A Joint Reinforcement-Learning Enabled Caching and Cross-Layer Network Code for Sum-Rate Maximization in F-RAN with D2D Communications

Mohammed S. Al-Abiad, Student Member, IEEE, Md. Zoheb Hassan, Student Member, IEEE, and Md. Jahangir Hossain, Senior Member, IEEE

Abstract—In this paper, we leverage reinforcement learning (RL) and cross-layer network coding (CLNC) for efficiently prefetching users’ contents to the local caches and delivering these contents to users in a downlink fog-radio access network (F-RAN) with device-to-device (D2D) communications. In the considered system, fog access points (F-APs) and cache-enabled D2D (CE-D2D) users are equipped with local caches for alleviating traffic burden at the fronthaul, while users’ contents can be easily and quickly accommodated. In CLNC, the coding decisions take users’ contents, their rates, and power levels of F-APs and CE-D2D users into account, and RL optimizes caching strategy. Towards this goal, a joint content placement and delivery problem is formulated as an optimization problem with a goal to maximize system sum-rate. For this NP-hard problem, we first develop an innovative decentralized CLNC coalition formation (CLNC-CF) algorithm to obtain a stable solution for the content delivery problem, where F-APs and CE-D2D users utilize CLNC resource allocation. By taking the behavior of F-APs and CE-D2D users into account, we then develop a multi-agent RL (MARL) algorithm for optimizing the content placements at both F-APs and CE-D2D users. Simulation results show that the proposed joint CLNC-CF and RL framework can effectively improve the sum-rate by up to 30%, 60%, and 150%, respectively, compared to: 1) an optimal uncoded algorithm, 2) a standard rate-aware NC algorithm, and 3) a benchmark classical NC with network-layer optimization.

Index Terms—Caching, D2D communications, F-RAN, NC, resource and power allocation, reinforcement learning.

I. INTRODUCTION

Fog radio access network (F-RAN) has recently given significant attention for beyond-5G era while leveraging the centralized processing of a cloud-RAN (C-RAN) and intelligence of the network edge. In addition, F-RAN takes advantage of a fast access of the contents through distributed local caches at the fog access points (F-APs) [1], [2]. Prior F-RAN systems, the particular popular contents were streamed from cloud base stations (CBSs) to network edge [3]. Essentially, these popular contents require duplicated downloads from the CBS, and such duplicate downloads severely degrade system performance. By caching the popular contents in the F-APs, the demands from users can be accommodated easily with minimum intervention of CBSs. Hence, F-RAN significantly alleviates traffic burden at the fronthaul and improves system performance [4]. However, in beyond 5G networks, caching the increased popular contents at F-APs, owing to their equipment cost and size issue, is a key concern.

In order to overcome this concern, distributed caching is envisioned, and caching at cache-enabled device-to-device (CE-D2D) users is employed [5], [6]. Therefore, the performance of an F-RAN can be further improved by implementing D2D communications [7], where caching at F-APs and CE-D2D users are leveraged [8]. With the distributed caching in an F-RAN, F-APs and CE-D2D users transmit their cached contents to the interested users via cellular and D2D links, respectively. However, pre-fetching popular contents to F-APs and CE-D2D users for effective delivery needs a careful optimization.

Evidently, a significant portion of these popular contents are delay-sensitive, and it is crucial to efficiently pre-fetch them to F-APs and CE-D2D users for immediate and effective delivery. Therefore, developing an innovative content placement and content delivery framework is imperative for harnessing the aforementioned benefits of F-RANs.

The joint content placement and content delivery optimization problem in F-RAN has been investigated separately. Particularly, the content delivery-based network coding (NC) problem was solved using Random Linear NC (RLNC) [10]–[12] and Instantly Decodable NC (IDNC) [13]–[25]. RLNC offers an optimal throughputs maximization [10], but it is not a suitable technique for delivering delay-sensitive contents for real-time applications that require instant content decoding. In contrast, IDNC [13]–[25] offers an immediate content delivery, and consequently, it provides fast and instantaneous decoding process that is affordable for real-time applications, e.g., streaming applications. In the contemporary literature, the works in [26]–[30] developed simple caching schemes for content placement optimization problem in small network settings. However, these works did not harvest the benefits of IDNC to multiplex many users to the same resource block. Unlike all the aforementioned works, our work considers a joint optimization of content placement and content delivery problem that will be referred as sum-rate maximization problem.

A. Related Works and Motivation

Most existing and relevant works on F-RANs focused on user scheduling problem in order to maximize sum-rate, e.g., [31]–[33]. Specifically, the study in [31] included power allocation optimization for the F-APs to further improve the sum-rate. However, all of these works viewed the network solely from the physical-layer perspective without taking into consideration upper-layer facts, e.g., combining users’ requests. As a
result, only a single user was assigned to each F-AP, and it is not affordable for large scale network. It has been noticed that users tend to stream a popular video, and consequently, users have a common interest in requesting same content within a small interval of time. This frequently happens in a hotspot, e.g., a playground, a public transport, a conference hall. Actually, transmitting requested contents to users without being combined severely degrades system performance. Therefore, IDNC[13] can wisely select a combination of contents (i.e., using the binary XoR combination) that can multiplex a subset of interested users to the same resource block.

In IDNC-based networks, the content delivery problem was investigated for various wireless networking scenarios, e.g., point-to-multipoint (PMP) [13], D2D networks [15–17], D2D F-RANs [18]. For example, in [18], the authors developed a centralized D2D F-RAN scheme for completion time reduction. Unfortunately, the above related works primarily relied on optimization at the network layer, and their main limitation is that the transmission rate of each F-AP is selected according to the user with the weakest channel quality. This is inefficient because the minimum selected transmission rate results in the prolonged file reception time and thus, consumes network time resources. To this end, two advancements aimed at developing promising techniques for improving content delivery, namely, (i) rate aware-IDNC (RA-IDNC) and (ii) cross-layer NC (CLNC) schemes.

Related works used RA-IDNC scheme: In RA-IDNC, the coding decisions depend on content combinations at the network layer and transmission rate at the physical layer. Such scheme was first proposed in [19] for completion time minimization in PMP system. In [21], the authors used RA-IDNC in a practical and promising paradigm of C-RANs. However, the authors assumed that all F-APs use a fixed transmit power level. Moreover, for synchronization purpose, the same transmission rate (i.e., the lowest transmission rate of all F-APs as in [21]) is selected, and it is impractical. In fact, it violates the QoS rate guarantee and leads to a longer time for content delivery. In addition, the aforementioned RA-IDNC works ignored the potentials of D2D communications. Addressing these RA-IDNC limitations is imperative for harvesting the benefits of next-generation F-RANs.

Related Works used CLNC scheme: The authors in [22], developed CLNC scheme to optimize the employed rates in RA-IDNC decisions using power control on each F-AP in C-RANs. Essentially, the coding decisions in CLNC scheme not only depends on NC and users’ rate, but also on the power levels of each F-AP. Accordingly, CLNC is a promising technique for significantly improving sum-rate [23], cloud offloading [24], and delay [25]. Particularly, the authors developed CLNC schemes for cloud offloading maximization in F-RAN [24] and for delay minimization in D2D-aided F-RAN [25]. However, all of these CLNC works ignored the content placement optimization problem that draws its importance from improving the delivery time of contents to users. More importantly, these works ignored the decision-making capability of F-APs and CE-D2D users, and proposed a graph-theoretical centralized solution. Note that in CLNC-based dense network, the CBS requires to generate a vertex for each possible NC combination, and thus, its complexity is exponentially increased [22]. Essentially, for a large scale network, these centralized CLNC solutions may not be affordable in practice.

The emerging distributed machine-learning (ML) based algorithms are postulated to address issues of the conventional optimization-based resource allocation algorithms. Specifically, by training a neural-network offline with large samples, ML-based algorithms allow the radio resource controller to rapidly determine resource allocation decisions while requiring low signaling overhead [34]. Recently, ML-based resource optimizations have been developed for caching and power allocation in F-RAN [35], [36]. In our considered optimization problem, the caching decision depends on the resource allocation at the physical-layer, and hence, without knowing that we can not determine caching optimally. Unfortunately, finding the resource allocation decision also depends on caching problem. Therefore, the overall optimization is very complex and can only be solved optimally via exhaustive search. However, this is not feasible. Therefore, distributed RL is a suitable platform because it does not need any prior information about the system (i.e., the resource allocation decision). To the best of our knowledge, a joint distributed framework of CLNC resource allocation, D2D communications, and RL caching optimization for F-RAN architecture has not been considered in state-of-the-art literature.

Motivation: The sum-rate maximization optimization problem is motivated by real-time applications, i.e., video streaming, where users need to immediately stream their requests while ensuring the minimum required quality-of-service (QoS). For this, CLNC can stream users’ requests from F-APs and CE-D2D users with the maximum possible sum-rate, while ensuring fast access to these requests. Essentially, streaming users’ requests can be done either by pre-loading them at F-APs and CE-D2D users at much lower rates or at off-peak times. Unlike these impractical scenarios, our work considers that users’ combined requests can be delivered with the maximum possible sum-rate, while users can progressively and instantly use them.

B. Contributions

The main contribution of this work is, thus, an innovative CLNC and RL framework jointly taking caching strategy, NC, D2D communications, power/rate optimization, and fronthaul capacity into account. Our key contributions include:

- In F-RAN with D2D communications, the F-APs and CE-D2D users exploit their cached files to use CLNC to maximize sum-rate. Hence, our framework jointly considers cache resources optimization and wireless edge communications for content delivery. To this end, a sum-rate maximization optimization problem is formulated with the constraints on content placement at local caches, NC, users scheduling, their limited coverage zones, transmission rate/power, and QoS rate guarantee. Such optimization problem is NP-hard and computationally intractable.
To solve this joint optimization problem, two-stage iterative approach is developed by decomposing user clustering and cache resource allocations using multi-leader-follower Stackelberg game. Specifically, we first develop an innovative decentralized CLNC coalition formation (CLNC-CF) algorithm to obtain a stable F-AP and CE-D2D user coalition formation for the followers, where F-APs and CE-D2D users are utilizing CLNC resource allocation. By taking the behavior of F-APs and CE-D2D users into account, we then develop a multi-agent RL (MARL) algorithm for the leaders. The aforementioned two-stage of CLNC-CF and MARL solutions is referred to a joint CLNC-CF-RL approach.

We rigorously analyze the convergence and the computational complexity of our joint proposed approach. Specifically, our proposed CLNC-CF algorithm is proved to be a Nash-stable and our proposed MARL algorithm is the best strategy for content placement.

Extensive simulations are conducted to verify performance gain of our proposed CLNC-CF-RL framework over several benchmark schemes. Simulation results show that our proposed joint CLNC-CF-RL framework can effectively improve the sum-rate by around 30%, 60%, and 150%, respectively, compared to: 1) an optimal uncoded algorithm, 2) a standard RA-IDNC algorithm, and 3) a benchmark classical IDNC with network-layer optimization.

The rest of this paper is organized as follows. The system model and the CLNC are described in Section II. We formulate the sum-rate maximization problem in Section III. In Section IV, we develop a joint CLNC-CF-RL approach for solving the problem. In Section V, we analyze the properties of the developed joint approach. Simulation results are presented in Section VI and in Section VII we conclude the paper.

II. SYSTEM OVERVIEW AND CLNC

A. System Model

We consider a downlink F-RAN system with D2D communications illustrated in Fig. 1 with one CBS, $K$ F-APs, and $N$ CE-D2D users. The sets of F-APs and CE-D2D users are denoted by $K = \{1, 2, \ldots, K\}$ and $N = \{1, 2, \ldots, N\}$, respectively, and they cooperate with each other to serve single-antenna $U$ users, denoted by the set $U = \{1, 2, \ldots, U\}$. Each F-AP is connected to the CBS by a fronthaul link of capacity $C_{ph}$ bits per second. We assumed that each user is equipped with a single antenna and uses half-duplex channel, and accordingly, each user can be served from CE-D2D user via D2D link or from F-AP using cellular channel. Moreover, the allocated channels for D2D communications are out-of-band to those used by F-APs, i.e., an overlay D2D communications model is adopted [8]. We adopt a partially connected D2D networks where CE-D2D users are low-complexity devices and they can transmit to users at a certain amount of power. Accordingly, each CE-D2D user has limited transmission range, denoted by $A_n$, which represents the service area of the $n$-th CE-D2D user to transmit data within a circle of radius $R$. The set of users within the transmission range of the $n$-th CE-D2D user is denoted by $W_{n,u}$. Each F-AP $k$ and each CE-D2D user $n$ are equipped with local caches, which can proactively cache a subset of files $F_k$ and $F_n$ that represent the $k$-th F-AP and the $n$-th CE-D2D user cache sets, respectively.

The file placement phase starts by proactively caching popular files to F-APs and CE-D2D users, subject to the aforementioned cache capacity constraints. After the placement phase, the system enters the file delivery phase, which is done in F-AP and CE-D2D user transmissions. We assume that each user is already downloaded a set of $F$ files, which is denoted by the Has set $H_u$ of the $u$-th user. At any transmission, each user is arbitrarily interested in streaming one of the $F$ files. The request of the $u$-th user in a given transmission is denoted by the Wants set $W_u$. For a given file delivery phase, F-APs and CE-D2D users exploit the users’ downloaded files to perform XOR encoding operation, respectively, when new files requested by users. At the users, users exploit the downloaded files to extract the requested files immediately by performing XOR decoding operation. The notations used throughout this paper are listed in Table II

B. Cross-layer Modeling: CLNC

1) Physical layer model: Let $P_k$ and $Q_n$ denote the transmission power of the $k$-th F-AP and the $n$-th CE-D2D user, respectively. To avoid complexity, we consider fixed transmission power for the CE-D2D users. We consider quasi-static fading channels between the $k$-th F-AP and the $u$-th user and between the $n$-th CE-D2D user and the $u$-th user, which are assumed to remain fixed in each transmission. The channel fading gain for the links between $k$-th F-AP and the $u$-th user and
between the $n$-th CE-D2D user and the $u$-th user are denoted by $|h_{k,u}|^2$ and $|h_{n,u}|^2$, respectively. These channel gains are summarized in the matrix $H$. The signal-to-interference-plus-noise (SINR) for the links between the $k$-th F-AP and the $u$-th user and the $n$-th CE-D2D user and the $u$-th user are given by, respectively, $\gamma_{k,u} = \frac{P_k|h_{k,u}|^2}{\sum_{n'\in\mathcal{K},n'\neq k}P_{n'}|h_{n',u}|^2 + N_0}$ and $\gamma_{n,u} = \frac{\sum_{n'\in\mathcal{N}}Q_{n'|u}^2|h_{n,u}|^2}{\sum_{n'\in\mathcal{N}}Q_{n'|u}^2|h_{n,u}|^2 + N_0}$, where $N_0$ denotes the additive white Gaussian noise variance.

For F-AP and CE-D2D user transmissions, files are transmitted via cellular and D2D links, respectively. The data rate of the $u$-th user assigned to the $k$-th F-AP can be expressed as $R_{k,u} = \log_2(1 + \gamma_{k,u})$ and the data rate of the $u$-th user assigned to the $n$-th CE-D2D can be expressed as $C_{n,u} = \log_2(1 + \gamma_{n,u})$. To ensure a successful delivery of files to users, the $k$-th F-AP and the $n$-th CE-D2D user can transmit at a rate which is at most equal to the minimum rate of their assigned users, i.e., $R_k \leq R_{k,u}$ and $C_n \leq C_{n,u}$, respectively. The set of achievable capacities of all users in all F-APs and CE-D2D users can be represented, respectively, as $\mathcal{R} = \bigotimes_{(k,u)\in\mathcal{K}\times\mathcal{U}} R_{k,u}$ and $\mathcal{C} = \bigotimes_{(n)\in\mathcal{N}\times\mathcal{U}} C_{n,u}$, where the symbol $\bigotimes$ represents the product of the set of the achievable capacities.

1) Network layer model: Let $\kappa_k$ and $\xi_n$ be the file combinations transmitted from the $k$-th F-AP and $n$-th CE-D2D user, respectively, to the set of targeted users $\tau(\kappa_k)$ and $u(\xi_n)$. Since each user can be scheduled to F-AP $k$ via cellular channel or to CE-D2D user $n$ via D2D link, we have $\tau(\kappa_k) \cap u(\xi_n) = \emptyset$. These file combinations are in fact elements of the power set (i.e., XOR combinations of the files) of the available files at the caches of the $k$-th F-AP and the $n$-th CE-D2D user. In other terms, $\kappa_k \in \mathcal{P}(\mathcal{F}_k)$ and $\xi_n \in \mathcal{P}(\mathcal{F}_n)$. For convenience, the term “targeted users” is referred to scheduled users who receive an instantly-decodable transmission. A transmission from the $k$-th F-AP is instantly decodable at the $u$-th user if it contains the requested file by the $u$-th user and the scheduled transmission rate at the $k$-th F-AP, $R_k$, is no larger than the channel capacity $R_{k,u}$. Mathematically, $u \in \tau(\kappa_k)$ will hold if and only if $\{u \in \mathcal{U} \mid |\kappa_k \cap \mathcal{W}_u| = 1 \text{ and } R_k \leq R_{k,u}\}$. Similarly, for D2D links, a transmission from the $n$-th CE-D2D user is instantly decodable at the $u$-th user if: i) it contains the requested file by the $u$-th user and the scheduled transmission rate at the $n$-th CE-D2D user, $C_n$, is no larger than the channel capacity $C_{n,u}$ and ii) the $u$-th user is in the coverage zone of the $n$-th CE-D2D user. Mathematically, the set of targeted users by the $n$-th CE-D2D user is expressed as $u(\xi_n) = \{u \in \mathcal{U} \mid |\xi_n \cap \mathcal{W}_u| = 1 \text{ and } u \in \mathcal{A}_n \text{ and } C_n \leq C_{n,u}\}$.

C. Caching Policy and File Delivery

1) Caching policy: Let $S$ denote the side information matrix which summarizes the Has and Wants sets of all the users in a binary $U \times F$ matrix wherein the entry $(u,f)$ represents that the $f$-th file is requested by the $u$-th user. Let $C$ denote the file caching matrix where $(f,k)$ represents file $f$ is cached at F-AP $k$ and $(f,n)$ represents file $f$ is cached at CE-D2D user $n$. Owing to their cache sizes, F-APs and CE-D2D users can cache a limited number of files. If some files are missing in the local caches, they need to be fetched from the CBS via capacity-limited fronthaul links. Therefore, the data rate for fetching the $f$-th file of the $u$-th user from the $k$-th F-AP is $R_{k,u} = \min(C_{fh}, R_{k,u})$. Accordingly, the set of users whose requests are being missed at the network is denoted by $\mathcal{U}_m$.

### Table I

**Main Symbols Used in the Paper**

| Notation | Description |
|----------|-------------|
| $\mathcal{U}, \mathcal{N}, \mathcal{K}, \mathcal{F}$ | Sets of $U$ users, $N$ CE-D2D users, $K$ F-APs, $F$ popular files |
| $\mathcal{U}_m$ | Set of users whose requests being missed at the network |
| $\mathcal{K}_m$ | Set of F-APs serve users in $\mathcal{U}_m$ |
| $\mathcal{R}$ | Set of all achievable capacities |
| $\mathcal{A}_n$ | Set of users in the transmission range of CE-D2D user $n$ |
| $\mathcal{W}_u(\mathcal{H}_u)$ | Set of wanted (received) files by user $u$ |
| $C_{fh}$ | Capacity of fronthaul link in bits per second |
| $h_{k}, C_k$ | Transmission rate of F-AP $k$ and CE-D2D user $n$ |
| $\kappa_{k,u}$, $\xi_n$ | The encoded file of F-AP $k$ and CE-D2D user $n$ |
| $\tau(\kappa_k), u(\xi_n)$ | Set of targeted users by F-AP $k$ and CE-D2D user $n$ |
| $R_{k}$ | Data rate of F-AP $k$ for fetching files from the CBS |
| $\Phi, \mathcal{S}, \mathcal{H}$ | Caching, side information, and channel gain matrices |
| $\mathcal{O}$ | XOR operation and complexity notation |
| $x_{u,f}$ | Binary variable representing user $u$ requests file $f$ |
| $y_{f,k}, y_{f,n}$ | Binary variables representing file $f$ is stored at the caches of F-AP $k$ and CE-D2D user $n$ |
| $\Phi, \mathcal{Z}$ | F-AP cluster and CE-D2D cluster |
| $\Psi, \mathcal{Z}$ | Sets of $\Phi$ and $\mathcal{Z}$ clusters |
| $\mathcal{G}_{F}, >_u, G$ | Game, preference, and graph notations |
| $\lambda, \alpha$ | Learning rates at iteration index $t$ |
| $a_{i,j}$ | Action state of agent $i$ for decision $j$ |
| $x_{i,j}^{C}, x_{i,j}^{D}$ | The estimated utility and probability of taking action $a_{i,j}$ |

4
The sum-rate maximization optimization problem in F-RAN system with D2D communications involves a joint optimization of instantaneous-optimization problem for file delivery and a long-term optimization problem for file placement.

A. Problem Formulation for File delivery

We introduce three binary variables $x_{u,f}$, $y_{f,k}$, $v_{f,n}$ as such $x_{u,f} = 1$ if the $u$-th user requests the $f$-th file, and $x_{u,f} = 0$ otherwise; $y_{f,k} = 1$ if the $f$-th file is stored in the local cache of the $k$-th F-AP, and $y_{f,k} = 0$ otherwise; and $v_{f,n} = 1$ if the $f$-th file is stored in the local cache of the $n$-th CE-D2D user, and $v_{f,n} = 0$ otherwise. Let $l_k$ denote the maximum number of missing files at the $k$-th F-AP that need to be fetched from the CBS. The instantaneous-optimization problem to be solved by the F-APs and the CE-D2D users is formulated as $P_1$ given at the top of the next page.

In $P_1$, C1 implies that the number of missing files at each F-AP is limited by $l_k$, due to the capacity-constrained fronthaul links; C2 implies that the set of scheduled users to all F-APs are disjoint, and similarly, the set of scheduled users to all CE-D2D users are disjoint; C3 makes sure that no user can be scheduled to F-AP and CE-D2D user at the same time instant; C4 ensures that all files to be combined using XOR operation at all the F-APs and the CE-D2D users are stored in their local caches, respectively; C5 satisfies the minimum transmission rates required to meet the QoS rate requirement $R_{th}$; and C6 bounds the maximum transmit power of each F-AP. $P_1$ is a mixed integer non-linear programming problem and it has NP-hard complexity.

To address the computational intractability of $P_1$, we consider the following two clustering strategies, namely, F-AP clustering and CE-D2D user clustering.

1) F-AP clustering: Users are grouped into disjoint clusters, where in each cluster there is only one F-AP. Let $\Phi$ denote the set of F-AP clusters and expressed as $\Phi = \{\psi_1, \ldots, \psi_m\}$, $\psi_i \subset K$, $\psi_i \cap \psi_j = \emptyset$, $\forall i, j \in \{1, \ldots, m\}$ and $i \neq j$. Within each cluster, CLNC scheme is adopted. To mitigate the inter-cluster interference due to the simultaneous transmissions of F-APs in the clusters, CLNC employs power allocation to efficiently allocate power levels to the F-APs. As such, a potential number of users can be targeted with NC combination with improved sum-rate.

2) CE-D2D user clustering: The considered realistic scenario of the partially connected D2D networks motivates us to naturally group CE-D2D users and their neighboring users into clusters. Thus, each CE-D2D user and its neighboring users can form a cluster $Z$. Let $Z$ denote the set of all disjoint clusters which can be expressed as $Z = \{Z_1, \ldots, Z_n\}$, $Z_i \subset N, \forall i, j \in \{1, \ldots, n\}$.

Therefore, the file delivery problem is solved in three steps. At first, we obtain the aforementioned clusters. Subsequently, we obtain the network-coded file delivery decision in each cluster. Finally, we optimize transmit power allocations of the FAPs to mitigate interference in the system.

B. Problem Formulation for File Placement

The sum-rate maximization in $P_1$ is not only related to $H$ and $S$, but also on the file placement matrix $C$, and consequently, $P_1$ is highly affected by $C$. For maximizing the benefit of caching at F-APs and CE-D2D users, the file placement problem taking $P_1$ into account have two objectives, namely, (i) maximize the system sum-rate and (ii) reduce the cost caused by cache deployments at F-APs and CE-D2D users and file pushing. Therefore, the sum-rate maximization optimization problem that includes $P_1$ is formulated as $P_2$ given at the top of the next page.

In $P_2$, $S^*$ is the NC and user scheduling of the F-APs and CE-D2D users, $\omega$ is the weight factor that represents the benefit of unit long-term sum-rate, and $\mu$ is the cost of caching a file. $C7$ and $C8$ make sure that the number of users’ requested files that are cached by each F-AP and each CE-D2D user, respectively, are limited, due to the constrained cache size. We can readily show that problem $P_2$ is NP-hard and intractable. Specifically, $P_2$ is a two-level problem formulation, namely, (i) upper-level caching problem and (ii) lower-level file delivery problem $P_1$ (i.e., NC, power allocation, resource scheduling). We emphasize that $P_2$ is computationally intractable since both $P_1$ and the caching optimization problem are NP-hard.

The caching decision $C$ depends on the resource allocation solution of $P_1$, and unfortunately, the solution of $P_1$ also depends on $C$. Therefore, the overall optimization is very complex and can only be solved optimally via exhaustive. However, this is not suitable for real-time applications. Moreover, the decision of $S^*$ and $C$ occurs in two different time scales, namely, (i) the objective of $P_1$ is to instantly improve the sum-rate and (ii) the objective of caching is to maximize the long term rate. Distributed RL is a suitable platform to solve problem $P_2$ whose goal is maximizing long term sum-rate, and its main idea is that agents do not have any prior information about the system (i.e., the resource allocation decision). Hence, agents react with environment, learn the instantaneous reward (i.e., instantaneous rate), and update their strategy. As such, agents can achieve benefit in the long run. To this end, capitalizing the distributed RL and game theory, we propose an effective and efficient framework to solve $P_2$ with the reduced computational complexity.

C. A Game Perspective of Problem $P_1$ and Problem $P_2$

1) Game description of $P_1$: The F-AP and CE-D2D user clustering process can be modeled as a coalition formation
can be modeled as a Stackelberg game. In Stackelberg game, if a given user
user in the new cluster of F-AP/CE-D2D user should join the new cluster is strictly improved; (ii) the sum data rate of
users involved in the new cluster of F-AP/CE-D2D user should join the new cluster is strictly improved; (ii) the sum data rate of
users involved in the new cluster of F-AP/CE-D2D user should join the new cluster.

\[ \max_{s_{\psi}} \Phi = \omega E_{H,S} \left[ s(S^*(C)) \right] - \mu \left( \sum_{f} \sum_{k} \gamma f_{k} + \sum_{f} \sum_{n} \gamma f_{n} \right) \]

\[ s_{u} = R_{k}, \]

\[ \psi \succ u \psi(u) \Leftrightarrow \left\{ \begin{array}{l}
|W_{u} \cap H(\psi)| = 1 \& s_{u}(\Phi_{u}) > s_{u}(\Phi_{u}), \\
\sum_{u' \in \psi} s_{u'}(\Phi_{u}) > \sum_{u' \in \psi(u)} s_{u'}(\Phi_{u}), \\
\sum_{f} \sum_{u' \in \psi} x_{u',f}(1 - y_{f,k}) \leq l_{k}, \forall k \in K,
\end{array} \right. \]

and

\[ Z \succ u Z(u) \Leftrightarrow \left\{ \begin{array}{l}
|W_{u} \cap H(\Phi)| = 1 \& s_{u}(Z_{u}) > s_{u}(Z_{u}), \\
\sum_{u \in Z_{u}} s_{u}(\Phi_{u}) > \sum_{u \in Z_{u}} s_{u}(\Phi_{u}), \\
\sum_{f} \sum_{u \in Z_{u}} x_{u,f}(1 - y_{f,k}) \leq l_{k}, \forall k \in K,
\end{array} \right. \]

In (4), \( \Phi_{u}, \Phi_{o} \) are the old and new F-AP clustering after user \( u \) switches from \( \Phi(u) \) to \( \Phi \in \Phi_{u} \); and in (5), \( \Phi_{o}, \Phi_{o} \) are the old and new CE-D2D clustering after user \( u \) switches from \( Z(u) \) to \( Z \in Z_{u} \). In (4) and (5), user \( u \) prefers another cluster over its current cluster if and only if:

(i) it can receive its requested file and its data rate for being in that new cluster is strictly improved; (ii) the sum data rate of users involved in the new cluster of F-AP/CE-D2D user should be increased; and (iii) after user \( u \) joins the new cluster, the fronthaul capacity constraint of the FAP/CE-D2D user in that cluster is still satisfied.

2) Game description of \( \mathcal{P}_{2} \): The two-level problems in \( \mathcal{P}_{2} \), as mentioned before, are coupled and have sequential decisions, and interestingly, the interaction between these decisions can be modeled as a Stackelberg game. In Stackelberg game, the F-APs and CE-D2D users are leaders and the users are followers. The equilibrium of such a game is expressed as \( \mathcal{S}_{E}^{*}(C^{*}, \Psi_{C}^{*}) \). In \( \mathcal{S}_{E}^{*}(C^{*}, \Psi_{C}^{*}) \), the F-AP and CE-D2D user clustering when the best strategy of caching is \( C^{*} \). \( \mathcal{S}_{E}^{*}(\cdot) \) is a Stackelberg equilibrium (SE) if and only if: \( s(C^{*}, \Psi_{C}^{*}) \geq s(C, \Psi_{C}), \psi^{*}(t)_{u} \succ_{u} \psi_{u}, Z^{*}(t)_{u} \succ_{u} Z_{u}(t), \forall u, \forall t, \forall \Phi_{u}(t) \in \Psi^{*}(t)/\{\Phi_{u}^{*}(t)\} \), where \( \psi_{u}(t) \in \Phi^{*}(t) \) and \( Z_{u}(t) \in Z^{*}(t) \) are the clusters to which users are assigned given the caching strategy is \( C^{*} \).

IV. SUM-RATE MAXIMIZATION: JOINT APPROACH

A. CLNC-CF Switch Algorithm

We consider two kinds of coalitions in our CFG because we have two kinds of clustering in our system, F-AP clustering and CE-D2D clustering. In our CFG, players switch coalitions in order to optimize overall system sum-rate. The players choose whether to switch coalitions or not according to a predefined preference order \( \succ \). Such preference order in our CFG is related to the system sum-rate, and each player should be able to obtain it by relying only on local network information. A framework of the proposed decentralized CLNC-CF switch algorithm is presented as follows.

Phase I: In this phase, we describe the switching process of players among F-AP and CE-D2D user coalitions.

Switching among F-APs: Given the initialized random coalition for F-APs, this step optimizes the selection of user-FAP assignment through several successive switch operations between F-AP coalitions. Specifically, players switch their coalitions based on preference order \( \succ \) that is related to the sum-rate. The switch operations are implemented by checking the switch possibilities of player \( u \) to each coalition \( \psi \). Initially, we have \( \Phi_{ini} \). Thus, a player \( u \) can decide to switch its current coalition \( \psi(u) \) to a new coalition \( \psi \) if the following holds: (i) the data rate of player \( u \) is strictly improved without affecting the utilities of all the remaining players in \( \psi \) and (ii) the requested file by player \( u \) is cached at the associated F-AP in \( \psi \) or the fronthaul capacity link of associated F-AP in \( \psi \) is still satisfied after joining player \( u \). Therefore, player \( u \) should
switch coalition \( \psi(u) \) and join coalition \( \psi \), and accordingly, \( \Phi_{mi} \) is updated.

**Switching among CE-D2D users:** Given the initialized random coalition for CE-D2D users, this step optimizes the selection of user-CE-D2D user assignment using many switch operations. Specifically, players switch their coalitions based on preference order \( \succ \) that is related to the sum-rate. The switch operations are implemented by checking the switch possibilities of player \( u \) to each coalition \( Z \). Initially, we have \( Z_{ini} \). Thus, a player \( u \) can decide to switch its current coalition \( Z(u) \) to a new coalition \( Z \) if the following holds: (i) the data rate of player \( u \) is strictly improved without affecting the utilities of all the remaining players in \( Z \) and (ii) the requested file by player \( u \) is cached at the associated CE-D2D user in \( Z \) or the fronthaul capacity link of associated CE-D2D user in \( Z \) is still satisfied after joining player \( u \). Hence, player \( u \) should switch coalition \( Z(u) \) and join coalition \( Z \), and consequently, \( Z_{ini} \) is updated.

**Remark 1:** Since the system constraints must be maintained, switching coalitions cannot take place if any of the constraints is violated. It can be seen that this switching process is completely decentralized and only local network information is required by the users. Specifically, player \( u \) only needs to know the data rates of its cellular and D2D links.

**Phase II:** In this phase, we develop a CLNC scheme that considers NC user scheduling and power level control of the resource allocation of F-APs and CE-D2D users. Given the resulting coalition formation process by the first phase of Algorithm 1, \( P_3 \) can be rewritten as

\[
P_3 : \max_{R_k, C_n, P_k, \tau_k, u_n} s = \sum_{k \in K} \sum_{u \in \tau_k} \min_{n,u} R_{k,u}^n + \sum_{u \in \tau_k} \min_{n} R_{u,n}^k + \sum_{k \in K} \min_{u \in \tau_k} R_{k,u}^n + \sum_{u \in \tau_k} \min_{n} R_{u,n}^k
\]

s.t. C4-C6.

\( P_3 \) is a NC user scheduling and non-convex power allocation problem which is NP-hard and computationally intractable. In this work, capitalizing the graph theory, we propose an iterative method to solve \( P_3 \) with the reduced computational complexity. For the given transmit power allocation of each F-AP and CE-D2D user, \( P_3 \) is solved using maximum weight clique (MWC) search method. Afterward, for the resulting NC user scheduling, the power allocation problem is solved numerically. These two steps are iterated until convergence.

**Graph description:** Let \( G_k(V_k, E_k) \) and \( G_n(V_n, E_n) \) denote the distributed-graphs of the \( k \)-th F-AP and the \( n \)-th CE-D2D user, wherein \( V \) and \( E \) are their corresponding vertices and edges, respectively. Each F-AP and each CE-D2D user generate their vertices based on their cached files and the scheduled users in their coalitions. A vertex \( v \in V_n \) in \( G_k \) is generated for each user \( u \in \psi_k \), its requested file \( W_u \), and for each achievable rate for that user \( r \in R_u = \{R | R \in R | R \leq R_{k,u} \} \). In a similar way, a vertex \( v \in V_n \) in \( G_n \) is generated for each user \( u \in \tau_n \), its requested file \( W_u \), and for each achievable rate for that user \( r \in C_u = \{C | C \leq C_{n,u} \} \). Two vertices in \( G_k \) or in \( G_n \) (representing the same transmitted file) are connected if they satisfy the instantly decodable NC constraints and their corresponding rates are equal \[22\]. Consequently, \( \bigcup_{k=1}^{K} V_k \) and \( \bigcup_{n=1}^{N} V_n \) give the entire graph \( G \) and its corresponding vertices are \( \forall v \in V = \bigcup_{l=\{1, \ldots, K, 1, \ldots, N\}} V_l \). The weight of each vertex \( v \) is the rate of the represented user as follows

\[
w(v) = \begin{cases} R_{k,u}^n & \text{if } v \in V_k, \forall k \in K, \\
C_{n,u} & \text{if } v \in V_n, \forall n \in N, \\
\min(C_{fkh}^k, R_{k,u}) & \text{if } v \in V_k, \forall k \in N.
\end{cases}
\]

Recall, \( w(v) = \min(C_{fkh}^k, R_{k,u}) \) represents the case of missing requested files at F-AP \( k \) that have to be fetched from the CBS via fronthaul link with capacity \( C_{fkh}^k \).

**MWC description:** With the designed distributed graph \( G \), \( P_3 \) is similar to MWC problems in several aspects. In MWC, two vertices must be connected in the graph, and similarly, in \( P_3 \), two different users must be scheduled to two different transmitters. Moreover, our objective in \( P_3 \) is to maximize the total sum-rate, and similarly, the goal of MWC is to maximize the weight of all vertices. Consequently, we have the following lemma.

**Lemma 1:** Using the distributed graph \( G \), \( P_3 \) can be equivalently transformed to the problem of determining the MWC in \( G \).

**Proof.** Assuming that \( X^* = \{v_1, v_2, \ldots, v_{|X|} \} \), \( \forall v \in G \) is the MWC. Let \( X \) be the set of all possible cliques in \( G \). For each \( v \in X^* \) that is associated with an association of (user, file, transmitter, and rate), the weight \( w(v) \) is the utility that the induced user in \( v \) can receive. Hence, the weight of the MWC \( X^* \) is mainly the objective function of \( P_3 \) and can be written as \( w(X^*) = \sum_{\forall v \in X^*} w(v) = \sum_{\forall v \in X^*} \min_{\forall u \in \tau_k} R_{k,u}^n + \sum_{\forall n \in N} \sum_{\forall u \in \tau_n} \min_{\forall u \in \tau_k} R_{u,n}^k + \sum_{\forall u \in \tau_k} \min_{\forall u \in \tau_n} R_{k,u}^n \). Since two vertices that represent different users are scheduled to different transmitters, all vertices are adjacent, constraints C2, C3 hold. Moreover, all the vertices are indeed generated based on the cached files of F-APs and CE-D2D users, and accordingly, constraint C4 holds. Finally, the set of targeted users and file combinations is obtained by combining the vertices of the MWC \( X^* \) corresponding to \( G_l, \forall l \in K \cup N \). Therefore, \( X^* = \bigcup_{l=1}^{K+N} X^*_l \).

**F-AP power control optimization:** For the stable coalition of F-APs that results in NC and F-AP user scheduling, obtained from solving \( P_3 \), the power allocation problem is written as

\[
P_4 : \max_{P_k} \sum_{k \in K} |\tau_k| \min_{u \in \tau_k} R_{k,u} + \sum_{s \in K_m} |\tau_n| \min_{u \in \tau_k} R_{s,u},
\]

s.t. \( 0 \leq P_k \leq P_{max} \).

To solve \( P_4 \) effectively, we derive optimal power allocations to maximize sum-rate for a given NC user scheduling. To this end, we introduce the following lemma.

\[1\]For different transmitters (F-APs or CE-D2D users), vertices always represent different users as this is already guaranteed from the clustering process in the first phase of Algorithm 1. Hence, such vertices are connected.
**Lemma 2:** Given the resulting NC and user-F-AP schedule by Algorithm 1, a converged power allocation is obtained by updating power \( \{ P_k \}_{k \in K} \) at the \((t+1)\)-th iteration, based on the following power update manipulation

\[
P_{k}^{t+1} = \left[ \frac{\tau_k \ast \hat{\nu}_{k,a}}{\sum_{l=1, l \neq k}^{K} \tau_l \ast \hat{\nu}_{l,a}} \right] P_{\text{max}},
\]

where \( \hat{\nu} = \arg \min_{u \in \mathcal{U}_k} R_{k,u} \), \( \hat{\nu} = \arg \min_{u \in \mathcal{U}_l} R_{l,u} \), \( k, l \in K \) and \( k \neq l \); \( \tau_k \) is the transmit power of the \( k \)-th F-AP at the \( t \)-th iteration.

**Proof.** The proof is analogous to [46, Proposition 2], and is omitted due to the space limitation.

The two-phase CLNC-CF switch algorithm is distributively executed at each iteration index and summarized in Algorithm 1.

### B. Multi-Agent Reinforcement Learning

After obtaining the resulting coalitions formation by Algorithm 1 for a given file placement strategy, our remaining task to obtain a SE for the formulated Stackelberg game is to near-optimally place files at the local caches, taking the behaviors of F-APs and CE-D2D users into account. Achieving a SE requires a global strategy of all F-APs’ and CE-D2D users’ actions in the system. Clearly, such strategy is unavailable to the F-APs and CE-D2D users in practice, and hence, achieving an SE solution is challenging. To address this challenge, we leverage multi-agent RL (MARL) that allows each F-AP and each CE-D2D users to construct its strategy without having global information.

Let \( (a_{i,j})_{i=1,\ldots,D,j \in \{1,2\}} \) is the \( j \)-th action of the \( i \)-th virtual agent, where \( D \) is the number of virtual agents in the network, i.e., \( D = FK + FN \). The action set of \( D \) virtual agents in the \( t \)-th index is defined as \( \mathcal{A}_t = \{(a_{1,1}, a_{1,2}), (a_{2,1}, a_{1,2}), \ldots, (a_{D,1}, a_{D,2})\} \), each action represents a possible cache matrix \( C \). Specifically, the action space of the \( i \)-th agent is \( (a_{i,1}, a_{i,2}) \) and when \( i = (k-1)F + f \), taking action \( a_{i,1} \) means F-AP \( k \) caches file \( f \), while taking \( a_{i,2} \) means F-AP \( k \) does not cache file \( f \). Similarly, when \( i = (n-1)F + f \), taking action \( a_{n,1} \) means CE-D2D user \( n \) caches file \( f \), while taking \( a_{n,2} \) means CE-D2D user \( n \) does not cache file \( f \).

Let \( \Gamma \) denote the history profile, \( \Gamma = \{H_{1}, S_{1}\}, \{H_{2}, S_{2}\}, \ldots, \{H_{T}, S_{T}\} \), where \( \{H_{i}, S_{i}\} \in \Gamma \) is the tuple of channel gain matrix and side information matrix in the \( t \)-th iteration index. By exploiting this history network data, the virtual agents learn and achieve sub-optimal caching strategy. Assuming that action \( a_{i}^{t} \) is chosen, Algorithm 1 is executed, and a feedback signal is provided to the virtual agent to guide the learning process. The feedback signal for the \( i \)-th virtual agent, denoted by \( v_{i}^{t+1} \), is given by

\[
v_{i}^{t+1} = ws(a_{i}^{t}, H_{i}, S_{i}) - \mu \sum_{j} \sum_{k} y_{f,k}(a_{i}^{t}).
\]

In [8], \( s(a_{i}^{t}, H_{i}, S_{i}) \) is the system sum-rate in the \( t \)-th iteration index defined in \( \mathcal{P}_{2} \) and \( y_{f,k}(a_{i}^{t}) \) represents the caching decision associated with file \( f \) at transmitter \( k \) under action \( a_{i}^{t} \).

**Algorithm 1 CLNC-CF Switch Algorithm for File Delivery**

1. **Input:** \( N, K, K_{u}, W_{u}, \forall u \in U, \forall n \in N. \)
2. **Initialize:** F-AP and CE-D2D user random clustering \( \Psi_{ini} = \{\Phi_{ini}, Z_{ini}\} = \{\psi_{1}, \cdots, \psi_{m}, Z_{1}, \cdots, Z_{n}\} \), \( \psi_{i} \cap \psi_{j} = \emptyset, Z_{k} \cap Z_{l} = \emptyset, \psi_{i} \cap Z_{l} = \emptyset \), \( \forall i \neq j, k \neq l, \psi_{i} = \Psi_{ini}. \)
3. **Phase I:** Coalition switch process
   4. **repeat**
   5. **Switching among F-APs:**
      6. **for** \( u \in \Phi_{ini} \) **do**
      7. **for** \( \psi_{i} \in \Psi_{ini}, \forall i = \{1,2,\cdots,m\} \) **do**
      8. **if** \( \psi_{i} \supset \psi(u) \) **then**
      9. **Update** \( \Phi_{ini} = \Phi_{ini} \setminus \psi(u) \)
      10. **end if**
      11. **end for**
      12. **end for**
5. **Switching among CE-D2D users:**
   13. **for** \( u \in Z_{ini} \) **do**
   14. **for** \( Z_{i} \in Z_{ini}, \forall i = \{1,2,\cdots,n\} \) **do**
   15. **if** \( Z \supset u \) **then**
   16. **Update** \( Z_{ini} = Z_{ini} \setminus Z(u) \)
   17. **end if**
   18. **end for**
   19. **end for**
20. **until** No further switch operation
22. **Phase II:** Solve \( \mathcal{P}_{2} \) using CLNC:
23. **MWC search method:**
24. Initialize \( X^{*} = \emptyset \) and \( G(X^{*}) \leftarrow G. \)
25. **∀v \in G, calculate** \( w(v) \) **as in** [9].
26. **while** \( G(X^{*}) \neq \emptyset \) **do**
27. **Choose the maximum weight** \( v^{*} = \arg \max_{v \in G(X^{*})} \{w(v)\}. \)
28. **Set** \( X^{*} = X^{*} \cup v^{*} \) and \( G(X^{*}) \leftarrow G(v^{*}). \)
29. **end while**
30. **Output** \( X^{*}. \)
31. **F-AP power allocation optimization:**
32. Using the F-AP-user association in \( X^{*} \), solve \( \mathcal{P}_{1} \) to compute the optimal power allocations.

After computing \( v^{*} \), each virtual agent implements two RL procedures to estimate the expected utility received from Algorithm 1 and to properly selects an action probability, respectively. Let \( x_{i,j}^{t} = [(x_{i,j}^{t,1}, x_{i,j}^{t,2}), \cdots, (x_{i,j,F+1}^{t,1}, x_{i,j,F+2}^{t,2})]_{i \in \{1,\ldots,D\}, j \in \{1,2\}} \) be the vector of estimated utilities and \( \rho_{i,j}^{t} = [(\rho_{i,j}^{t,1}, \rho_{i,j}^{t,2}), \cdots, (\rho_{i,j,F+1}^{t,1}, \rho_{i,j,F+2}^{t,2})]_{i \in \{1,\ldots,D\}, j \in \{1,2\}} \) be the vector of action probability of all agents, respectively, at the \( t \)-th iteration index. Consequently, to strike a balance between exploration and exploitation, we introduce a variable \( \sigma \), and thus, we have \( \beta(x_{i,j}^{t}) = \frac{\exp(\sigma x_{i,j}^{t,1})}{\exp(\sigma x_{i,j}^{t,1}) + \exp(\sigma x_{i,j}^{t,2})}. \) The utility estimation and action probability update equations are
expressed as, respectively,  
\[ x_{i,j}^{t+1} = \alpha_t (x_{i}^t - x_{i,j}^t) + x_{i,j}^t, \quad \forall i \in \{1, \ldots, D\}, \]
\[ \rho_{i,j}^{t+1} = \rho_{i,j}^t + \lambda_t \beta(x_{i,j}^t - \rho_{i,j}^t), \quad \forall i \in \{1, \ldots, D\}. \]
In (9) and (10) \( \lambda_t \) and \( \alpha_t \) are learning rates and \( I_{a_i=a_j} = 1 \) if action \( a_i \) is taken in the \( t \)-th iteration index interval \( t \) and 0 otherwise. Similar to [39], \( a_i \) follows the same definition that all virtual agents share the same feedback signal, and, \( \lambda_t \) and \( \alpha_t \) should satisfy the following conditions: \( \sum_{t=1}^\infty \lambda_t = \infty \), \( \sum_{t=1}^\infty \alpha_t = \infty \), \( \sum_{t=1}^\infty \alpha_t^2 \leq \infty \), \( \lim_{t \to \infty} \lambda_t = 0 \). To calculate (9) and (10) the \( i \)-th agent initializes a utility vector and an equal probability action vector, given by, \[ x_{i,j}^t = [(0,0)]_{j=1,2} \] and \( \rho_{i,j}^t = [(0.5,0.5)]_{j=1,2}, \forall i \). With the increase in the \( t \)-th index, the parameter \( \sigma \) in (10) continuously grows, and correspondingly, each virtual agent intends to explore the two actions and choose the action associated with the highest expected utility. To perform this, we consider the well-known \( \epsilon \)-greedy policy that allows the agents to choose exploration and exploitation with probability \( \epsilon \) and \( 1 - \epsilon \), respectively. To summarize, each virtual agent iteratively repeats the following three steps, (i) action selection, (ii) learning calculation, and (iii) update step that are presented in Algorithm 2.

V. THE PROPERTIES OF THE PROPOSED APPROACH

In this section, we analyze the properties of the two-algorithms proposed approach, in terms of convergence, optimality and complexity.

A. CLNC-CF Switch Algorithm

1) Convergence: The switching process of users in stage 1 and stage 2 is converged because there is an infinite number of F-APs and CE-D2D users and thus, infinite number of switch operations. It can be observed that the achieved data rate by user \( u \) in any coalition formation of CE-D2D users is not influenced by the coalition formation of F-APs outside the coalition to which it belongs. Therefore, when the clustering of F-APs or CE-D2D users is updated based on the preference order, the system sum-rate will strictly increase. Otherwise, users would not switch. Since sum-rate is bounded, after finite repeats of Algorithm 1's steps, the clustering of F-APs and CE-D2D users will finally not change. Hence, Algorithm 1 will converge.

2) Stability: The stability of the final clustering depends on the existence of a Nash-stable solution for our CF game [45]. Let \( \Psi_{fin} = \{ \Psi_{fin}, Z_{fin} \} \) denote the final clustering obtained by Algorithm 1. Once Algorithm 1 converges to \( \Psi_{fin} \), it can be readily concluded that \( \psi \succ \psi(u) \) for any user \( u \) and any \( \psi \in \Phi_{fin}/\{ \psi(u) \} \), and similarly, \( Z \succ Z(u) \) for any user \( u \) and any \( Z \in Z_{fin}/\{ Z(u) \} \). Such a stable state is Nash-stable [45].

3) Complexity: At each iteration, Algorithm 1 needs to check at most \( |F| \) sum-rate constraints for switching process in phase I, and consequently, the computational complexities of phase I execution is \( O(|K| + |N|) \). For \( L \) iterations until convergence, the computational complexity is \( O(N|K| + |N|) \). Next, we analyze the computational complexity of phase II. For finding the NC user scheduling, each transmitter (i.e., F-AP and CE-D2D user) needs to build its distributed graph and find the corresponding MWC. The computational complexity of constructing the graph and finding its MWC of transmitter \( t \) is bounded by \( O(|F_t|^2|U|^3) \) and \( O(|U|^2|F_t|^2) \) [13]. This gives a total complexity of \( O(|F_t|^2|U|^2) \) that is associated with finding the NC combinations in the system [13]. For solving the power allocation problem in \( \mathcal{P}_1, \) Algorithm 1 needs \( C_p = O(|\mathcal{S}|) \) to \( \mathcal{P}_1 \), where \( |\mathcal{S}| \) represents the NC user scheduling of the first transmitter. Therefore, the overall worst-case computational complexity of Algorithm 1 is \( O(|N| |F_t|^2|U|^2) \), which can be approximated to a polynomial computational complexity of \( O(|N| |F_t|^2|U|^2) \).

B. MARL based Caching Algorithm

1) Convergence: At each iteration, Algorithm 2 updates the virtual agents’ probability of selecting the actions. Thereby, at the convergence of Algorithm 2, the action selection probability of the virtual agent become stable. Such a fact is demonstrated by the following theorem.

**Theorem 1:** The \( i \)-th virtual agent’s action selection probability is converged as \( \lim_{t \to \infty} \rho_{i,j}^{(t)} \rightarrow \rho_{i,j}^* \), \( \forall i \in \{1,2, \ldots, D\} \) and \( j \in \{1,2\} \). Here
\[
\rho_{i,j}^* = \frac{\exp (\sigma \pi_{i,j})}{\sum_{j=1}^2 \exp (\sigma \pi_{i,j})}
\]
where $\tau_{i,j}$ is the $i$-th virtual agent's expected reward from the $j$-th action.

**Proof.** According to [40] Theorem 4, under the condition of $\lim_{t \to \infty} \frac{\lambda}{\alpha} = 0$, we obtain that the following two conditions are satisfied. 

1. $\lim_{t \to \infty} \left| x_{i,j}^t - \tau_{i,j}^t \right| = 0$, $\forall i, j$; and
2. As $t \to \infty$, the update equation for the action selection probability of the virtual agent's action converges to the following ordinary differential equation (ODE):

$$\dot{\rho}_{i,j}^t = \beta(x_{i,j}^t) - \rho_{i,j}^t$$

(12)

Note that, as $\lim_{t \to \infty} \frac{\lambda}{\alpha} = 0$, the action selection probabilities are updated relatively slowly compared to the estimated utility. On the other hand, as per (11), the estimated utilities are converged as $t \to \infty$. Substituting the converged value of the estimated utility in (12) and using the fact that at the stationary point of (12), $\dot{\rho}_{i,j}^t = 0$, the value of the stationary action selection probability is obtained as

$$\rho_{i,j}^* = \beta(\tau_{i,j})$$

(13)

Next, we justify that (13) provides the converged action selection probability for the virtual agents. Based on [41], eq. (2), the optimal action selection probability is obtained by maximizing the expected reward of the virtual agents augmented by the entropy function. Accordingly, for the $i$-th virtual agent, $\forall i \in \{1, 2, \ldots, D\}$, we obtain

$$\left\{ \rho_{i,1}^*, \rho_{i,2}^* \right\} = \arg \max_{\rho_{i,1} \geq 0, \rho_{i,2} \geq 0} \sum_{j=1}^{2} \rho_{i,j} \tau_{i,j}$$

$$+ \frac{1}{\sigma} \sum_{j=1}^{2} \rho_{i,j} \ln \left( \frac{1}{\rho_{i,j}} \right), \text{ s.t.} \sum_{j=1}^{2} \rho_{i,j} = 1.$$  

(14)

In (14), the factor $\frac{1}{\sigma}$ controls the relative importance of maximizing the total expected reward and entropy functions. Applying the Lagrangian optimization technique, we can readily obtain

$$\rho_{i,j}^* = \frac{\exp \left( \sigma \tau_{i,j} \right)}{\sum_{j=1}^{2} \exp \left( \sigma \tau_{i,j} \right)} \equiv \beta(\tau_{i,j}).$$

(15)

Evidently, as $t \to \infty$, the action selection probability of virtual agents becomes stable, and such a stable action selection probability also coincides with the optimal action selection probability for the virtual agents. Therefore, Algorithm 2 converges as the number of iterations is increased.

2) Optimality: The optimality of Algorithm 2 is confirmed from the following theorem.

**Theorem 2:** Algorithm 2 provides a near-optimal solution to $\mathcal{P}_2$ optimization problem.

**Proof.** Since (13) is obtained at the maximal point of (14), we can readily justify

$$\tau_{i,j}(\rho_{i,j}^*, \rho_{-i}) + \frac{1}{\sigma} \rho_{i,j}^* \ln \left( \frac{1}{\rho_{i,j}^*} \right) \geq \tau_{i,j}(\rho_{i,j}, \rho_{-i})$$

$$+ \frac{1}{\sigma} \rho_{i,j} \ln \left( \frac{1}{\rho_{i,j}} \right) \Rightarrow \tau_{i,j}(\rho_{i,j}, \rho_{-i}) - \tau_{i,j}(\rho_{i,j}^*, \rho_{-i})$$

$$\leq \frac{1}{\sigma} \left( \rho_{i,j}^* \ln \left( \frac{1}{\rho_{i,j}^*} \right) - \rho_{i,j} \ln \left( \frac{1}{\rho_{i,j}} \right) \right), \forall j$$

(16)

where $\rho_{-i}$ is the action selection probability of all the virtual agents except the $i$-th virtual agent. Using the upper bound of the entropy function, we can justify

$$\frac{1}{\sigma} \left( \rho_{i,j}^* \ln \left( \frac{1}{\rho_{i,j}^*} \right) - \rho_{i,j} \ln \left( \frac{1}{\rho_{i,j}} \right) \right) \leq \frac{1}{\sigma} \ln 2.$$  

Therefore we obtain

$$\tau_{i,j}(\rho_{i,j}^*, \rho_{-i}) - \tau_{i,j}(\rho_{i,j}, \rho_{-i}) \leq \frac{1}{\sigma} \ln 2 \triangleq \epsilon.$$  

(17)

Note that $\sigma \to \infty$, $\tau_{i,j}(\rho_{i,j}^*, \rho_{-i}) \leq \tau_{i,j}(\rho_{i,j}, \rho_{-i})$ is satisfied. In addition, as $\sigma \to \infty$, $\rho_{i,j}^* = 1$ where $j^* = \arg \max_{j \in \{1,2\}} \tau_{i,j}$ and $\rho_{i,j \neq j^*} = 0$. Later, we demonstrate that $\sigma$ is an increasing function of the iteration index $t$. As a result, as $t \to \infty$, on the one hand, the action selection of the virtual agents becomes deterministic, i.e., each agent adheres to the action having the maximum expected payoff. On the other hand, with such an action selection strategy, the expected reward of each agent becomes larger than any given action selection strategy. Consequently, as $t \to \infty$, Algorithm 2 guarantees the optimal convergence of each virtual agent's achievable reward. Moreover, we can readily that for the deterministic action selection of the virtual agents, the statistical expected value of (8) and objective function of $\mathcal{P}_2$ optimization problem are same. Consequently, as $t \to \infty$, Algorithm 2 converges to the joint caching decisions that provides a higher objective value of $\mathcal{P}_2$ optimization problem than any other joint caching decisions. In other words, Algorithm 2 can converge to the optimal solution to $\mathcal{P}_2$ optimization problem as $t \to \infty$.

However, in practice, we can consider a finite number of iterations for Algorithm 2 and accordingly, from (17), the virtual agents may have certain incentive to deviate from the converged action selection strategy. Nevertheless, eq. (17) depicts that with a finite number of iterations, the virtual agents cannot increase their expected payoff more than $\epsilon$ by deviating from the converged action selection strategy, and the value of $\epsilon$ decreases as the number of iterations is increased. Hence, we conclude that Algorithm 2 provides a near-optimal solution to $\mathcal{P}_2$ optimization problem.

3) Complexity: Since each agent needs to make one action, the computational complexity associated with selecting all the actions in Algorithm 2 is $O(D)$. Further, the update equations in (2) and (9) involve fixed number of operations, thus the complexity for each virtual agent is $O(1)$. Essentially, these corresponding complexities are negligible as compared to the complexity of Algorithm 1 that is one step in Algorithm 2. Therefore, the overall worst-case computational complexity of Algorithm 2 is $O(D)$, which is $O(T \mid \mathcal{F}_i \mid^2 12 K)$.

4) Designing Parameter $\sigma$: Algorithm 2 can achieve a local optimal caching strategy. However, this result is established on the convergence when $t \to \infty$, where we should collect infinite amount of history network data. For limited data, the parameter $\sigma$ should be selected carefully to achieve a good performance. To ease the analysis, we first proof that the sum of action selection probabilities for each virtual agent is always equal to 1.

**Theorem 3:** For each virtual agent $i$, if $\sum_{j=1}^{2} \rho_{i,j} = 1$, the update equation in (10) guarantees that $\sum_{j} \rho_{i,j} = 1$ with $t \geq 2$. 

10
Proof. From (10), we have \( \sum_j \rho_{i,j}^{t+1} = \sum_j \rho_{i,j}^t + \lambda t \sum_j \beta(x_{i,j}^t - \rho_{i,j}^t) = \sum_j \rho_{i,j}^t + \lambda \left( 1 - \sum_j \rho_{i,j}^t \right) \). Thus, it can be noted that \( \sum_j \rho_{i,j}^t = 1 \Rightarrow \sum_j \rho_{i,j}^t = 1 \ldots \), and thus \( \sum_j \rho_{i,j}^t = 1 \) with \( t \geq 2 \).

The following theorem analyzes the impact of \( \sigma \) on the strategy evolution.

**Theorem 4:** For virtual agent \( i \), consider that action \( a_{i,1} \) has a larger utility estimation than action \( a_{i,2} \), and consequently, \( \beta(x_{i,1}^t) \) monotonically increases with \( \sigma \).

Proof. We take the derivative of \( \beta(x_{i,1}^t) \) with respect to \( \sigma \) as follows
\[
\frac{d\beta(x_{i,1}^t)}{d\sigma} = \beta(x_{i,1}^t) \frac{\exp(\sigma x_{i,1}^t) (x_{i,1}^t - x_{i,2}^t)}{\exp(\sigma x_{i,2}^t) + \exp(\sigma x_{i,2}^t)}. \tag{18}
\]
Since \( x_{i,1} > x_{i,2} \), we have \( \frac{d\beta(x_{i,1}^t)}{d\sigma} > 0 \). Hence, when \( \sigma \) becomes larger, \( \beta(x_{i,1}^t) \) will increase, and the probability of selecting action \( a_{i,1} \) will have a higher chance to increase as in (10). Consequently, the probability of selecting action \( a_{i,2} \) decreases as in Theorem 3.

A practical way of setting \( \sigma \) is \( \sigma_t = \frac{t}{b} \) [39, 42], where \( b \) is a constant. This setting adjusts the value of \( \sigma \) with iteration index \( t \). Particularly, \( \sigma \) is initially set to a small value to ease the full estimation of utility associated with each action. When \( t \) increases, \( \sigma \) increases, and the virtual agent selects the action with the largest utility estimation, and this improve its own performance.

VI. NUMERICAL RESULTS
In this section, the effectiveness of the proposed CLNC-CF-RL approach is demonstrated, and corresponding numerical results are conducted.

**A. Simulation Setting and Comparison Schemes**

1) Simulation setting: We consider an D2D-aided F-RAN where F-APs and CE-D2D users have fixed locations and users are distributed randomly within a hexagonal cell of radius 1500m. Unless otherwise stated, we set the radius of the CE-D2D users’ transmission range \( R \) to 500m and the numbers of F-APs \( K \) and CE-D2D users is set to 3, 6, respectively. In addition, each user requests one file in each transmission by following Zipf distribution, and the probability of file \( f \) to be requested is given by \( \frac{f^{-\gamma}}{\sum_{j=1}^{F} j^{-\gamma}} \), where \( \gamma \) is the Zipf parameter that governs the popularity distribution skewness which is equal to 0.5. To further emphasize on the discrepancies among files popularity, each user requests a file with a probability given in the range \([0 \text{ \ldots } 0.25]\). The channel model of both F-RAN and D2D communications follow the standard path-loss model, which consists of three components: 1) path-loss of \( 128.1 + 37.6 \log_{10} \text{dis.[km]} \) for F-RAN transmissions and path-loss of \( 148 + 40 \log_{10} \text{dis.[km]} \) for D2D communications; 2) log-normal shadowing with 4dB standard deviation; and 3) Rayleigh channel fading with zero-mean and unit variance. The cellular and D2D channels are assumed to be perfectly estimated. The noise power and the maximum’ F-AP and CE-D2D user power are assumed to be \(-174 \text{ dBm/Hz} \) and \( P_{\text{max}} = Q = -42.60 \text{ dBm/Hz} \), respectively. The link bandwidth is 10 MHz. Unless otherwise stated, we set \( l_k = 5, \forall k \) and \( C_{f,th} = 30 \text{ Mbps} \). For the update equations in (9) and (10), the learning parameters are set to \( \sigma = \frac{1}{t^{0.5}}, \alpha_t = \frac{1}{(1 + t)^{0.5}}, \lambda_t = \frac{1}{(1 + t)^{0.5}} \). For the learning rate \( r(t) \) of each agent \( i \) in (8), we set \( \omega = 1 \) and \( \mu = 0.8 \). To use MARL to learn caching strategy, 1000 tuples of \( \{H, S\} \) are generated and utilized to construct history network environment, and in these history data generations, the locations of F-APs, CE-D2D users, and users are fixed.

2) Comparison schemes: First, to assess the performance of our proposed CLNC-RL approach with different thresholds \( (R_{th1} = 0.5 \text{ and } R_{th2} = 5) \), we simulate various scenarios with different number of users and number of files. These thresholds represent the minimum transmission rates required for QoS. The performances of our proposed solution for \( R_{th1} \) and \( R_{th2} \) are shown in solid red line and dash red line, respectively. For the sake of comparison, our proposed schemes are compared with the following baseline schemes.

- **Optimal Uncoded**: The transmission strategy in this scheme is performed irrespective of the available information at the network layer, i.e., prior download files. The user-F-AP/CE-D2D user association in this scheme is proposed in [31].
- **Classical IDNC**: For F-AP and CE-D2D user transmis-
sions, this scheme focuses on network layer optimization, where the coding decisions depends solely on the file combinations.
- **RA-IDNC**: This scheme was studied in [21].
- **All Caching**: In this scheme, each F-AP and each CE-
D2D user cache all the files requested by the users.
- **No Caching**: In this scheme, there is no popular file cached at F-APs and CE-D2D users.
- **0.5 Caching**: In this scheme, F-APs and CE-D2D users cache files with probability of 0.5.
- **Distributed Q-learning**: This scheme was proposed in [43].

**B. Simulation Discussions**

1) Simulation of Algorithm 1: In Figs. 2 and 3, we present average sum-rate vs. number of users \( U \) for 30 files and vs. number of files \( F \) for 40 users, respectively. Fig. 2 shows the impact of multiplexing many users to the F-APs and CE-D2D users using NC on the physical layer performance, and Fig. 3 considers different sizes of popular contents. We can see from Figs. 2 and 3 that the proposed CLNC-CF algorithm offers an improved performance in terms of sum-rate as compared to contemporary coded and uncoded schemes. These improved performances are becuase our proposed scheme judiciously schedules users, adopts the transmission rate of each F-APs and optimizes the transmission power of each F-AP, and selects potential users for transmitting coded files over D2D

2) Each presented value in Fig. 2, 3, 4, and 5 is calculated by averaging sum-rate over 1000 realizations of \( \{H, S\} \). For changing the network topology, the locations of users are randomly generated in each realization.
links. Specifically, the classical IDNC scheme suffers from scheduling many users to the FAPs and the CE-D2D users by adopting their transmission rates to the minimum rates of all their scheduled users. Optimal uncoded schedule schemes users to the F-APs and the CE-D2D users based on their strong channel qualities, but it suffers from scheduling few users that is equal to the number of F-APs and CE-D2D users in the network. On the other hand, RA-IDNC scheme offers an improved performance compared to uncoded and classical IDNC schemes since it effectively balances between the number of scheduled users and the transmission rate of F-APs/CE-D2D users. However, selecting one transmission rate (the minimum rate) for all the F-APs and CE-D2D users degrades the sum-rate performance of the RA-IDNC scheme. This is a clear limitation of the RA-IDNC scheme as it does not fully exploit the typical variable channel qualities of the different F-APs and CE-D2D users to their scheduled users. Our proposed CLNC-CF-RL scheme fully utilize the FAPs and CE-D2D users’ resources to choose their own transmission rates, file combinations, and scheduled users. Consequently, a better performance of our proposed schemes compared to the RA-IDNC scheme is achieved. Moreover, the joint scheme optimizes the employed rates using power control on each F-AP. Thus, it works better than our proposed coordinated scheme. It is worth remarking that the increase in the number of files in Fig. 3 leads to a slight increase in the sum-rate for all coded schemes. This is due to the fact that the number of non-feasible edges between vertices increases as the number of files increases. This results in a smaller IDNC opportunities to combine files, thus leading to a slight improvement in terms of the sum-rate of all NC schemes.

In Figs. 4 and 5, we study the impacts of fronthaul capacity and cache size on sum-rate performance of our proposed scheme with different cache-levels, respectively, under various values of $\gamma$. From Fig. 4 it can be seen that a smaller $\gamma$ leads to a higher sum-rate and larger $\gamma$ leads to a reduction in sum-rate performance. For smaller $\gamma$, most of users will request popular files, and hence their requests have higher probabilities to be satisfied locally from the caches of F-APs and CE-D2D users. Consequently, the need for fronthaul capacity is less stringent. From Fig. 5 we note that a larger cache size makes F-APs and CE-D2D users have more chances to cooperate and thus, system performance is raised. The improved sum-rate performances in Figs. 4 and 5 demonstrate the pronounced role of our proposed
In this paper, a joint cache and radio resource allocation optimization problem has been evaluated and studied for F-RAN system with D2D communications, which includes a long-term caching optimization problem and an instantaneous distributed Q-learning scheme. This is due to the fact that our proposed scheme optimizes the agents’ actions and the cluster formation behavior of F-APs and CE-D2D users by learning from the feedback signal of the history environment, which helps making proper caching decisions. Particularly, the all caching scheme caches all the requested popular files at F-APs and CE-D2D users, and accordingly, it fully exploits the cooperation between F-APs and CE-D2D users that significantly improves system sum-rate. However, each F-AP and CE-D2D user in this scheme needs to cache all the files, which is impractical. For no caching scheme, the system sum-rate is the worst since there are less F-AP and CE-D2D users’ cooperation chances under a limited fronthaul capacity. Compared to the distributed Q-learning scheme, our proposed scheme improves the sum-rate by 10.3%, which again confirms its superiority. Finally, in Fig. 9, we show that our proposed scheme is also efficient in terms of the total number of cached files as compared to all other schemes.

**VII. CONCLUSION**

In this paper, a joint cache and radio resource allocation optimization problem has been evaluated and studied for F-RAN system with D2D communications, which includes a long-term caching optimization problem and an instantaneous...
cluster formation problem. The considered joint problem has been modeled as a Stackelberg game between the multi-agent leaders, and the clustering behavior of F-APs and CE-D2D users is captured by a CFG. To achieve Stackelberg equilibrium, a distributed CLNC clustering formation algorithm is first developed for F-APs to reach a stable solution under any fixed caching strategy. Then, to tackle the challenges incurred by no closed form and binary caching decision variables, an innovative MARL algorithm is developed to achieve a local optimal caching strategy. Simulation results showed that the proposed joint CLNC-CF-RL framework can effectively improve the sum-rate by around 30%, 60%, and 150%, respectively, compared to: 1) an optimal uncoded algorithm, 2) a standard RA-IDNC algorithm, and 3) a benchmark classical IDNC with network-layer optimization.

REFERENCES

[1] Y. Cai, F.-R. Yu, and S. Bu, “Cloud computing meets mobile wireless communications in next generation cellular networks,” IEEE Netw., vol. 28, no. 6, pp. 54-59, Nov. 2014.

[2] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, “Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks,” IEEE Trans. Signal Process., vol. 61, no. 22, pp. 5646-5658, Nov. 2013.

[3] R. Tandon and O. Simeone, “Harnessing cloud and edge synergies: Toward an information theory of fog radio access networks,” IEEE Commun. Mag., vol. 54, no. 8, pp. 44-50, Aug. 2016.

[4] E. Bastug, M. Bennis, and M. Debbah, “Living on the edge: The role of proactive caching in 5G wireless networks,” IEEE Commun. Mag., vol. 52, no. 8, pp. 82-89, Aug. 2014.

[5] K. Kaneva, N. Aboutorab, S. Sorour, and M. C. Reed, “Cellular fronthaul offloading using device fogs, caching, and network coding,” IEEE Trans. Mobile Comput., vol. 19, no. 2, pp. 276-287, Feb. 2019.

[6] K. Kaneva, N. Aboutorab, S. Sorour, and M. C. Reed, “Energy-aware cross-layer offloading in fog-RANs using network coded device cooperation,” IEEE Access, vol. 8, pp. 169930-169943, Sept. 2020.

[7] A. Asadi, Q. Wang, and V. Mancuso, “A survey on device-to-device communication in cellular networks,” IEEE Commun. Surveys Tus., vol. 16, no. 4, pp. 1801-1819, 2014.

[8] Roy Karasik, O. Simeone, and S. Shamai, “How much can D2D communication reduce content delivery latency in fog networks with edge caching?,” IEEE Trans. on Commun., Early Access, Dec. 2019.

[9] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, “Network information flow,” IEEE Transactions on Information Theory, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[10] T. Ho et al., “A random linear network coding approach to multicast,” IEEE Trans. Inf. Theory, vol. 52, no. 10, pp. 4413-4430, Oct. 2006.

[11] Z. Dong, S. H. Dau, C. Yuen, Y. Gu, and X. Wang, “Delay minimization for relay-based cooperative data exchange with network coding,” in IEEE/ACM Trans. on Netw., vol. 23, no. 6, pp. 1890-1902, Dec. 2015.

[12] X. Wang, C. Yuen, T. T. Li, W. Song, and Y. Xu, “Minimizing transmission cost for third-party information exchange with network coding,” in IEEE Trans. on Mobile Comput., vol. 14, no. 6, pp. 1218-1230, Jun. 2015.

[13] S. Sorour and S. Valaee, “Completion delay minimization for instantly decodable network codes,” IEEE J. Sel. Areas Commun., vol. 33, no. 5, pp. 970-984, Jul. 2015.

[14] M. S. Al-Abiad, M. Z. Hassan, A. Douik, and M. J. Hossain, “Low-complexity power allocation for network-coded User scheduling in Fog-RANs,” IEEE Trans. Commun., vol. 67, no. 12, pp. 8435-8449, Jun. 2019.

[15] M. S. Al-Abiad and M. J. Hossain, “Completion time minimization in F-RANs using D2D communications and rate-aware network coding,” in IEEE Trans. on Wireless Commun., Early Access, Jan. 2021.

[16] S. Gitzenis, G. S. Paschalidis, and T. Tullius, “Asymptotic laws for joint content replication and delivery in wireless networks,” IEEE Trans. Inf. Theory, vol. 59, no. 5, pp. 2760-2776, May 2013.

[17] A. Douik, H. Dahrouj, T.-Y. Al-Naffouri, and M.-S. Alouini, “Rate aware network codes for relay-based cooperative data exchange with network coding,” in IEEE Trans. on Information Theory, vol. 59, no. 12, pp. 8402-8413, December 2013.

[18] A. Douik and S. Sorour, “Data dissemination using instantly decodable binary codes in fog radio access networks,” IEEE Trans. on Commun., vol. 66, no. 5, pp. 2052-2064, May 2018.

[19] A. Douik, S. Sorour, T.-Y. Al-Naffouri, and M.-S. Alouini, “Rate aware instantly decodable network codes,” IEEE Trans. on Wireless Commun., vol. 16, no. 3, pp. 998-1011, Feb. 2017.

[20] X. Wang, C. Yuen, and Y. Xu, “Coding based data broadcasting for time critical applications with rate adaptation”, IEEE Trans. on Vehicular Tech., vol. 63, no. 5, pp. 2429-2442, Jun. 2014.

[21] M.-S. Al-Abiad, A. Douik, and S. Sorour, “Rate aware network codes for cloud radio access networks,” IEEE Trans. on Mobile Comput., vol. 18, no. 10, pp. 1898-1910, Aug. 2019.

[22] M. S. Al-Abiad, A. Douik, S. Sorour, and Md. J. Hossain, “Throughput maximization in cloud-radio access networks using rate-aware network Coding,” IEEE Trans. Mobile Comput., Early Access, Aug. 2020.

[23] M. S. Al-Abiad, M. Z. Hassan, A. Douik, and M. J. Hossain, “Low-complexity power allocation for network-coded User scheduling in Fog-RANs,” IEEE Commun. Letters, Early Access, Dec. 2020.

[24] M. S. Al-Abiad, M. J. Hossain, and S. Sorour, “Cross-layer cloud offloading with quality of service guarantees in Fog-RANs,” in IEEE Trans. on Commun., vol. 67, no. 12, pp. 8435-8449, Jun. 2019.

[25] S. E. Tajbakhsh and P. Sadeghi, “Coded cooperative data exchange in fog radio access networks,” in IEEE/ACM Trans. Netw., vol. 23, no. 5, pp. 1658-1670, Sep. 2015.

[26] Y. Cui, M. Peng, and H. Vincent Poor, “A distributed approach to cache allocation in wireless networks,” in Proc. of IEEE ISIT, pp. 1490-1494, Jun. 2015.

[27] Y. Sun, M. Peng, and H. Vincent Poor, “A distributed approach to cache allocation in wireless networks,” in Proc. of IEEE ISIT, pp. 1490-1494, Jun. 2015.
[43] W. Wang, R. Lan, J. Gu, A. Huang, H. Shan, and Z. Zhang, “Edge caching at base stations with device-to-device offloading,” IEEE Access, vol. 5, pp. 6399-6410, Mar. 2017.

[44] Y. Sun, M. Peng, and S. Mao, “Deep reinforcement learning based mode selection and resource management for green fog radio access networks,” IEEE Internet Things J., vol. 6, no. 2, pp. 1960-971, Apr. 2019.

[45] E. Hossain, D. I. Kim, and V. K. Bhargava, Cooperative Cellular Wireless Networks. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[46] M. S. Alabiad, M. Z. Hassan, and M. J. Hossain, “Cross-layer network codes for content delivery in cache-enabled D2D networks,” IEEE Trans. on Vehicular Tech. [Online].

Mohammed S. Al-Abiad received the B.Sc. degree in computer and communications engineering from Taiz University, Taiz, Yemen, in 2010, the M.Sc. degree in electrical engineering from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, in 2017, and the Ph.D. degree in electrical engineering from the University of British Columbia, Kelowna, BC, Canada, in 2020. He is currently a Postdoctoral Research Fellow with the School of Engineering at the University of British Columbia, Canada. His research interests include cross-layer network coding, optimization and resource allocation in wireless communication networks, machine learning, and game theory. He is a student member of the IEEE.

Md. Zoheb Hassan received the Ph.D. degree from the University of British Columbia, Vancouver, BC, Canada, in 2019. He is a Research Fellows with the École de technologie supérieure (ETS), University of Quebec, Canada. His research interests include wireless optical communications, optimization and resource allocation in wireless communication networks, and digital communications over fading channels. He was the recipient of Four-Year Doctoral Fellowship of the University of British Columbia in 2014. He serves/served as a Member of the Technical Program Committee of IEEE IWCMC 2018, IEEE ICC 2019, and IEEE ICC 2020.

Md. Jahangir Hossain (S’04, M’08, SM’18) received the B.Sc. degree in electrical and electronics engineering from the Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, the M.A.Sc. degree from the University of Victoria, Victoria, BC, Canada, and the Ph.D. degree from The University of British Columbia (UBC), Vancouver, BC. He was a Lecturer with BUET. He was a Research Fellow with McGill University, Montreal, QC, Canada, the National Institute of Scientific Research, Quebec, QC, and the Institute for Telecommunications Research, University of South Australia, Mawson Lakes, Australia. His industrial experience includes a Senior Systems Engineer position with Redline Communications, Markham, ON, Canada, and a Research Intern position with Communication Technology Lab, Intl., Hillsboro, OR, USA. He is currently an Associate Professor with the School of Engineering, UBC Okanagan campus, Kelowna, BC. His research interests include designing spectrally and power-efficient modulation schemes, applications of machine learning for communications, quality-of-service issues and resource allocation in wireless networks, and optical wireless communications. He regularly serves as a member of the Technical Program Committee of the IEEE International Conference on Communications (ICC) and the IEEE Global Telecommunications Conference (Globecon). He has been serving as an Associate Editor for IEEE COMMUNICATIONS SURVEYS AND TUTORIALS and an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS. He previously served as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.