Max–Min Fairness Optimization for D2D Communications Coexisting with Cellular Networks

Hoai Giang Nguyen 1, Xuan Tung Nguyen 2, Van Son Nguyen 1, Trinh Van Chien 2, Tien Hoa Nguyen 2,* and Soonghwan Ro 3

1 Faculty of Electronics Telecommunication & Information Technology, Hanoi Open University, Hanoi 100000, Vietnam; giangnh@hou.edu.vn (H.G.N.); sonnv.feit@gmail.com (V.S.N.)
2 School of Electronics and Telecommunications, Hanoi University of Science and Technology, Hanoi 100000, Vietnam; tung.nguyenvuan1310@gmail.com (X.T.N.); trinh.van.chien@liu.se (T.V.C.)
3 Department of Information and Communication, Kongju National University, Cheonan 31080, Korea; rosh@kongju.ac.kr
* Correspondence: hoa.nguyentien@hust.edu.vn

Received: 30 July 2020; Accepted: 29 August 2020; Published: 2 September 2020

Abstract: This paper considers a system consisting of a nonorthogonal multiple access (NOMA)-based device-to-device (D2D) communication system within a cellular network, in which the time and frequency resources are shared by everyone. In particular, D2D groups that constitute different pairs are assigned to the subchannels that the cellular users occupy. A max–min fairness optimization problem with power budget constraints is formulated and solved in this paper to reduce the mutual interference between the cellular users and D2D devices that substantially impacts that with the worst channel condition. For a low computational complexity solution, we propose the use of the bisection method together with the solution of a system of linear equalities. The proposed algorithm can provide uniformly good service to all of the cellular users and D2D devices in the coverage area by utilizing the minimal total transmit power. The simulation results indicate the effectiveness of the proposed algorithm in the improvement of the spectral efficiency of the worst user under the different widely used subchannel assignments and pairing techniques.

Keywords: NOMA; D2D communications; max-min fairness optimization; linear programming

1. Introduction

The 5G-and-beyond networks must cope with the drastic increases in mobile data traffic and the number of wireless devices from different services foreseen in the near future [1,2]. When considering next-generation wireless access technologies, a major concern is achieving a thousandfold increase in the spectral efficiency as compared with that of contemporary wireless networks while providing new applications with uniform coverage and high reliability [3,4]. This is indeed a vital goal that is considered by both academia and industry. Specifically, in recent years, wireless local area networks have become increasingly popular, requiring massive connections not only between many devices and a base station, but also between devices, which is called device-to-device communication (D2D) [5]. Spectrum resources that are allocated for cellular users are limited and require more local facilities. As a solution, incorporating D2D links into an existing cellular infrastructure can potentially enhance the spectrum and energy efficiency, extend the network coverage, and improve the communication reliability [6]. Spectrum sharing and frequency selection have attracted much research attention, for example, [7] and the references therein.

For better performance in future wireless networks, various techniques have been integrated into a D2D system, such as cell-free massive multiple-input multiple-output (MIMO) [8,9],...
intelligent reflection surface (IRS) [10], nonorthogonal multiple access (NOMA) [11], and mmWave communications [12]. These breakthrough technologies facilitate multiple access to keep pace with the high data demands of multimedia services with low latency [13]. Because of the excessive potentials, many of the initiatives have progressed with work on those technologies and they have achieved significant improvements in the spectral efficiency.

Among them, NOMA has recently received considerable attention as a promising multiple access candidate through the use of superposition coding and successive interference cancellation [14]. It is defined as NOMA in the power domain, and the network can serve multiple users with the same time and frequency resources as current systems. Alternatively, this technology can solve problems that are related to limitations in bandwidth, where a wireless network serves a number of users that exploit the frequency resource at the same time while requiring no additional physical facilities. NOMA is capable of upgrading the total channel capacity, especially when the number of user requests for service exceeds the dimension of each coherence interval either under perfect or imperfect channel state information (CSI) [15,16]. In order to overcome this issue, the subset of users is divided into pairs, for which the users with a better channel quality exploit the successive interference cancellation technique to mitigate the mutual interference that originates from the messages intended for the other users before decoding their desired signals. Such a activity results in much better the capacity of users with strong channel conditions [17]. In addition, NOMA concepts are considered in conjunction with other techniques, such as MIMO or mmWave [4,18]. Many previous works have investigated the potential contributions of NOMA technology for 5G-and-beyond wireless networks [19–21].

Power control is of paramount importance in managing the power budget and upgrading the system performance [22,23]. It has been shown in [11] that carefully allocating power to each user improves the channel capacity based on perfect CSI. Apart from this, the authors in [18] showed superior improvements in the sum spectral efficiency under different linear beamforming techniques. In addition, a power efficient resource allocation scheme subject to the quality of the service constraints was proposed in [24]. However, in the case of multiple access, the optimization problem is infeasible when at least one user does not satisfy the quality of the service requirement due to an empty feasible set. The authors in [25] proposed the joint allocation of subchannels and powers to maximize the total rate of the D2D-NOMA system. A similar study was expanded in [26], where the joint base station-user association and power allocation are used to improve the quality for an user with the worst channel condition. One thing in common between these studies is the authors sequentially modeled the non-convex power allocation to a convex problem for a homogeneous system.

In contrast to the previous studies, this paper considers heterogeneous users with different characteristics. To the best of our knowledge, there are no related works focusing on the max–min fairness optimization problem for a heterogeneous network where both the cellular users and D2D devices are active. We investigate a NOMA-based D2D communication system, in which both cellular users and D2D devices are at arbitrary locations in the coverage area, and they have their own transmit power level [11,27,28]. Each D2D group includes a number of devices that are paired, so that the NOMA technique can be used to mitigate co-subchannel interference. The network targets a provision of uniform quality of service (QoS) to every cellular user and D2D device, regardless the channel quality. Our study provides a comprehensive optimization framework, instead of introducing fixed QoS thresholds. Particularly, we focus on the max–min fairness optimization problem. The proposed optimization framework has been proven by using and evaluating in the same models given in [11,29]. It is worth to noting that these models only optimize the total transmit power subjective the required QoSs for both cellular users and D2D devices.

When compared with previously published works, our main contributions are summarized, as follows:

- We formulate a max-min fairness optimization problem, which aims at providing a uniformly good QoS to everyone in the network where both cellular users and D2D users coexist, subject to the limited power budgets. In particular, a number of NOMA-based D2D device groups are allowed
to reuse the bandwidth from that dedicated to the cellular users, which is controlled by a base station (BS). For a given QoS level, we then transfer the max–min fairness optimization problem into a minimum total transmit power consumption problem under the QoS requirements and the power constraints.

- We observe the quasilinear convexity of the proposed max-min fairness optimization problem and utilize the bisection method to obtain the global optimum. We rigorously find an upper bound on the QoSs that the system can provide to the cellular users and D2D devices. This upper bound offers a reduction of the computational complexity and also confirms that the proposed algorithm can be implemented in polynomial time. Our proposed algorithm balances between maximizing the worst spectral efficiency performance and minimizing the total transmit power consumption.

- The numerical results demonstrate the effectiveness of the proposed optimization framework in providing equally good QoS to all cellular users and D2D devices under the different subchannel assignments (either random assignment or a heuristic assignment that minimizes the mutual interference between cellular users and D2D devices), pairing techniques (either random paring or a heuristic pairing that reduces the total transmit power), and channel conditions.

The rest of this paper is organized, as follows: Section 2 presents the considered network model and the spectral efficiency analysis for each user. Section 3 formulates and solves the max–min QoS optimization problem subject to the power constraints. Finally, the numerical results are analyzed in Section 4, while the main conclusions are given in Section 5.

Notations: we use the upper and lower bold letters to denote matrices and vectors, respectively. The circularly symmetric Gaussian distribution is $\mathcal{CN}(\cdot, \cdot)$, while $E\{\cdot\}$ is the expectation of a random variable. The Hermitian transpose is denoted $(\cdot)^H$ and $\mathbb{C}$ is the complex field. Finally, $I_M$ is the identity matrix of size $M \times M$.

2. System Model and Performance Analysis

This section presents the considered system model and the computation of the performance for each user under uncorrelated Rayleigh fading channels.

2.1. System Model and Assumptions

We consider a system where a number of downlink communication D2D groups operate within the uplink cellular network, as schematically illustrated in Figure 1. The base station is equipped with $M$ antennas, managing $L$ subchannels and serving $M_l$ single-antenna cellular users on each subchannel. The system admits $K$ D2D groups in the same geographical area to reuse the spectrum resources. All of the D2D devices, i.e., D2D transmitters and D2D receivers, are equipped with a single antenna. In each D2D group, the devices are assigned in pairs and they utilize successive interference cancellation in the scheme of NOMA.

The cellular users are assumed to be randomly distributed in a disc, with radius $r_C$. Similarly, the D2D devices communicate in an area with radius $r_D$. In the uplink scenario of the cellular network, the channel between user $m$ using subchannel $l$ and the base station is denoted by the vector $\mathbf{h}_{l,m} \in \mathbb{C}^M$. The channel between the $k$-th D2D transmitter and the BS on the $l$-th is $\mathbf{g}_{l,k} \in \mathbb{C}^M$. At the downlink D2D communication, the channels between the D2D transmitter and the two receivers in the $n$-th pair are denoted by $\mathbf{g}_{l,k,n,1}, \mathbf{g}_{l,k,n,2} \in \mathbb{C}$, respectively. In addition, the channels between the cellular user $m$ using subchannel $l$ for the two users in the $n$-th pair are $h_{l,m,k,n,1} \in \mathbb{C}$ and $h_{l,m,k,n,2} \in \mathbb{C}$.
2.2. Channel Model

In this network, all of the propagation channels follow uncorrelated Rayleigh fading that is suitable for non-line-of-sight (NLoS) environments with full scatterers [30]. Mathematically, these propagation channels are distributed by:

\[ h_{l,m} \sim \mathcal{CN}(0, \beta_{l,m} I_M), \forall m, l, \]  

\[ g_{l,k} \sim \mathcal{CN}(0, \beta_{l,k} I_M), \forall l, k, \]  

\[ h_{l,m,k,n,1} \sim \mathcal{CN}(0, \beta_{l,m,k,n,1}), \forall l, m, k, n, \]  

\[ h_{l,m,k,n,2} \sim \mathcal{CN}(0, \beta_{l,m,k,n,2}), \forall l, m, k, n, \]  

\[ g_{l,k,n,1} \sim \mathcal{CN}(0, \beta_{l,k,n,1}), \forall l, k, n, \]  

\[ g_{l,k,n,2} \sim \mathcal{CN}(0, \beta_{l,k,n,2}), \forall l, k, n, \]  

where \( \beta_{l,m}, \beta_{l,k}, \beta_{l,m,k,n,1}, \beta_{l,m,k,n,2}, \beta_{l,k,n,1}, \) and \( \beta_{l,k,n,2} \) represent the large-scale fading coefficients modeling the path loss and shadow fading that, for instance, represent the attenuation of signals because of diffraction around large obstacles, such as high towers, and because of propagation over a long distance between the transmitter and receiver. In particular, they are respectively defined as:

\[ \beta_{l,m} = \nu_{l,m} 10^{\frac{\sigma l m}{10}}, \forall m, l, \]  

\[ \beta_{l,k} = \nu_{l,k} 10^{\frac{\sigma l k}{10}}, \forall l, k, \]  

\[ \beta_{l,m,k,n,1} = \nu_{l,m,k,n,1} 10^{\frac{\sigma l m k n,1}{10}}, \forall l, m, k, n, \]  

\[ \beta_{l,m,k,n,2} = \nu_{l,m,k,n,2} 10^{\frac{\sigma l m k n,2}{10}}, \forall l, m, k, n, \]  

\[ \beta_{l,k,n,1} = \nu_{l,k,n,1} 10^{\frac{\sigma l k n,1}{10}}, \forall l, k, n, \]  

\[ \beta_{l,k,n,2} = \nu_{l,k,n,2} 10^{\frac{\sigma l k n,2}{10}}, \forall l, k, n, \]  

where \( \sigma \) is the standard deviation of the shadow fading, while the random variables \( z_{l,m}, z_{l,k}, z_{l,m,k,n,1}, z_{l,m,k,n,2}, z_{l,k,n,1}, \) and \( z_{l,k,n,2} \) are distributed as \( \mathcal{CN}(0, 1) \). For the path loss coefficients \( \nu_{l,m}, \nu_{l,k}, \nu_{l,m,k,n,1}, \nu_{l,m,k,n,2}, \nu_{l,k,n,1}, \) and \( \nu_{l,k,n,2} \), we use the dual-slope model [31], which is given by

\[ \nu = \begin{cases} 
-47.85 - 20 \log 10(d), & \text{if } 0 < d \leq d_c(m), \\
-47.85 - 20 \log_{10}(d_c) - 40 \log_{10}(d/d_c), & \text{otherwise}, 
\end{cases} \]
where \( v \) (dBm) \( \in \{ v_{l,m}, v_{l,k}, v_{l,m,k,n,1}, v_{l,m,k,n,2}, v_{l,k,n,1}, v_{l,k,n,2} \} \) and \( d \) (measured in meters) represents the distance between the transmitter and receiver, i.e., \( d \in \{ d_{l,m}, d_{l,k}, d_{l,m,k,n,1}, d_{l,m,k,n,2}, d_{l,k,n,1}, d_{l,k,n,2} \} \). For an arbitrary communication link between the base station and a cellular user or a D2D device, the threshold \( d_e \) is set to 152 (m), while \( d_c = 25 \) (m) for a communication link between two D2D devices or that from a cellular user to any D2D device.

### 2.3. Channel Capacity of D2D Devices and Cellular Users

This subsection computes the uplink channel capacity for each cellular user, exploiting a detection technique to decode the desired signals. We also compute the downlink channel capacity for each D2D device under the exploitation of the successive interference cancellation technique.

#### 2.3.1. Uplink Cellular Network Transmission

The base station sends a message \( x_{l,m} \) to user \( m \) while using the \( l \)-th subchannel with \( \mathbb{E}\{|x_{l,m}|^2\} = 1 \). The received signal at the \( l \)-th subchannel is given by

\[
y_l = \sum_{m=1}^{M_l} \sqrt{P_{l,m}} h_{l,m} x_{l,m} + \sum_{k=1}^{K} \sum_{n_1=1}^{N_k} \eta_{l,k} b_{l,k} \left( \sqrt{P_{l,k,n,1}} t_{l,k,n,1} + \sqrt{P_{l,k,n,2}} t_{l,k,n,2} \right) + \xi_l, \tag{14}
\]

where \( P_{l,m} \) denotes the allocated power from the base station to user \( m \) using the \( l \)-th subchannel. At the \( n \)-th pair in the \( k \)-th group, the two complex signals \( t_{l,k,n,1} \) and \( t_{l,k,n,2} \) are transmitted from the transmitter with \( \mathbb{E}\{|t_{l,k,n,1}|^2\} = 1 \) and \( \mathbb{E}\{|t_{l,k,n,2}|^2\} = 1 \). The corresponding power levels \( P_{l,k,n,1} \) and \( P_{l,k,n,2} \) are allocated to these signals. \( \xi_l \) is additive noise, which is distributed as \( \xi_l \sim \mathcal{CN}(0, \sigma_{\xi_l}^2) \).

Here, \( \eta_{l,k} \) is a binary variable and it is defined as

\[
\eta_{l,k} = \begin{cases} 1, & \text{if group } k \text{ uses the } l \text{-th subchannel}, \\ 0, & \text{Otherwise}. \end{cases}
\tag{15}
\]

For the practical power constraints, the transmitted power levels from the base station, and the D2D transmitters satisfy

\[
0 \leq P_{l,m} \leq P_{\max,l,m}, \tag{16}
\]
\[
0 \leq P_{l,k,n,1} \leq P_{\max,l,k,n,1}, \tag{17}
\]
\[
0 \leq P_{l,k,n,2} \leq P_{\max,l,k,n,2}. \tag{18}
\]

To decode the desired signal from user \( m \) using subchannel \( l \), we assume that the base station utilizes a detection vector \( w_{l,m} \in \mathbb{C}^M \) as

\[
\hat{x}_{l,m} = w_{l,m}^H y_l = \sqrt{P_{l,m}} w_{l,m}^H h_{l,m} x_{l,m} + \sum_{m' \neq m}^{M_l} \sqrt{P_{l,m'}} w_{l,m'}^H h_{l,m'} x_{l,m'} + \sum_{k=1}^{K} \sum_{n_1=1}^{N_k} \eta_{l,k} w_{l,m}^H b_{l,k} \left( \sqrt{P_{l,k,n,1}} t_{l,k,n,1} + \sqrt{P_{l,k,n,2}} t_{l,k,n,2} \right) + w_{l,m}^H \xi_l, \tag{19}
\]

\[
\text{Desired Signal} \quad \text{Cellular Interference} \quad \text{D2D Interference}
\]

In Equation (19), the first term contains the desired signal and the second term denotes mutual interference from other cellular users. The last terms are mutual interference from the D2D devices and thermal noise. Therefore, the spectral efficiency of user \( m \) in the \( l \)-th subchannel is computed as

\[
R_{l,m} = \log_2 (1 + \gamma_{l,m}) \quad (\text{b/s/Hz}), \tag{20}
\]
where the signal-to-interference-and-noise (SINR) ratio, denoted by $\gamma_{l,m}$, is expressed as

$$\gamma_{l,m} = \frac{P_{l,m} |w_{l,m}^H h_{l,m}|^2}{\sum_{m'=1, m'=m}^{M_l} P_{l,m'} |w_{l,m'}^H h_{l,m'}|^2 + \sum_{k=1}^{K} \sum_{n=1}^{N_k} \eta_{l,k} |w_{l,m}^H g_{l,k}|^2 (P_{l,k,n,1} + P_{l,k,n,2}) + \|w_{l,m}\|^2 \sigma^2_{DL}}. \quad (21)$$

The spectral efficiency expression that is obtained in Equation (20) is generally applied to all of the channel models and detection techniques. In particular, the numerator in Equation (21) manifests the effectiveness of the utilized detection method by increasing the received power at the base station. Additionally, the denominator contains both the mutual interference and noise that degrades the spectral efficiency of any cellular user $m, \forall m$.

2.3.2. Downlink D2D Transmission

For D2D communications at the $n$-th pair in the $k$-th group, the device with smaller channel gain is called the weaker device. Without loss of generality, we state that in the $n$-th pair, device 1 has the weaker channel gain (let us call the weaker device), and device 2 has the stronger channel gain (let us call the stronger device), for which it is shown that

$$|g_{l,k,n,1}|^2 \leq |g_{l,k,n,2}|^2. \quad (22)$$

In each pair, we assume that the perfect successive interference cancellation only applies to the stronger device. This suggests that the weaker device employs a conventional decoding method. We then model the received signal for each of the two D2D devices while using the $l$-th subchannel. In particular, the received signal of device 1 in the $n$-th pair is formulated as

$$y_{l,k,n,1} = \sqrt{P_{l,k,n,1}} x_{l,m} + \sum_{m=1}^{M_l} \sqrt{P_{l,m}} h_{l,m,k,n,1} x_{l,m} + \sum_{k'=1, k' \neq k}^{K} \sum_{n=1}^{N_k'} \eta_{l,k'} g_{l,k',n,1} \left( \sqrt{P_{l,k',n,1}} t_{l,k',n,1} + \sqrt{P_{l,k',n,2}} t_{l,k',n,2} \right) + \xi_{l,k,n,1}, \quad (23)$$

where $\xi_{l,k,n,1} \sim CN(0, \sigma^2_{DL})$. In Equation (23), the first term denotes the desired signal and the second term refers to the intrapair interference. The interference from the cellular users and other D2D transmitter devices using the $l$-th subchannel are displayed by the third and fourth terms. The last term represents thermal noise. The spectral efficiency of this device is given as

$$R_{l,k,n,1} = \log_2 \left(1 + \gamma_{l,k,n,1}\right) \quad (b/s/Hz), \quad (24)$$

where the SINR value, denoted by $\gamma_{l,k,n,1}$, is

$$\gamma_{l,k,n,1} = \frac{P_{l,k,n,1} |g_{l,k,n,1}|^2}{\sum_{m=1}^{M_l} P_{l,m} |h_{l,m,k,n,1}|^2 + \sum_{k'=1, k' \neq k}^{K} \sum_{n=1}^{N_k'} \eta_{l,k'} |g_{l,k',n,1}|^2 (P_{l,k',n,1} + P_{l,k',n,2}) + \sigma^2_{DL}}. \quad (25)$$

In Equation (25), the numerator indicates the essential level of the power control and how much signal is depleted over the dissemination environment. In contrast, device 2 in the $n$-th pair utilizes perfect successive interference cancellation for subtracting intrapair interference; therefore, we obtain the spectral efficiency as

$$R_{l,k,n,2} = \log_2 \left(1 + \gamma_{l,k,n,2}\right) \quad (b/s/Hz), \quad (26)$$
where the SINR value of the device 2 of the $n$-th pair in the $k$-th group is computed as

$$
\gamma_{l,k,n,2} = \frac{P_{l,k,n,2} |g_{l,k,n,2}|^2}{\sum_{m=1}^{M} P_{l,m} |h_{l,m,k,n,2}|^2 + \sum_{k'=1,k'\neq k}^{K} \sum_{n=1}^{N'} \eta_{l,k',n} |g_{l,k',n,2}|^2 (P_{l,k',n,1} + P_{l,k',n,2}) + \sigma^2_{DL}}.
$$

Equation (27) illustrates that the intrapair interference from device 1 is completely mitigated using perfect successive interference cancellation. Hereby, we use these obtained spectral efficiencies to formulate and solve an optimization, which provides uniformly good service for every user in the system, regardless of the instantaneous CSI conditions.

3. Max–Min QoS Optimization

This section aims at maximizing the QoS level for all cellular users and D2D devices by utilizing the minimal total transmit power consumption in the condition of the radio channels that are impacted by a propagation environment. Indeed, the provision of a concrete level to every cell user and D2D device is hardly standardized. Therefore, we propose a method to determine the common good QoS level for both D2D devices and cellular users. In this model, we assume that all D2D groups have been assigned into specific subchannels by the diversity-based method in [11]; therefore, the set of $\eta_{l,k}$ has been defined. In the following, we consider the optimization problem of maximizing the lowest QoS level, corresponding to that with the worst channel condition in the network, as

$$
\begin{align*}
\text{maximize} & \quad \{ P_{l,k,n,1}, P_{l,k,n,2}, P_{l,m} \geq 0 \} \quad \{ \frac{R_{l,k,n,1}}{w_{l,k,n,1}}, \frac{R_{l,k,n,2}}{w_{l,k,n,2}}, \frac{R_{l,m}}{w_{l,m}} \} \\
\text{min} & \quad \{ \frac{R_{l,k,n,1}}{w_{l,k,n,1}}, \frac{R_{l,k,n,2}}{w_{l,k,n,2}}, \frac{R_{l,m}}{w_{l,m}} \} \\
\text{subject to} & \quad P_{l,k,n,1} \leq P_{\text{max},l,k,n,1} \forall l, k, n, \\
& \quad P_{l,k,n,2} \leq P_{\text{max},l,k,n,2} \forall l, k, n, \\
& \quad P_{l,m} \leq P_{\text{max},l,m} \forall l, m,
\end{align*}
$$

where $w_{l,k,n,1}, w_{l,k,n,2}, w_{l,m} > 0$ are the weights that are associated with each corresponding user and they demonstrate the users’ priority. By introducing a new optimization $\xi$ as a consequence of utilizing the epigraph representation [32,33], we convert problem (28) into the equivalent form of

$$
\begin{align*}
\text{maximize} & \quad \xi \\
\text{subject to} & \quad R_{l,k,n,1} \geq \xi w_{l,k,n,1} \forall l, k, n, \\
& \quad R_{l,k,n,2} \geq \xi w_{l,k,n,2} \forall l, k, n, \\
& \quad R_{l,m} \geq \xi w_{l,m} \forall l, m, \\
& \quad P_{l,k,n,1} \leq P_{\text{max},l,k,n,1} \forall l, k, n, \\
& \quad P_{l,k,n,2} \leq P_{\text{max},l,k,n,2} \forall l, k, n, \\
& \quad P_{l,m} \leq P_{\text{max},l,m} \forall l, m.
\end{align*}
$$

Here, $\xi$ is the minimum QoS parameter for all users and devices, which we aim to maximize in order to provide a uniformly good service to all users regardless of the channel conditions. For proceeding further, we set up $\xi_{l,k,n,1} = 2^{\xi_{l,k,n,1}} - 1$, $\xi_{l,k,n,2} = 2^{\xi_{l,k,n,2}} - 1$, and $\xi_{l,m} = 2^{\xi_{l,m}} - 1$, respectively.
and reformulate the optimization problem subject to the QoS requirements in Equation (29) to the corresponding one with the SINR constraints, as

$$\begin{align*}
\text{maximize} & \quad \xi \\
\text{subject to} & \quad \gamma_{l,k,n,1} \geq \xi_{l,k,n,1}, \forall k, n, \\
& \quad \gamma_{l,k,n,2} \geq \xi_{l,k,n,2}, \forall k, n, \\
& \quad \gamma_{l,m} \geq \xi_{l,m}, \forall l, m, \\
& \quad P_{l,k,n,1} \leq P_{\max,l,k,n,1} \forall k, n, \\
& \quad P_{l,k,n,2} \leq P_{\max,l,k,n,2} \forall k, n, \\
& \quad P_{l,m} \leq P_{\max,l,m} \forall l, m.
\end{align*}$$

(30)

We stress that jointly optimizing the power coefficients and the max–min SINR level makes problem (30) a geometric program, since the SINR constraints can be reformulated as posynomials, while the power budget constraints are monomials. Because a geometric program is nonconvex in standard form, the global optimum to problem (30) is obtained by exploiting a hidden convex structure, but this structure has high computational complexity. Consequently, as a main contribution of this paper, we propose an algorithm to solve Equation (30) with lower cost. For this purpose, the following observations are attained for a given $\xi$, as in Lemma 1.

**Lemma 1.** For a given nonnegative value of $\xi$, problem (30) is equivalent to

$$\begin{align*}
\text{maximize} & \quad 0 \\
\text{subject to} & \quad \gamma_{l,k,n,1} \geq \xi_{l,k,n,1}, \forall l, k, n, \\
& \quad \gamma_{l,k,n,2} \geq \xi_{l,k,n,2}, \forall l, k, n, \\
& \quad \gamma_{l,m} \geq \xi_{l,m}, \forall l, m, \\
& \quad P_{l,k,n,1} \leq P_{\max,l,k,n,1} \forall l, k, n, \\
& \quad P_{l,k,n,2} \leq P_{\max,l,k,n,2} \forall l, k, n, \\
& \quad P_{l,m} \leq P_{\max,l,m} \forall l, m,
\end{align*}$$

(31)

which is a linear program and, therefore, the global optimum can be obtained in polynomial time.

**Proof.** The main proof is to verify that problem (31) is a linear program and is, therefore, convex. Indeed, the objective function is a constant, which aligns with a special case of a linear function. Additionally, the power budget constraints are affine functions. By utilizing the expressions in Equation (21), we now reformulate the SINR constraints as

$$\begin{align*}
\sum_{m' = 1, m' \neq m}^{M_I} & \xi_{l,m} P_{l,m'} |\mathbf{w}_{l,m'}^H \mathbf{h}_{l,m'}|^2 + \sum_{k = 1}^{K} \sum_{n = 1}^{N_I} \xi_{l,m} \eta_k |\mathbf{w}_{l,m}^H \mathbf{g}_k|^2 (P_{l,k,n,1} + P_{l,k,n,2}) + \xi_{l,m} |\mathbf{w}_{l,m}|^2 \sigma_{DL}^2
\\
- P_{l,m} |\mathbf{w}_{l,m}^H \mathbf{h}_{l,m}|^2 & \leq 0, \\
\xi_{l,k,n,1} P_{l,k,n,2} |g_{l,k,n,1}|^2 + \sum_{m = 1}^{M_I} \xi_{l,k,n,2} P_{l,m} |h_{l,m,k,n,2}|^2 + \sum_{k' = 1}^{K} \sum_{k' \neq k}^{K} \sum_{n = 1}^{N_I} \xi_{l,k,n,1} \eta_{k,k',n} |g_{l,k',n,1}|^2 (P_{l,k',n,1} + P_{l,k',n,2})
\\
+ \xi_{l,k,n,2} P_{l,k,n,1} |g_{l,k,n,1}|^2 & \leq 0, \\
\xi_{l,k,n,2} \sum_{m = 1}^{M_I} P_{l,m} |h_{l,m,k,n,2}|^2 + \sum_{k' = 1}^{K} \sum_{k' \neq k}^{K} \xi_{l,k,n,2} \eta_{k,k',n} |g_{l,k',n,2}|^2 (P_{l,k',n,1} + P_{l,k',n,2}) + \xi_{l,k,n,2} \sigma_{DL}^2
\\
- P_{l,k,n,2} |g_{l,k,n,2}|^2 & \leq 0,
\end{align*}$$

(32)

which are affine constraints. Therefore, problem (31) is a linear program on a standard form, and its global optimum can be obtained, as demonstrated in the lemma. \qed
We have proven that the global solution to problem (31) can be obtained for a fixed $\xi$ by solving a linear program. However, this problem does not minimize the total transmit power consumption, which is a goal of beyond 5G networks for improving energy efficiency. Therefore, we recast Equation (31) into a new optimization problem that can also provide an energy-efficient solution while guaranteeing the SINR requirements, as:

$$\begin{align*}
\text{minimize} & \quad \sum_{k=1}^{K} \sum_{n=1}^{N_k} (P_{l,k,n,1} + P_{l,k,n,2}) + \sum_{l=1}^{L} \sum_{m=1}^{M_l} P_{l,m} \\
\text{subject to} & \quad \gamma_{l,m} \geq \xi_{l,m}, \forall l, m, \\
& \quad \gamma_{l,k,n,1} \geq \xi_{l,k,n,1}, \forall l, k, n, \\
& \quad \gamma_{l,k,n,2} \geq \xi_{l,k,n,2}, \forall l, k, n, \\
& \quad P_{l,k,n,1} \leq P_{\text{max},l,k,n,1} \forall l, k, n, \\
& \quad P_{l,k,n,2} \leq P_{\text{max},l,k,n,2} \forall l, k, n, \\
& \quad P_{l,m} \leq P_{\text{max},l,m} \forall l, m.
\end{align*}$$

Following a methodology that is similar to that in Lemma 1, problem (33) is proven to be a linear program; therefore, the global optimum can be obtained by utilizing a general optimization toolbox with the interior-point method [34]. From the constraints about the SINR requirements of all users and devices, we observe that the solution of all SINR constraints when they are equal is the solution of Equation (33) into a new optimization problem that can also provide an energy-efficient solution.

$$\begin{align*}
\sum_{l=1}^{L} M_m \text{ SINR constraints of the cellular users can be rewritten, as:}
\end{align*}$$

$$\begin{align*}
P_{l,m} |w_{l,m}^H h_{l,m}|^2 - \sum_{m' = 1, m' \neq m}^{M_l} \xi_{l,m'} P_{l,m} |w_{l,m'}^H h_{l,m'}|^2 - \sum_{k=1}^{K} \sum_{n=1}^{N_k} \xi_{l,m} \eta_{l,k}|w_{l,m}^H g_{l,k,n}|^2 (P_{l,k,n,1} + P_{l,k,n,2}) \\
- \xi_{l,m} ||w_{l,m}||_2^2 \sigma_{UL}^2 = 0, \forall m,
\end{align*}$$

which contain the $\sum_{l=1}^{L} M_m$ power variables that are associated with all of the cellular users. Similarly, the $2K \sum_{k=1}^{K} N_k$ SINR constraints of D2D devices are reformulated as:

$$\begin{align*}
P_{l,k,n,1} |g_{l,k,n,1}|^2 - \xi_{k,n,1} P_{l,k,n,2} |g_{l,k,n,2}|^2 - \sum_{m=1}^{M_l} \xi_{l,m} P_{l,m} |h_{l,m,k,n,1}|^2 - \sum_{k' = 1, k' \neq k}^{K} \sum_{n=1}^{N_k} \xi_{l,m} \eta_{l,k'}|g_{l,k',n,1}|^2 (P_{l,k',n,1} + P_{l,k',n,2}) \\
- \xi_{k,n,2} \sigma_{DL}^2 = 0, \forall k, n,
\end{align*}$$

which contain the power variables for the D2D devices. With a nonempty feasible set, we obtain the optimal solution to problem (33) by solving the system of linear equations in (34) and (35). Otherwise, problem (33) is infeasible if the solution contains one power coefficient that does not satisfy the power budget constraints. In addition, the solution has been obtained for a given set of $\{\xi_{k,n,1}, \xi_{k,n,2}, \xi_{l,m}\}$. To find the optimal max–min SINR values for all cellular users and D2D devices, we propose the use of the bisection method with the search range that is defined in Theorem 1.

**Theorem 1.** The optimal value of Equation (29) can be obtained by checking the feasibility of Equation (33) over the SINR search range $\mathcal{S} = [0, \xi_{0}^{\text{upper}}]$, where $\xi_{0}^{\text{upper}}$ is selected to make Equation (33) infeasible. One option to calculate $\xi_{0}^{\text{upper}}$ is defined as:

$$\xi_{0}^{\text{upper}} = \min_{(l,k,n,1), (l,k,n,2), (l,m)} \left\{ \frac{P_{l,k,n,1}}{P_{l,k,n,2}} \frac{|g_{l,k,n,1}|^2}{\sigma_{DL}^2}, \frac{P_{l,m} \| h_{l,m} \|^2}{\sigma_{UL}^2} \right\}.$$
Proof. The main proof is to compute the upper bound stated in Equation (36). Particularly, for an arbitrary cellular user \( m \) utilizing the subchannel \( l \), the SINR value is upper bounded as

\[
\gamma_{l,m} \leq \frac{P_{l,m} |w_{l,m}^H h_{l,m}|^2}{\|w_{l,m}\|_2^2 C_{UL}} \leq \frac{P_{l,m} \|w_{l,m}\|^2 \|h_{l,m}\|^2}{\|w_{l,m}\|_2^2 C_{UL}} = \frac{P_{l,m} \|h_{l,m}\|^2}{C_{UL}},
\]

where \((a)\) is obtained due to the inherent interference and thermal noise in a multi-user system. \((b)\) is obtained by the Cauchy–Schwarz inequality, i.e.,

\[
|w_{l,m}^H h_{l,m}|^2 \leq \|w_{l,m}\|^2 \|h_{l,m}\|^2.
\]

With the optimistic assumption that there is no thermal noise or mutual interference from the cellular users or from the other D2D devices, the upper bound of the SINR value of the weak D2D device in pair \( k \)

\[
\gamma_{l,k,n,1} \leq \frac{P_{l,k,n,1} |g_{l,k,n,1}|^2}{P_{l,k,n,2} |g_{l,k,n,1}|^2} = \frac{P_{l,k,n,1}}{P_{l,k,n,2}}.
\]

For the strong D2D device in pair \( k \), we ignore all the mutual interference to obtain an upper bound on the SINR value of

\[
\gamma_{l,k,n,2} \leq \frac{P_{l,k,n,2} |g_{l,k,n,2}|^2}{C_{DL}^2}.
\]

Combining Equations (37)–(40) and obtaining an upper bound that is applied for all users, we obtain the result that is shown in the theorem.

From Theorem 1, problem (33) is solved in an iterative manner by reducing the search range, and then the maximum QoS level can be obtained. One such line search procedure is the well-known bisection method [35], which we exploit in this paper. For more detail, we define two parameters, \( \xi_{\text{candidate}} \) and \( \delta \). While \( \xi_{\text{candidate}} \) represents the selected QoS value at each iteration search, \( \delta \) is the line-search accuracy value. After determining the upper and lower bounds, \( \xi_{\min}, \xi_{\max} \), of the SINR requirement, the feasibility of Equation (33) is demonstrated for the value

\[
\xi_{\text{candidate}} = (\xi_{\max} + \xi_{\min})/2.
\]

If Equation (33) is feasible, then its solution to the power variables is saved in \( P_{\text{save}} \), \( \xi_{\min} \) is changed to the middle value \( \xi_{\text{candidate}} \), and the value \( \Delta \), which is the gap between the upper and lower bound of \( \xi \), is compared with \( \delta \). The algorithm is terminated when the gap \( \Delta \) is smaller than the line-search accuracy \( \delta \). Otherwise, if problem (33) is infeasible, \( \xi_{\max} \) is labeled by \( \xi_{\text{candidate}} \). The search range declines by half after each iteration, since either its lower or upper bound is assigned to \( \xi_{\text{candidate}} \). Algorithm 1 summarizes the proposed max–min QoS optimization. In each iteration, the main cost of Algorithm 1 spends on solving problem (33) or alternatively solving a system of linear equations in (34) and (35). This is dominated my a matrix inverse computation. By exploiting the Cholesky decomposition [36], the computational complexity of matrix inversion is in the order of

\[
\mathcal{O} \left( \left( \sum_{l=1}^{L} M_m + 2 \sum_{k=1}^{K} N_k \right)^3 \right).
\]

Moreover, the number of iteration in need for the bisection method is \( \lceil \log_2 (\xi_{\text{upper}}/\delta) \rceil \), where \( \lceil \cdot \rceil \) is the ceiling function. As a consequence, the total computational complexity of Algorithm 1 is

\[
\mathcal{O} \left( \left( \sum_{l=1}^{L} M_m + 2 \sum_{k=1}^{K} N_k \right)^3 \right).
\]
which can be performed in the polynomial time and also demonstrate the effectiveness of selecting the initial value $\xi^{\text{upper}}_0$ as shown in Theorem 1.

Algorithm 1: Max–min quality of service (QoS) based on the bisection method.

1. **Result:** solve the optimization in (33).
2. **Input:** Initial upper bound $\xi^{\text{max}}_0$ and line-search accuracy $\delta$;
3. Set $\xi^{\text{min}} = 0, \xi^{\text{max}} = \xi^{\text{max}}_0$;
4. Calculate: $\Delta = \xi^{\text{max}} - \xi^{\text{min}}$
5. while $\Delta > \delta$ do
6. Set $\xi^{\text{candidate}} = (\xi^{\text{max}} + \xi^{\text{min}})/2$;
7. if (33) is infeasible for $\hat{c} = w_k \xi^{\text{candidate}}$, $\forall k$, then
8. Set $\xi^{\text{max}} = \xi^{\text{candidate}}$;
9. Update $\Delta$;
10. Set $\xi^{\text{min}} = \xi^{\text{candidate}}$;
11. else
12. Update $\Delta$
13. Set $\xi^{\text{min}}_{\text{final}} = \xi^{\text{min}}$ and $\xi^{\text{max}}_{\text{final}} = \xi^{\text{max}}$.

Output: Final interval $[\xi^{\text{min}}_{\text{final}}, \xi^{\text{max}}_{\text{final}}]$ and $\{P_{\text{save}}\}$.

4. Numerical Results

In this section, we analyze the system performance by utilizing the bisection method in order to maximize the worst spectral efficiency in the network subject for the given power budgets. To clarify the effectiveness of the proposed approach, we compare the four scenarios of the subchannel assignment and D2D device pairing that have been proposed in [11], consisting of:

1. **The diversity-based method (DBM) combined with heuristic pairing (denoted DBM + Heuristic Pairing in the figures):** All cellular users have already been assigned to the subchannels. Subsequently, these subchannels are allocated to the D2D devices via channel diversity. In each D2D group, the devices are paired by the heuristic algorithm.
2. **The DBM using random pairing (denoted DBM + Random Pairing in the figures):** the system utilizes the heuristic algorithm to allocate the subchannels for the D2D groups; however, the devices in each group are paired randomly.
3. **The random channel assignment using heuristic pairing (denoted NonDBM + Heuristic Pairing in the figures):** the D2D groups are assigned randomly to the subchannels, and the devices in each group are paired by the heuristic algorithm.
4. **The random channel assignment using heuristic pairing (denoted NonDBM + Random Pairing in the figures):** the D2D groups are assigned randomly into subchannels, and the devices in each group are paired randomly.

The general parameters of the system setting and propagation channels are listed in Table 1, while several specific parameters are presented in each figure.

Figure 2 depicts the relationship between the number of D2D groups and the max–min fairness QoS (bit/s/Hz) of each cellular user and D2D device. Generally, when the number of D2D groups in the system increases, it leads to an increase in interference, and thus, the achievable rate obviously declines as the number of groups increases. However, the system utilizing the bisection method still demonstrates the good performance of the heuristic algorithm. In particular, the highest QoS rate of all users and devices is obtained when the system only allows one D2D group. In the case where the
heuristic algorithm is used to pair the D2D devices, the max-min QoS is approximately 3.8 bit/s/Hz. Additionally, the QoS of the system that does not use the heuristic assignment algorithm and uses random pairing only reaches 3.4 bit/s/Hz. Subsequently, all four scenarios show downward trends when the number of D2D groups is increased due to the growth of mutual interference. When the system accepts 10 D2D groups reusing the subchannels, the QoS provided to everyone in the case where the heuristic algorithm and heuristic pairing are used still reaches 2.1 bit/s/Hz. This spectral efficiency is significantly higher than that of the system that uses random pairing only, with approximately 1.8 bit/s/Hz.

Table 1. Simulated network parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellular radius</td>
<td>250 (m)</td>
</tr>
<tr>
<td>Base station height</td>
<td>6 (m)</td>
</tr>
<tr>
<td>Beamforming technology</td>
<td>Maximum Ratio Combining (MRC)</td>
</tr>
<tr>
<td>D2D group radius</td>
<td>10 (m)</td>
</tr>
<tr>
<td>Device and cellular user height</td>
<td>1.5 (m)</td>
</tr>
<tr>
<td>Operating frequency</td>
<td>1.9 (GHz)</td>
</tr>
<tr>
<td>Maximum transmit power of devices and cellular users</td>
<td>23 (dBm)</td>
</tr>
<tr>
<td>Noise power</td>
<td>−96 (dBm)</td>
</tr>
</tbody>
</table>

Figure 2. The max-min fairness QoS (bit/s/Hz) versus the number of D2D groups.

Figure 3 manifests the efficiency of the bisection method for the system. The simulated scenario includes 5 cellular users, which have been assigned to the corresponding subchannels, and the 6 D2D groups are allowed to access the system and reuse the subchannels. With the heuristic algorithm, the system can provide up to 80% of all users and devices with a medium rate over 1.5 bit/s/Hz. In contrast, for the system without the heuristic algorithm, only 65% users reach the data rate of 1.5 bit/s/Hz. The effectiveness of the network that applies the bisection method and the heuristic algorithm for assigning the D2D devices yields better QoS than that with random pairing since approximately 18% users and devices can transmit signals at the pace of 3 bit/s/Hz, while that of the users without the heuristic algorithm is approximately 8%.
Figure 3. The cumulative distribution function of the max-min fairness QoS (bit/s/Hz).

Figure 4 compares the total transmit power of all the D2D devices as the number of groups increases. The total transmit power of the D2D devices is directly proportional to the number of groups. The figure clearly illustrates the efficiency of the network in four cases. First, in the system with the heuristic algorithm, the total transmit power of all the devices is approximately 7–20% lower when the proposed pairing is used. Particularly, the total transmit power with the heuristic algorithm and proposed pairing is 100 mW, while that with random pairing is 80 mW for the case where only one group accesses the network. Subsequently, the total transmit power values of both the DBM with the proposed pairing and random pairing increase substantially, especially when the number of groups exceeds the number of subchannels, up to 280 and 305 mW, respectively. Second, without the heuristic algorithm, the system reaches power consumption values as high as 190 mW when using random pairing and reaches the peak level of 560 mW with 10 groups accessing the network. The figure also shows that, when the number of groups exceeds the number of subchannels, the total transmit power increases more sharply. However, from Figure 2, the spectral efficiency of each user when the heuristic algorithm is used is still higher than that when this algorithm is not used. In a similar trend, the total transmit power of all the D2D devices when the heuristic algorithm is used is lower than that when this algorithm is not used.

In contrast with the trend of the power allocation for D2D devices, Figure 5 indicates that the total transmit power (measured in mW) of all the cellular users tends to decline over the number of groups allowed to reuse spectrum resources in all four studied schemes. The rationale is that the base station does not apply the NOMA technique and, thus, when the maximum spectral efficiency decreases, the total transmit power decreases as well. A comparison of the four scenarios shows that the bisection method perfectly conjugates with the model, since the system that uses the heuristic algorithm consumes more power than that without this algorithm and allows all of the users and devices to connect to the network at a higher data rate. Particularly, the total transmit power of all the cellular users with the heuristic algorithm and heuristic pairing uses 3900 mW to 2600 mW when the system serves 2 to 10 groups. However, the total transmit power of the heuristic algorithm and random pairing decreases to 3500 mW to 2300 mW.
Figure 4. The total transmit power of the device-to-device communication (D2D) devices versus the number of D2D groups.

Figure 5. The total transmit power of cellular users versus the number of D2D groups.

Figure 6 shows the feasible probability that computes the fraction of the number of cellular users/D2D devices who are served by the QoS not less than 1.5 bit/s/Hz and the total number of existing cellular users and D2D devices among the 2000 realizations of different locations. The fixed QoS was proposed in [11,29]. Those results demonstrate the significant improvements of the proposed max–min fairness optimization than a fixed QoS assignment, which provides the better feasible probability. For example, the utilization of max–min fairness for the DBM combined with heuristic pairing gives 78% of cellular users and D2D devices with a higher rate than 1.5 bit/s/Hz. This is 28% better than the corresponding method in [29] with the fixed QoS.
Figure 6. The feasible probability of the different benchmarks with the QoS 1.5 bit/s/Hz.

5. Conclusions

We investigated a NOMA-based D2D communication system coexisting with a cellular network, where the power resource needs to be carefully allocated due to the shortage in practice. We formulated a max–min fairness optimization problem with energy budget constraints to provide a good QoS to all devices and cellular users in the network while still limiting the total transmit power of the whole system. The bisection method combined with solving linear equalities is proposed in this model for the sake of reducing the computational complexity. Consequently, the numerical results demonstrate that all cellular users and devices acquire better spectral efficiency with the limited power budget.

Author Contributions: All authors contributed to this paper. Conceptualization, H.G.N. and methodology, X.T.N.; Software, V.S.N.; Validation, H.G.N.; Formal analysis and investigation, T.H.N. and T.V.C.; Resources, X.T.N.; Data curation S.R.; Writing—Original draft preparation, T.H.N. and H.G.N.; Writing—Review and editing, T.V.C. and S.R.; Funding acquisition, V.S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Vietnam’s Ministry of Education and Training (MOET) Project under Grant B2019-MHN-05.

Conflicts of Interest: The authors declare no conflict of interest.

References


© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).