-9, 2012. Article (CrossRef Link)

[14] X. Xiao, X. Tao and J. Lu, "A QoS-aware power optimization scheme in OFDMA systems with integrated device-to-device (D2D) communications," in *Proc. of IEEE 74th Vehicular Technology Conference Fall (VTC-F)*, pp. 1-5, Sep. 5-8, 2011. Article (CrossRef Link)

[15] M.-H. Han, B.-G. Kim and J.-W. Lee, "Subchannel and transmission mode scheduling for D2D communication in OFDMA networks," in *Proc. of IEEE 76th Vehicular Technology Conference Fall (VTC-F)*, pp. 1-5, Sep. 3-6, 2012. Article (CrossRef Link)

[16] C.-P. Chien, Y.-C. Chen and H.-Y. Hsieh, "Exploiting spatial reuse gain through joint mode selection and resource allocation for underlay device-to-device communications," in *Proc. of International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 80-84, Sep. 24-27, 2012. Article (CrossRef Link)

[17] F. H. Khan, Y.-J. Choi and S. Bahk, "Opportunistic mode selection and RB assignment for D2D underlay operation in LTE network," in *Proc. of IEEE 75th Vehicular Technology Conference Spring (VTC-S)*, pp. 1-5, May 18-21, 2014. Article (CrossRef Link)

[18] G. Yu, L. Xu, R. Yin and et al, "Joint mode selection and resource allocation for device-to-device communications," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3814-3824, Nov., 2014. Article (CrossRef Link)

[19] C. Gao, X. Sheng, J. Tang and et al, "Joint mode selection, channel allocation and power assignment for green device-to-device communication," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 178-183, June 10-14, 2014. Article (CrossRef Link)

[20] H. Zhou, Y. Ji, J. Li and B. Zhao, "Joint mode selection, MCS assignment, resource allocation and power control for D2D communication underlaying cellular network," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1667-1672, April 6-9, 2014. Article (CrossRef Link)

[21] D. Feng, L. Lu, Y.-Y. Wu and et al, "Device-to-Device communications underlaying cellular networks," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3541-3551, Aug., 2013. Article (CrossRef Link)

[22] W. Zhong, Y. Fang, S. Jin and et al, "Joint resource allocation for device-to-device communications underlaying uplink MIMO cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 1, pp. 41-54, Jan., 2015. Article (CrossRef Link)

[23] X.-Y. Li, J. Li, W. Liu and et al, "Group-sparse-based joint power and resource block allocation design of hybrid device-to-device and LTE-Advanced networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 41-57, Jan., 2016. Article (CrossRef Link)

[24] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Mobile Relay for E-UTRA," *TR36.836, V12.0.0*, June, 2013. Article (CrossRef Link)

[25] J.-M. Chung, C. H. Kim and D. Lee, "A minimum energy consuming mobile device relay scheme for reliable QoS support," *KSII Transactions on Internet and Information Systems*, vol. 8, no. 2, pp. 618-633, Feb., 2014. Article (CrossRef Link)

[26] K. J. R. Liu, A. K. Sadek, W. Su and A. Kwasinski, *Cooperative Communications and Networking*, Cambridge University Press, New York, NY, USA, 2009. Article (CrossRef Link)

[27] X. Ma, R. Yin, G. Yu and Z. Zhang, "A distributed relay selection method for relay assisted device-to-device communication system," in *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1020-1024, Sep. 9-12, 2012. Article (CrossRef Link)

[28] M. Miao, J. Sun and S. Shao, "A cross-layer relay selection algorithm for D2D communication system," in *Proc. of International Conference on Wireless Communication and Sensor Network (WCSN)*, pp. 448-452, Dec. 13-14, 2014. Article (CrossRef Link)

[29] Y. Ni, S. Jin, K.-K. Wong and et al, "Outage performances for device-to-device communication assisted by two-way amplify-and-forward relay protocol," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 502-507, April 6-9, 2014. Article (CrossRef Link)

[30] Z. Lin, Y. Li, S. Wen and et al, "Stochastic geometry analysis of achievable transmission capacity for relay-assisted device-to-device networks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 2251-2256, June 10-14, 2014. Article (CrossRef Link)

[31] J. M. B. da Silva Jr, G. Fodor and T. F. Maciel, "Performance analysis of network-assisted two-hop D2D communications," in *Proc. of IEEE Globecom Workshop*, pp. 1050-1056, Dec. 8-12, 2014. Article (CrossRef Link)

[32] M. Jo, T. Maksymyuk, B. Strykhalyuk and C.-H. Cho, "Device-to-Device based heterogeneous radio access network architecture for mobile cloud computing," *IEEE Wireless Communications Magazine*, vol. 22, no. 3, pp. 50-58, June, 2015. Article (CrossRef Link)

[33] H. A. Chae, J. Gu, B.-G. Choi and M. Y. Chung, "Radio resource allocation scheme for device-to-device communication in cellular networks using fractional frequency reuse," in *Proc. of 17th Asia-Pacific Conference on Communications (APCC)*, pp. 58-62, Oct. 2-5, 2011. Article (CrossRef Link)

[34] M. Hasan, E. Hossain and D. I. Kim, "Resource allocation under channel uncertainties for relay-aided device-to-device communication underlaying LTE-A cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 2322-2338, April, 2014. Article (CrossRef Link)

[35] S.-Y. Kim, C.-H. Lim and C.-H. Cho, "Performance analysis of a dense device to device network," *KSII Transactions on Internet and Information Systems*, vol. 8, no. 9, pp. 2967-2981, Sep., 2014. Article (CrossRef Link)

[36] R. Tang, J. Dong, Z. Zhu and et al., "Resource allocation for underlaid Device-to-Device communication by incorporating both channel assignment and power control," in *Proc. of IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 432-436, April 4-6, 2015. Article (CrossRef Link)

[37] A. Gjendemsjø, D. Gesbert, G. Øien and S. Kiani, "Optimal power allocation and scheduling for two-cell capacity maximization," in *Proc. of the 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pp. 1-6, Feb. 26-Mar. 2, 2006. Article (CrossRef Link)

[38] S. Sakai, M. Togasaki and K. Yamazaki, "A note on greedy algorithms for the maximum weighted independent set problem," *Elsevier Discrete Applied Mathematics*, vol. 126, no. 2-3, pp. 312-322, Mar., 2003. Article (CrossRef Link)
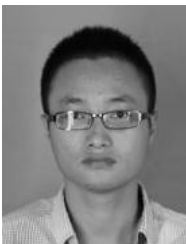
Rui Tang received his B.E. degree in Information and Communication Engineering from Xi'an Jiaotong University, Xi'an, Shaanxi, P. R. China, in June 2011. He is currently pursuing a doctoral degree at the School of Electronics and Information Engineering in Xi'an Jiaotong University. His current research interests include the design of radio resource management in device-to-device communication and heterogeneous networks, and the study of future radio communications systems, i.e., 5G. He currently also serves as a reviewer for a few magazines including *Wireless Personal Communications* and *China Communications*.

Jihong Zhao is a professor at both Xi'an Jiaotong University and Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi, P. R. China. She received her B.E. degree from Huazhong University of Science and Technology, Wuhan, Hubei, P. R. China, and a PhD degree from Xi'an Jiaotong University, Xi'an, Shaanxi, P. R. China. Her current research interests include broadband communications technologies, network protocol design and control strategies for supporting emerging applications in next generation communication systems, and network function virtualization and software defined networking.

Hua Qu is a professor at Xi'an Jiaotong University, Xi'an, Shaanxi, P. R. China. He received his B.E. degree from Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, P. R. China, and a PhD degree from Xi'an Jiaotong University, Xi'an, Shaanxi, P. R. China. His current research interests include IP-based network, mobile Internet, network protocol design and control strategies for supporting emerging applications in ubiquitous networks, and radio resource management in future cellular networks. He is a senior member of the China Institute of Communications and serves as an expert committee member for *China Communications* Magazine and *China Telecommunications Science* Magazine.

Zhengcang Zhu received his B.E. degree in Information and Communication Engineering from Anhui University of Technology, Ma'anshan, Anhui, P. R. China, in June 2013. He is currently pursuing a master's degree at the School of Electronics and Information Engineering in Xi'an Jiaotong University. His current research interests include the implementation of device-to-device communication in network offloading and future mmWave transmissions.

Yanpeng Zhang received his B.E. degree in Software Engineering from Xi'an Jiaotong University, Xi'an, Shaanxi, P. R. China, in June 2013. He is currently pursuing a doctoral degree at the School of Software Engineering in Xi'an Jiaotong University. His current research interests include machine learning and its applications in future radio communications systems, e.g., cross-layer design, handoff and load balancing in multi-tier heterogeneous networks.