Hindawi
Mobile Information Systems
Volume 2021, Article ID 7995090, 10 pages
https://doi.org/10.1155/2021/7995090

Research Article

Resource Allocation Scheme for Fog-Enabled Wireless Access Networks under the QoS of Users

Huilin Jiang,[1,2] **Lili Chen**,[1] **Xiang Song**,[1] **and Xueming Liu**[1]

[1]*School of Electronic Engineering, Nanjing Xiaozhuang University, Nanjing, China*
[2]*Jiangsu Key Construction Laboratory of IoT Application Technology, Wuxi, China*

Correspondence should be addressed to Huilin Jiang; huilin.jiang@njxzc.edu.cn

Received 13 September 2021; Revised 2 November 2021; Accepted 12 November 2021; Published 25 November 2021

Academic Editor: Fazlullah Khan

With the complexity of the network architecture, the diversity of network slicing, and the introduction of advanced techniques such as device to device (D2D), it is difficult for the next-generation (5G+ or 6G) networks to comprehensively consider the requirements of users from different slices and jointly allocate wireless resources to improve network energy efficiency. This paper studies the energy efficiency optimization problem for D2D-enabled fog radio access networks (FRANs). A resource allocation algorithm is proposed to maximize the network energy efficiency by jointly optimizing the beamforming vector, resource block allocation, and transmission power of the remote radio heads (RRHs), fog access point (FAP), and D2D users. The developed algorithm is based on nonlinear programming, convex optimization, and Lagrangian duality. Simulation results show that, by applying the proposed algorithm, the system throughput is significantly improved, and the network energy consumption is greatly reduced, which can ultimately improve the network energy efficiency obviously.

1. Introduction

In the context of the ever-increasing digitization and globalization of the society, people's demand for mobile communications is growing dramatically. With the large-scale deployment of 5G communication networks, mobile communication has entered the 5G+ era and has become popular throughout the society [1]. Due to high transmission rate and high reliability of mobile communications, the development space of mobile communications has expanded incredibly. Its advantages have become an important support for the highly developed software and even hardware industries and have become an important cornerstone for the rapid development of the society.

Of course, correspondingly, wireless communication is developing towards the next-generation network with higher speed and higher reliability. The focus of mobile communication development is shifting towards the development of state-of-the-art technologies such as massive multiple-input and multiple-output (massive MIMO), intelligent reflective surface (IRS), millimeter-wave and terahertz communication, advanced network architectures and scenarios such as fog radio access networks (FRANs) and coexistence of diversified transmission modes in communications, and advanced wireless resource allocation schemes related to beamforming, power allocation, and transmission mode selection [2]. Under the pressure of the increasing demand for wireless networks, cloud computing and fog computing have emerged. Compared with cloud computing, fog computing has several obvious characteristics: lower latency and more energy saving. Benefited from these advantages, fog-enabled network architecture, that is, FRAN with strong caching, communication, and computing capabilities, is widely adopted by 5G and the next-generation (6G) networks and believed to be the most promising network technology for the future networks [3]. In the FRAN, while providing high spectrum efficiency of the network, it is also necessary to ensure that the quality of service (QoS) for users requesting real-time service is above a certain level. The key role of wireless resource allocation is to further improve network performance and ensure the QoS. Therefore, in the FRAN, it is of great significance to adopt

resource allocation methods that can guarantee the users' QoS to further improve the network performance.

Many studies have been done on resource allocation in the FRAN in recent years. In [4], the inefficiency of the current contributions in computational offloading and resource allocation was pointed out, and a deep reinforcement learning-based resource allocation algorithm was proposed to minimize latency by joint optimization of mode selection, resource allocation, and power allocation. Ai et al. [5] focused on the network slicing-based FRAN and developed a joint resource allocation and admission control scheme to maximize the number of users in the hotspot slice that can be supported with the desired quality of service. Admission control and beamforming were performed in the hotspot slice, while subchannel and power allocation were performed in the Internet of Things (IoT) slice. In [6], sequential resource allocation of the fog nodes was discussed, where the access decision of the IoT users was made by reinforcement learning-based methods to improve the system utility in terms of latency. Although advanced schemes were proposed and resources other than wireless resources, such as computing resources, were used, most of the previous research studies focused on the latency and association number of the users, which did not take the network energy efficiency and throughput into consideration [7]. Higher energy efficiency is also one of the expected achievements for next-generation mobile systems such as beyond 5G and 6G [8]. To maximize network energy efficiency, Dinh et al. [9] proposed a joint optimization algorithm based on user association and beamforming as well as a heuristic low-complexity strategy under the constraint of local edge processing capability. The scenarios of global but outdated user channel state information (CSI) and perfect but local CSI were both considered. In [10], energy efficiency and cross-tier interference problem of the network was studied for the FRAN, where the user association, caching, and power allocation were jointly optimized based on an alternating direction method of multipliers. Energy efficiency, spectrum efficiency, and interference were jointly maximized in [11] by optimizing the content distribution from fog nodes and cloud to users.

In the FRAN, device-to-device (D2D) communication is a technology that allows direct communication between two fog users that are relatively close by multiplexing remote radio header (RRH) spectrum resources [12]. Benefiting from the multiplex gain, distance gain, and offloading gain of D2D communication, D2D technology can improve the spectrum efficiency and energy efficiency of the FRAN and, to a large extent, alleviate the problem of the transmission capacity limitation of the fronthaul link. Therefore, D2D scenario is an important case that the resource allocation should take into account, and D2D users should be included into the research of resource allocation-based energy efficiency optimization in the FRAN [13]. However, as far as we know, there is very little research in this area. This paper focuses on the resource allocation algorithm under the users' QoS requirements to maximize the network energy efficiency in the FRAN where RRH users, IoT users, and D2D users all exist. The data rate and delay are formulated as the constraints of users' QoS requirements, and the energy efficiency is finally improved by joint optimization of the transmission mode, beamforming vector, and the allocation of power and resource blocks. By a reasonable design of the joint optimization algorithm, the corresponding energy-efficient resource allocation algorithm for the coexistence of multiple modes is proposed. By building a MATLAB simulation platform, system simulation is performed to verify the performance of the proposed algorithm. A resource allocation method for the application of the FRAN architecture in current and future mobile communication networks to ensure real-time service requirements is designed. The research not only provides theoretical analysis and performance reference for the FRAN to better support delay-sensitive services but also has great significance for promoting the commercial deployment of the FRAN.

However, further paper is arranged in an order such as in Section 2, a brief introduction about the system model is given, and a network energy efficiency maximization problem for the sliced FRAN is formulated. To facilitate the resolution of the formulated problem, in Section 3, the original problem is transformed into 3 solvable subproblems. An energy-efficient joint optimization algorithm on beamforming, resource block allocation, and power control is proposed in Section 4. To prove the effectiveness of the algorithm, simulation configurations and the corresponding results are given in Section 5. Conclusion is presented in Section 6.

2. System Model

2.1. Network Scenario. D2D-enabled FRAN is studied in this paper. To meet the requirements of high-speed and low-latency users, respectively, as shown in the system model in Figure 1, both high-speed slicing and low-latency slicing are considered in the network scenario, and both high-speed slicing and low-latency slicing are realized by the network slicing technology based on NFV (network function virtualization) and SDN (software-defined network). Users in the network can be divided into RRH users, IoT users, and D2D users, which are served by RRH, FAP, and directly connected devices, respectively. Each user has only a single antenna, the total bandwidth that can be utilized is denoted as B, and OFDM (orthogonal frequency division multiplexing) technology is adopted by the system.

2.1.1. High-Speed Slice. In the high-speed slice, M RRHs which provide service to G RRH users with high-speed requirements are connected to the baseband unit (BBU) pool through capacity-limited downlink. The RRHs can participate in collaborative communication under high-speed slicing. The set of the RRHs is represented by $M = \{1, 2, m, \ldots, M\}$, and the number of antennas for each RRH is T. The set of RRH users in the high-speed slicing is denoted as $G = \{1, 2, g, \ldots, G\}$. Different from the IoT users in the low-latency slice, RRH users requesting high-speed services can choose to use the D2D function to provide the corresponding high-speed services. Once the D2D function is

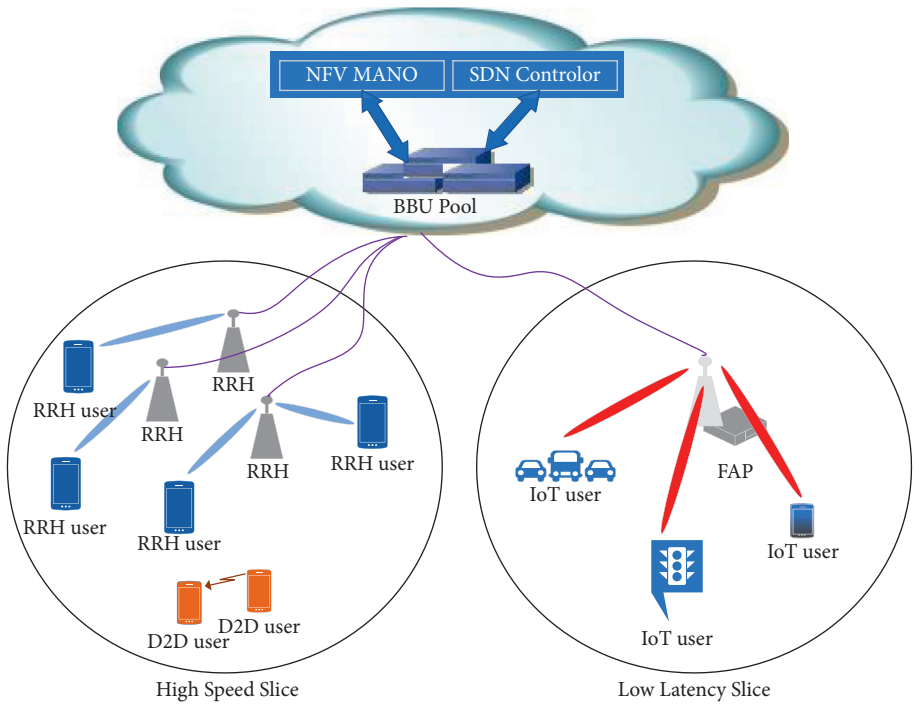

Figure 1: System model.

enabled, the user is called a D2D user. Using $x_g \in X = \{0, 1\}$ to indicate the D2D mode selection of user g, then the user is a D2D user if $x_g = 1$, and it is an RRH user if $x_g = 0$. For an RRH user g, its serving RRH is decided by the RRH association indicator $D_{m,g}$; if $D_{m,g} = 1$, then RRH user g is associated with and served by RRH m; otherwise, $D_{m,g} = 0$.

At time t, the received signal of the gth RRH user served by the BBU pool can be expressed by

$$y_g^R(t) = h_g(t)w_g(t)s_g(t) + \sum_{l\in\mathcal{G},l\neq g} h_g(t)w_l(t)s_l(t) + n_g(t), \tag{1}$$

where $h_g \in C^{1\times MS}$ is the CSI (channel state information) vector from the RRHs to RRH user g, w_g is the beamforming matrix of the RRHs to RRH user g, where $w_g = [w_{1,g}^T, w_{2,g}^T, \ldots, w_{M,g}^T]^T \in C^{MS\times 1}$ and $w_{m,g} \in C^{S\times 1}$ represents the beamforming vector from RRH m to RRH user g, s_g is the data symbol of RRH user g with $E\{|s_g(t)|^2\} = 1$, and n_g is the noise received by RRH user g, and $n_g \sim CM(0, \sigma^2)$. According to Shannon's formula, the data transmission rate of RRH user g can be expressed by

$$R_g^R(t) = B\log_2\left(1 + \frac{|h_g(t)w_g(t)|^2}{\sum_{l\in\mathcal{G},l\neq g}|h_g(t)w_l(t)|^2 + \sigma^2}\right). \tag{2}$$

For D2D users, the received signal of D2D user d served by D2D user r can be expressed by

$$y_d^D(t) = P_r(t)|h_{r,d}(t)|^2 + n_d(t), \tag{3}$$

where P_r denotes the transmission power of D2D user r and $|h_{r,d}|^2$ is the path loss between the D2D pair. The data transmission rate of D2D user d is represented by

$$R_d^D(t) = B\log_2\left(1 + \frac{P_r|h_{r,d}(t)|^2}{\sum_{I\in\mathcal{D},I\neq r,d}P_I|h_{I,d}(t)|^2 + \sigma^2}\right), \tag{4}$$

where $\mathcal{D}$ is the set of D D2D users and I is one of the interfering D2D users to user g.

2.1.2. Low-Latency Slice. In the low-latency slice, Q IoT users with low-latency requirements are served by the fog access point (FAP) equipped with S antennas. FAP in the low-latency slice can effectively implement resource allocation and power control to IoT users. The set of IoT users is $\psi = \{1, 2, \ldots, q, \ldots, Q\}$. At time t, K resource blocks can be allocated to the IoT users, the set of resource blocks is $\nabla = \{1, 2, \ldots, k, \ldots, K\}$, and $Q \leq K$. Due to the limited signal processing capabilities of IoT users, an IoT user can only be allocated one resource block at most. If at time t, resource block k is allocated to IoT user q, then resource block allocation indicator $a_{k,q} = 1$, and the received signal of the IoT user q on resource block k is

$$y_{q,k}^F(t) = P_{q,k}(t)|h_{q,k}(t)|^2 + n_q(t), \tag{5}$$

where $P_{q,k}$ is the transmission power of the FAP to IoT user q on the resource block k, $|h_{q,k}|^2$ refers to the path loss between the FAP and IoT user q on resource block k, and n_q is the received noise power of IoT user q.

According to Shannon's formula, the data transmission rate of IoT user q can be represented by

$$R_q^F(t) = B_k \log_2\left(1 + \frac{P_{q,k}|h_{q,k}(t)^2}{\sigma^2}\right). \tag{6}$$

Network energy efficiency is defined as the ratio of the average total throughput of the network to the average power consumption, in bit/Joule, and the expression is as follows:

$$\eta_{EE} = \frac{R_{tot}}{PC_{tot}} = \frac{\lim_{t\longrightarrow\infty} 1/T\sum_{t=0}^{T-1} E\{R_{tot}(t)\}}{\lim_{t\longrightarrow\infty} 1/T\sum_{t=0}^{T-1} E\{PC_{tot}(t)\}}, \tag{7}$$

where $R_{tot}(t)$ and $PC_{tot}(t)$ are the total throughput and power consumption of the network at time t, respectively, which are

$$R_{tot} = \sum_{g=1}^{G} R_g^R(t) + \sum_{d=1}^{D} R_d^D(t) + \sum_{q=1}^{Q} R_q^F(t), \tag{8}$$

and

$$PC_{tot} = \sum_{m=1}^{M} PC_m^R(t) + PC^F(t), \tag{9}$$

respectively. In equation (9), $PC_m^R = \sum_{g\in G}\|w_g\|_2^2/\eta_m^R$ is the power consumption of RRH m, where η_m^R is the amplifier efficiency of RRH m. $PC^F = \sum_{k=1}^{K}\sum_{q=1}^{Q} a_{k,q}P_{q,k}/\eta^F$ is the power consumption of the FAP, where η^F is the amplifier efficiency of the FAP.

2.2. Problem Formulation. According to equation (7), there are two ways to improve the energy efficiency of the network: one is to increase throughput R_{tot}, and the other is to reduce total energy consumption PC_{tot}. Considering the relationship among equations (2), (4), (6), (9), and (7), the energy efficiency can be improved by jointly optimizing the beamforming vector $w_{m,g}$ of each RRH m, resource block allocation k of the FAP, D2D mode selection indicator x_g, RRH association indicator $D_{m,g}$, transmission power $P_{q,k}$ of the FAP to IoT user q on the resource block k, and the transmission power P_r of the D2D users.

Using $\mathbf{K} = [K_1, \ldots K_q, \ldots K_Q]$ to denote the resource block allocation indicator vector of the IoT users, $\mathbf{W} = [w_{1,1}, \ldots, w_{m,g}, \ldots, w_{M,G}]$ to denote the beamforming vector of the network, $\mathbf{P}^F = [P_{1,1}, \ldots, P_{q,k}, \ldots, P_{Q,K}]$ to denote the transmission power vector of the FAP, and $\mathbf{P}^D = [P_1, \ldots, P_d, \ldots, P_D]$ to represent the transmission power vector of the D2D users, considering network stability constraints, QoS requirements of users, and restriction conditions such as transmission power limitation and backhaul link capacity limitation, the resource allocation-based energy efficiency maximization problem in the FRAN is defined as follows:

$$\begin{aligned}
&\max_{K,W,P^F,P^D} \eta_{EE} = \frac{R_{tot}}{PC_{tot}},\\
&C1:\ \overline{D}_g, \overline{D}_q, \overline{D}_d < \infty, \quad \forall g \in \mathscr{G}, \forall q \in \mathscr{Q}, \forall d \in \mathscr{D},\\
&C2:\ R_g^R(t) \geq R_g^{\min}, \quad \forall g \in \mathscr{G},\\
&C3:\ R_d^D(t) \geq R_d^{\min}, \quad \forall d \in \mathscr{D},\\
&C4:\ \Pr\left\{\overline{D}_q \geq D_q^{\max}\right\} = \varepsilon, \quad \forall q \in \mathscr{Q},\\
&C5:\ \sum_{g=1}^{G}\left\|D_{m,g}w_{m,g}\right\|_2^2 \leq P_{RRH}^{\max}, \quad \forall g \in \mathscr{G}, \forall m \in \mathscr{M},\\
&C6:\ 0 \leq \sum_{k\in\mathscr{K}} P_{q,k} \leq P_{FAP}^{\max}, \quad \forall q \in \mathscr{Q},\\
&C7:\ D_{q,k} \in \{0,1\}, \quad \forall q \in \mathscr{Q}, \forall k \in \mathscr{K},\\
&C8:\ \sum_{q\in\mathscr{Q}} D_{q,k} \leq 1, \quad \forall k \in \mathscr{K},\\
&C9:\ \sum_{g=1}^{G} D_{m,g}R_g^R \leq C_m, \quad \forall m \in \mathscr{M},\\
&C10:\ 0 \leq P_r \leq P_D^{\max}, \quad \forall r \in \mathscr{D},
\end{aligned} \tag{10}$$

where $\overline{D}_g$, $\overline{D}_d$, and $\overline{D}_q$ are the traffic queues of the RRH user g, D2D user d, and IoT user q, respectively. $C1$ constitutes the stability condition of the network. $C2$ and $C3$ are the QoS constraints for RRH users and D2D users in the high-speed slice, respectively. $C4$ is to guarantee the probability that the delay of IoT users does not exceed the given threshold $1-\varepsilon$. $C5$ and $C6$ are the maximum transmission power constraints of each RRH and FAP, respectively. $C7$ and $C8$ mean each IoT user can only be allocated to one resource block. $C9$ is the capacity limitation on the backhaul link of the BBU pool. $C10$ is the maximum transmission power constraint of D2D user d.

3. Transformation of the Original Problem

Using Lyapunov's theory, it is easy to prove that the proposed optimization problem can be transformed into the following form:

$$\begin{aligned}
\min_{K,W,P^F,P^D}\quad & \sum_{m=1}^{M} V\eta_{EE}PC_m^R - \sum_{g=1}^{G}\left(V + D_g^R\right)R_g^R + V\eta_{EE}PC^F\\
& -\sum_{k=1}^{K}\left(V + D_q^F\right)R_q^F - \sum_{d=1}^{D}\left(V + D_d^R\right)R_d^R \quad ,\\
s.t.\quad & C2 - C9
\end{aligned} \tag{11}$$

where V is the Lyapunov control parameter, which is a nonnegative real number. The larger the value of V is, the more important the wireless resource allocation strategy will be to optimize the network energy efficiency and the worse the delay performance will be. D_m^R and D_k^F are the traffic

queues of RRH m and resource block k of the FAP, respectively.

From (11), it can be found that the beamforming vector set $\mathbf{W}$ is only related to PC_m^R, R_m^R, $C2$, $C5$, and $C9$. Resource block indicator variable set $\mathbf{K}$ and transmission power vector $\mathbf{P}^F$ of the FAP are only related to PC^F, R_k^F, $C4$, $C6$, $C7$, and $C8$. Transmission power vector $\mathbf{P}^D$ of the D2D users is only related to R_d^D and $C3$. Therefore, the optimization problem in equation (11) can be decomposed into 3 independent subproblems: the optimization problem of beamforming design in high-rate slices, the optimization problem of joint resource block allocation and power control in low-latency slices, and the optimization problem of power control in the D2D mode.

3.1. Subproblem 1: Optimization Problem of Beamforming Design in High-Rate Slices. According to the above analysis, the optimization problem of beamforming design in high-rate slices can be rewritten as

$$\min_{W} \quad \sum_{m=1}^{M} \frac{V\eta_{EE}}{\eta_m^R} \sum_{g\in G} \left\|w_g\right\|_2^2 - \sum_{g=1}^{G}\left(V + D_g^R\right)R_g^R \quad . \tag{12}$$

$$s.t. \quad C2, C5, C9$$

To transform optimization problem (12) into convex, constraints $C2$ and $C9$ should be transformed. $C2$ can be rewritten as

$$C2\colon \sqrt{\sum_{l\in\mathcal{G},l\neq g}\left|h_g w_l\right|^2 + \sigma^2} \le \frac{\mathrm{Re}\left\{h_g w_g\right\}}{2^{R_g^{\min}/B} - 1}, \quad \forall g \in \mathcal{G}. \tag{13}$$

$C9$ can be reconstructed into

$$C9\colon \sum_{g=1}^{G} \frac{\left\|w_{m,g}\right\|_2^2}{\left\|w_{m,g}\right\|_2^2 + \xi} R_g^R \le C_m, \quad \forall m \in \mathcal{M}, \tag{14}$$

where ξ is a regular factor close to 0.

3.2. Subproblem 2: Optimization Problem of Joint Resource Block Allocation and Power Control in Low-Latency Slices. The optimization problem of joint resource block allocation and power control in low-latency slices can be transformed into

$$\min_{K,P^F} \quad \frac{V\eta_{EE}}{\eta^F} \sum_{k=1}^{K}\sum_{q=1}^{Q} a_{k,q}\left[P_{q,k} - \left(V + D_q^F\right)B_k \log_2\left(1 + \frac{P_{q,k}|h_{q,k}^2}{\sigma^2}\right)\right].$$

$$s.t. \quad C4, C6, C7, C8 \tag{15}$$

To transform (15) into convex, constraints $C4$ and $C7$ should be transformed. Since each IoT UE can only be allocated to one resource block, the minimum data rate requirement of each IoT UE at time t is equivalent to its minimum transmit power requirement on the resource block. And furthermore, taking into consideration the relationship between traffic queue and data rate, constraint $C4$ can be transformed into

$$C4\colon \left(2^{R_{\min}^F/W} - 1\right)\left|h_{q,k}\right|^2/\sigma^2 \le P_{q,k}, \quad \forall q \in \mathcal{Q}, \forall k \in \mathcal{K}. \tag{16}$$

To make the problem easier to solve, discrete constraint $C7$ can be transformed into

$$C7\colon 0 \le D_{q,k} \le 1, \quad \forall q \in \mathcal{Q}, \forall k \in \mathcal{K}. \tag{17}$$

3.3. Subproblem 3: Optimization Problem of Power Control in the D2D Mode. The optimization problem of power control in the D2D mode can be represented by

$$\max_{P^D} \quad \sum_{d=1}^{D}\left(V + D_d^R\right)B \log_2\left(1 + \frac{P_r\left|h_{r,d}(t)\right|^2}{\sum_{I\in\mathcal{D},I\neq r,d} P_I\left|h_{I,d}(t)\right|^2 + \sigma^2}\right).$$

$$s.t. \quad C3, C10 \tag{18}$$

4. Proposed Joint Optimization Algorithm

4.1. Beamforming Design for RRH Users in High-Rate Slices. With the help of the WMMSE method [14], the original subproblem 1 in (12) can be transformed into the following WMMSE problem with the same optimal solution:

$$\min_{W} \quad \sum_{m=1}^{M} \frac{V\eta_{EE}}{\eta_m^R} \sum_{g=1}^{G} \left\|w_g\right\|_2^2 + \sum_{g=1}^{G}\left(V + D_g^R\right)\left(\rho_g e_g - \log \mu_g\right) \quad ,$$

$$s.t. \quad C2, C5, C9 \tag{19}$$

where ρ_g, e_g, and μ_g are the mean square error weight, mean square error, and best receiving weight of the received signal of RRH user g, respectively. The definition of the mean square error is as follows:

$$\begin{aligned} e_g &= \mathbb{E}\left\{\left(\mu_g^H y_g^R - s_g\right)^2\right\} \\ &= \mu_g^H\left(\sum_{l=1}^{M} h_g w_l w_l^H h_g^H + \sigma^2\right)\mu_g - 2\mathrm{Re}\left\{\mu_g^H h_g w_g\right\} + 1. \end{aligned} \tag{20}$$

For any RRH user g, when w_g and μ_g are given, the calculation of the optimal mean square error weight ρ_g is

$$\rho_g = e_g^{-1}. \tag{21}$$

Given w_g, the best receiving weight μ_g can be calculated according to

$$\mu_g = \frac{h_g w_g}{\sum_{l=1}^{M} h_g w_l w_l^H h_g^H + \sigma^2}. \tag{22}$$

Given the set of ρ_g and μ_g for all the RRH users, the WMMSE optimization problem of $\mathbf{W}$ in (19) is

$$\begin{aligned} \min_{W} & \sum_{g=1}^{G} w_g^H \left(\sum_{m=1}^{M} \frac{V\eta_{EE}}{\eta_m^R} + \sum_{l=1}^{G} \left(V + D_l^R\right) \rho_l \mu_l^H h_l^H h_l \mu_l \right) w_g \\ & -2 \sum_{g=1}^{G} \left(V + D_g^R\right) \rho_g \mathrm{Re}\left\{\mu_g^H h_g w_g\right\} \\ s.t. & C2, C5, C9 \end{aligned} . \tag{23}$$

$C2$ in optimization problem (23) is a second-order conical constraint, and $C5$ and $C9$ are both quadratic constraints; therefore, problem (23) is a QCQP problem which can be solved by the convex optimization toolbox CVX [15]. Using toolbox CVX, the result of (23) in one iteration can be obtained. Through multiple iterations, the beamforming vector will approach the optimal solution.

4.2. Joint Resource Block Allocation and Power Control Algorithm for IoT Users in Low-Latency Slices. The objective function in problem (15) can be minimized by optimizing the transmission power $\mathbf{P}^F$ for IoT users first and then optimizing the resource block indicator $\mathbf{K}$. Besides, the solution obtained by solving the corresponding Lagrangian dual problem of (15) is at least a local optimal solution of (15). Therefore, optimization problem (15) can be reformed as the following optimization problems about $\mathbf{P}^F$ and $\mathbf{K}$, respectively.

Optimization problem (15) can be reformed as the following optimization problem about $\mathbf{P}^F$:

$$\begin{aligned} \min_{P^F} \quad & \frac{V\eta_{EE}}{\eta^F} \sum_{k=1}^{K} \sum_{q=1}^{Q} a_{k,q} \left[P_{q,k} - \left(V + D_q^F\right) B_k \log_2 \left(1 + \frac{P_{q,k} |h_{q,k}^2|}{\sigma^2} \right) \right] \\ s.t. \quad & \left(2^{R_{\min}^F/W} - 1\right) \left|h_{q,k}\right|^2 / \sigma^2 \le P_{q,k} \le P_{FAP}^{\max}, \forall q \in \mathcal{Q}, \forall k \in \mathcal{K} \end{aligned} . \tag{24}$$

The objective function of (24) has a continuous first-order derivative, and the constraint gives the range of the transmit power; therefore, the defined bound set is obviously a convex set. According to the KKT condition [16], to obtain the optimal transmit power, it is only necessary to let the first partial derivative of the objective function in (24) equal to 0. The optimal transmit power $P_{q,k}^*$ can be expressed as follows:

$$P_{q,k}^* = \max\left\{ \left(2^{R_{\min}^F/W} - 1\right) \left|h_{q,k}\right|^2 / \sigma^2, \min\left[\frac{\left(V + D_q^F\right)\eta^F W}{V\eta_{EE}} - \frac{\sigma^2}{\left|h_{q,k}\right|^2}, P_{FAP}^{\max} \right] \right. . \tag{25}$$

From (25), it can be found that when the quality of the channel is good or the traffic queue of the user is long, the resource management module of the FAP will increase the transmit power to serve the IOT user, and vice versa. Substituting $P_{q,k}^*$ into optimization problem (15), the optimization problem about $\mathbf{K}$ is

$$\begin{aligned} \min_{K} \quad & \frac{V\eta_{EE}}{\eta^F} \sum_{k=1}^{K} \sum_{q=1}^{Q} a_{k,q} \left[P_{q,k}^* - \left(V + D_q^F\right) B_k \log_2 \left(1 + \frac{P_{q,k}^* |h_{q,k}^2|}{\sigma^2} \right) \right], \\ s.t. \quad & C7, C8, \end{aligned} \tag{26}$$

which is a linear optimization problem. The optimal resource block allocation strategy can be obtained by

$$a_{k,q} = \begin{cases} 1 & \text{if } g_{k,q} < 0 \text{ and } k = \arg\min_{l} g_{l,q} \\ 0, & \text{otherwise} \end{cases}, \tag{27}$$

where $g_{k,q} = V\eta_{EE} P_{q,k}^* / \eta^F - (V + D_q^F) B_k \log_2 (1 + P_{q,k}^* |h_{q,k}|^2 / \sigma^2)$.

4.3. Power Control Algorithm for D2D. The objective function of problem (18) is a strictly convex function, and constraints $C3$ and $C10$ are linear and convex, respectively. Therefore, Lagrangian method can be used to solve problem (18). The Lagrangian dual problem corresponding to (18) is as follows:

$$\begin{aligned} & \max_{\delta,\theta} \min_{P^D} \left\{ L\left(P^D, \delta, \theta\right) \right\} \\ & s.t. \quad \delta \ge 0, \theta \ge 0, \end{aligned} \tag{28}$$

where $L(\mathbf{P}^D, \delta, \theta)$ is the Lagrangian function defined as

$$\begin{aligned} L\left(\mathbf{P}^D, \delta, \theta\right) = & \sum_{d=1}^{D}\left(V + D_d^R\right) B \log_2\left(1 + \frac{P_r\left|h_{r,d}(t)\right|^2}{\sum_{I\in\mathcal{D}, I\neq r,d} P_I\left|h_{I,d}(t)\right|^2 + \sigma^2}\right) \\ & + \delta\left(P_r - \left(\sum_{I\in\mathcal{D}, I\neq r,d} P_I\left|h_{I,r}(t)\right|^2 + \sigma^2\right)\left(2^{R_g^{\min}} - 1\right)\right) \\ & + \theta\left(P_D^{\max} - P_r\right). \end{aligned} \tag{29}$$

Problem (28) is a convex optimization problem about δ and θ. The subgradient method in [17] can be used to update the optimal Lagrangian multipliers δ and θ iteratively according to

$$\delta^* = \max\left\{0, \delta^* + \Delta_\delta\left(P_r - \left(\sum_{I\in\mathcal{D}, I\neq r,d} P_I|h_{I,d}|^2 + \sigma^2\right)\left(2^{R_g^{\min}} - 1\right)\right)\right\}, \tag{30}$$

and

$$\theta^* = \max\{0, \theta^* + \Delta_\theta\left(P_D^{\max} - P_r\right)\}, \tag{31}$$

respectively, where Δ_δ and Δ_θ are the step size of searching for optimal δ and θ, and they are positive and small real numbers close to zero. Then, the optimal transmission power can be obtained by

$$\frac{\partial L\left(P^D, \delta, \theta\right)}{\partial P_r}\Big|_{\delta=\delta^*, \theta=\theta^*} = 0, \quad \forall r \in \mathcal{D}. \tag{32}$$

Finally, calculate the transmission power of D2D user r to D2D user d according to

$$P_r^* = \max\{0, P_r\}. \tag{33}$$

4.4. Overall Algorithm. First, each user reports the reference signal received power (RSRP), signal-to-interference-noise ratio, channel state information reference signal (CSI-RS), location, and other parameters to the cloud.

Then, the cloud accepts the service requests from users. If the distance between two users is less than threshold d_{th} and the relative moving speed is less than threshold v_{th}, the two users release the communication link with the RRU and start the D2D communication mode. Based on this, x_g can be determined.

After that, the cloud calculates and assigns the beamforming vectors to RRH users for each RRH based on the result of CVX provided in Section 4.1. The configuration of beam weights ultimately serves the energy utility function, that is, the objective function. Given the beamforming vectors of the RRH users, the transmission power of the FAP and D2D transmitter can be determined by equations (25) and (33), respectively. At the same time, the cloud calculates and determines the resource block allocation according to equation (27) for each IoT user in the FAP.

Finally, the cloud detects the status of users in real time, determines the users' service status and the distance between the users performing communication services, and updates access and departure of the users to enable the updates of the D2D communication mode.

5. Simulation and Results

5.1. Simulation Configurations. To verify the effectiveness of the algorithm, system-level simulations are conducted. Related simulation parameters and settings are listed in Table 1.

5.2. Simulation Results. Figure 2 gives the relationship between the number of users in the D2D mode and the overall throughput. Simulation results show that when the number of users in the D2D transmission mode increases, the overall throughput of the system also increases. This proves that the D2D mode can alleviate the pressure on the fronthaul links, reduce interference between UEs, and increase the SINR, thereby increasing throughput. On the contrary, users who choose the D2D transmission mode do not occupy base station resources and release a large amount of system capacity, thereby greatly improving the overall throughput of the system.

Figure 3 is the relationship between the number of iterations of each algorithm and the overall throughput. The simulation results show that, as the number of iterations increases, the global solutions of the two algorithms are approaching the optimal solution. At the same time, the overall throughput of the fusion algorithm is greater than that of a single beamforming algorithm. This is because compared with a single beamforming algorithm, the system throughput can only be improved by adjusting the beam to improve the signal-to-noise ratio, and the fusion algorithm

Table 1: Simulation parameters and settings.

Parameter name	Parameter setting
Bandwidth	5 MHZ
Subcarrier interval	10 kHz
Downlink path loss of RRHs	$105.7 + 28.6\log(D)$, D in km
Downlink path loss of the FAP	$105.7 + 28.6\log(D)$, D in km
Maximum transmission power of RRHs	32 dBm
Maximum transmission power of the FAP	14 dBm
Capacity constraints of fronthaul links	10 Mbps
Power amplifier efficiency of RRHs	0.7
Power amplifier efficiency of the FAP	0.5
Small-scale fading model	Rayleigh fading
Lognormal shadow	5 dB
Noise power density	−147 dBm/Hz
Regularization factor	20^{-13}

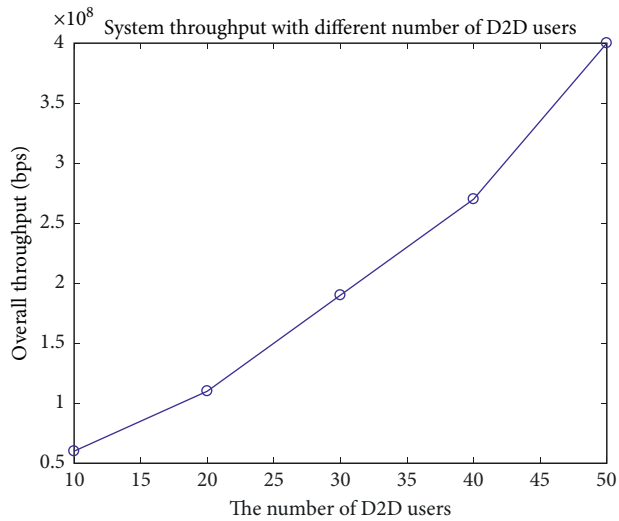

Figure 2: Overall throughput of the system with different numbers of D2D users.

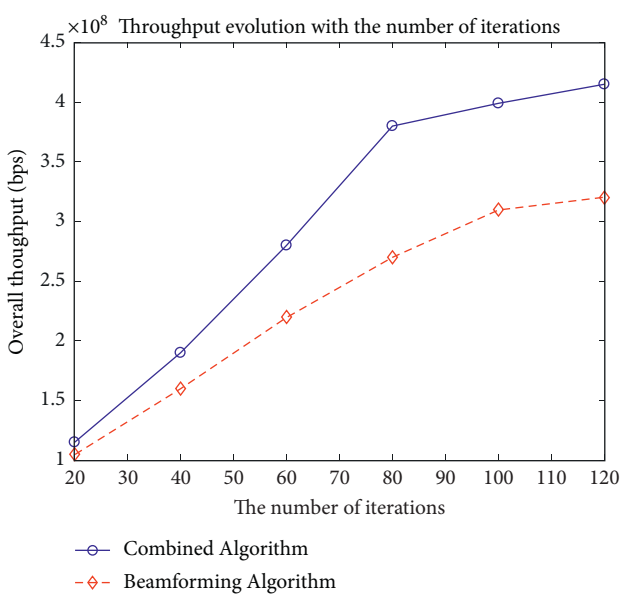

Figure 3: Evolution of system throughput with the number of iterations.

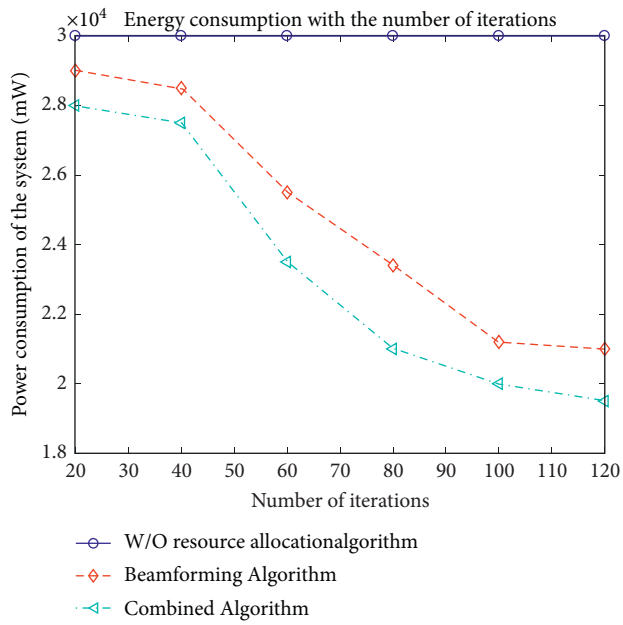

Figure 4: Energy consumption of the network with the number of iterations.

can reduce the base station load and expand the system capacity on this basis.

Figure 4 shows the relationship between the number of iterations of each algorithm and the overall energy of the system. The simulation results confirm the effectiveness of the beamforming algorithm and fusion algorithm for system energy saving. The beamforming algorithm adjusts the gain in each direction of the beam according to the distribution of the UE to achieve the effect of energy saving. Because of the direct communication between terminals, the fusion algorithm does not occupy base station resources, so the beamforming effect is more significant, and the system energy is further reduced.

6. Conclusion

In this paper, the energy efficiency optimization problem for D2D-enabled fog radio access networks (FRANs) was studied. An energy efficiency maximization algorithm based on beamforming, resource block allocation, and power control was proposed to solve the energy efficiency optimization problem. Beamforming vector, resource block allocation, and transmission power of the RRHs, FAP, and D2D users were jointly optimized with the help of nonlinear programming, convex optimization, and Lagrangian duality. Simulation results indicated that, by using the proposed optimization algorithm, the network energy efficiency and throughput can be significantly improved, and the network energy consumption can be greatly reduced, which finally proved the effectiveness of the proposed algorithm.

6.1. Future Work. This work can be further enhanced and expanded on a huge and massive scale. It can be used for highly scalable wireless access networks where the resource allocation becomes miserable. This allocation is based on the quality of service provided and acquired by the user. Similarly, it can be dug out that when and which kind of resources shall be needed when the number of users generally increases.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (no. 61801227), Qinglan Project of Jiangsu Province (nos. QLGC2021 and QLGC2020), Future Network Scientific Research Fund Project (no. FNSRFP-2021-YB-28), Open Research Fund of Jiangsu Key Construction Laboratory of IoT Application Technology (no. 19WXWL04), and Industry-University Cooperation Collaborative Education Project, Department of Higher Education, Ministry of Education (no. 202101220012).

References

[1] M. Asad, A. Basit, S. Qaisar, and M. Ali, "Beyond 5G: hybrid end-to-end quality of service provisioning in heterogeneous IoT networks," *IEEE Access*, vol. 8, pp. 192320–192338, 2020.

[2] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: a comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[3] M. Kaneko, I. Randrianantenaina, H. Dahrouj, H. Elsawy, and M.-S. Alouini, "On the opportunities and challenges of

NOMA-based fog radio access networks: an overview," *IEEE Access*, vol. 8, pp. 205467–205476, 2020.

[4] G. M. S. Rahman, T. Dang, and M. Ahmed, "Deep reinforcement learning based computation offloading and resource allocation for low-latency fog radio access networks," *Intelligent and Converged Networks*, vol. 1, no. 3, pp. 243–257, 2020.

[5] Y. Ai, G. Qiu, C. Liu, and Y. Sun, "Joint resource allocation and admission control in sliced fog radio access networks," *China Communications*, vol. 17, no. 8, pp. 14–30, 2020.

[6] A. Nassar and Y. Yilmaz, "Reinforcement learning for adaptive resource allocation in fog RAN for IoT with heterogeneous latency requirements," *IEEE Access*, vol. 7, pp. 128014–128025, 2019.

[7] N. N. Khumalo, O. O. Oyerinde, and L. Mfupe, "Reinforcement learning-based resource management model for fog radio access network architectures in 5G," *IEEE Access*, vol. 9, pp. 12706–12716, 2021.

[8] C. Miranda, G. Kaddoum, J.-Y. Baek, and B. Selim, "Task allocation framework for software-defined fog v-RAN," *IEEE Internet of Things Journal*, vol. 8. , 2021 in press.

[9] T. H. L. Dinh, M. Kaneko, E. H. Fukuda, and L. Boukhatem, "Energy efficient resource allocation optimization in fog radio access networks with outdated channel knowledge," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 1, pp. 146–159, 2021.

[10] H. Zhang, X. Liu, K. Long, A. Nallanathan, and V. C. M. Leung, "Energy efficient resource allocation and caching in fog radio access networks," in *Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM)*, December 2018.

[11] Y. Xu, Y. Chu, F. Jiang, Y. Guo, and D. Gong, "SVMs classification based two-side cross domain collaborative filtering by inferring intrinsic user and item features," *Knowledge-Based Systems*, vol. 141, pp. 80–91, 2018.

[12] R. Karasik, O. Simeone, and S. Shamai Shitz, "How much can D2D communication reduce content delivery latency in fog networks with edge caching?" *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2308–2323, 2020.

[13] M. Z. Hassan, M. J. Hossain, J. Cheng, and V. C. M. Leung, "Device-clustering and rate-splitting enabled device-to-device cooperation framework in fog radio access network," *IEEE Transactions on Green Communications and Networking*, in press, 2020.

[14] J. Lin, Q. Shi, Q. Li, and D. Zhao, "Joint mode selection and transceiver design for device-to-device communications underlaying multi-user MIMO cellular networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 6, pp. 3312–3328, 2019.

[15] X. Yu, D. Zhan, L. Liu, H. Lv, L. Xu, and J. Du, "A privacy-preserving cross-domain healthcare wearables recommendation algorithm based on domain-dependent and domain-independent feature fusion," *IEEE Journal of Biomedical and Health Informatics*, vol. 1, 2021.

[16] Y. Xu, J. Yang, and Z. Xie, "Training SVMs on a bound vectors set based on Fisher projection," *Frontiers of Computer Science*, vol. 8, no. 5, pp. 793–806, 2014.

[17] Y. Mo, M. Peng, H. Xiang, Y. Sun, and X. Ji, "Resource allocation in cloud radio access networks with device-to-device communications," *IEEE Access*, vol. 5, pp. 1250–1262, 2017.