IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 May 2023, accepted 24 May 2023, date of publication 29 May 2023, date of current version 20 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3280604

RESEARCH ARTICLE

D2D Cooperative Communication Network Resource Allocation Algorithm Based on Improved Monte Carlo Tree Search

XINZHOU LI AND GUIFEN CHEN
School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China

Corresponding author: Guifen Chen (2019100482@mails.cust.edu.cn)

This work was supported in part by the Special Project on Industrial Technology Research and Development of Jilin Province under Grant 2022C047-8; in part by the ''Thirteenth Five-Year Plan'' Science and Technology Research Project of Jilin Provincial Department of Education, Research on Large-Scale Device-to-Device Access and Traffic Balancing Technology for Heterogeneous Wireless Networks under Grant JJKH20181130KJ; in part by the National Natural Science Foundation of China under Grant 61540022; and in part by the Key Research and Development Projects of Changchun Science and Technology Bureau under Grant 21ZGM43.

ABSTRACT In recent years, with the rapid development of mobile communication, D2D (Device-to-Device, D2D) cooperative communication network has become the main component of future communication network, which greatly improves the spectrum efficiency of the network and the quality of user communication. However, the existing D2D network resource allocation schemes have some problems, such as weak dynamic resource allocation capability and low user communication quality. In view of this challenge, this paper proposes a resource allocation algorithm for D2D cooperative communication networks based on improved Monte Carlo tree search. First, a double-chain deep deciduous Monte Carlo tree search (Dcdd-MCTS) resource allocation model is established, Then, the loss function composed of deciduous MCTS and parallel convolution network is used to update the parameters of the deep neural network model of Dcdd-MCTS. Then, the theory of optimal classification is used to solve the user's transmit power. Finally, the optimal scheme of dynamic output resource allocation is output. The simulation results show that Dcdd-MCTS has good convergence. In the research on the distance between devices, compared with single-chain deep MCTS and joint optimization algorithm, the proposed algorithm in this paper increases the system throughput by 5%, 2%, respectively, and reduces the outage probability by 33%,18%.

INDEX TERMS D2D cooperative communication network, Monte Carlo tree search, double chain parallel neural network, resource allocation, interference management.

I. INTRODUCTION

With the rapid development of mobile communication, people have entered an information-based intelligent society, and communication network has become an in-dispensable part of human social life [1]. In recent years, with the rapid development from 1G to 5G, mobile communication has undergone tremendous changes. A large number of intelligent machine communication terminals and personal intelligent communication devices have emerged [2]. The demand for efficient and reliable transmission of application data between devices has sharply increased. At the same time, the business types of mobile communication networks based on direct connection of devices are more diversified, involving smart farming [3], smart grid [4], smart transportation [5], smart city [6], [7] and many other fields.

According to statistics, the number of global smartphone users will reach 10 billion by 2025, and it is expected to exceed 12 billion by 2030. Other intelligent terminals such as tablet computers and mobile robots will reach about 3 billion by 2025 and 5 billion by 2030 [8]. The machine-to-machine (M2M) equipment is expected to exceed 20 billion in 2025 and reach about 90 billion by 2030. The accompanying mobile communication traffic and M2M traffic will explode

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

exponentially. By 2025, the generated video traffic will be 20 times the amount of video traffic, and 60 times the amount of video traffic by 2030. The proportion of M2M traffic will expand from 7% to 30% [9], [10].

However, the spectrum resources that the traditional cellular communication network can provide for data communication are very limited. It can not afford the access of massive mobile devices in the future, let alone meet the exponential growth of data transmission demand in the future [11]. In addition, the current battery capacity cannot meet the energy supply of future communications [12], [13]. Therefore, how to solve the contradiction between the growing demand for network data transmission and the shortage of wireless spectrum resources and power consumption has become the focus of current research.

In the face of these challenges, the International Communications Development Research Institute has carried out a series of standardization work to improve the utilization of spectrum resources and user communication quality in cellular communication networks, such as building heterogeneous networks [14]. However, the con-struction of base stations will generate a lot of resource consumption. Communication operators are more inclined to the continuous optimization of the system. D2D coop-erative communication technology has become an important part of mobile development [15]. In the traditional cellular communication network, the data transmission between users must pass through the base station in the whole process, and cannot be carried out point-to-point communication. The specific communication process is as follows: first, the transmitting user sends the data to the base station, that is, the uplink, and then the base station sends the data to the receiving user, that is, the downlink. Although the user improves the ability of the base station to manage the spectrum re-sources through this data information transmission mode, at the same time, it also brings many problems, such as low utilization of the cell spectrum resources, high energy consumption, and heavy burden on the communication network.

D2D communication is a technology that directly transmits data between two users in a short distance without the control of the base station. The combination of this technology and traditional cellular network will reduce the pressure of the base station to process data, improve the utilization of wireless spectrum resources, and improve the total throughput in the communication system and the communication quality of users. Adding D2D communication and D2D relay communication technology to the traditional cellular network will greatly improve the communication capability of the traditional cellular network. It is of great social significance to study D2D cooperative communication [16]. It is widely used in smart city construction, smart community construction, smart transportation development, energy sustainable development and many other fields [17], [18]. On the one hand, it can communicate directly without passing through the base station, which greatly alleviates the demand pressure of current communication spectrum resources and meets the development needs of 5G and future mobile communication. On the other hand, the existence of D2D relay communication in the D2D cooperative network fills the technical defects of the traditional communication network that the communication channel conditions are poor and unable to communicate due to various factors such as communication construction and bad environment. However, due to the existence of spectrum resource reuse in the network, it has greatly increased the difficulty of interference management in the net-work within the system. The problem of interference management in D2D cooperative network has become a technical bottleneck in the development of communication field, and it has attracted the attention of relevant domestic institutions, experts and scholars [19].

A. RELATED WORKS

In recent years, many experts in the field of communication have carried out technical research on interference management in D2D cooperative networks, which can be divided into four directions. Through game theory, matching theory, convex optimization and machine learning.

Article [20] proposes a two-step auction D2D cooperative network resource allocation scheme based on sealed bidding. First, the cellular user broadcast identifies the user group, then estimates and prices based on the throughput data and interference near the base station, and finally conducts an auction game. This method greatly improves the network throughput. Article [21] proposes a game theory method, which uses Nash equilibrium bargaining mechanism to model multi-hop routing, and then determines the participants of information sharing and transmission relay, reducing the energy consumption in the system. Article [22] proposes a reverse auction game D2D cooperative network resource management scheme from relay node to destination node, which realizes user power control and greatly improves the energy efficiency of the system. Article [23] proposes a relay-assisted D2D network interference management method based on game theory. In the method, the optimization problem is first transformed into a non-convex nonlinear programming problem, and then the user transmission rate selection and power control are realized using the idea of dynamic game. The simulation results show that the scheme reduces the transmission energy consumption of relay-assisted D2D links and cellular links. In [24], starting from the user transmission time scale, based on the instantaneous channel information transmitted, a matching game scheme is proposed and the optimal threshold strategy is determined. The simulation results show that the algorithm has good convergence and robustness. Article [25] proposes an incentive based Stackelberg game D2D cooperative network resource allocation scheme, and discusses relay selection and power allocation in single-source and multi-relay D2D networks. In the case of incomplete relay channel information, the interference management capability of the system is greatly improved. Article [26] proposes a communication

network resource allocation algorithm based on hybrid routing game in multi-hop D2D network, which uses adaptive amplification and forwarding factor to achieve relay selection and resource allocation, maximizing the spectrum efficiency and capacity efficiency of the system.

Article [27] proposes an interference coordination scheme for D2D cooperative network based on three-part three-dimensional matching. First, the priority of cellular users and D2D users is determined, and then the interference signal is adaptively received and detected. Finally, the user's closed optimal power matching expression is derived in the interference and noise limited scenario. The proposed algorithm improves the security ability of the system. Article [28] proposes a graph coloring D2D cooperative network resource allocation scheme, which uses the weighted priority of spectrum resources in the system to realize multiple D2D users multiplexing a single cellular user resource. This algorithm reduces channel interference in the system. Article [29] proposes a D2D dual-relay network resource allocation scheme based on energy collection. The proposed algorithm achieves the optimal matching of resources in the system, the optimal allocation of user power, and improves the transmission rate and energy efficiency of the D2D link in the system. Article [30] proposes a resource allocation scheme for downstream D2D cooperative network based on quantum coral reef optimization algorithm. The use of idle users as relays to assist D2D link communication greatly improves the interference coordination capability of the system. This paper [31] proposes a resource matching algorithm for two-way relay D2D network based on improved particle swarm optimization, establishes a two-way relay-assisted D2D communication model, and maximizes the transmission rate and energy efficiency of the two-way relay D2D link in the system. Article [32] proposes a stable matching D2D network optimization algorithm. First, the interference minimization clustering model based on physical proximity and social attributes is established, and then the user communication ability based on social and physical proximity is evaluated under this model. The stable matching theory algorithm is optimally used to achieve one-to-one resource allocation and maximize the transmission rate of the system. In [33], considering the asymmetry of energy consumption and spectrum re-sources of D2D users and cellular users in traditional communication networks, a D2D communication resource allocation scheme based on maximum weighted bipartite matching is proposed, which improves the overall throughput and energy efficiency of full-duplex D2D communication systems. Article [34] proposes a one-to-one stable matching resource allocation scheme for D2D cooperative network, uses nonlinear energy collection model to model energy, and carries out optimal relay selection and optimal power allocation under the premise of ensuring the quality of service of users in the system. The proposed algorithm effectively improves the throughput and energy efficiency of D2D links. Article [35] Aiming at the resource allocation problem in the downlink full-duplex cooperative cellular communication of D2D communication, a bilateral stable many-to-one channel matching algorithm based on Pareto improvement is proposed, which greatly improves the spectral efficiency of the system. Literature [36] proposes an energy-saving resource allocation scheme for joint uplink/downlink D2D. This scheme considers the many-to-one matching standard of channel reuse between users, and conducts according to the service quality satisfaction of cellular users. The optimized performance measurement improves the energy efficiency of the system. Literature [37] proposed a centralized channel allocation algorithm based on the well-known bilateral preference Gale-Shapley algorithm, using suboptimal distributed power control to optimize the uplink and downlink communication links, reducing system cost and improving system throughput.

Article [38] proposes an alternate iterative algorithm based on block alternate de-scent and continuous convex approximation for the safe transmission of the D2D UAV relay network, which optimizes the UAV trajectory and transmission power. The algorithm has good convergence and improves the security rate of the system. Article [39] proposes a convex optimization scheme of D2D cooperative UAV communication network based on energy collection. The non-convex problem of the combination of radio resource allocation and flight altitude is converted into a convex optimization problem by using variable relaxation and variable replacement methods, and then the optimal solution of resource allocation is derived by using Lagrange duality theory. The simulation results show that the algorithm is effective. Article [40] uses the fractional programming theory to transform the energy consumption problem in the D2D cooperative network into a standard convex optimization problem, and uses the iterative algorithm to find the optimal solution. The proposed algorithm greatly improves the energy efficiency of the system. In [41], a robust D2D communication resource al-location scheme based on convex optimization is proposed. The problem is converted into a convex optimization problem by using the worst case limit and chance constraint method. Then, the closed expression of power and channel allocation is derived by using the Lagrange dual method. The iterative algorithm based on distributed sub-gradient is optimally used to achieve the optimal robust resource allocation. Article [42] proposes a joint optimization relay selection and resource allocation algorithm based on convex optimization. First, the joint optimization problem is transformed into a convex optimization problem by using convex optimization technology, and then the system resource allocation is carried out by using Lagrange method, which greatly im-proves the energy efficiency and network capacity of the system.

Literature [43] applies the combined multi-arm bandit form in machine learning theory to D2D communication network, and dynamically adaptive relay selection, and at the same time realizes the resource allocation of D2D cooperative network under unknown channel state information.

Article [44] proposes a D2D communication interference management method based on reinforcement learning, which optimizes channel allocation and relay selection, and greatly improves the network capacity of the system. Article [45] proposes a resource allocation scheme for D2D energy collection network based on non-orthogonal multi-access technology. Kuhn – Munkres algorithm is used to complete channel allocation and relay selection in the system, and then reinforcement learning is used to perform offline power allocation for D2D users. Finally, neural network learning is used to obtain the optimal power allocation model, which reduces the outage probability of users in the system. Article [46] A D2D collaborative network resource sub-scheme based on the deep learning framework greatly improves the throughput of the system through the learning and training of the deep neural network. Article [47] proposes an emotion-driven online learning relay selection and channel allocation algorithm, which realizes the relay selection and channel selection of the system by continuously learning the user's real-time transmission rate and the change trend of transmission speed. This algorithm greatly improves the total transmission rate of the system. Literature [48] proposed an IoT resource allocation optimization scheme using federated learning to develop efficient integration of joint edge intelligent nodes, effectively optimize the computing frequency allocation, and reduce the energy consumption of IoT devices. Literature [49] proposed a new dual-depth Q network communication network resource intelligent management scheme, which reduces the power consumption and communication delay of devices in the system. Literature [50] proposed a multi-agent D2D communication resource allocation algorithm based on Advantage Actor Critic (A2C), which dynamically and adaptively outputs the optimal channel allocation scheme, which improves the throughput of the system. Literature [51] proposed a deep reinforcement learning communication resource optimization scheme for refined generative adversarial networks, which improves the reliability of the system and reduces the delay of the system.

B. MOTIVATION AND CONTRIBUTIONS

To sum up, these four types of technologies improve the performance of D2D collaborative systems to a certain extent. Inspired by the above-mentioned literature, this paper introduces a new joint optimization resource allocation scheme based on deep reinforcement learning for D2D collaborative networks, which solves the key gap in resource allocation in existing D2D collaborative networks, and in terms of system transmission rate and user outage probability provides significant improvements, and to the best of our knowledge this is the first study combining MCTS with parallel deep residual networks and successfully applying it in a D2D collaborative network. Overall, the research work in this paper represents a significant contribution to the field of D2D communication and provides important insights for future communication research in this field.

The main contributions and the structure of this paper are as follows:

1) Aiming at the problem that the influence of channel selection of cellular users on system performance is ignored in the traditional D2D communication research, this paper studies the influence of channel selection of cellular users on system performance.

2) The four aspects of resource allocation in D2D cooperative network are modeled as MDP (Markov Decision Process, MDP), and the resource allocation model of double-chain deep deciduous Monte Carlo tree is proposed, as well as the solution of user transmission power using optimization theory.

3) The composition of double-chain neural network, the mechanism of deciduous MCTS and the updating process of deep neural network are described in detail.

4) The system simulation experiment proves the reliability and effectiveness of the algorithm proposed in this paper. The comparison with similar algorithms shows that in the research on the distance between devices, the algorithm proposed in this paper is compared with single-chain deep MCTS and joint optimization algorithm [52], the proposed algorithm in this paper increases the system throughput by 5%, 2%, respectively, and reduces the outage probability by 33%,18%.

The second part introduces the D2D cooperative communication system. In the third part, the resource allocation algorithm of double-chain deep defoliation MCTS and the solution of user transmit power are introduced in detail. The fourth part carries out the convergence experiment of the algorithm, as well as the comparative experiment of similar algorithms and the analysis of experimental results. The fifth part summarizes the full text and introduces the future work plan.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. SYSTEM MODEL

This paper studies the resource reuse scenario of the cellular D2D uplink communication link. As shown in Figure 1, a single-cell D2D cooperative communication network system model is established. The system includes 1 base station, M cellular users (CUE), N D2D user pairs (DUE) (each pair of D2D users includes a D2D transmitter and a D2D receiver), and Q relay equipment (RUE). The CUE set $C = \{c_1, \ldots c_m, \ldots, c_M\}$, DUE set $D = \{d_1, \ldots, d_n, \ldots, d_N\}$, RUE set, channel number set $Q = \{r_1, \ldots, r_l, \ldots, r_Q\}$. CUE, DUE, RUE and BS are all configured with a single antenna. It is assumed that the channels of all links obey the Rayleigh distribution. Wherein, the base station allocates independent orthogonal channel resources for each cellular user. In this system, D2D users communicate by multiplexing the channel resources of cellular users. Each D2D user can occupy at most one channel resource.

The interference existing in the above system mainly includes: in D2D direct transmission mode, interference from cellular users to D2D receiver; interference from D2D transmitter to base station; interference from cellular users to relay

in D2D relay transmission mode; interference from D2D to base station and so on.

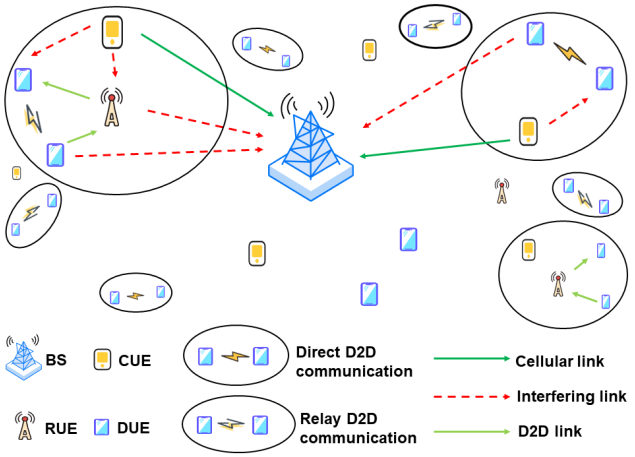

FIGURE 1. D2D cooperative communication network system model [52], [53].

B. PROBLEM FORMULATION

It can be seen from Figure 1 that there are three user communication modes in the D2D collaborative network, the cellular user direct communication mode; the D2D multiplexing cellular user channel resource direct transmission mode; and the D2D relay forwarding communication mode. Wherein, the D2D relay forwarding communication method includes two stages. In the first stage, the D2D sender multiplexes the CUE channel resources to send to the relay, and in the second stage, the RUE multiplexes the same channel resources to send to the D2D receiver. Since cellular direct communication and D2D direct communication do not have the problem of multiplexing interference, this paper focuses on the latter two communication methods.

1) When cellular user m occupies channel resource k alone, the SINR and transmission rate of CUE can be expressed as:

$$\gamma_{c_m}^{k(0)} = \frac{p_{c_m}^k \left|h_{c_m}^k\right|^2}{\sigma^2} \tag{1}$$

$$V_{c_m,k}^0 = Blog_2\left(1+\gamma_{c_m}^{k(0)}\right) \tag{2}$$

$P_{c_m}^k$ represents the transmission power of the m-th CUE on channel k; $h_{c_m}^k$ Indicates the channel status of the m-th CUE and BS on channel k; σ^2 Indicates the power of Gaussian noise; B Indicates channel bandwidth.

2) When cellular user m and D2D user n share channel k, the SINR and transmission rate of CUE user and DUE user can be expressed as:

$$\gamma_{C_m}^{k(d)} = \frac{p_{c_m}^k \left|h_{c_m}^k\right|^2}{\sigma^2 + p_{d_n}^k \left|h_{d_n,b}^k\right|^2} \tag{3}$$

$$V_{c_m,k}^d = B\log_2\left(1+\gamma_{c_m}^{k(d)}\right) \tag{4}$$

$p_{d_n}^k$ indicates the transmission power of the n-th D2D transmitter on channel k; $h_{d_n,b}^k$ Indicates the channel status of the n-th D2D and BS on channel k.

$$\gamma_{d_n,d_n}^{k(d)} = \frac{p_{d_n}^k \left|h_{d_n}^k\right|^2}{\sigma^2 + p_{c_m}^k \left|h_{c_m,d_n}^k\right|^2} \tag{5}$$

$$V_{d_n,k}^d = Blog_2\left(1+\gamma_{d_n,d_n}^{k(d)}\right) \tag{6}$$

$h_{d_n}^k$ indicates the channel status between the n-th D2D transmitter and receiver; h_{c_m,d_n}^k indicates the channel status of the m-th CUE and n-th D2D receiver on channel k.

3) When D2D user n shares the channel with cellular user m in relay mode: In the first stage, on channel k, the SINR received by CUE and RUE can be expressed as:

$$\gamma_{c_m}^{k(l1)} = \frac{p_{c_m}^k \left|h_{c_m}^k\right|^2}{\sigma^2 + p_{d_n}^k \left|h_{d_n,b}^k\right|^2} \tag{7}$$

$$\gamma_{d_n\to r_{l1}}^{k(c)} = \frac{p_{d_n}^k \left|h_{d_n,r_{l1}}^k\right|^2}{\sigma^2 + p_{c_m}^k \left|h_{c_m,r_{l1}}^k\right|^2} \tag{8}$$

$h_{d_n,r_{l1}}^k$ indicates the channel status of the n-th D2D transmitter and the 1st relay on channel k; $h_{c_m,r_{l1}}^k$ indicates the channel status of the m-th CUE and relay l on channel k.

Similarly, in the second stage of transmission, the signal-to-interference ratio of the CUE and the signal-to-noise ratio forwarded by the RUE can be expressed as:

Similarly, in the second phase of transmission, the signal-to-interference ratio of CUE user m received by the base station and the signal-to-noise ratio of RUE forwarding received by the D2D receiver can be expressed as:

$$\gamma_{c_m}^{k(l_2)} = \frac{p_{c_m}^k \left|h_{c_m,b}^k\right|^2}{\sigma^2 + p_{r_{l2}}^k \left|h_{r_{l2},b}^k\right|^2} \tag{9}$$

$$\gamma_{r_{l2}\to d_n}^{k(c)} = \frac{p_{r_{l2}}^k \left|h_{r_{l2},d_n}^k\right|^2}{\sigma^2 + p_{c_m}^k \left|h_{c_m,d_n}^k\right|^2} \tag{10}$$

$h_{r_{l2},b}^k$ indicates the channel status of relay l and BS on channel k; $P_{r_{l2}}^k$ indicates the transmission power of the first relay on channel k; h_{r_{l2},d_n}^k indicates the channel state on channel k from relay l to the n-th D2D receiver; h_{c_m,d_n}^k indicates the channel status of the m-th CUE to n-th D2D receiver on channel k.

The transmission rate of CUE and DUE users in the relay forwarding mode can be expressed as:

$$V_{c_m,k}^r = \frac{B}{2}\log\left(1+\gamma_{c_m}^{k(l1)}\right) + \frac{B}{2}\log\left(1+\gamma_{c_m}^{k(l2)}\right) \tag{11}$$

$$V = \sum_{k=1}^{K}\sum_{l=1}^{R}\sum_{n=1}^{N}\sum_{m=1}^{M}$$

$$X_1 = \left(1 - b_{d_n}^k\right) V_{c_m,k}^0$$
$$X_2 = b_{d_n}^k \left[\left(1 - e_{r_l}^k\right) V_{c_m,k}^d + e_{r_l}^k V_{c_m,k}^r\right]$$
$$X_3 = b_{d_n}^k \left[\left(1 - e_{r_l}^k\right) V_{d_n,k}^d + e_{r_l}^k V_{d_n,k}^r\right] \quad (12)$$

$\left(1 - b_{d_n}^k\right) V_{c_m,k}^0$ indicates the exclusive channel of CUE; $b_{d_n}^k \left[\left(1 - e_{r_l}^k\right) V_{c_m,k}^d + e_{r_l}^k V_{c_m,k}^r\right]$ indicates that CUE and DUE share the channel; $\left(1 - e_{r_l}^k\right) V_{d_n,k}^d$ indicates the DUE direct transmission scenario; $e_{r_l}^k V_{d_n,k}^r$ indicates the RUE forwarding scenario.

To sum up, the problem expression of the maximum network capacity of the system and related constraints can be expressed as:

$$\max_{s^1,s^2,s^3,P} V \quad (13)$$

$$s.t.\ C_1: \sum_{k=1}^{K} s_k^m \le 1, \quad \forall m; \sum_{m=1}^{M} s_k^m \le 1, \ \forall k \quad (14)$$

$$C_2: \sum_{k=1}^{K} s_k^n \le 1, \quad \forall d; \sum_{n=1}^{N} b_k^n \le 1, \ \forall k \quad (15)$$

$$C_3: \sum_{k=1}^{K} s_k^l \le 1, \quad \forall l; \sum_{l=1}^{R} s_k^l \le 1, \ \forall k \quad (16)$$

$$C_4: \gamma_{c_m,b}^{k(0)}, \quad \gamma_{c_m,b}^{k(d)}, \ \gamma_{c_m,b}^{k(l1)}, \ \gamma_{c_m,b}^{k(l2)} \ge, \ \gamma_c^{min} \quad (17)$$

$$C_5: \gamma_{d_n,d_n}^{k(d)}, \quad \gamma_{d_n \to r_l}^{k(c)}, \ \gamma_{r_l \to d_n}^{k(c)} \ge, \ \gamma_{d_n}^{min} \quad (18)$$

$$C_6: 0 \le p_{c_m} \le p_{c_m}^{max}, \quad 0 \le p_{d_n} \le p_{d_n}^{max}, \quad 0 \le p_{r_l} \le p_{r_l}^{max} \quad (19)$$

s^1 indicates the channel allocation vector of CUE $s^1 = \left[s_k^m\right]_{k\in K, m\in C}$, $s_k^m \in \{0,1\}$, $s_k^m = 1$ indicates that the m-th CUE occupies channel k, otherwise $s_k^m = 0$; s^2 indicates the channel allocation vector of DUE $s^2 = \left[s_k^n\right]_{k\in K, n\in D}$, $s_k^n \in \{0,1\}$, $s_k^n = 1$ indicates that the nth DUE occupies channel k, otherwise $s_k^d = 0$; s^3 indicates the channel allocation vector of RUE $s^3 = \left[s_k^l\right]_{k\in K, l\in R}$, $s_k^l \in \{0,1\}$, $s_k^l = 1$ indicates that the first RUE is transmitted on channel k, otherwise, $s_k^l = 0$; $P = \{P_c, P_d, P_l\}$ indicates the power distribution vector; The power vector of CUE is: $P_c = \left[P_{c_1}, \ldots, P_{c_M}\right]$, $P_{c_m} = \left[P_{c_m}^k\right]_{k\in K}$ indicates CUE user c_m power vector on different channels; DUE user's power vector is: $P_d = \left[P_{d_1}, \ldots, P_{d_N}\right]$, $P_{d_n} = \left[P_{d_n}^k\right]_{k\in K}$ indicates DUE user d_n, power vectors on different channels; The power distribution vector of RUE is $P_l = \left[P_{l_1}, \ldots, P_{l_R}\right]$, $P_{l_r} = \left[P_{l_r}^k\right]_{k\in K}$ indicates RUE user l_r Power vector on each subchannel; C_1 indicates that a CUE can only occupy one channel and there is at most one CUE on one channel;C_2 indicates that one DUE can only occupy one channel and there is at most one DUE on one channel; C_3 indicates that one RUE can only occupy one channel and there is at most one RUE on one channel. C_4, C_5, It means that CUE and DUE meet the minimum signal-to-noise ratio limit $\gamma_c^{\min}$ and $\gamma_{d_n}^{min}$ respectively. $P_{c_m}^{max}$, $P_{d_n}^{max}$, $P_{l_r}^{max}$ represent the maximum transmit power of CUE, DUE and RUE respectively.

III. D2D COLLABORATIVE NETWORK RESOURCE ALLOCATION ALGORITHM

A. PROBLEM TRANSFORMATION - MARKOV DECISION PROCESS

In this paper, the problems of user communication mode selection, channel allocation and relay selection in each time slot are modeled as finite MDP quads $\{S, A, T, R\}$. User's channel status $S = \left[s^1, s^2, s^3\right]$, The CUE channel status is recorded as $s^1 = \left[x_{k,i}^1\right]_{k\in K, i\in M}$, DUE channel status is marked as $s^2 = \left[x_{k,i}^2\right]_{k\in K, i\in N}$, RUE channel status is marked as $s^3 = \left[x_{k,i}^3\right]_{k\in K, i\in R}$, $x_{k,i}^1 = 1$ indicates that the i-th cellular user is selected for allocation to the kth channel; $x_{k,i}^2 = 1$ indicates that the i-th D2D user is selected to be assigned to the k-th channel; $x_{k,i}^3 = 1$ indicates that the i-th relay has been selected and he participates in D2D cooperation on the k-th channel, $x_{k,i}^3 = 0$ indicates that although the i-th relay has been selected, it does not participate in D2D cooperation on the k-th channel. The composition and transition process of state s are shown in Figure 2.

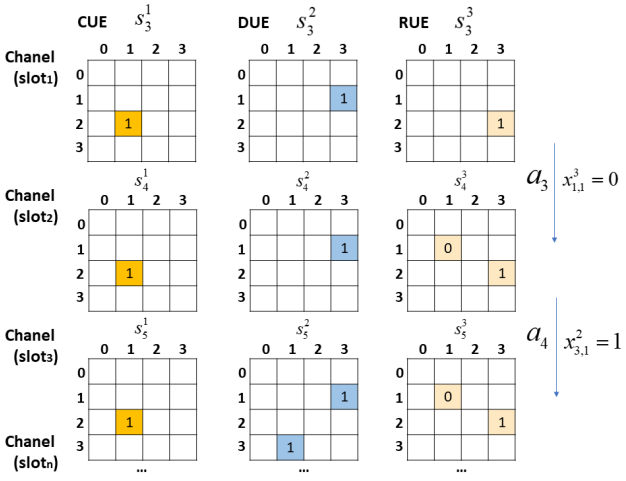

FIGURE 2. Status S composition and conversion process.

In $t \in [0, T]$, System based on policy $\pi(a_t, s_t)$, that is, the prior probability of the system $Pr(s_{t+1}|s_t)$, execute actions, $a_t \in A$, get reward and punishment value $r(s_t, a_t)$. Then enter the next state. Loop down to get multiple states and action tracks in the system $E(s_t) = \{s_t, a_t, \ldots, s_{N+M+R-1}, a_{N+M+R-1}, s_{N+M+R-1}\}$, In addition, we can also get the expected reward $Q(s, a)$ of starting from the state and taking action a.

$$Q(s,a) = E\left\{\sum_{\tau=t}^{N+M+R} r_t \,|s_t = s, a_t = a\right\} \quad (20)$$

B. RESOURCE ALLOCATION MODEL OF DOUBLE-CHAIN DEEP DECIDUOUS MONTE CARLO TREE

The resource allocation model of double-chain deep deciduous Monte Carlo tree under D2D cooperative network mainly includes double-chain deep residual network and deciduous MCTS module. The D2D collaboration network resource allocation diagram is shown in Figure 3.

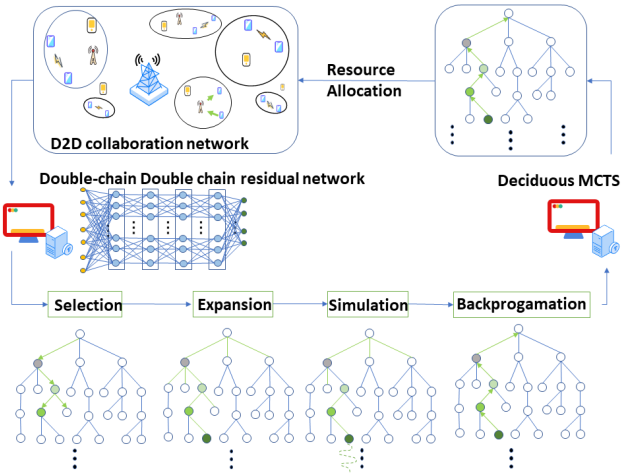

FIGURE 3. D2D collaboration network resource allocation diagram.

1) DOUBLE CHAIN DEPTH RESIDUAL MODULE

The deep residual network is expressed as $(P_r, v) = f_\theta\left(s_t, \tilde{H}\right)$, Take the current resource allocation status matrix s_t and the channel gain $\tilde{H}$ of the corresponding timeslot as input, The elements of $\tilde{H}$ can be converted into dB by converting the unit of each channel gain, and then normalize them to make each element become the value of zero mean and unit variance.

According to the form of formula (4) in [54], the channel gain between any two users (user i and user j) can be obtained. The calculation formula of $\tilde{H}$ can be expressed as:

$$h_{i,j}^k = \frac{log_{10}\left(h_{i,j}^k\right) - E\left[log_{10}\left(h_{i,j}^k\right)\right]}{\sqrt{E\left[\left(log_{10}\left(h_{i,j}^k\right) - E\left[log_{10}\left(h_{i,j}^k\right)\right]\right)^2\right]}} \tag{21}$$

The dual-chain deep residual network adopts a three-layer structure and is composed of a parallel dual-chain neural network structure. The model includes input, feature extraction module, feature blending module and output. Among them, L^A, L^B in the feature extraction module represents two parallel residual networks, and G represents the feature blending module. The output layer is divided into two full-connection layer branches, followed by a softmax and a tanh, respectively, to output the prior probability vector and state-action value of the action.

The specific operation of the module is shown in Figure 4. The input of the model is set to x_ Input, first complete the first feature extraction through a convolution layer, and the output is y_0. Then enter the feature extraction module, which is used as the input of two parallel networks, and carry out a series of convolution, batch normalization and other processes respectively to obtain the output y_n^A, y_n^B. The two feature data are combined to get y_n^C. The output is the action probability vector P_r and the state-action value v, v is the evaluation scalar to estimate the probability of the system winning from the state s_t, and P_r represents the probability of selecting each action from s_t.

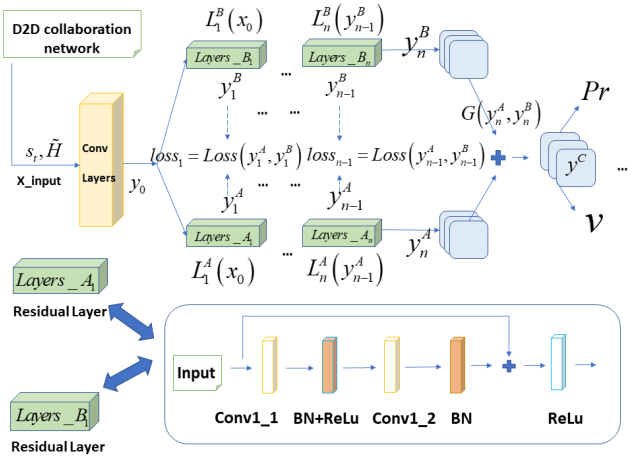

FIGURE 4. Double-chain deep residual network.

2) DECIDUOUS MCTS MODULE

The traditional MCTS module uses the action probability P_r and state-action value v output by the deep residual network to evaluate and select the action, and generates self-game data through simulation to train the subsequent deep residual network. Considering the multi-thread situation of MCTS, this paper adds the defoliation mechanism in the traditional MCTS module to improve the system performance. The basic idea is that deciduous MCTS adds the judgment and execution module of defoliation mechanism to the original MCTS module to determine whether the current search tree needs defoliation and execute defoliation when needed.

According to the above improved MCTS idea, the specific steps of deciduous MCTS are as follows:

(1) Determine the upper and lower bounds of the leaves. The total number of threads is the upper bound and marked as Node_upbound; The lower bound of memory is marked as Node_downbound

(2) According to the fact that each thread inherits the same Monte Carlo search tree from the beginning to the end, record the total number of nodes in the Monte Carlo search tree since the game of each model, and record it as tree_total_count.

(3) This step is divided into external loop judgment and internal loop judgment.

External loop judgment to determine the current tree_total_ is count greater than Node_upbound. If so, execute the defoliation process, and then continue the self-game; If no, do not execute the defoliation, and directly continue the self-game.

Internal loop judgment: 1) Set the access threshold of the node and record it as Node_Limit, and use it as the threshold value to determine whether a node is defoliated. In the initial model, Node_The limit starts from 0 and increases by 1 each iteration. 2) Set the total number of sub-node accesses obtained by traversing the entire Monte Carlo search tree from the root node, and record it as Node_cur. 3) If Node_Limit is less than Node_Cur, the node is retained. On the contrary, the node is culled and the node in the parent node of the node is removed_The cur value subtracts 1 from itself to determine whether the parent node at this time is a leaf node and update its leaf node label. 4) Delete the deciduous nodes that need to be removed from the data structure storing the entire tree and continue to traverse until all nodes have been traversed.

(4) Judge whether the total number of nodes on the current search tree (tree_total_count) is less than (Node_downbound). If so, it will end falling leaves. If not, increase the node access threshold in the search by 1, and then continue with step (2) (3).

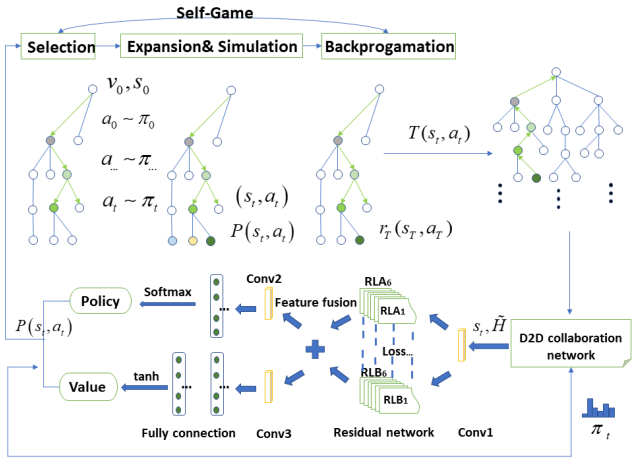

FIGURE 5. Working process of double-chain deep defoliation MCTS.

3) WORKING PROCESS OF DOUBLE-CHAIN DEEP DEFOLIATION MCTS

First, configure the initial parameters of the deep residual network θ_0. In each subsequent iteration $i \geq 1$, in each state s_t, the MCTS tree searches for the action probability P_{rt} that uses the output of the previous iteration of the deep residual network $\pi_t = \alpha_{\theta_{i-1}}(s_t)$, and selects an action by using the MTCS value function P_{rt} until step $t = M+N+R$ is reached or all devices (including CUE, DUE, RUE) are arranged on the channel, and stops the tree search. In particular, for the search tree based on the deep residual network f_{θ_i}, the MCTS tree search value function a_{θ_i} is constantly updated to predict the winner of the simulated self-game. The best value function a_{θ_i} so far is generated into the self-game data in the format $\left(\tilde{H}, s, \pi, Q\right)$, which is used to train the following deep residual network.

As shown in Figure 5, in each state s_t, an MCTS tree a_θ is executed, and the action search is guided by the latest depth residual network f_θ. The action is selected according to the search strategy probability $a_t \sim \pi_t$ calculated by MCTS. Start from the root state s_0, then select an action $a_t \sim \pi(s_t)$ from the strategy of deep residual network prediction, and update the state $s_{t+1} = T(s_t, a_t)$ until the complete tree is reached.

4) TRAINING AND IMPLEMENTATION OF DOUBLE-CHAIN DEEP DECIDUOUS MONTE CARLO TREE RESOURCE ALLOCATION MODEL

First, randomly initialize the weight of depth residuals θ_0. Then, the parameters of the residual network are updated in the form of self-supervised learning, so that the action probability and state-action value $(P_r, v) = f_\theta\left(s, \tilde{H}\right)$ match the action strategy obtained from MCTS more closely. Specifically, it is to adjust the parameters of the residual network to minimize the gap between the predicted state-action value v and the stored data Q^*, and maximize the similarity between the output action probability P_r of the deep residual network and the MTCS strategy probability π. Parameter θ is updated by gradient descent of loss function composed of mean square error part and cross entropy part:

$$L = -\pi^T \log P_r + \left(v - Q^*\right) + \Delta \|\theta\|^2 + loss_p \qquad (22)$$

Multiply the loss value of each layer by the constant coefficient μ, The loss function $loss_p$ of the whole parallel network can be obtained.

$$loss_p = \mu \sum_i loss_i = \mu \sum_i \left(\left|y_i^A - y_i^B\right|\right)^2 \qquad (23)$$

Q^* is the state-action value of the self-play winner obtained from MCTS simulation, and Δ is the weight regularization factor to prevent over-fitting. The size of the constant coefficient μ will determine the coupling degree of the two parallel branches. The smaller the coefficient μ, the lower the coupling degree. The parameters of the depth residual network can be obtained by minimizing the loss function.

The training data set of the residual network is stored as $\left\{\tilde{H}, s, \pi, v\right\}$. The pseudo-code of double-chain residual network training is shown in Table 1.

5) POWER CONTROL

After using the MCTS method to obtain the binary variables allocated by CUE, D2D and RUE channels, the original objective function is converted into a linear function of the continuous variable power P.

TABLE 1. Pseudocode of double-chain deep residual network.

Input:
Channel status of cellular users, D2D users and relay users (s_t^1, s_t^2, s_t^3) and the normalized channel gain of users (N_{slot}); The minimum number of data samples meeting data training (N_b); Number of data samples for a network training (N_g); Network data set ($D_{net} \neq \phi$).
Initialization: learning rate $\alpha = 10^{-4}$; regularization $\Delta = 10^{-4}$; Double-chain depth residual network parameters θ_i .
For $slot = 0,1,2...N_{slot} - 1 \quad do:$
Read the channel status of a time slot H
Generate a root node v_0
$\pi \leftarrow MCTS(v_0)$
$D_{net} \cup \{H, s, \pi, v\}$
If $slot^* (N + M + R) \geq N_b$;
Randomly extract N_g data samples from data set D_{net}
$(v, P_r) \leftarrow f_\theta (s, H)$
$\theta \leftrightarrow L = -\pi^T log P_r + (v - Q^*)^2 + \Delta \|\theta\|^2 + loss_p$
End if
End for
Output: double-chain deep residual network $f_\theta (s, H)$.

This section uses the power solution method in [54] to obtain the optimal transmission power of D2D users, cellular users and relay users in the D2D collaborative network as follows:

1) When DUE adopts direct transmission mode, the objective function is simplified as follows:

$$f_1(P_c, P_d) = \max_{\{P_c, P_d\}} \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=1}^{N} \left(V_{c_m,k}^d + V_{d_n,k}^d \right) \tag{24}$$

$$\frac{P_{c_m}^{\max} \left| h_{c_m}^k \right|^2}{\sigma^2 + P_{d_n}^{\max} \left| h_{d_n,b}^k \right|^2} = \tilde{\gamma}_{c_m}^{k(d)} \tag{25}$$

$$\frac{P_{d_n}^{\max} \left| h_{d_n}^k \right|^2}{\sigma^2 + P_{c_m}^{\max} \left| h_{c_m,d_n}^k \right|^2} = \tilde{\gamma}_{d_n}^{k(d)} \tag{26}$$

$$\tilde{\gamma}_{c_m}^{k(d)} \geq \gamma_c^{\min} \tag{27}$$

$$\tilde{\gamma}_{d_n}^{k(d)} \geq \gamma_{d_n}^{\min} \tag{28}$$

The optimal solution of the objective function:

$$f_1 \left(P_{c_m}^{k*}, P_{d_n}^{k*} \right) = \max_{\left(p_{c_m}^k, p_{d_n}^k \right)} \begin{bmatrix} f_1 \left(P_{c_m}^{\max}, \frac{\left(P_{c_m}^{\max} \left| h_{c_m,d_n}^k \right|^2 + \sigma^2 \right) \gamma_{d_n}^{\min}}{\left| h_{d_n}^k \right|^2} \right), \\ f_1 \left(P_{c_m}^{\max}, P_{d_n}^{\max} \right), \\ f_1 \left(\frac{\left(P_{d_n}^{\max} \left| h_{d_n,b}^k \right|^2 + \sigma^2 \right) \gamma_c^{\min}}{\left| h_{c_m}^k \right|^2}, P_{d_n}^{\max} \right) \end{bmatrix} \tag{29}$$

$$\left(P_{c_m}^{k*}, P_{d_i}^{k*} \right) = \arg\max_{\left(P_{c_m}^k, P_{d_i}^k \right)} f_1 \left(P_{c_m}^{k*}, P_{d_i}^{k*} \right) \tag{30}$$

When CUE multiplexes channels with DUE users in relaying mode, the objective function is simplified as follows:

$$f_2(P) = \max_P \sum_{k=1}^{K} \sum_{l=1}^{R} \sum_{n=1}^{N} \sum_{m=1}^{M} \left(V_{c_m}^r + V_{d_n,k}^c \right) \tag{31}$$

Power optimization is carried out in two stages. In the first stage, the transmission power of CUE and DUE is optimized. The objective of optimization is $f_2^1(P_c, P_d)$: Take c_m and cooperation d_n multiplexing channel k as an example, assuming that there is an optimal solution in the definition domain of $P_{c_m}^k, P_{d_i}^k$. When $P_{c_m}^k = P_{c_m}^{\max}$ and $P_{d_i}^k = P_{d_i}^{\max}$.

$$\frac{P_{c_m}^{\max} \left| h_{c_m}^k \right|^2}{\sigma^2 + P_{d_n}^{\max} \left| h_{d_n,b}^k \right|^2} = \tilde{\gamma}_{c_m}^{k(l_1)} \tag{32}$$

$$\frac{P_{d_n}^{\max} \left| h_{d_n,r_{l1}}^k \right|^2}{\sigma^2 + P_{c_m}^{\max} \left| h_{c_m,r_{l1}}^k \right|^2} = \tilde{\gamma}_{d_n \to r_{l1}}^{k(c)} \tag{33}$$

$$\tilde{\gamma}_{c_m}^{k(1)} \geq \gamma_c^{min} \tag{34}$$

$$\tilde{\gamma}_{d_n \to r_l}^{k(c)} \geq \gamma_{d_n}^{min} \tag{35}$$

The optimal solution of the objective function is:

$$f_2^1 \left(P_{c_m}^{k*}, P_{d_n}^{k*} \right) = \max_{\left(P_{c_m}^k, P_{d_n}^k \right)} \begin{bmatrix} f_2^1 \left(P_{c_m}^{\max}, \frac{\left(P_{c_m}^{\max} \left| h_{c_m,r_l}^k \right|^2 + \sigma^2 \right) \gamma_{d_n}^{\min}}{\left| h_{d_n,r_l}^k \right|^2} \right), \\ f_2^1 \left(P_{c_m}^{\max}, P_{d_n}^{\max} \right), \\ f_2^1 \left(\frac{\left(P_{d_n}^{\max} \left| h_{d_n,b}^k \right|^2 + \sigma^2 \right) \gamma_c^{\min}}{\left| h_{c_m}^k \right|^2} \right), P_{d_n}^{\max} \end{bmatrix} \tag{36}$$

$$\left(P_{c_m}^{k*}, P_{d_n}^{k*} \right) = argmax_{\left(P_{c_m}^k, P_{d_n}^k \right)} f_2^1 \left(P_{c_m}^{k*}, P_{d_n}^{k*} \right) \tag{37}$$

In the second stage, the transmission power of CUE and RUE is optimized. The optimization objectives are $f_2^2(P_c, P_r)$. If channel k is multiplexed c_m, r_l with d_n,

if $P_{c_m}^k = P_{c_m}^{\max}$ and $P_{r_l}^k = P_{r_l}^{\max}$:

$$\frac{P_{c_m}^{\max}\left|h_{c_m}^k\right|^2}{\sigma^2 + P_{r_l}^{\max}\left|h_{r_l,b}^k\right|^2} = \tilde{\gamma}_{c_m}^{k(2)} \tag{38}$$

$$\frac{P_{r_l}^{\max}\left|h_{r_l,d_n}^k\right|^2}{\sigma^2 + P_{c_m}^{\max}\left|h_{c_m,d_n}^k\right|^2} = \tilde{\gamma}_{r_l\to d_n}^{k(c)} \tag{39}$$

$$\tilde{\gamma}_{c_m}^{k(2)} \geq \gamma_c^{min} \tag{40}$$

$$\tilde{\gamma}_{\gamma_l\to d_n}^{k(c)} \geq \gamma_{d_n}^{min} \tag{41}$$

The optimal solution of the objective function is:

$$f_2^2\left(P_{c_m}^{k*}, P_{r_l}^{k*}\right) = \max_{\left(P_{c_m}^k, P_{r_l}^k\right)} \begin{bmatrix} f_2^2\left(P_{c_m}^{\max}, \frac{\left(P_{cm}^{\max}\left|h_{c_m,d_n}^k\right|^2+\sigma^2\right)\gamma_{d_n}^{\min}}{\left|h_{r_l,d_n}^k\right|^2}\right), \\ f_2^2\left(P_{c_m}^{\max}, P_{r_l}^{\max}\right), \\ f_2^2\left(\frac{P_{r_l}^{\max}\left|h_{r_l,d_n}^k\right|^2-\gamma_{r_l}^{\min}\sigma^2}{\gamma_{d_n}^{\min}\left|h_{c_m,d_n}^k\right|^2}, P_{r_l}^{\max}\right) \end{bmatrix} \tag{42}$$

$$\left(P_{c_m}^{k*}, P_{r_l}^{k*}\right) = \underset{\left(P_{c_m}^k, P_{r_l}^k\right)}{argmax}\, f_2^2\left(P_{c_m}^{k*}, P_{r_l}^{k*}\right) \tag{43}$$

The process flow of D2D cooperative network channel allocation and power control scheme is shown in the following figure 6:

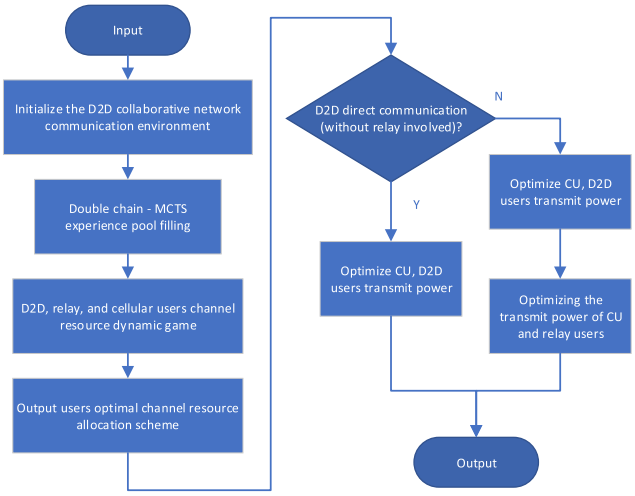

FIGURE 6. Flow chart of D2D communication resource allocation algorithm based on double-chain-MCTS.

IV. SIMULATION RESULTS AND ANALYSIS

A. SIMULATION EXPERIMENT PARAMETER SETTING

The simulation environment in this article consists of a base station, a cellular user, a D2D user, and a relay user. The cellular user, a D2D user, and a relay user are randomly distributed within the base station, with two cellular users, two D2D users, and two relay users. The main parameters in other simulation experiments are shown in Table 2.

TABLE 2. System simulation parameter.

Simulation parameters	Parameter value(unit)
Base station coverage radius	500 m
Distance between D2D devices	50 m – 200 m
Channel bandwidth B	1 Hz
Gaussian white noise power	−20 dBm
Path attenuation factor of channel	3
Cellular users transmit power	23 dB
Minimum data sample size N_b	400
Neural network training batch N_g	64
Double chain network learning rate	10^{-4}
Double chain network regularization	10^{-4}
Single chain network learning rate	10^{-4}
Single chain network regularization	10^{-4}
Double Chain network Residual Layers	6 pieces per branch
Single Chain network Residual Layers	8 pieces
Double chain network hidden layer (Activation function)	Tanh
Single chain network hidden layer (Activation function)	Relu
Double chain network output layer (Activation function)	Softmax and Tanh
Single chain network output layer (Activation function)	Softmax and Tanh

B. ANALYSIS OF SIMULATION EXPERIMENT RESULTS

1) COMPUTATIONAL COMPLEXITY

Next, this section will introduce the algorithm time complexity of MCTS, single-chain-based deep residual network MCTS, and double-chain-based deep residual network MCTS. The algorithmic complexity of MCTS is $O(D * |A| * C_p * N_p)$. Among them, D: the depth of the search tree; A: represents the action set of each state; C_p represents the search parameters of the MCTS algorithm; N_p represents the number of visits to this node.

The algorithmic complexity of the single-chain deep residual network MCTS is: the sum of the time complexities of the MCTS and the single-chain deep residual network.

MCTS:

$$O\left(D * |A| * C_p * N_P\right) \tag{44}$$

single-chain:

$$O\left(N_{\{ep\}} * k * H * W * C * \log_2 c + N_{\{ResNet\}} * T_{\{ResNet\}}\right) \tag{45}$$

Among them, $N_{\{ep\}}$ represents the number of training rounds; k represents the size of the convolution kernel; H, W, C represents the channel status of D2D users, relays, and cellular users; $\log_2 c$ Indicates the number of available channels; $N_{\{ResNet\}}$ represents the time complexity of forward propagation and back propagation of the ResNet network; T_{ResNet} represents the number of layers of the ResNet network.

The algorithmic complexity of the double-chain deep residual network MCTS is: the sum of the time complexities of MCTS and double -chain deep residual network.

MCTS:

$$O\left(D * |A| * C_{p'} * N_{P'}\right) \tag{46}$$

double -chain:

$$O\left(N_{\{ep\}} * k * H * W * C * \log_2 c + 2 * N_{\{DPN\}} * T_{\{DPN\}}\right) \tag{47}$$

Among them, $C_{p'}$ represents the search parameters of the deciduous MCTS algorithm; $N_{P'}$ represents the visit times of the deciduous MCTS algorithm nodes. $N_{\{ep\}}$ represents the number of training rounds; represents the size of the convolution kernel; represents the channel status of D2D users, relays, and cellular users respectively; $\log_2 c$ indicates the number of available channels; $N_{\{DPN\}}$ represents the time complexity of forward propagation and back propagation of the DPN network; $T_{\{DPN\}}$ represents the number of layers of the DPN network.

Although the time complexity of the parallel double-chain deep residual network proposed in this paper is relatively high, the overall algorithm time complexity is not particularly high due to the existence of the leaf removal mechanism, so the algorithm proposed in this paper can be better applied to small-scale in the D2D collaboration network.

Algorithm complexity is an integral part of D2D communication resource allocation. In order to prove the advantages of the proposed algorithm in terms of algorithm complexity and algorithm reliability, this paper conducts comparative experiments on algorithm loss functions (double-chain deep deciduous MCTS, single-chain deep MCTS, and MCTS) from the perspective of algorithm convergence speed and convergence value. In the experiment, the transmit power of the D2D user is 200 mW, and 2500 iterations of the three algorithms are tested. Figure 6 is a comparison chart of the convergence trend of the loss function. It can be seen from the trend graph that the double-chain deep leaf-deleting MCTS (Dcdd-MCTS) algorithm converges successfully, and the final convergence value is better than that of single-chain deep MCTS (Scd-MCTS) and MCTS, which proves the reliability of the algorithm proposed in this paper. On the other hand, the trend graph shows that Dcdd-MCTS is better than MCTS in algorithm convergence speed, and slightly slower than Scd-MCTS. At the same time, it can be seen that Dcdd-MCTS, Scd-MCTS, and MCTS converged 500, 450, and 800 times, respectively. This is because the algorithm proposed in this paper has a double-chain structure and a leaf removal mechanism, which reduces the complexity of algorithm operations. At the same time, it can be found that the convergence value of the algorithm proposed in this paper is the smallest, because this paper adopts the double-chain deep leaf-leaf MCTS algorithm to improve the system's data feature mining ability and learning ability.

Figures 8, 10, and 12 in this paper show the simulation experiments of system throughput, and Figures 9, 11, and 13 show the simulation experiments of outage probability. Combining the simulation experiments of system throughput and outage probability, compared with Scd-MCTS and joint Optimization algorithm and MCTS-related algorithms, it can be found that the algorithm proposed in this paper has a faster algorithm convergence speed, a lower outage probability, and the ability to output higher throughput. The specific analysis is shown in Figure 8-13.

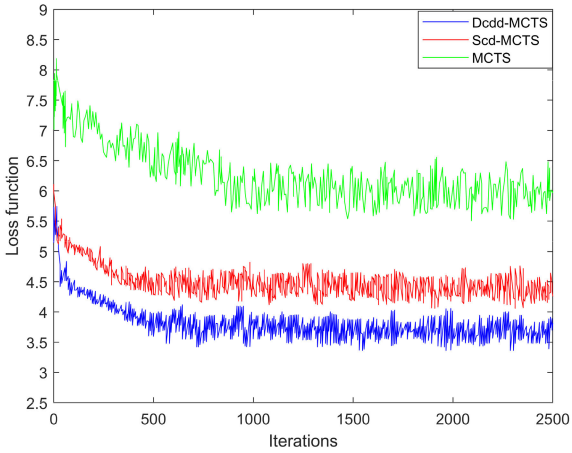

FIGURE 7. Comparison chart of loss function convergence trend.

Figure 7 and Figure 8 show the relationship between the D2D maximum transmission power and the total system throughput, and the relationship between the D2D maximum transmission power and the outage probability. In order to improve the effectiveness of the algorithm proposed in this paper, linear programming, a joint optimization algorithm for resource allocation based on graph coloring (Join-algorithm) [52], MCTS, and single-chain deep MCTS (Scd-MCTS) are used as a comparison of double-chain leaf depth MCTS algorithm.

It can be seen from Figure 7 that with the increase of D2D transmission power, the system throughput of the four algorithms continues to grow steadily. In terms of system throughput, Dcdd-MCTS > Scd-MCTS > MCTS > Join-algorithm > Linear -program, this is because the double-chain deep deciduous MCTS has good data learning ability; at the same time, it can be seen that in 0.05- Between 0.75w, the output curve of the algorithm proposed in this paper is relatively steep, which is due to the existence of the deciduous mechanism to accelerate the convergence of the algorithm.

It can be seen from Figure 8 that as the D2D transmission power increases, the outage probability of the four algorithms decreases continuously, Dcdd-MCTS < Scd-MCT < MCTS < Join-algorithm < linear-programming. This is because except for the linear programming algorithm, other algorithms include the idea of joint optimization. At the same time, it can be seen that in the range of 0.125-0.2w, the interruption probability of the proposed algorithm curve in this paper is close to a uniform decrease, and the value is the

lowest. On the one hand, this performance proves that the algorithm proposed in this paper is more suitable for actual communication experiments, and also proves the effectiveness of the algorithm.

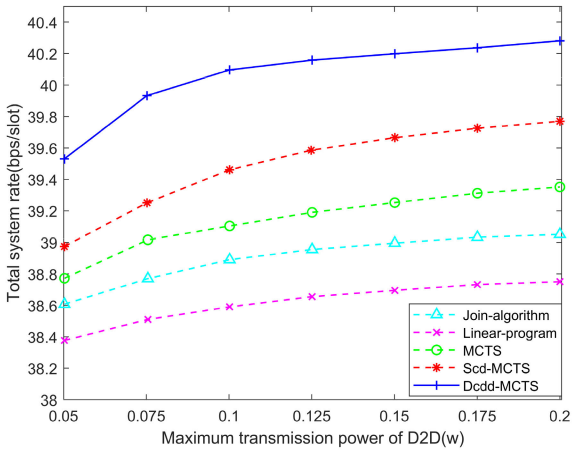

FIGURE 8. Relation curve between D2D maximum transmission power and total system throughput.

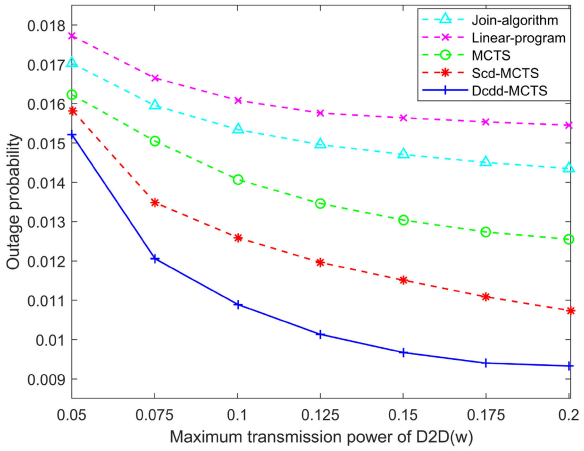

FIGURE 9. Relation curve between D2D maximum transmission power and outage probability.

Figure 9 and Figure 10 show the relationship between the D2D minimum communication rate limit and the total throughput of the system, as well as the relationship between the D2D minimum communication rate limit and the outage probability.

It can be seen from Figure 9 that as the minimum communication rate limit of D2D increases, the system throughput of the four algorithms continues to decrease, but Dcdd-MCTS > Scd-MCTS > MCTS > Join-algorithm > Linear-program. This is because compared to linear programming, Join-algorithm only has the ability to optimize resource allocation from multiple angles, MCTS and its related algorithms have reverse self-learning ability, and the algorithm proposed in this paper not only has deep learning ability but also has leaf search mechanism and moderate The leaf search mechanism and the idea of joint optimization, so the performance is the best.

It can be seen from Figure 10 that with the increase of the D2D minimum communication rate limit, the outage probability decreases continuously, Dcdd-MCTS < Scd-MCTS < MCTS < Join-algorithm < Linear-program. At the same time, it can be seen that at 0.5bps-2bps, the trend curves of other algorithms are more stable than those of the Linear-program, because the Join-algorithm and MCTS related algorithms are both in the D2D collaborative network from multiple perspectives. This method greatly improves the interference coordination and resource allocation capabilities of the network.

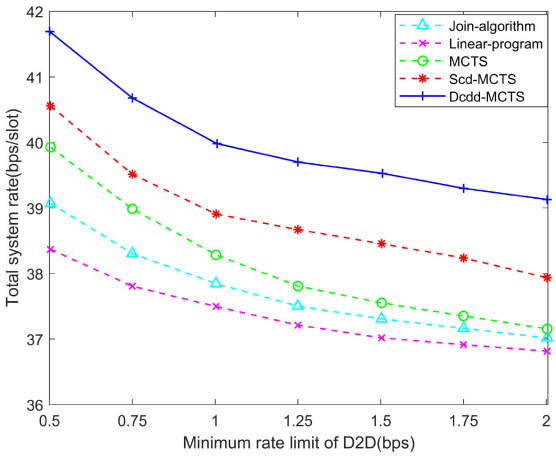

FIGURE 10. Relation curve between D2D minimum communication rate limit and total system throughput.

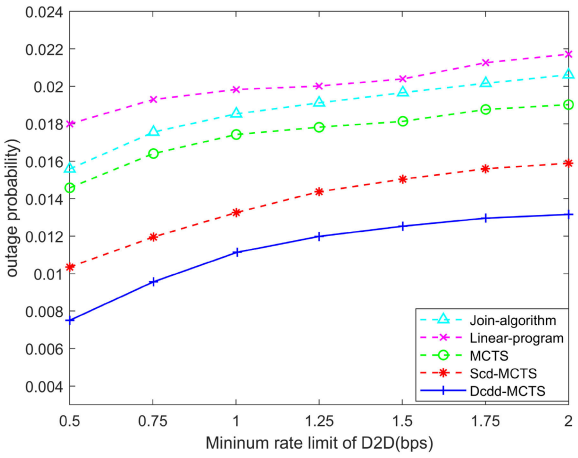

FIGURE 11. Relation curve between D2D minimum transmission rate limit and outage probability.

Figure 11 and Figure 12 are experimental results of the relationship between the distance between D2D devices and the total throughput of the system, as well as the relationship between D2D devices and outage probability.

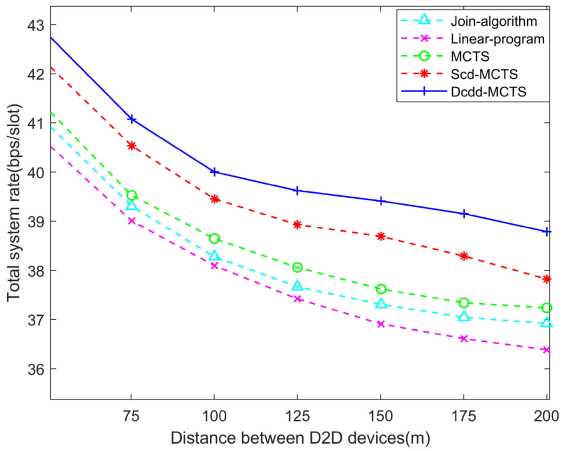

FIGURE 12. Relation curve between distance between devices and total system throughput.

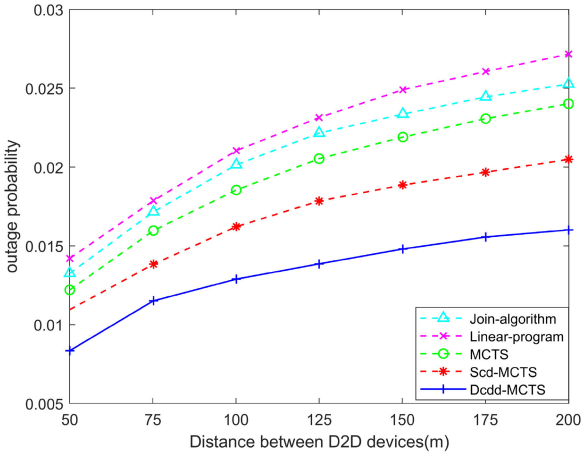

FIGURE 13. Relation curve between distance between D2D equipment and outage probability.

It can be seen from Figure 11 that as the distance between D2D devices increases, the system throughput of the four algorithms continues to decrease, but Dcdd-MCTS > Scd-MCTS > MCTS > Join-algorithm > Linear-program. This is because the self-learning ability of the double-chain deep learning algorithm is better than that of the single-chain deep learning algorithm and MCTS. Although joint optimization has the idea of joint optimization, it does not have self-learning ability, so the output curve is between Linear-program and MCTS. At the same time, it can be seen from Figure 12 that as the distance between D2D devices increases, the outage probability also increases, Dcdd-MCTS < Scd-MCTS < MCTS < Join-algorithm < Linear-program. This is because Linear-program and Join-algorithm do not have data feature mining and self-learning capabilities, while MCTS and its related algorithms have deep data feature mining and learning capabilities. The algorithm proposed in this paper contains more complex network structures and feature fusion mechanisms. so it has better system performance optimization capability.

In addition, it can be found that the closer the distance between D2D, the higher the throughput and the lower the interruption probability. This is due to Shannon's theorem. It can be seen from Shannon's theorem that the distance between devices is positively correlated with the user's interference value and negatively correlated with the user's throughput.

V. CONCLUSION

This paper studies the problem of resource allocation in D2D collaborative networks. A D2D cooperative communication network resource allocation algorithm based on improved Monte Carlo tree search is proposed. First, the optimization problem is modeled as a finite MDP process, and then a resource allocation model for a double-chain deep deciduous MCTS network is constructed. Using the prior probability and action value evaluation generated by deciduous MCTS to select the optimal action, and then use the optimal value in the MCTS self-game as the label to continuously update and train the double-chain deep residual network. The simulation results show that compared with single-chain deep MCTS and joint optimization algorithm [52], the proposed algorithm in this paper increases the system throughput by 5%, 2%, respectively, and reduces the outage probability by 33%, 18%.

However, on the one hand, this paper greatly improves the system throughput, but the complexity of the algorithm is slightly higher, which is not suitable for large-scale D2D collaborative networks. At the same time, with the rapid development of mobile networks, the degree of network heterogeneity will be further deepened. The next step will be to carry out research on multi-D2D collaborative network resource allocation technology in multi-cell scenarios.

REFERENCES

[1] M. Ashraf, B. Tan, D. Moltchanov, J. S. Thompson, and M. Valkama, "Joint optimization of radar and communications performance in 6G cellular systems," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 1, pp. 522–536, Mar. 2023.

[2] I. Bilbao, L. Fanari, E. Iradier, P. Angueira, and J. Montalban, "Sparse vector coding for short-packet transmission on industrial communications: Reference architecture and design challenges," *IEEE Open J. Ind. Electron. Soc.*, vol. 4, pp. 1–13, 2023.

[3] B. Xiong, Z. Zhang, and H. Jiang, "Reconfigurable intelligent surface for mmWave mobile communications: What if LoS path exists?" *IEEE Wireless Commun. Lett.*, vol. 12, no. 2, pp. 247–251, Feb. 2023.

[4] Y. Liu, M. Li, J. Zhang, M. Wu, and L. Li, "Deep learning aided two-stage multi-finger beam training in millimeter-wave communication," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 26–30, Jan. 2023.

[5] A. Dogra, R. K. Jha, and K. R. Jha, "Intelligent routing for enabling haptic communication in 6G network," in *Proc. 15th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Bengaluru, India, Jan. 2023, pp. 419–422.

[6] Y. Tang, N. Zhou, Q. Yu, D. Wu, C. Hou, G. Tao, and M. Chen, "Intelligent fabric enabled 6G semantic communication system for in-cabin scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 1, pp. 1153–1162, Jan. 2023.

[7] M. Tian, Z. Zhang, Q. Xu, and L. Yang, "A personalized solution for deep learning-based mmWave beam selection," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 183–186, Jan. 2023.

[8] B. Zhao, M. Wang, Z. Xing, G. Ren, and J. Su, "Integrated sensing and communication aided dynamic resource allocation for random access in satellite terrestrial relay networks," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 661–665, Feb. 2023.

[9] Z. Wei, H. Qu, Y. Wang, X. Yuan, H. Wu, Y. Du, K. Han, N. Zhang, and Z. Feng, "Integrated sensing and communication signals towards 5G-A and 6G: A survey," *IEEE Internet Things J.*, early access, Jan. 9, 2023, doi: 10.1109/JIOT.2023.3235618.

[10] J. Zou, C. Wang, Y. Liu, Z. Zou, and S. Sun, "Vision-assisted 3-D predictive beamforming for green UAV-to-vehicle communications," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 1, pp. 434–443, Mar. 2023.

[11] E. Andrianopoulos et al., "Real-time sub-THz link enabled purely by optoelectronics: 90–310 GHz seamless operation," *IEEE Photon. Technol. Lett.*, vol. 35, no. 5, pp. 237–240, Mar. 1, 2023.

[12] P. Singh, H. Jeon, S. Yun, B. W. Kim, and S.-Y. Jung, "Vehicle positioning based on optical camera communication in V2I environments," *Comput., Mater. Continua*, vol. 72, no. 2, pp. 2927–2945, 2022.

[13] P. Singh, B. W. Kim, and S.-Y. Jung, "DS-OOK for terahertz band nanonetworks," *Nat. Acad. Sci. Lett. USA*, vol. 44, no. 1, pp. 43–46, Feb. 2021.

[14] T. Yu, S. Zhang, X. Chen, and X. Wang, "A novel energy efficiency metric for next-generation green wireless communication network design," *IEEE Internet Things J.*, vol. 10, no. 2, pp. 1746–1760, Jan. 2023.

[15] N. S. Saatchi, H. Yang, and Y. Liang, "Novel adaptive transmission scheme for effective URLLC support in 5G NR: A model-based reinforcement learning solution," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 109–113, Jan. 2023.

[16] A. Etemadi, M. Farahnak-Ghazani, H. Arjmandi, M. Mirmohseni, and M. Nasiri-Kenari, "Abnormality detection and localization schemes using molecular communication systems: A survey," *IEEE Access*, vol. 11, pp. 1761–1792, 2023.

[17] J.-J. Kao, C.-L. Lin, and J. Yang, "Adaptive wireless power transfer system with relay transmission and communication," *IEEE Trans. Power Electron.*, vol. 38, no. 3, pp. 4110–4123, Mar. 2023.

[18] K. B. Devika, G. Rohith, and S. C. Subramanian, "Impact of V2V communication on energy consumption of connected electric trucks in stable platoon formation," in *Proc. 15th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Bengaluru, India, Jan. 2023, pp. 42–47.

[19] D. F. Cotton, "Intelligent resource management with deep reinforcement learning in device-to-device communication," Ph.D. dissertation, 2022.

[20] P. R. Teja and P. K. Mishra, "Sealed bid single price auction model (SBSPAM)-based resource allocation for 5G networks," *Wireless Pers. Commun.*, vol. 116, no. 3, pp. 2633–2650, Feb. 2021.

[21] D. Zhang, Y. Fang, Y. Zhou, J. He, and Y. Zhang, "Game theoretic multihop D2D content sharing: Joint participants selection, routing, and pricing," *IEEE Trans. Mobile Comput.*, vol. 21, no. 6, pp. 2013–2028, Jun. 2022.

[22] M. V. S. Aditya, H. Pancholi, P. Priyanka, and G. S. Kasbekar, "Beyond the VCG mechanism: Truthful reverse auctions for relay selection with high data rates, high base station utility and low interference in D2D networks," *Wireless Netw.*, vol. 26, no. 5, pp. 3861–3882, Jul. 2020.

[23] Z. Zhang, Y. Wu, X. Chu, and J. Zhang, "Energy-efficient transmission rate selection and power control for relay-assisted device-to-device communications underlaying cellular networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 8, pp. 1133–1136, Aug. 2020.

[24] Y. Yuan, T. Yang, Y. Hu, H. Feng, and B. Hu, "Two-timescale resource allocation for cooperative D2D communication: A matching game approach," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 543–557, Jan. 2021.

[25] A. K. Lamba, R. Kumar, and S. Sharma, "A robust Stackelberg game approach for joint relay selection and optimal power allocation for cooperative device-to-device communication under channel uncertainties," *Wireless Pers. Commun.*, vol. 110, no. 1, pp. 169–183, Jan. 2020.

[26] S. Selmi and R. Bouallègue, "Energy and spectral efficient relay selection and resource allocation in mobile multi-hop device to device communications," *IET Commun.*, vol. 15, no. 14, pp. 1791–1807, Aug. 2021.

[27] Y. Liu, W. Wang, H. Chen, L. Wang, N. Cheng, W. Meng, and X. Shen, "Secrecy rate maximization via radio resource allocation in cellular underlaying V2V communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7281–7294, Jul. 2020.

[28] A. K. Hamid, F. N. Al-Wesabi, N. Nemri, A. Zahary, and I. Khan, "An optimized algorithm for resource allocation for D2D in heterogeneous networks," *Comput., Mater. Continua*, vol. 70, no. 2, pp. 2923–2936, 2022.

[29] M. M. Salim, D. Wang, H. A. E. A. Elsayed, Y. Liu, and M. A. Elaziz, "Joint optimization of energy-harvesting-powered two-way relaying D2D communication for IoT: A rate–energy efficiency tradeoff," *IEEE Internet Things J.*, vol. 7, no. 12, pp. 11735–11752, Dec. 2020.

[30] H. Gao, S. Zhang, Y. Su, and M. Diao, "Joint resource allocation and power control algorithm for cooperative D2D heterogeneous networks," *IEEE Access*, vol. 7, pp. 20632–20643, 2019.

[31] M. M. Salim, H. A. Elsayed, M. A. Elaziz, M. M. Fouda, and M. S. Abdalzaher, "An optimal balanced energy harvesting algorithm for maximizing two-way relaying D2D communication data rate," *IEEE Access*, vol. 10, pp. 114178–114191, 2022.

[32] W. Zhuang, M. Chen, X. Wei, and H. Li, "Social-aware resource allocation based on cluster formation and matching theory in D2D underlaying cellular networks," *KSII Trans. Internet Inf. Syst.*, vol. 14, no. 5, pp. 1984–2002, 2020.

[33] D. Feng, X. Huang, W. Jiang, Y. Sun, S. Xiao, C. He, and F. Zheng, "Power-spectrum trading for full-duplex D2D communications in cellular networks," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 4, pp. 2016–2026, Dec. 2021.

[34] O. M. El-Nakhla, M. I. Obayya, and S. E. Kishk, "Stable matching relay selection (SMRS) for TWR D2D network with RF/RE EH capabilities," *IEEE Access*, vol. 10, pp. 22381–22391, 2022.

[35] A. Amer, A.-M. Ahmad, and S. Hoteit, "Resource allocation for downlink full-duplex cooperative NOMA-based cellular system with imperfect Si cancellation and underlaying D2D communications," *Sensors*, vol. 21, no. 8, p. 2768, Apr. 2021.

[36] R. Gour and A. Tyagi, "Joint uplink–downlink resource allocation for energy efficient D2D underlaying cellular networks with many-to-one matching," *Phys. Commun.*, vol. 58, Jun. 2023, Art. no. 102016.

[37] R. Gour and A. Tyagi, "Semi-distributed resource management for underlay D2D communication with user's cooperation," *Int. J. Commun. Syst.*, vol. 33, no. 4, Mar. 2020, Art. no. e4243.

[38] J. Ji, K. Zhu, D. Niyato, and R. Wang, "Joint trajectory design and resource allocation for secure transmission in cache-enabled UAV-relaying networks with D2D communications," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1557–1571, Feb. 2021.

[39] Y. Xu, Z. Liu, C. Huang, and C. Yuen, "Robust resource allocation algorithm for energy-harvesting-based D2D communication underlaying UAV-assisted networks," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 17161–17171, Dec. 2021.

[40] G. Feng, X. Qin, Z. Jia, and S. Li, "Energy efficiency resource allocation for D2D communication network based on relay selection," *Wireless Netw.*, vol. 27, no. 5, pp. 3689–3699, Jul. 2021.

[41] Y. Ma, T. Liu, L. Cui, X. Yin, and Q. Liu, "Robust resource allocation with power outage guarantees for energy harvesting aided device-to-device communication," *IEEE Access*, vol. 8, pp. 124563–124578, 2020.

[42] Y. Li, G. Xu, K. Yang, J. Ge, P. Liu, and Z. Jin, "Energy efficient relay selection and resource allocation in D2D-enabled mobile edge computing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15800–15814, Dec. 2020.

[43] V. Hakami, H. Barghi, S. Mostafavi, and Z. Arefinezhad, "A resource allocation scheme for D2D communications with unknown channel state information," *Peer-to-Peer Netw. Appl.*, vol. 15, no. 2, pp. 1189–1213, Mar. 2022.

[44] X. Zhong, Y. Guo, N. Li, and Y. Chen, "Joint optimization of relay deployment, channel allocation, and relay assignment for UAVs-aided D2D networks," *IEEE/ACM Trans. Netw.*, vol. 28, no. 2, pp. 804–817, Apr. 2020.

[45] N. Su and Q. Zhu, "Outage performance analysis and resource allocation algorithm for energy harvesting D2D communication system," *Wireless Netw.*, vol. 26, no. 7, pp. 5163–5176, Oct. 2020.

[46] Z. Ali, G. A. S. Sidhu, F. Gao, J. Jiang, and X. Wang, "Deep learning based power optimizing for NOMA based relay aided D2D transmissions," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 3, pp. 917–928, Sep. 2021.

[47] X. Zhong, Y. Guo, N. Li, and S. Li, "Joint relay assignment and channel allocation for opportunistic UAVs-aided dynamic networks: A mood-driven approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15019–15034, Dec. 2020.

[48] A. Salh, R. Ngah, L. Audah, K. S. Kim, Q. Abdullah, Y. M. Al-Moliki, K. A. Aljaloud, and H. N. Talib, "Energy-efficient federated learning with resource allocation for green IoT edge intelligence in B5G," *IEEE Access*, vol. 11, pp. 16353–16367, 2023.

[49] A. Salh, R. Ngah, G. A. Hussain, L. Audah, M. Alhartomi, Q. Abdullah, R. Alsulami, S. Alzahrani, and A. Alzahmi, "Intelligent resource management using multiagent double deep *Q*-networks to guarantee strict reliability and low latency in IoT network," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 2245–2257, 2022.

[50] X. Li, G. Chen, G. Wu, Z. Sun, and G. Chen, "Research on multi-agent D2D communication resource allocation algorithm based on A2C," *Electronics*, vol. 12, no. 2, p. 360, Jan. 2023.

[51] A. Salh, L. Audah, K. S. Kim, S. H. Alsamhi, M. A. Alhartomi, Q. Abdullah, F. A. Almalki, and H. Algethami, "Refiner GAN algorithmically enabled deep-RL for guaranteed traffic packets in real-time URLLC B5G communication systems," *IEEE Access*, vol. 10, pp. 50662–50676, 2022.
[52] M. Liu and L. Zhang, "Graph colour-based resource allocation for relay-assisted D2D underlay communications," *IET Commun.*, vol. 14, no. 16, pp. 2701–2708, Oct. 2020.
[53] Y. He, H. U. Khan, K. Zhang, W. Wang, B. J. Choi, A. A. Aly, B. F. Felemban, N. S. Sani, Q. A. Tarbosh, and Ö. Aydogdu, "D2D-V2X-SDN: Taxonomy and architecture towards 5G mobile communication system," *IEEE Access*, vol. 9, pp. 155507–155525, 2021.
[54] W. Lee, "Resource allocation for multi-channel underlay cognitive radio network based on deep neural network," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1942–1945, Sep. 2018.

XINZHOU LI received the bachelor's degree in communication engineering from the Changchun University of Science and Technology, where he is currently pursuing the Ph.D. degree. His current research interests include resource allocation, D2D communication, interference management, and heterogeneous networks.

GUIFEN CHEN received the B.S. and M.S. degrees in information and communication engineering from the Jilin University of Technology, China, in 1986 and 1991, respectively, and the Ph.D. degree in optical engineering from the Changchun University of Science and Technology, China, in 2009.

She is currently a Professor of information and communication engineering with the Changchun University of Science and Technology. Her research interests include optical information and wireless communication technology, the Internet of Things, and sensor networks.