Energy-Aware Cross-Layer Offloading in Fog-RANs Using Network Coded Device Cooperation

KAMELIYA KANEVA [1], **(Student Member, IEEE)**,
NEDA ABOUTORAB [1], **(Senior Member, IEEE), SAMEH SOROUR** [2], **(Senior Member, IEEE),**
AND MARK C. REED [1], **(Senior Member, IEEE)**

[1] School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2612, Australia
[2] School of Computing, Queen's University, Kingston, ON K7L 3N6, Canada

Corresponding author: Neda Aboutorab (n.aboutorab@unsw.edu.au)

ABSTRACT This paper studies a Fog Radio Access Network (F-RAN) architecture that utilises the increasing storage and device-to-device communication capacities of users' smart devices (referred to as F-UEs) in order to reduce the time that the central processing unit (referred to as BBU) has to spend to serve these F-UEs and thus, increases the system's capacity. Indeed, these F-UEs can employ these capacities in cooperatively serving each other's file requests, as long as the F-UEs have the files in their storage, rather than always receiving them through the BBU. In addition, network coding (NC) can be employed to minimize the communications among the F-UEs and from the BBU to further offload the BBU resources and reduce the energy consumed in F-UEs' cooperation. This paper develops an algorithm for scheduling the file encoding and transmissions both among half-duplex F-UEs and from the BBU to minimise the consumed time-frequency resources from the BBU given constraints on the energy consumed by each F-UE for their cooperation to help preserve their battery life. The influence of the energy constraint on BBU offloading is investigated and the performance of the system under a variety of settings is evaluated through extensive simulations. The benefit of using NC is shown and the ability of the F-UEs to transmit with different power levels is also studied and shown to considerably impact the number of cooperating F-UEs and the BBU offloading.

INDEX TERMS Device to device communications, energy constraint, fog radio access networks, Fronthaul offloading, network coding.

I. INTRODUCTION

A. MOTIVATION

The ever increasing demand for popular online content like video and audio puts a strain on the network as more network bandwidth and infrastructure is needed to serve all user requests. According to Deloitte [2], the end users prefer mobile devices as the medium for consuming online media. These consumer trends define the technical challenges today. The traditional user-server approach where a user sends a request to the server and the server then streams the content to the user cannot scale to accommodate the current online media consumption and provide the expected quality, mainly

due to bandwidth limitations. Thus, a new architecture is required.

A promising solution to these challenges is Fog Radio Access Network (F-RAN), which aims to bring the data content physically closer to the end users, thus, reducing the download time. In order to achieve this, the F-RAN's central baseband processing unit (BBU) caches popular files on either enhanced remote radio heads (eRRHs) that are installed closely to the end users, or directly on the user devices (denoted by F-UEs).

The work in this paper considers the case when the BBU caches some popular files on the F-UEs during off-peak times. The reason for this is not just to save equipment and installations costs, but also to take advantage of F-UEs' prior downloads. Due to the F-UEs' cooperative and storage capabilities of previously downloaded popular files, they naturally

---

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

become sources of content. In other words, we take advantage of the F-UEs primary purpose, downloading content, and reuse this content in order to minimise the BBU bandwidth. Thus, these F-UEs can collaborate with each other during peak times to retrieve their requested files from one another within a very short time period, with hopes to fully relieving the BBU from their delivery. This increases the system's capacity and allows the BBU to serve the increasing number of end devices. Yet, due to the quality of experience requirements of next generation networks, requests that are not *immediately* honoured by the F-UEs must still be served from the BBU. By "immediately", we mean within the minimum duration in which a group of half-duplex devices can collaborate to help each other, namely two transmission phases. In such scenarios, the decisions on which entity will serve which request(s) in each of the two phases must be done so as to minimally involve the BBU and consume the least amount of its time-frequency radio resources.

This work also considers the fact that this collaborative approach to achieve BBU offloading comes at the cost of using F-UEs' resources that, otherwise, could be saved to prolong their working hours and capacity. While providing some incentive mechanisms to motivate the F-UEs to contribute is important [3], [4], this paper addresses the reality that these devices will most probably impose limits on the energy they will spare for this collaboration, and studies how these energy limits will influence the offloading the BBU's physical resources.

In order to improve the network throughput, preserve BBU bandwidth, and abide by F-UEs' energy constraints, all F-UEs are allowed to use Network Coding (NC). Originally introduced in [5], NC takes advantage of the files already known to the F-UEs by mixing (i.e., encoding) some of the F-UEs' requested files and sending them together instead of the traditional method of using multiple uncoded unicast transmissions. This allows each transmitting device to target more than one receiver. We use this property of the NC to maximize the service from the F-UEs and minimise the used BBU resources. In this work, we use Opportunistic Network Coding (ONC) because it is computationally inexpensive and allows the F-UEs to decode their required files as soon as the encoded packet is received, making it available for use momentarily.

B. RELATED WORK
1) NETWORK CODING (NC)
In general, Network Coding (NC) allows the nodes not simply to route packets but instead, to compute and transmit functions of those packets. The NC has been shown to increase the throughput of the network [6], improve the latency [7] and the quality of service. A popular type of NC used in wireless networks is Opportunistic Network Coding (ONC). The ONC-enabled transmitter performs a mathematical operation (e.g. XOR) over two or more packets and sends the encoded packet to multiple receivers [8], [9] in one

transmission (instead of several unicast transmissions). The instantly decodable NC (IDNC) is a very useful and well-known subclass of ONC which we use in this work.

As the name suggests, IDNC encodes packets in a way that allows the receivers to instantly decode their requested packets by simply performing an XOR operation [10]. Our work takes advantage from utilising IDNC in the following ways. By benefiting several receivers with one transmission, the IDNC allows us to maximize the number of served F-UEs with each transmission. Moreover, the IDNC has been shown to reduce the completion time of delivering a group of packets [10]. This property ensures that if the BBU needs to serve some F-UEs, it will do so in minimum time. In conclusion, maximizing the number of receiving F-UEs increases their cooperation capabilities which ultimately, decreases the BBU's working time and increases the system's capacity.

2) COLLABORATIVE DATA EXCHANGE
The area of Collaborative Data Exchange (CDE) studies different scenarios where the smart end devices collaborate with each other to retrieve some files. Although this body of work seems to be very close to our scenario, the objectives are quite different. In [11], [12], the aim is to develop algorithms that decrease the decoding delay or completion time of the system, whereas our algorithm aims to organise the communication between the F-UEs in a way that shortens the BBU's completion time serving these F-UEs, while preserving the F-UEs consumed energy below their specified thresholds. Moreover, in [13]–[15] all the receivers require all files (a broadcast scenario) or a subset of them [11], [12] (a multicast scenario), whereas in this work the F-UEs require to immediately (with very low latency) receive only one popular file, at least once every two transmission phases.

Since in the F-RAN, the F-UEs are capable of communicating with each other, the area of CDE is very relevant to this problem. However, the main difference between the CDE-related works and this paper is that they study the communications between the F-UEs regardless of the ways they affect the BBU. Furthermore, it is assumed in the system model of this paper that more than one F-UE can be selected to transmit in any of the two possible transmission phases, which further complicates the scheduling of F-UE transmissions given their half-duplex nature.

3) F-RANs AND BBU OFFLOADING
To achieve F-RAN's full potential, the BBU must cache popular files smartly so that metrics like connectivity, energy consumption, delay and BBU offloading are optimised [16]–[19]. Although the authors in [20] did not utilize NC transmissions, they showed that significant amount of offloading can be achieved even when the energy cost of the collaborative F-UEs is kept low. It is also worth mentioning that a significant body of work like [21]–[23] explores the computation offloading and resource allocation in the mobile edge. However, in those papers the devices have some tasks to be computed as opposed to receiving files. Thus, the

offloading has a different objective. A number of papers have proposed algorithms for downloading a file through content helpers [24]–[26] or relaying mobile devices [27] with the aim of BBU offloading, minimising energy consumption, or maximising success rate. However, those papers did not consider NC nor device collaboration, but rather employed simple file broadcasts to or between F-UEs. The work in [28] considered the offloading of the BBU's physical resources using network coded transmission from eRRHs. It thus differs from this work as it did not explore file delivery cooperation between half-duplex devices nor energy considerations. Although the work in [29] considers the same system model of this paper (referred to as device-based F-RAN), the developed upper-layer algorithms in [29] minimised the number of fronthaul channels that the BBU must use in order to complete the service of all F-UEs requests. In this paper, we show that minimising this metric does not guarantee the minimum BBU involvement in terms of actual physical resources. Indeed, the BBU can transmit coded packets on fewer channels (good channel offloading) but with low download rates (due to the poor physical channels from the BBU to the served F-UEs), which may consume more BBU resources as opposed to serving requests of other F-UEs using slightly more transmissions (i.e., more frequency channels) but significantly higher download rates on each of them.

C. OUR CONTRIBUTIONS

The aim of this paper can be summarised in the following question: "What is the optimal schedule of network coded transmissions from half-duplex F-UEs that can maximise the BBU's physical offloading while delivering all requested files from all F-UEs within two transmission phases and conforming with the F-UEs' energy constraints?" To answer this question, the paper achieves the following main contributions:

1) The problem of physical BBU offloading in a device-based F-RAN is formulated as a constrained joint independent set subtraction and chromatic number identification problem over an NC graph, and its NP-hardness is proved.
2) To reduce the computational complexity, a greedy vertex search algorithm on an extended half-duplex ONC graph is proposed. This graph shows all possible conflicts in the system while complying to all energy constraints.
3) Extensive simulations have been performed to both exhibit the merits of our proposed solution in offloading the BBU's physical resources and study how this offloading is influenced by the energy limits that are imposed by the F-UEs in order to preserve their batteries.

D. PAPER ORGANIZATION

The rest of the paper is organised as follows. The system model and parameters are introduced in Section II.

In Section III, the problem is formulated for the general case where each F-UE is under a different energy constraint. The problem is shown to be NP-hard, thus Section IV introduces the proposed heuristic algorithm. Simulation results and discussion are presented in Section V. Finally, Section VI concludes the paper.
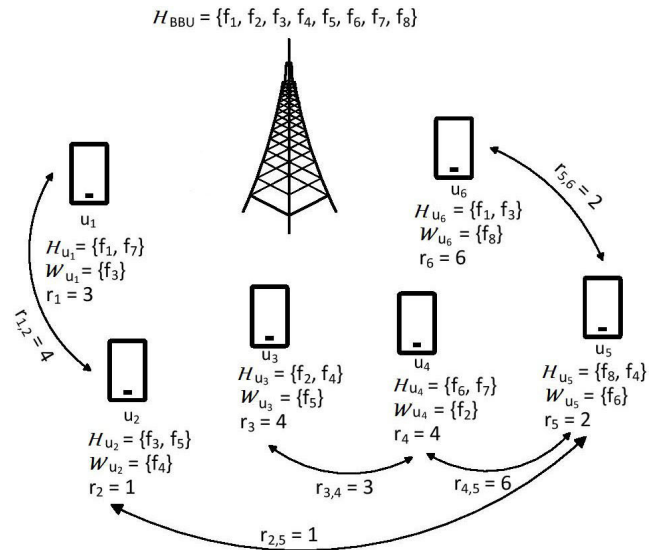
FIGURE 1. The example architecture.

II. SYSTEM MODEL

This paper studies a type of F-RAN architecture that uses the F-UEs as caches. As shown in Fig. 1, the BBU connects to U F-UEs via a wireless fronthaul. The BBU holds the complete library of files $\mathcal{H}_{BBU} = \{f_1, \ldots, f_F\}$ and during off-peak times, caches a subset of them onto each F-UE $u_i \in \mathcal{U} = \{u_1, \ldots, u_U\}$. The files cached onto each F-UE together with any previously downloaded files form the F-UE's Has set $\mathcal{H}_{u_i}, \forall u_i \in \mathcal{U}$. It is assumed that the F-UEs' Has sets collectively possess all the files in the BBU's library, $\mathcal{H}_{BBU}$. Typically, each F-UE is interested in downloading only one file at a time (e.g. watch one popular video at a time). The requested file forms the F-UE's Wants set $\mathcal{W}_{u_i}, \forall u_i \in \mathcal{U}$ for this transmission epoch (two transmission phases). A transmission epoch starts with a scheduling process followed by two transmission phases. The scheduling process is the process of BBU running the algorithm and determining the schedules for the two transmission phases. During the two transmission phases, the F-UEs exchange packets and/or receive packets from the BBU. Please note that we use two transmission phases and an epoch interchangeably despite the slight difference.

It is reasonable to assume that the signal fading on the channels between the F-UEs follow the Rician distribution, as they typically have strong line-of-sight components to one another. On the other hand, the signal fading on the channels between the BBU and each F-UE will follow the Rayleigh distribution to properly represent the dense scattering of these

signals before reaching the F-UEs. As shown in Fig. 1, the quality of the channels between the BBU and each F-UE is represented by rate r_{u_i} defined as [28]:

$$r_{u_i} = B \cdot \log_2 \left(1 + \frac{P_t |h_i|^2}{\sigma^2} \right), \quad (1)$$

where B is the channel bandwidth, P_t is the transmit power of the transmitting device, h_i is the complex channel gain from the BBU to the F-UE u_i, and σ^2 is the additive white Gaussian noise (AWGN) power. The rate between two F-UEs u_i and u_j is noted as $r_{i,j}$. For a successful transmission to F-UE u_i, the BBU should not send with a rate exceeding the rate r_{u_i}. To ensure this, if a transmitter is simultaneously sending to two or more F-UEs, it must choose the lowest rate. Thus, the transmitter would need $(F_s)/(min_{u_i \in S} r_{u_i})$ seconds to send an encoded (combined) file to the set $\mathcal{S}$ of targeted receivers, where F_s is the size of the file in Mbits (equal size files is assumed). Thus, the rate determines the duration of the transmission t which also means that it influences the amount of energy E_{u_i} that the transmitter has to use. We define E_{u_i} as the energy an F-UE u_i uses for a single transmission and E_m as the maximum energy each F-UE is allowed to spend on a single transmission. Although the F-UEs use energy for functions like coding, decoding, being in idle state, etc., our energy constraint is related to the energy used solely for transmissions. The XOR operation used for coding/decoding is computationally inexpensive, and the necessary amount of energy is negligible.

Realistically, the F-UEs are assumed to be half-duplex (they can either transmit or receive at a given time). This assumption together with the aim to keep high quality of experience necessitate the delivery of the requested files in at most two transmission phases (one transmission epoch), $t \in \{1, 2\}$. Scheduling transmissions on these two phases enables the opportunity for F-UEs' to collaborate with each other while still receiving their requests (from another F-UE or the BBU). It is assumed that some of the F-UEs cannot transmit directly to each other, however, the BBU can reach all F-UEs. All F-UEs in the system can use network coded transmissions and combine several files to serve more F-UEs with fewer transmissions. They thus need to ensure that each served F-UE by a coded file can immediately decode its requested file from it. Therefore, the ONC concept is suitable for this purpose as it not only ensures the immediate usability of the received information but also, is computationally inexpensive.

It is assumed that the BBU has huge processing capabilities, all the information about the channels' quality, and the Has $\mathcal{H}_{u_i}$ and Wants $\mathcal{W}_{u_i}$ sets of all F-UEs. Thus, it is the unit that computes the optimal transmission schedule between the F-UEs and itself for the two transmission phases. It then uses control channels to instruct the F-UEs of their roles (transmitter or receiver) in each of the two transmission phases, the files they have to transmit, and the devices they have to target or tune to. The F-UEs can send on one channel

at a time, whereas the BBU can utilise multiple channels, where all channels are assumed to be orthogonal. The sum of the times consumed by the BBU across all the channels involved in delivering requests to the F-UEs is defined as the consumed time of the BBU.

III. PROBLEM FORMULATION

The aim of this paper is to find the best schedule for transmitting/receiving F-UEs, files, and possibly the BBU in two transmission phases so as to minimise the consumed time the BBU needs to serve the remaining F-UEs while obeying the F-UEs energy constraints. In this section, we first look at an example that shows the nature of the problem and the potential benefit an efficient solution can bring. Then, we formulate the problem and discuss the effect of the transmit power on the performance. Having energy constraints would help the F-UE devices have longer battery life and potentially convince more F-UEs to participate in cooperative file delivery.

A. MOTIVATING EXAMPLE

The network used for this example is shown in Fig. 1. The BBU has the whole library of 8 files (each file is 12Mbits) and has already cached some of them on the F-UEs' caches. The BBU knows the channel rate r_{u_i} to each F-UE and the rates between all F-UEs $r_{i,j}$. For simplicity, only some of the rates between the F-UEs are shown in Fig. 1.

The optimal solution for this example results in 3 seconds of BBU consumed time on one BBU channel. However, in order to illustrate the problem's nature and the benefit the smart scheduling on the physical layer could bring, two other possible solutions are listed here.

Solution 1: In the first transmission phase, u_2 sends f_3 to u_1, u_3 sends f_2 to u_4 and u_5 sends f_8 to u_6; in the second transmission phase, u_4 transmits file f_6 to u_5, the BBU sends f_4 and f_5 to u_2 and u_3 in one coded transmission in the form of $f_4 \oplus f_5$.

Solution 2: In the first transmission phase, u_2 sends f_3 to u_1, u_4 sends f_6 to u_5; in the second transmission phase, u_5 sends f_4 to u_2 and the BBU sends f_5 to u_3, f_2 to u_4 and f_8 to u_6 in three different transmissions.

Thus, Solution 1 requires only one BBU channel and Solution 2 requires three BBU channels. According to the offloading strategy in our previous work [29], [30], Solution 1 is the better solution because it minimises the number of the utilised BBU channels. Yet, this solution may not minimise the actual physical time that the BBU spends (whether on one or more multiple channels) to serve the F-UEs. Therefore, the download rates need to be also considered to properly assess the actual BBU offloading. As explained in Section II, having the file size F_s and the rate r_{u_i} between the BBU and the receiving F-UE, allows us to easily calculate the duration of the transmission. In Solution 1, the BBU would spend $(12Mbits)/(min\{1Mbits, 4Mbits\}) = 12$ seconds for the single transmission, whereas, in Solution 2, the BBU

would need $(12Mbits)/(4Mbits) + (12Mbits)/(4Mbits) + (12Mbits)/(6Mbits) = 8$ seconds[1] for the three transmissions. Despite utilising more channels, the BBU sends over better ones in Solution 2 and thus, its resources are occupied for less total time.

Let us now put a limit on the energy that the serving F-UEs are allowed to use for one transmission. For simplicity, we assume that the energy constraint is the same for all F-UEs and set it to $0.5J$. Also, we assume all F-UEs transmit with power $P_t = 0.1W$. The duration of each transmission can be calculated just like in the first part of this example. Then, we know the energy E for a particular transmission can be calculated as [31]:

$$E = P_t \cdot t \qquad (2)$$

and we can compare it to the threshold of $0.5J$. Let us examine again the two solutions and calculate the required energy for each of the cooperative transmissions.

Solution 1: In the first transmission phase, u_2 sends f_3 to u_1. The channel rate $r_{1,2}$ between u_1 and u_2 is 4, thus $E_{u_2} = P_t \cdot t = 0.1 \times (12/4) = 0.3J$. Doing similar calculations reveals that the transmission between u_3 and u_4 takes $0.4J$, between u_5 and u_6 requires $0.6J$, and between u_4 and u_5 needs $0.2J$. Only u_5 exceeds the energy limit and consequently, this transmission is not allowed. The BBU has to serve u_6 and use $t_{BBU} = (12/6) = 2$ seconds (the channel rate between the BBU and u_6 is 6). As mentioned before, the BBU also has to serve u_2 and u_3 in one coded transmission that uses 12 seconds. The total time that the BBU spends for this solution is thus 14 seconds.

Solution 2: In the first transmission phase, u_2 sends f_3 to u_1, which requires $0.3J$ of energy from u_2. The transmission from u_4 to u_5 requires $0.2J$ and from u_5 to u_2 requires $1.2J$ (the channel rate is 1). As the last transmission exceeds the energy limit, the BBU has to serve u_2. Luckily, the u_2's request can be coded together with the u_3's request and send in 12 seconds. As mentioned before, the BBU also has to serve u_4 (in 3 seconds) and u_6 (in 2 seconds). Thus, Solution 2 requires the BBU to spend $12 + 3 + 2 = 17$ seconds. Consequently, in this case, Solution 1 would be the preferred one.

In conclusion, the introduction of F-UEs' energy constraints can prevent some transmissions and restrain the cooperative capabilities of the F-UEs. Therefore, we need to find the optimal solution of the system given the F-UEs' energy constraints and channel conditions.

B. ENERGY-CONSTRAINED COOPERATIVE D2D TRANSMISSIONS

We extend the NC conflict graph $\mathcal{G}$ [32], [33] to model the energy-aware cross-layer BBU offloading problem. Each vertex $v_{i,k}$ of the graph represents a user u_i that requires a file f_k. It is well-known that the conditions for connecting two vertices $v_{i,k}$ and $v_{i',k'}$ by an edge are: 1) the requested files k and k' are different, and 2) at least one of the F-UEs i or i' does not know the file requested by the other F-UE (i.e., $f_k \neq f_{k'}$ AND either $f_k \notin H_{u_{i'}}$ OR $f_{k'} \notin H_{u_i}$). If both conditions are satisfied and thus, there is an edge between the vertices, the coded packet identified by the vertices would not be immediately decodable by at least one of these F-UEs. This means that each edge indicates a conflict between the vertices it connects. Consequently, each independent set (IS)[2] in the graph shows that the files indicated by the vertices can be encoded together and served in one transmission by an F-UE or the BBU.
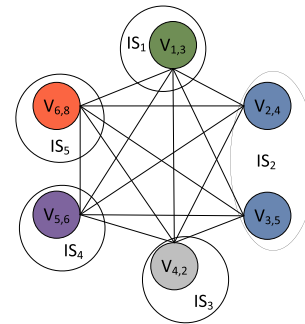

FIGURE 2. Traditional NC graph used to model the network in Fig. 1.

Fig. 2 shows how the traditional NC conflict graph can be used to model the network in Fig. 1. In this particular example, the only coding opportunity can be utilised by the BBU because no F-UE possesses the files identified by the IS_2.

In this work, our aim is to ensure that each serving F-UE does not exceed its energy threshold for a single transmission. The well-known relationship between energy, transmit power and duration of the transmission is given in Eq. (2). As explained in Section II, if the transmitted file size F_s and the rate r_{u_i} between the BBU and u_i are known, t can be simply calculated as: (F_s/r_{u_i}) (the same argument is valid for the rates between the F-UEs $r_{i,j}$). In this paper, we assume that the file size F_s is constant and that the BBU has knowledge about the F-UEs' transmit powers and the rates on all channels. By performing simple calculations, the BBU can determine if a transmitting F-UE to any other given F-UE will exceed its energy constraint. While guaranteeing this constraint, we should find the transmission schedule among F-UEs that will result in the shortest time from the BBU to serve the remaining F-UEs' requests.

In the traditional NC graph, the ISs $I(u_1^{T,(t)}), \ldots, I(u_U^{T,(t)})$ should be served by $u_1^{T,(t)}, \ldots, u_U^{T,(t)}$. After the eligible F-UEs serve each other in a time epoch, there might be some remaining vertices that can form a set of ISs. We denote this

---

[1]Please note that as mentioned earlier, the BBU consumed time is defined as the sum of the times consumed by the BBU on all channels.

[2]An IS is a set of vertices in a graph no two of which are connected by an edge.

set as $\mathcal{Q}$:

$$\mathcal{Q} = \mathcal{G} \setminus \coprod_{t=1}^{2} \left(\coprod_{i=1}^{U^{T,(t)}} I(u_i^{T,(t)}) \right),$$

subject to: $(1)\ \mathcal{U}^{T,(1)} \cap \mathcal{U}^{T,(2)} = \phi,$

$\quad\quad (2)\ \mathcal{U}^{R,(t)} \cap \mathcal{U}^{T,(t)} = \phi, \quad \forall t \in \{1, 2\},$ (3)

where $\mathcal{U}^{T,(t)} = \left\{ u_1^{T,(t)}, u_2^{T,(t)}, \ldots, u_{U^{T,(t)}}^{T,(t)} \right\}$ is the set of F-UEs transmitting (serving ISs) in the transmission phase t, $U^{T,(t)}$ is the cardinality of the set $\mathcal{U}^{T,(t)}$, $\mathcal{U}^{T,(t)} = \mathcal{U}^{T,(1)} \cup \mathcal{U}^{T,(2)}$, where T represents transmitting F-UEs; $\mathcal{U}^{R,(t)} = \left\{ u_1^{R,(t)}, u_2^{R,(t)}, \ldots, u_{U^{R,(t)}}^{R,(t)} \right\}$ is the set of receiving devices in transmission phase t, and $U^{R,(t)}$ is the cardinality of the set, where R represents receiving F-UEs. Moreover, $\mathcal{U}^{R,(t)} = \coprod_{i=1}^{U^{T,(t)}} C(I(u_i^{T,(t)}))$, where $C(I(u_i^{T,(t)}))$ represents the F-UEs served by the IS $I\left(u_i^{T,(t)}\right)$. The first constraint means that each F-UE can transmit only in one of the transmission phases because it must receive its requested packet during the other transmission phase (within the considered time epoch). The second constraint caters for the half-duplex nature of the F-UEs and thus, ensures that each F-UE can either transmit or receive in one transmission phase, but cannot do both simultaneously. We denote a partitioning in $\mathcal{Q}$ as Q and let $S(Q)$ be the set of ISs in Q. Thus, the problem can be formulated as:

$$(I(u_1^{T,(t)}), \ldots, I(u_U^{T,(t)}), Q)$$

$$= \arg \min_{\substack{I(u_i^{T,(t)}) \in \mathcal{G} \\ Q \in \mathcal{Q}}} \sum_{I \in S(Q)} \frac{B}{min_{u_j \in \mathcal{U}(\mathcal{I})} r_{u_j}},$$

subject to: $(1)\ \mathcal{F}(I(u_i^{T,(t)})) \subseteq \mathcal{H}_{u_i^{T,(t)}}, \quad \forall t \in \{1, 2\}$ and

$\quad\quad i \in \left\{ 1, \ldots, U^{T,(t)} \right\},$

$\quad\quad (2)\ E_{u_i} \leq E_m, \quad \forall i \in \left\{ 1, \ldots, U^{T,(t)} \right\}$

$\quad\quad$ and $\quad \forall t \in \{1, 2\}$ (4)

where $\mathcal{F}\left(I\left(u_i^{T,(t)}\right)\right)$ is the set of files corresponding to the vertices in IS $I\left(u_i^{T,(t)}\right)$. The first constraint ensures that the serving F-UE $u_i^{T,(t)}$ of the IS $I(u_i^{T,(t)})$ possesses in its Has set $H_{u_i^{T,(t)}}$ all the files identified in that IS because these are the files the serving F-UE must transmit to the F-UEs indicated by the IS. The min function ensures that all F-UEs are able to receive the file successfully because the minimum F-UE rate r_{u_j} is selected in each IS. The last condition guarantees that the energy each F-UE uses for a single transmission E_{u_i} should not be more than the maximum allowed energy E_m. In essence, the problem formulation in ((4)) says that we need to find the transmission schedule between the F-UEs that will result in the least amount of consumed time spent by the BBU while guaranteeing the fulfilment of each F-UE's energy constraint.

Finding the optimal solution of the problem in Eq. (4) involves two steps. In the first step, Eq. (3) describes the selection of the IS comprising F-UE vertices only for the two transmission phases and their removal from graph $\mathcal{G}$. It is known from Eq. (4) that the optimal selection of such IS in this first step must leave a subgraph of BBU vertices which would result in the lowest consumed time from the BBU in the second step of the solution. This means that all possible ISs comprising of F-UEs vertices only, must be listed in the first step to find the one that achieves the best BBU outcome in the second step. This IS listing itself is known to be an NP-hard problem [34]. Moreover, as the IS consists of F-UE vertices representing transmissions for two phases, each of the feasible combinations of phase one and phase two vertices must be examined to check on which of them will result in the remaining BBU subgraph with the shortest consumed time. This is clearly a very complex combinatorial problem. Finally, in each of these combinatorial steps, all ISs in the BBU subgraph must be listed and their resulting consumed time must be computed, to be compared with that of the other combinations. This results in an internal NP-hard IS listing in each of these combinatorial steps. The above description clearly shows that finding the optimal solution for this problem is NP-hard.

Furthermore, if equal channel rates are assumed for all F-UEs, the problem can be reduced to the one studied in [30] where the BBU offloading is done in terms of minimizing the used channels, rather than minimizing the physical resources. The problem in that work has been proven to be NP-hard and is a special case of the problem studied here.
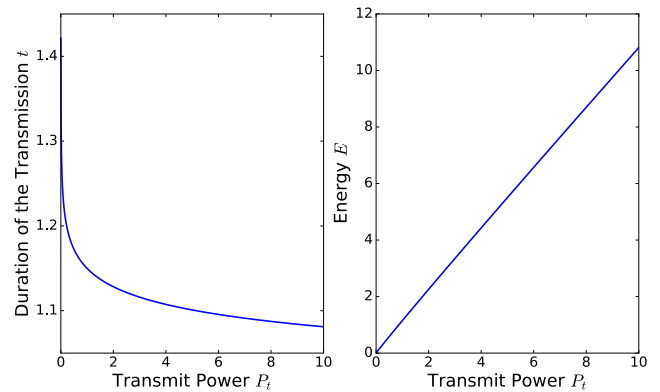

FIGURE 3. The relationship between the transmit power P_t (in the range 0.01 to 10W) and the duration of the transmission t (left). The relationship between P_t and the used energy E (right).

C. EFFECT OF POWER LEVEL ON THE PERFORMANCE

Before starting to develop a solution to the formulated problem, we first discuss here some trivial but important observations. In Fig. (3), we show the relationship between the transmit power P_t and the duration of the transmission t. As mentioned before:

$$E = P_t \cdot t = P_t \cdot \frac{F_s}{B \cdot \log_2 \left(1 + \frac{P_t |h_i|^2}{\sigma^2} \right)}$$ (5)

Parameter	Value
B	200 kHz [35]
N_0	-174 dBm/Hz [36]
σ^2	$1.59e^{-15}$ dBm

$$V_{i, p, j, k, l}$$
$$\downarrow\ \downarrow\ \downarrow\ \downarrow\ \downarrow$$
$$\text{TX PL RX F T}$$

FIGURE 4. Example of a vertex' indices. The transmitter (TX) *i* uses power level (PL) *p* to send the *k*th file (F) to the *j*th receiver (RX) in the *l*th transmission phase (T).

The logarithmic relationship between P_t and t and the linear relationship between P_t and E shows that the P_t term is the main contributor in the equation. As shown in Table 1,[3] the parameters used in Eq. (5) result in a relatively small change in t. Thus, a unit change in P_t induces almost the same change in E (for our realistically chosen parameters). The nearly linear relationship between P_t and E becomes even stronger when P_t increases because t saturates. For example, when the $P_t = 10W$, the duration of the transmission is 1.08 second and thus, the energy for the transmission is 10.8J.

When the objective is to only minimise the BBU consumed time with no energy constraint on the F-UEs, we have shown in [1] that we need to find the optimal (in terms of minimising the BBU transmission time) cooperation between the F-UEs. However, when an energy constraint is set, the F-UEs' collaboration abilities are suppressed. For example, if the energy constraint is very strict and the F-UEs use high transmit power P_t, they will easily exceed the constraint and not be allowed to collaborate, which will result in more BBU transmission time (i.e., less offloading). The main factor for the F-UEs to reach the energy limit is the transmit power P_t they are using. Thus, in order to achieve best collaboration and minimum BBU time, the F-UEs should use minimum P_t. Though this will slow down the transmissions between the F-UEs, the transmission rate degradation is logarithmic.

IV. PROPOSED HEURISTIC ALGORITHM
A. EXTENDED NC GRAPH
The proposed algorithm[4] is classified as a greedy weighted vertex search similar to the algorithms in works like [10] and [37]. As shown in Section III, the vertices of the traditional NC graph show only the file each F-UE requests. As explained in [1], it is not convenient to perform vertex search on this graph because the vertices need to also show information about the possible transmitter that will serve the request and the transmission phase in which the communication should happen. Thus, in this paper, we use the extended NC graph that was first introduced in [30]. However, here we further extend the graph to cater for the addition of the energy constraint that influences the creation of the vertices. As shown in Fig. 4, every vertex acquires an additional index (the second one) that represents the used transmit power level. This value is needed for the calculation of the energy that the transmitter will have to spend. If the calculated energy is more
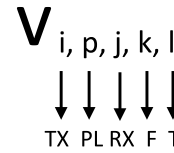
than the allowed one, the communication is not allowed and the vertex is deleted.

In this paper, the graph creates two vertices v_{ipjkl} (for transmission phase $l = 1$ and $l = 2$) for each F-UE u_i that 1) possesses in its Has sets file f_k that is requested by u_j and 2) needs less energy than E_m for transmitting to u_j. The design of the graph should well match our two-phased, multi-transmitter, and energy-constrained system, where all requirements and constraints are embedded and every possible conflict represented as an edge between every two conflicting vertices. The conflicts in the system can be categorized as follows:

- Coding Conflict: occurs if two F-UEs u_i and u_j receive a coded packet that contains more than one file which is not known to at least one receiving F-UE making it undecodable for that F-UE. This can mathematically be expressed as: $i = i'$ and $j \neq j'$ and $k \neq k'$ and $l = l'$ but ($f_k \notin H_{u_{j'}}$ or $f_{k'} \notin H_{u_i}$).
- Transmission Conflict: occurs if a receiving F-UE is served by more than one transmitting F-UEs in the same transmission phase. This can be written as: $i \neq i'$ and $j = j'$ and $l = l'$.
- Half-Duplex Conflict: occurs if an F-UE is scheduled to transmit and receive in the same transmission phase. This conflict is written as: $l = l'$ but ($i = j'$ or $j = i'$).
- Temporal Conflict: occurs if the same F-UE is scheduled to transmit or receive in both transmission phases. This conflict can be expressed as: ($i = i'$ and $l \neq l'$) OR ($j = j'$ and $l \neq l'$).

Note that, initially, the coding conflicts between the vertices for which the BBU is the transmitter will be removed from the graph. Otherwise, any IS-based solution might be incomplete in the sense that it will allow only one transmission from the BBU. However, the BBU is capable of using multiple channels to serve the conflicting requests of F-UEs that were not served by other F-UEs in the two phases. Not including the conflicts between the BBU's vertices thus, allows the inclusion of all such vertices representing non-served requests by F-UEs in any IS. Designed in this way, a maximum IS shows a complete energy-abiding transmission scheme to serve all requested packets by coded transmissions from a subset or all the F-UEs and the BBU[5] in both transmission phases. The question is how to find the IS that, after re-adding the coding conflicts between the BBU vertices, will result in the minimum total transmission time from the BBU.

---

[3]The channel is assumed to be Rayleigh.
[4]Although our greedy algorithm is not guaranteed to find the optimal solution, it provides close to optimal offloading gains with tractable complexity.

[5]For the BBU, the IS shows the packets it needs to send without specifying their conflict-free coded combinations.
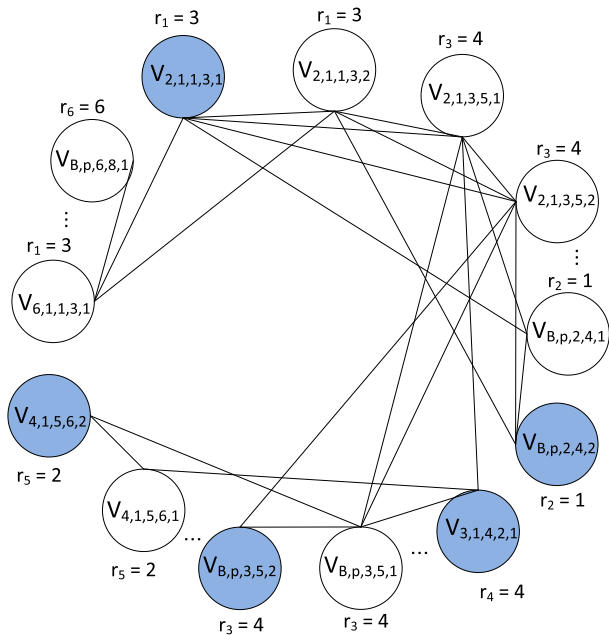
FIGURE 5. This extended NC graph shows only some of the F-UEs' and BBU vertices for the example architecture in Fig. 1.

B. EXAMPLE WITH THE EXTENDED NC GRAPH

Fig. 5 shows the extended NC graph for the motivating example in Section III-A. For clarity, Fig. 5 only depicts some of the F-UEs' vertices and some of the BBU vertices.[6] It was shown earlier that the transmissions from u_5 to both u_2 and u_6 exceed the energy limit and thus, those vertices were not created in the graph. Because the extended graph consists of both the F-UEs' and BBU's vertices, each maximal IS gives a complete transmission scheme (solution) and describes the role of each device in each of the transmission phases. The blue vertices correspond to Solution 1 (after the introduction of the energy limit). It should be noted that at this stage, the BBU vertices are not subjected to the coding conflict. Once the final solution is chosen, in this case Solution 1 (the blue vertices), the coding conflict is applied on the BBU vertices. In this particular case, the BBU can transmit an instantly decodable packet to both u_2 and u_3. However, if Solution 2 was chosen, the coding conflict would induce edges between the BBU verices indicating that the BBU must transmit the uncoded packets on separate channels.

C. PROPOSED CROSS-LAYER HEURISTIC ALGORITHM

The proposed heuristic uses the extended NC graph to greedily search for the vertices that form a maximum IS including either no BBU vertices at all if possible or BBU vertices that result in the shortest BBU transmission time. The latter case is equivalent to a search for the maximum IS with the maximum number of F-UE vertices that serve the F-UEs

[6]All devices can transmit with more than one power level p. However, for readability and clear presentation, we show only one power level for the F-UEs and keep the BBU power level just as p.

having the lowest channel rates to the BBU. The general strategies employed in the proposed heuristic algorithm to achieve this goal can be summarized as follows:

1) Create the extended NC graph that includes only the vertices that do not result in energy constraint violation.
2) Find the IS that contains the maximum number of vertices v_{ipjkl} with worst channels from the BBU to u_j (the F-UE requesting the file) and whose u_i (the transmitting node) is an F-UE.
3) If some requests remain after finding this IS, the BBU will serve them while minimising the amount of consumed time spent in serving those.
4) Whenever the impact on offloading is similar, the algorithm should try to choose F-UE transmissions that are shorter in time to improve the QoS of the receiving F-UEs and further reduce the energy consumption of the transmitting ones.

The proposed algorithm starts by building the graph according to Section IV-A. Each F-UE u_j is assigned a rate r_{u_j} that represents the channel conditions between the receiving F-UE u_j and the BBU. Moreover, each F-UE is assigned a second rate $r_{u_{ij}}$ representing the channel conditions between the transmitting F-UE u_i and the receiving F-UE u_j. Calculating the channel rates between the cooperative F-UEs allows the algorithm to calculate the time needed in Eq. (5) to check if any of the possible transmissions violates the energy limit. The first strategy can thus be fulfilled by not creating any vertices that violate this limit.

To implement the second strategy, a weighted vertex selection approach will be employed. For each of the created vertices, a preliminary weight can be calculated according to the following equation:

$$w_{ipjkl} = \frac{B}{r_{u_j}}. \tag{6}$$

Clearly, the weight of vertex v_{ipjkl} has a larger value if the channel between the BBU and u_j is bad. This means that the rate from the BBU to user u_j is low and thus a transmission from the BBU to this user will consume a long time from the former. Since our aim is to find the solution (i.e., the IS) with the maximum number of maximum weighted F-UE vertices reflecting the above property, we need to introduce more tights of the non-adjacent vertices of vertex v_{ipjkl} into its weights. If we rely only on the weights in Eq. (6), we may select one F-UE vertex that is really bad in terms of channel but is not adjacent to any other bad-channel F-UE vertices. This will either conclude the IS search quickly or will add good-channel F-UE vertices to this IS, leaving a large number of F-UE vertices with very bad-channels to the BBU to be served by the BBU, clearly resulting in very poor performance. That is why we re-weight the vertices to avoid picking the vertex in the above example, and alternatively pick sequentially the ones that are not only bad in terms of channel quality to the BBU but are also non-adjacent with many bad F-UE vertices. This could be done

as follows:

$$w'_{ipjkl} = w_{ipjkl} \times \sum_{v_{i'p'j'k'l'} \in N(v_{ipjkl})} w_{i'p'j'k'l'}, \qquad (7)$$

where $N(v_{ipjkl})$ represents the set of vertices that are not adjacent to v_{ipjkl}. Designing the weights this way means that a vertex will accumulate a large weight if it both has a bad channel to the BBU and is not connected (non-adjacent) to many vertices with bad channels to the BBU. Thus, the final solution will contain large number of vertices with slow transmission channels from the BBU. All the vertices in this IS will be served cooperatively by other F-UEs.

The algorithm operates in an iterative manner, adding to the solution IS at each iteration the maximum weight vertex that both represents a transmitting F-UE and is non-adjacent to all others previously selected vertices in the IS. To improve the QoS for F-UEs and shorten their transmission times (thus fulfilling the fourth strategy), if two or more vertices have the same highest weight in a particular iteration, the algorithm chooses the vertex with the best channel rate between the transmitting and receiving F-UEs. At the end of this process, the IS only contains vertices representing transmitting F-UEs. If the number of these vertices is the same as the number of requests, the solution is complete and the algorithms stops. This solution will definitely be the best possible scenario as it involves no BBU transmissions at all.

On the other hand, if the number of vertices in the IS is smaller than the number of all F-UE requests, this means that some requests have not been served, and thus the BBU has to serve them. Given the design of the graph, all these requests will be represented by all the remaining non-adjacent BBU vertices in the graph to all the vertices of the chosen IS. The algorithm then adds the coding conflicts to this sub-graph of BBU vertices and employs the algorithm in [28] to partition the BBU subgraph in the smallest number of ISs. As such, the third aforementioned strategy will be fulfilled and the algorithm operation is completed.

The pseudo code of all the aforementioned steps is presented in Algorithm 1.

D. COMPLEXITY ANALYSIS

1) VERTEX GENERATION

The extended NC graph creates a vertex for each requested file that is also in an F-UE's or BBU's cache. Each device is able to transmit the file with one of its allowed power levels p.[7] Thus, the total number of the created vertices depends on the number of the F-UEs in the system, the number of files in each F-UE's Has set and the number of transmitting power levels p. Consider the case when none of the transmissions exceed the energy constraint. If each F-UE has cached all files and all files are wanted, then the induced number of vertices would be $2p(U - 1)U$. Moreover, the BBU will induce $2pU$ vertices making the total number of vertices

---

[7]To ease the calculations, we assume that the number of power levels p of the F-UEs and the BBU is the same.

Algorithm 1 Proposed Energy-Aware Cross-Layer (EACL) Algorithm

1: Require: $\mathcal{U}, H_{u_i}, W_{u_i} \; \forall i \in [1, U], r_j, r_{ij}$.
2: Create the v_{ipjkl} using H_{u_i} and E_{u_i}.
3: **if** E_m exists, **then**
4:      using Eq. (5), check if v_{ipjkl} is allowed to exist in $\mathcal{G}$.
5: **end if**
6: Construct $\mathcal{G}$ using v_{ipjkl} and the 4 conflicts.
7: Compute w'_{ijkl} according to Eq. (7).
8: Initialize $k^* = \emptyset$ and numCount $= 0$.
9: **for** each $v_{ipjkl} \in \mathcal{G}$ non-conflicting with k^* and $i \neq B$: **do**
10:      Select $v^*_{ipjkl} = \arg\max_{v_{ipjkl} \in \mathcal{G}} w'_{ipjkl}$
11:      Set $k^* \leftarrow k^* \cup v^*_{i,j,k,l}$
12:      numCount $=$ numCount $+ 1$
13:      **if** several v^*_{ipjkl} with equal w'_{ipjkl} **then**
14:          Select $v^*_{ipjkl} = \arg\max_{v_{ipjkl} \in \mathcal{G}} r_{ij}$
15:      **end if**
16: **end for**
17: **while** numCount $< U$: **do**
18:      Select v^*_{Bpjkl} non-conflicting with k^* such that $v^*_{Bpjkl} = \arg\min_{v_{Bpjkl} \in \mathcal{G}} w_{Bpjkl}'$
19:      Set $k^* \leftarrow k^* \cup v^*_{Bpjkl}$
20:      numCount $=$ numCount $+ 1$
21: **end while**
22: Add the coding conflicts between the v_{Bpjkl}.
23: Partition the v_{Bpjkl} in the smallest number of ISs and calculate the time BBU spends for each transmission.
24: Output k^* and the BBU consumed transmission time.

$2p(U - 1)U + 2pU$. Thus, the complexity of building the graph's vertices in the worst-case scenario of having no energy constraint violation cases is of order $O(U^2)$.

2) ADDING CONFLICT EDGES

The process of adding conflict edges to the graph consists of testing all conflict conditions for each pair of vertices. As shown in [29], checking all vertices against the conflict conditions requires $\sum_{i=1}^{V-1}(V - i) = 1/2(V^2 - V) = O(V^2)$ operations, where V is the number of vertices. In the worst complexity case, $V = O(U^2)$, which leads to a complexity of $O(U^4)$ for adding the edges.

3) FINDING THE F-UEs IS

To find the solution IS, the greedy vertex search process requires $O(V) = O(U^2)$ weight computations and one vertex selection per each of maximum of U steps (as the maximum size of an IS is U). The complexity of finding the IS is thus of order $O(U.V) = O(U^3)$.

4) IDENTIFYING BBU TRANSMISSIONS

The BBU vertices that are part of the final solution are sorted and greedily coloured. The sorting step has a complexity of $O(V \log V)$ [28] and the colouring step has a complexity of

$O(V^2)$ in the worst-case scenario. This results in an overall complexity of $O(V^2)$ or $O(U^4)$.

5) OVERALL COMPLEXITY

Given the above components, the overall worst-case complexity of the algorithm is $O(U^2) + O(U^4) + O(U^3) + O(U^4) = O(U^4)$.

It is worth mentioning that depending on the presence and strictness of the F-UEs' energy constraints, some of the vertices will not be allowed to exist in the graph and will not participate in building the edges. Therefore, the actual complexity of the algorithm will be much lower than the above worst-case complexity.

V. SIMULATION RESULTS

The simulation results presented in this section quantify the offloading gains achieved by the proposed cross-layer offloading approach in this paper in comparison to the upper-layer network coding schemes in [29], [30]. They also illustrate the benefit of utilizing NC and the impact of changing the energy constraints and transmission power levels on the consumed BBU time and the duration of the cooperative F-UE transmissions. It is assumed that there are 50 files in the BBU library and each F-UE knows 10 of them. During off-peak times, the BBU randomly caches 10 of those 50 popular files on the F-UEs' caches. We assume that the BBU has a large number of orthogonal channels to allocate to the F-UEs so there is no interference between them. Also, the F-UEs are realistically assumed not to be able to communicate with all other F-UEs at a time. For simplicity, we assume that all F-UEs have to comply to the same energy constraint. As mentioned in Section 3, the channels between the BBU and each of the F-UEs follow a Rayleigh distribution, whereas the channels between each pair of cooperating F-UEs follow a Rician distribution. We assume equal packet size of 12 MB,[8] a bandwidth of 200 kHz per channel [35], and a noise power spectral density of -174dBm/Hz [36].

A. COMPARISON WITH UPPER-LAYER OFFLOADING

Here, our algorithm is compared to the upper layer algorithm proposed in [30]. The upper layer algorithm aims to reduce the number of orthogonal channels used by the BBU without considering the physical channel conditions. As it can be seen on Fig. 6, the upper-layer algorithm performs worse compared to the no energy constraint case and even to most of the cases where there is a constraint on the energy that the F-UEs are allowed to use for the collaborative transmissions. Because the upper layer algorithm does not consider the channel rates, the BBU spends more time transmitting. The upper layer catches up with the cross layer performance only at a large number of cooperating users which is not

---

[8]We assume that the requested popular files are of equal size. However, even if the packets are not of equal size, they can be zero-padded. The chosen packet size is typical for applications like sending animated GIFs, profile photos for social media or logo images. In case of streaming video, the data flow can be divided into smaller equal size packets.
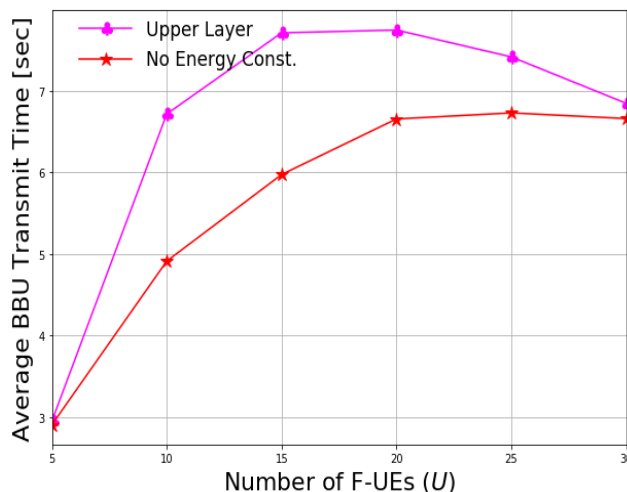

FIGURE 6. Comparison of the BBU consumed time versus the number of F-UEs between the upper layer algorithm and the proposed algorithm for physical offloading without energy constraints.
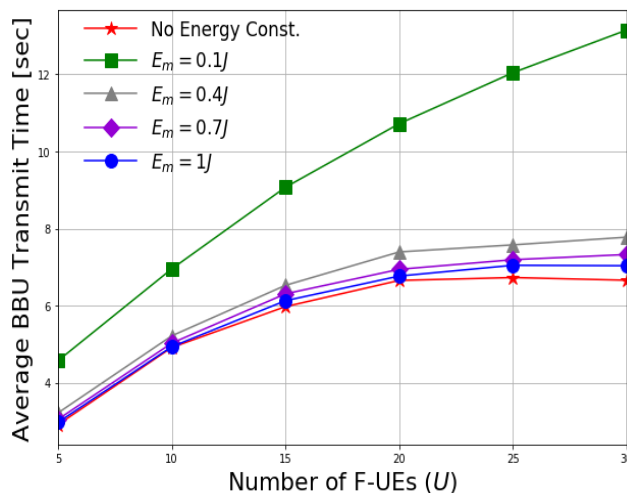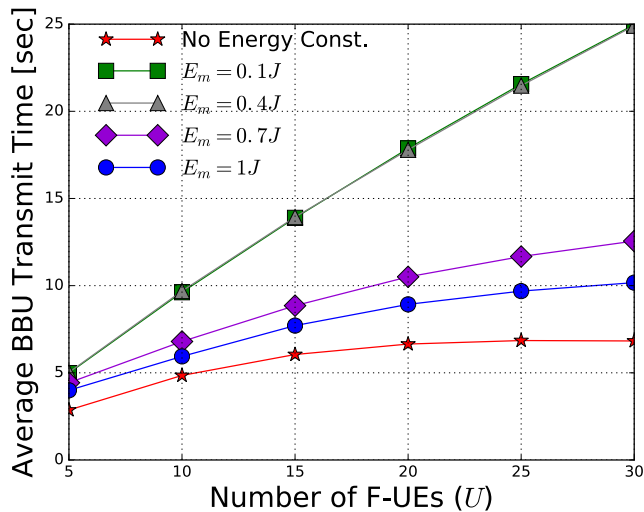

FIGURE 7. Comparison of the BBU consumed time versus the number of F-UEs for different energy constraint requirements and an F-UEs' transmit power P_t of 0.1W.

a typical scenario. The increased number of F-UEs means an increased number of potential transmitters, which allows the BBU to utilise only a small number of channels and improve its performance even if the the channel rates are low. Figures 6 and 7 suggests that the two strategies should work together but for different number of F-UEs.

B. EFFECT OF ENERGY CONSTRAINTS AND POWER LEVELS

Figures 7 and 8 examine the effect that the energy constraint has on the BBU consumed time for transmit powers of 0.1W and 0.7W, respectively. Clearly, stringing the energy constraint decreases the F-UEs' cooperative capabilities because more F-UEs are likely to exceed the energy limit when sending to one another. This results in an increased consumed time

FIGURE 8. Comparison of the BBU transmit time versus the number of F-UEs for different energy constraint requirements and an F-UEs' transmit power P_t of 0.7W.
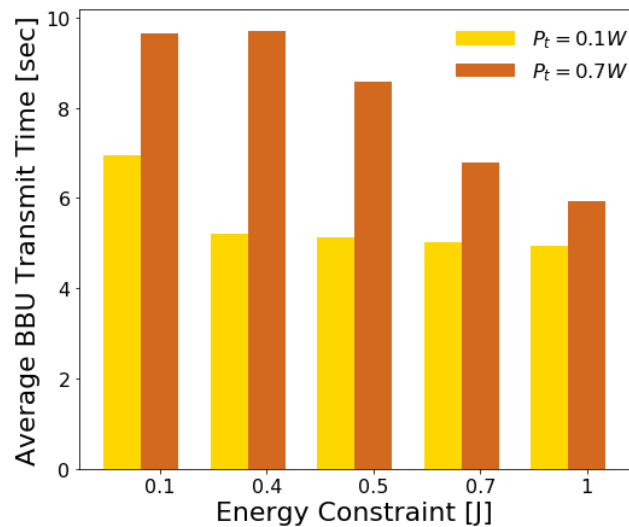

FIGURE 9. Comparison of consumed times from the BBU between different transmit powers P_t and energy constraints E_m for 10 F-UEs (i.e., $U = 10$).

from the BBU as shown in both figures. The figures show that, when there is an energy constraint in place, the BBU's consumed time increases with the number of F-UEs but with a slower slope as more and more F-UEs are involved in the system which increases the collaboration opportunities between the F-UEs.

As stated above, the main difference between Fig. 7 and Fig. 8 is that the F-UE transmit power P_t is increased from 0.1W to 0.7W. By comparing each pair of corresponding energy-constrained curves (i.e., curves for the same energy constraint) between both figures, the impact of this power increase is clearly exhibited by an increase in the consumed time from the BBU (less offloading of physical resources). To further clarify this fact, Fig. 9 highlights the values of the

energy-constrained curves of both Figures 7 and 8 (plus an added energy constraint level of 0.5) for the case of 10 F-UEs. We can clearly notice from this focused figure the afore-mentioned effect of increased BBU consumed time when the F-UEs transmit power is increased. This effect can be easily interpreted from Eq. (1) and (5). Indeed, both equations show that a linear increase in P_t is accompanied by a logarithmic increase in the bit rate, and thus a logarithmic reduction in the transmission time. The combination of these two factors is a sub-linear increase in the consumed energy in each F-UE transmission. Consequently, the increase in P_t causes more F-UEs to exceed their energy constraint and reduce their capacities to serve any other F-UEs, thus resulting in more BBU consumed time spent on serving more unserved F-UEs. For example, at energy constraint of 0.4J and 30 F-UEs, the BBU spends only 8 sec (Fig. 7) when $P_t = 0.1W$, whereas it needs about 25 sec to serve the F-UEs when $P_t = 0.7W$.

Please note that the energy constraints of 0.4J and 0.1J in Fig. (8) exhibit the same BBU consumed time results because they both cannot be fulfilled by any of the F-UEs when $P_t = 0.7W$, and thus all communication is accommodated by the BBU. This is not the case when $P_t = 0.1W$ where some F-UEs can still collaborate at the same energy constraint levels (as illustrated in Fig. 10). Another interesting observation from Fig. 9 is that, when $P_t = 0.1W$, relaxing the energy constraint beyond 0.4J does not contribute much to BBU offloading. This suggests that with a correctly chosen energy limit (in this case 0.4J), we can achieve a good BBU offloading and at the same time prolong the F-UEs' battery life.
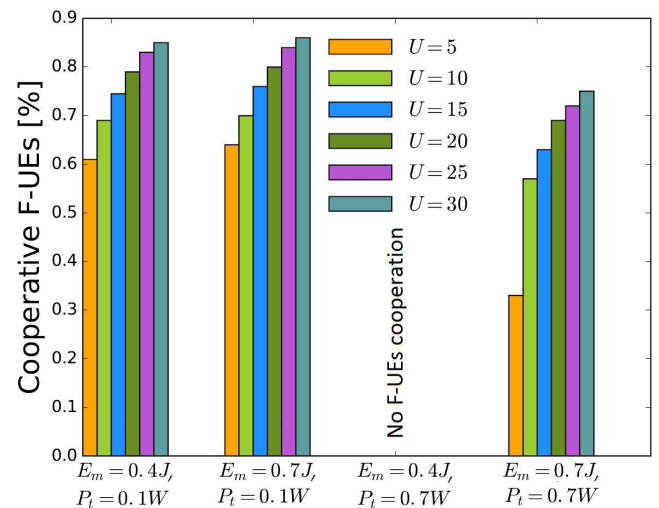
FIGURE 10. Comparison between the number of cooperative F-UEs for different energy constraints E_m and transmit powers P_t.

Fig. 10 confirms that the percentage of cooperating F-UEs decreases with the increase of the transmit power. When $P_t = 0.1W$ the F-UEs achieve very similar level of coop-eration for both energy constraints which results in similar BBU offloading for those cases as it can be seen in Fig. 9. On the other hand, when $P_t = 0.7J$ the F-UEs' cooperation

is very different for the two considered energy constraints: no cooperation for $E_m = 0.4J$ and some cooperation for the more relaxed constraint of $E_m = 0.7J$. The reason for this drastic difference is the significant contribution of the transmit power P_t in the energy equation (5), which leads to many F-UEs surpassing the energy threshold. It can also be noted that the increase in the transmit power, P_t has a more significant impact when there are very few F-UEs. For example, when $U = 5$ and $E_m = 0.7J$, the difference between the cooperation for the two transmit power levels is about 30%. The gap decreases with the increase of the F-UEs and for $U = 30$, it reaches about 10%. This closing gap can be explained as follows. No matter how many F-UEs there are in the system, they are all subject to the same energy threshold and roughly the same percentage of F-UEs for the different cases pass the constraint. However, the larger number of F-UEs in the system gives the potential transmitting F-UEs more opportunities to serve. On the other hand, when there are few F-UEs in the network, the potential transmitters that pass the energy limit might be hindered by obstacles like not being able to physically communicate with the receiving F-UEs (due to not being in the coverage area of the transmitting F-UE) or no F-UE requesting a file in the potential transmitter's Has set.
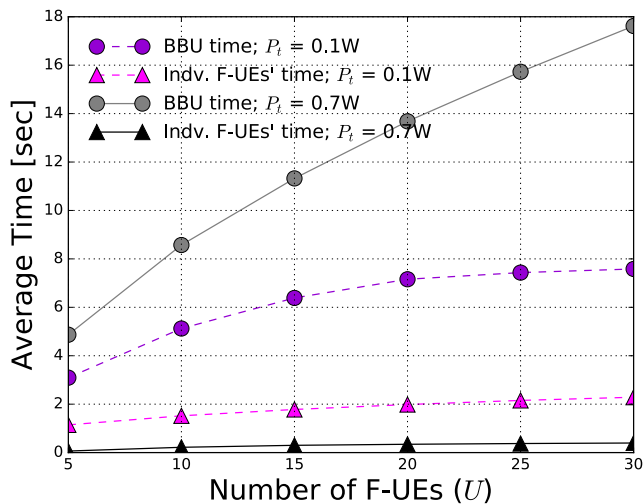

FIGURE 11. Comparison between the time spent by the BBU and the individual F-UE's time for two F-UE transmit power levels of 0.1W and 0.7W. The energy constraint E_m is set to 0.5J.

Fig. 11 compares the time the BBU needs to spend to the time the F-UEs collectively spend for serving each other. In the case of $P_t = 0.1W$, the F-UEs are able to cooperate better but slower and this results in longer average transmission times for the F-UEs and shorter transmission time for the BBU (the BBU's transmit power does not change). However, when the transmit power increases to 0.7W, the F-UEs have limited cooperative capabilities which increases the BBU's working time. This figure shows that in times when the BBU is busy (i.e. peak times), a better offloading of the BBU can be achieved by decreasing the transmit power of the F-UEs
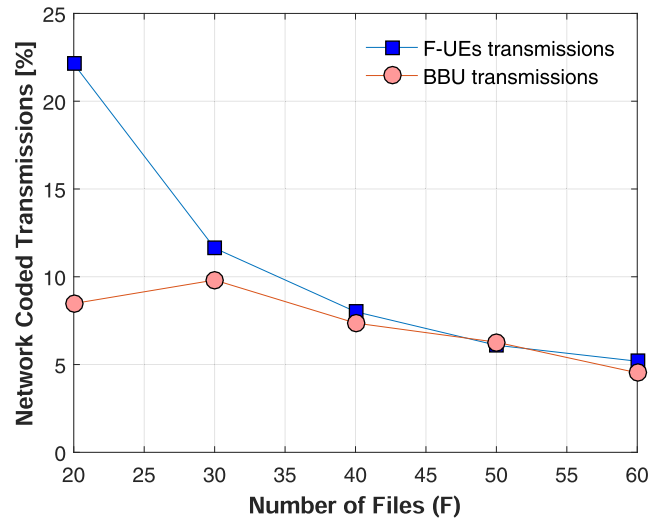
FIGURE 12. Percentage of the network coded transmissions to all transmissions (no energy constraint).
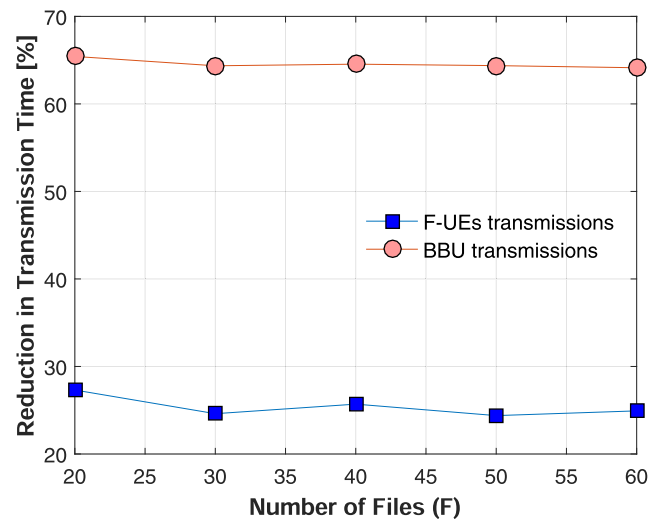
FIGURE 13. Percentage of reduction in transmission time due to using network coded transmissions (no energy constraint).

while still meeting the same energy constraint. For example, for 30 F-UEs an offloading of about 10 sec can be achieved by decreasing the F-UEs' P_t from 0.7W to 0.1W.

The last two figures examine the effect that NC has on the transmission time. In Fig. 12, the number of the F-UEs is 10 and each of the F-UEs knows 10 files meaning that the size of the Has sets is also constant. As the number of files increases, expectedly the network coded opportunities decrease. This has a similar effect as keeping the number of files F constant but decreasing the number of F-UEs U in the system. The decrease in the number of NC opportunities is due to the fact that having more files in the system decreases the probability of an F-UE wanting a file that is known by another F-UE that also has the first F-UE's wanted file in its Has set. Although the number of the network coded opportunities diminishes, Fig. 13 shows that employing NC when

possible is still beneficial. The figure depicts the percentage of reduction in the transmission time among the F-UEs and BBU which is calculated as $(t_{noNC} - t_{NC})/t_{noNC}$, where t_{NC} is the average time the devices spend on NC transmissions and t_{noNC} is also the average time the devices spend on the same transmissions but if NC was not allowed. In other words, tNC and tnoNC represent the same transmission time but with and without NC respectively. The figure quantifies the benefit of utilizing NC and shows in percentage how much time is spared by the F-UEs thanks to using NC compared to the traditional uncoded transmissions.

VI. CONCLUSION

In this paper, we developed an energy-constrained and cross-layer network coding device cooperation algorithm for F-RANs aiming to minimise the BBU's consumed physical resources in delivering cached popular files. Our algorithm ensures this minimization of the BBU's consumed physical time while complying with any energy limitations enforced on the devices. We investigate the change on the BBU load under a variety of circumstances and discuss what factors influence the system's performance. Our simulations show the superior performance of our proposed cross-layer network coding approach compared to its conventional upper-layer counterpart in minimising the BBU's consumed physical resources. They also show that having a strict energy constraint and high F-UE transmit power will increase the BBU's consumed time. We finally illustrate the different levels of BBU offloading that can be achieved with different parameters and constraints. Finally, we showed the value of using network coded transmissions in the system.

REFERENCES

[1] K. Kaneva, N. Aboutorab, S. Sorour, and M. C. Reed, "Cross-layer offloading in fog-RANs using device cooperation and network coding," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[2] *Digital media: Rise of on-demand content*, Deloitte, London, U.K., 2000.

[3] P. Li and S. Guo, "Incentive mechanisms for D2D communications," *IEEE Netw.*, vol. 29, no. 4, pp. 75–79, Jul. 2015.

[4] Y. Zhang, L. Song, W. Saad, Z. Dawy, and Z. Han, "Contract-based incentive mechanisms for Device-to-Device communications in cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2144–2155, Oct. 2015.

[5] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[6] A. Agarwal and M. Charikar, "On the advantage of network coding for improving network throughput," in *Proc. Manuf. Eng.*, 2004, pp. 247–249.

[7] V. N. Swamy, P. Rigge, G. Ranade, A. Sahai, and B. Nikolic, "Network coding for high-reliability low-latency wireless control," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, pp. 1–7.

[8] K. Alic and A. Svigelj, *Opportunistic Network Coding*. Cham, Switzerland: Springer, 2018, pp. 319–338.

[9] B. Muniswamy and N. Geethanjali, "A review on network coding and its applications in wired and wireless networks," *Int. J. Res. Eng. Technol.*, vol. 2, no. 5, pp. 179–186, 2014.

[10] S. Sorour and S. Valaee, "On minimizing broadcast completion delay for instantly decodable network coding," in *Proc. IEEE Int. Conf. Commun.*, May 2010, pp. 1–5.

[11] N. Aboutorab, P. Sadeghi, and S. E. Tajbakhsh, "Instantly decodable network coding for delay reduction in cooperative data exchange systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 3095–3099.

[12] N. Aboutorab and P. Sadeghi, "Instantly decodable network coding for completion time or decoding delay reduction in cooperative data exchange systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1212–1228, Mar. 2016.

[13] S. E. Rouayheb, A. Sprintson, and P. Sadeghi, "On coding for cooperative data exchange," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Jan. 2010, pp. 1–5.

[14] A. Sprintson, P. Sadeghi, G. Booker, and S. E. Rouayheb, "Deterministic algorithm for coded cooperative data exchange," in *Proc. Int. Conf. Heterogeneous Netw. Qual., Rel., Secur. Robustness*, 2010, pp. 282–289.

[15] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb, "A randomized algorithm and performance bounds for coded cooperative data exchange," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 1–4.

[16] M. Ji, G. Caire, and A. F. Molisch, "Wireless Device-to-Device caching networks: Basic principles and system performance," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176–189, Jan. 2016.

[17] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of distributed caching in D2D wireless networks," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2013, pp. 1–5.

[18] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.

[19] Y. N. Shnaiwer, S. Sorour, N. Aboutorab, P. Sadeghi, and T. Y. Al-Naffouri, "Network-coded content delivery in femtocaching-assisted cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–8.

[20] B. Chen, C. Yang, and A. F. Molisch, "Cache-enabled Device-to-Device communications: Offloading gain and energy cost," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4519–4536, Jul. 2017.

[21] P.-Q. Huang, Y. Wang, K. Wang, and Z.-Z. Liu, "A bilevel optimization approach for joint offloading decision and resource allocation in cooperative mobile edge computing," *IEEE Trans. Cybern.*, early access, Jun. 19, 2019, doi: 10.1109/TCYB.2019.2916728.

[22] Z. Hong, W. Chen, H. Huang, S. Guo, and Z. Zheng, "Multi-hop cooperative computation offloading for industrial IoT–Edge–Cloud computing environments," *IEEE Trans. Parallel Distrib. Syst.*, vol. 30, no. 12, pp. 2759–2774, Dec. 2019.

[23] J. Feng, F. Richard Yu, Q. Pei, X. Chu, J. Du, and L. Zhu, "Cooperative computation offloading and resource allocation for blockchain-enabled mobile-edge computing: A deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6214–6228, Jul. 2020.

[24] L. Wang, H. Wu, Y. Ding, W. Chen, and H. V. Poor, "Hypergraph-based wireless distributed storage optimization for cellular D2D underlays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2650–2666, Oct. 2016.

[25] L. Wang, H. Wu, Z. Han, P. Zhang, and H. V. Poor, "Multi-hop cooperative caching in social IoT using matching theory," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2127–2145, Apr. 2018.

[26] Y. N. Shnaiwer, S. Sorour, P. Sadeghi, N. Aboutorab, and T. Y. Al-Naffouri, "Network-coded macrocell offloading in femtocaching-assisted cellular networks," in *Proc. TVT*, 2016, p. 6.

[27] Y. Wu, J. Chen, L. P. Qian, J. Huang, and X. S. Shen, "Energy-aware cooperative traffic offloading via Device-to-Device cooperations: An analytical approach," *IEEE Trans. Mobile Comput.*, vol. 16, no. 1, pp. 97–114, Jan. 2017.

[28] Y. N. Shnaiwer, S. Sorour, T. Y. Al-Naffouri, and S. Al-Ghadhban, "Cross-layer cloud offloading using fog radio access networks and network coding," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–10.

[29] 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. 2020.

[30] K. Kaneva, N. Aboutorab, S. Sorour, and M. C. Reed, "On offloading fog radio access networks fronthaul using device caching and cooperation," in *Proc. IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–6.

[31] C. Zhang, F. Vahid, J. Yang, and W. Najjar, "A way-halting cache for low-energy high-performance systems," *ACM Trans. Archit. Code Optim.*, vol. 2, no. 1, p. 34–54, Mar. 2005.

[32] M. A. R. Chaudhry and A. Sprintson, "Efficient algorithms for index coding," in *Proc. IEEE INFOCOM - IEEE Conf. Comput. Commun. Workshops*, Apr. 2008, pp. 1–4.

[33] S. Y. El Rouayheb, M. A. R. Chaudhry, and A. Sprintson, "On the minimum number of transmissions in single-hop wireless coding networks," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2007, pp. 120–125.

[34] R. M. Garey and S. D. Johnson, *Computers and Intractability; A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co., 1990.

[35] Y. Chen, E. Sun, and Y. Zhang, "Joint optimization of transmission and processing delay in fog computing access networks," in *Proc. 9th Int. Conf. Adv. Infocomm Technol. (ICAIT)*, Nov. 2017, pp. 155–158.

[36] A. Rabih, A. Ghandour, Y. N. Shnaiwer, and S. Al-Ghadhban, "Rate-aware instantly decodable network codes for heterogeneous cellular networks," in *Proc. IEEE 89th Veh. Technol. Conf. (VTC-Spring)*, Apr. 2019, pp. 1–5.

[37] S. Sorour and S. Valaee, "Minimum broadcast decoding delay for generalized instantly decodable network coding," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2010, pp. 1–5.

KAMELIYA KANEVA (Student Member, IEEE) received the bachelor's degree from the Technical University of Varna, Bulgaria, in 2010, and the master's degree from Harbin Engineering University, Harbin, China, in 2014. She is currently pursuing the Ph.D. degree in electrical engineering with the University of New South Wales Canberra, Australia. Her research interests include wireless networking, device-to-device communication, and distributed networking.
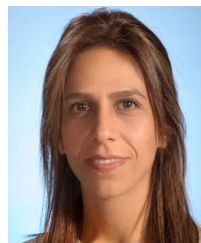
NEDA ABOUTORAB (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from The University of Sydney, Sydney, NSW, Australia, in 2012. She is currently a Senior Lecturer with the School of Engineering and Information Technology, University of New South Wales Canberra, Australia. From 2012 to 2015 and before joining the University of New South Wales Canberra, she was a Postdoctoral Research Fellow at the Research School of Engineering, The Australian National University. Her research interests include network coding, wireless communications, data caching and storage systems, compressive sensing, and signal processing.

SAMEH SOROUR (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Alexandria University, Egypt, in 2002 and 2006, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Toronto, Canada, in 2011. After two postdoctoral fellowships at the University of Toronto and the King Abdullah University of Science and Technology (KAUST), he joined the King Fahd University of Petroleum and Minerals (KFUPM), in 2013, before moving to the University of Idaho, in 2016. He is currently an Assistant Professor with the School of Computing, Queen's University, Canada. His research interests include advanced communications/networking/computing/learning technologies for smart cities applications, including cyber physical systems, the Internet of Things (IoT) and IoT-enabled systems, cloud and fog networking, network coding, device-to-device networking, autonomous driving and autonomous systems, intelligent transportation systems, and mathematical modeling and optimization for smart systems.

MARK C. REED (Senior Member, IEEE) received the B.Eng. degree (Hons.) in electronic engineering from the Royal Melbourne Institute of Technology, in 1990, and the Ph.D. degree in communication engineering from the University of South Australia, Australia, in 2000. He is currently a Senior Director with the CTOs Office, Dolby Laboratories Inc., and an Adjunct Associate Professor with UNSW Canberra. He previously was the Founder and the CEO of InterfereX Communications Pty. Ltd. He has published numerous journal and conference papers, where his research interests include applications of iterative techniques to signal processing problems, fundamental limits of heterogeneous wireless networks, modem signal acquisition and tracking techniques, asynchronous logic, coding techniques, and other communication systems which help increase data throughput by 100 times per decade.

• • •