This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMC.2014.2314107



# Optimal Scheduling for Multi-radio Multi-channel Multi-hop Cognitive Cellular Networks

Ming Li, *Student Member, IEEE,* Sergio Salinas, *Student Member, IEEE,* Pan Li, *Member, IEEE,* Xiaoxia Huang, *Member, IEEE,* Yuguang Fang, *Fellow, IEEE,* and Savo Glisic, *Senior Member, IEEE*

***Abstract*—Due to the emerging various data services, current cellular networks have been experiencing a surge of data traffic and already overloaded, thus not able to meet the ever exploding traffic demand. In this study, we first introduce a Multi-radio Multi-channel Multi-hop Cognitive Cellular Network ($\mathrm{M}^3\mathrm{C}^2\mathrm{N}$) architecture to enhance network throughput. Under the proposed architecture, we then investigate the minimum length scheduling problem by exploring joint frequency allocation, link scheduling, and routing. In particular, we first formulate a maximal independent set based joint scheduling and routing optimization problem called Original Optimization Problem (OOP). It is a Mixed Integer Non-Linear Programming (MINLP) and generally NP-hard problem. Then, employing a column generation based approach, we develop an $\epsilon$-bounded approximation algorithm which can obtain an $\epsilon$-bounded approximate result of OOP. Noticeably, in fact we do not need to find all the maximal independent sets in the proposed algorithm, which are usually assumed to be given in previous works although finding all of them is NP-complete. We also revisit the minimum length scheduling problem by considering uncertain channel availability. Simulation results show that we can efficiently find the $\epsilon$-bounded approximate results and the optimal result as well, i.e., when $\epsilon = 0\%$ in the algorithm.**

***Index Terms*—Cognitive cellular networks; multi-radio multi-channel; cross-layer optimization; minimum length scheduling.**

## I. Introduction

Due to the emerging various data services, current cellular networks have been experiencing a surge of data traffic and already overloaded, thus not able to meet the ever exploding traffic demand. Even the new generation LTE or WiMAX cellular networks may still suffer from low per-user throughput because of a large number of network users sharing limited frequency bandwidth as well as poor cellular signals in certain areas like obstructed or suburban areas. Although Wi-Fi networks may provide high data rates, they have serious shortcomings as well. First, wireless local area networks (WLANs) or hot spots have poor coverage and can easily get overcrowded. Second, citywide Wi-Fi networks like mesh networks have not been widely deployed yet, thus requiring additional deployment cost, and may interfere with existing WLANs, hot spots, and other ISM (Industrial, Scientific and Medical) band users (e.g., cordless phones, RFID systems, wireless telemetric systems like smart meter networks).

In this paper, we first introduce a Multi-radio Multi-channel Multi-hop Cognitive Cellular Network ($\mathrm{M}^3\mathrm{C}^2\mathrm{N}$) architecture to meet the fast-growing traffic demand in cellular networks. In particular, both cellular base stations and network users are equipped with multiple cognitive radios. Thus, we can exploit the greatly under-utilized licensed spectrums, i.e., white spaces/spectrum holes, for communications, and hence enhance network throughput. Moreover, instead of delivering all the traffic between base stations and users in one hop like that in traditional cellular networks, we propose to carry such traffic in hybrid mode, i.e., either in one-hop or via multiple hops depending on the local available frequency channels and the corresponding channel conditions. In so doing, we can further take advantage of local available channels, frequency reuse, and link rate adaptivity to provide higher network throughput. Note that a couple of works such as [1], [2] investigate the capacity of such multihop cellular networks and have shown that such hybrid mode communications can improve the network capacity a lot compared to one-hop communications. However, these works only consider the case where nodes share the cellular frequency channels and have not exploited the local available channels or multi-radio as we propose in this study. Besides, although asymptotic capacity bounds have been studied, the exact optimal throughput value remains unknown. Generally, the proposed $\mathrm{M}^3\mathrm{C}^2\mathrm{N}$ architecture can enhance network performance and adapt to dynamic traffic distribution, yet relieving service providers from any significant additional infrastructure costs.

Under the $\mathrm{M}^3\mathrm{C}^2\mathrm{N}$ architecture, we investigate the minimum length scheduling problem by exploring joint frequency channel allocation, link scheduling, and routing. Specifically, by constructing a two-layer conflict graph, we first formulate a maximal independent set based joint scheduling and routing optimization problem called Original Optimization Problem (OOP). It is a Mixed Integer Non-Linear Programming (MINLP) and generally NP-hard problem. We notice that finding all the maximal independent sets in a conflict graph is NP-complete, and most previous research just assumes that they are given [3]–[5]. In this study, we do not make such assumptions. Instead, we decompose OOP into a sequence of Linear Programming (LP) problems, which

This work was supported by the U.S. National Science Foundation under grants CNS-1149786 (CAREER Award), ECCS-1128768/1129062, and CNS-1147851/1147813. The work of Huang was supported by the National High Technology Research and Development Program of China (863 program) under grant No. 2011AA010503.

M. Li, S. Salinas, and Pan Li are with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762. E-mail: {ml845@, sas573@, li@ece.}msstate.edu.

X. Huang is with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. E-mail: xx.huang@siat.ac.cn.

Y. Fang is the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32608. Email: fang@ece.ufl.edu.

S. Glisic is with the Department of Communications Engineering, University of Oulu, Finland. Email: savo.glisic@ee.oulu.fi.

we name Master Problems (MPs). After that, employing a column generation based approach, we further decompose each MP into a Restricted Master problem (RMP) and a Pricing Problem (PP), which are a small-scale LP problem and a Binary Integer Programming (BIP) problem, respectively. The basic idea is that RMP starts with some initial independent sets, while PP updates the set of independent sets in each iteration. Notice that RMP can be solved in polynomial time, but PP is still a problem with high complexity. Therefore, we design a sequential-fix (SF) algorithm which can find a suboptimal solution to PP in polynomial time. Although SF is suboptimal, we can still find the optimal solution to MPs and hence OOP due to the intrinsic iterative nature of column generation. Besides, it has been observed in the context of column generation algorithms [6], [7] that one can usually determine solutions that are at least $95\% - 99\%$ of the global optimality fairly quickly. Subsequently, we develop an $\epsilon$-bounded approximation algorithm, which can obtain upper and lower bounds that are less than $(1+\epsilon)$ and larger than $(1-\epsilon)$ of the optimal result of each MP, respectively, and an $\epsilon$-bounded approximate result of the OOP. Simulation results show that upper and lower bounds converge quickly and thus we can efficiently find the optimal result as well, i.e., when $\epsilon = 0\%$ in the algorithm. In other words, we are able to solve OOP very efficiently without having to find all the maximal independent sets. Furthermore, although most previous research on network optimization assumes constant channel bandwidth, in practice, the vacancy/occupancy of licensed channels can be uncertain and dynamic at different times, due to the unpredictable activities of the primary users. In this study, we also revisit the minimum length scheduling problem by taking uncertain channel availability into consideration.

Our main contributions can be summarized as follows.

- We introduce a Multi-radio Multi-channel Multi-hop Cognitive Cellular Network ($M^3C^2N$) architecture and a new hybrid mode communication scheme to enhance network throughput.
- We explore the minimum length scheduling problem by joint frequency channel allocation, link scheduling, and routing. Most previous works only obtain suboptimal results that are either unbounded or still far from the optimal results, and many works based on conflict graphs also assume that all the maximal independent sets are given. In this paper, we develop a column generation based $\epsilon$-bounded approximation algorithm, which relaxes this assumption and is able to find tight $\epsilon$-bounded approximate solutions and the optimal solutions as well. The developed algorithm can also be applied to cross-layer optimization problems in other networks.
- We consider heterogeneous networks and take uncertain channel availability into account when studying the minimum length scheduling problem, which is an intrinsic feature of cognitive radio networks but has rarely been studied before.
- We conduct extensive simulations to validate the efficiency of the proposed algorithms.

The rest of this paper is organized as follows. We discuss some related work in Section VII and briefly introduce our system models in Section II. We then formulate a minimum length scheduling problem for $M^3C^2$Ns in Section III. After that, we propose in Section IV a column generation based $\epsilon$-bounded approximation algorithm which can efficiently find $\epsilon$-bounded approximate solutions and the optimal solution when $\epsilon = 0$. Subsequently, we revisit the minimum length scheduling problem by considering uncertain channel availability in Section V. Simulations results are presented in Section VI to evaluate the performance of the proposed algorithm. We finally conclude this paper in Section VIII.

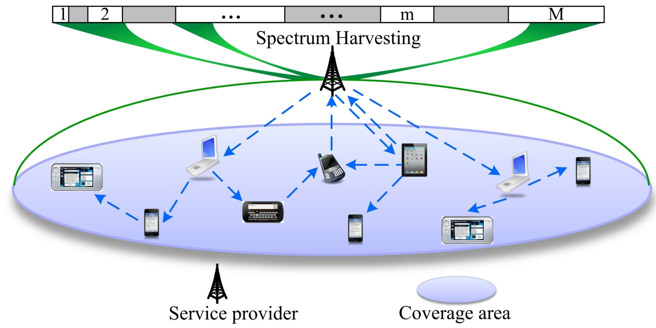


Fig. 1. The architecture of a multi-radio multi-channel multi-hop cognitive cellular network ($M^3C^2N$).

## II. System Models

### A. Network Architecture

As shown in Fig. 1, we propose a novel Multi-radio Multi-channel Multi-hop Cognitive Cellular Network ($M^3C^2N$) architecture. Specifically, an $M^3C^2N$ is a cellular network in which both the service provider and network users can access multiple channels with multiple cognitive radios. For example, base stations and more powerful terminals (e.g., laptops and tablets) can have higher cognitive capabilities and span a larger range of frequency spectrum (e.g., from MHz spectrum to GHz spectrum), while less powerful devices (e.g., smart phones and cellular phones) may just access only several typical frequency spectrum, such as the cellular spectrum, the 2.4 GHz ISM spectrum, and the TV spectrum which has large bandwidth and good penetration and propagation performances. We call cellular spectrum "the basic channel", and other spectrums "the secondary channels". The service provider uses the basic channel for signaling, controlling, handling handoffs, accommodating users' voice traffic, etc., and uses all the available channels to support users' data traffic. As a central coordinator, the service provider performs network optimization to find out the optimal radio and frequency allocation, link scheduling, and routing schemes for satisfying users' traffic demand based on the the observed, collected, and predicted channel information [8]–[10] in the coverage area.

Besides,instead of delivering all data traffic in one hop like that in traditional cellular networks, we propose to carry such traffic either in one-hop or via multiple hops, depending on the available channels and the corresponding channel conditions. In addition, since downlink transmissions from base stations to users will likely outweigh uplink transmissions, we focus on downlink transmissions in this study. The analysis for uplink transmissions simply follows the same process presented herein.

### B. Network Model

Consider a cell in an M$^3$C$^2$N consisting of $\underline{\mathcal{N}} = \{1, 2, \cdots, i, \cdots, N\}$ users and a set of available secondary channels $\underline{\mathcal{M}} = \{1, 2, \cdots, b \cdots, M\}$ with different bandwidths[1]. We denote the base station by $B$ and the basic channel by 0, and consequently let $\mathcal{N} = \underline{\mathcal{N}} \cup \{B\}$ and $\mathcal{M} = \underline{\mathcal{M}} \cup \{0\}$. The bandwidth of channel $b$ is denoted by $W^b$. Moreover, we denote the set of radios at node $i \in \mathcal{N}$ by $\mathcal{R}_i = \{1, 2, ..., R_i\}$ where $R_i$ is the number of radios that node $i$ has. Suppose there are a set of $\mathcal{L} = \{1, 2, ..., l, ...L\}$ downlink sessions from the base station to network users. We let $s(l)$ and $d(l)$ denote the source node and the destination node of session $l \in \mathcal{L}$, respectively. Thus, $s(l) = B$ and $d(l_1) \neq d(l_2)$ for any $l, l_1, l_2 \in \mathcal{L}$. We also denote by $r(l)$ the throughput demand of session $l$. Besides, due to their different geographical locations, users in the network may have different available channels. Let $\mathcal{M}_i \subseteq \mathcal{M}$ represent the set of available channels at node $i \in \mathcal{N}$. Then $\mathcal{M}_i$ might be different from $\mathcal{M}_j$, where $j$ is not equal to $i$, i.e., possibly $\mathcal{M}_i \neq \mathcal{M}_j$. Note that the local available channels can be determined by spectrum sensing, which can be performed in several different ways, such as centralized sensing, distributed sensing, and external sensing [16], [17]. There has been a lot of work in the literature studying this problem and is out of the scope of this paper.

Some important notations are summarized in Table I.

TABLE I
IMPORTANT NOTATIONS

| Symbol | Definition |
|---|---|
| $\mathcal{N}$ | Set of users and the BS in the network |
| $\mathcal{M}$ | Set of the secondary channels and the basic channel |
| $\mathcal{L}$ | Set of downlink sessions |
| $A_1/A_2/A_3$ | Infrastructure/Proxy/Ad hoc region |
| $\mathcal{K}/\overline{\mathcal{K}}$ | Set of all maximal independent Sets/Independent Sets |
| $w_q$ | Time share of IS $\mathcal{I}_q$ being active |
| $c^b_{ij,mn}(\mathcal{I}_q)$ | Data rate on the LRC tuple $((i,j),(m,n),b)$ |
| $f_{ij}(l)$ | Flow rate of session $l$ over link $(i,j)$ |
| $\epsilon$ | Approximation factor |
| $\Psi$ | Result of OOP |
| $\psi_g/\overline{\psi_g}$ | Result of MP/RMP when the proxy region is $S_g$ |
| $\psi^*_g/\psi^l_g/\psi^u_g$ | Optimal/Low bound/Upper bound result of MP when the proxy region is $S_g$ |
| $v_g$ | Result of PP when the proxy region is $S_g$ |

### C. Transmission/Interference Range and Link Capacity

Suppose the power spectral density of node $i$ on channel $b$ is $P^b_i$. A widely used model [18], [19] for power propagation gain between node $i$ and node $j$, denoted by $g_{ij}$, is

$$g_{ij} = C \cdot [d(i,j)]^{-\gamma},$$

where $i$ and $j$ also denote the positions of node $i$ and node $j$, respectively, $d(i,j)$ refers to the Euclidean distance between $i$ and $j$, $\gamma$ is the path loss factor, and $C$ is a constant related to the antenna profiles of the transmitter and the receiver, wavelength, and so on. We assume that the data transmission is successful only if the received power spectral density at the receiver exceeds a threshold $P^b_T$. Meanwhile, we assume interference becomes non-negligible only if it produces a power spectral density over a threshold of $P^b_I$ at the receiver[2]. Thus, the transmission range for a node $i$ on channel $b$ is $R^{i,b}_T = (CP^b_i/P^b_T)^{1/\gamma}$, which comes from $C(R^{i,b}_T)^{-\gamma} \cdot P^b_i = P^b_T$. Similarly, based on the interference threshold $P^b_I (P^b_I < P^b_T)$, the interference range for a node is $R^{i,b}_I = (CP^b_i/P^b_I)^{1/\gamma}$, which is larger than $R^{i,b}_T$. Thus, different nodes may have different transmission ranges/interference ranges on different channels with different transmission power.

In addition, according to the Shannon-Hartley theorem, if node $i$ sends data to node $j$ on link $(i,j)$ using channel $b$, the capacity of link $(i,j)$ on channel $b$ is[3]

$$c^b_{ij} = W^b \log_2 \left(1 + \frac{g_{ij}P^b_i}{\eta}\right), \tag{1}$$

where $\eta$ is the thermal noise at the receiver. Note that the denominator inside the log function only contains $\eta$. This is because of one of our interference constraints, i.e., when node $i$ is transmitting to node $j$ on channel $b$, all the other neighbors of node $j$ within its interference range are prohibited from using this channel. We will address the interference constraints in detail in the following section.

## III. Minimum Length Scheduling for M$^3$C$^2$Ns

In this section, we investigate the minimum length scheduling problem for M$^3$C$^2$Ns by joint frequency allocation, link scheduling, and routing. Traditional cellular networks employ one-hop transmissions to support the traffic between base stations and network users, which we call the "*infrastructure mode*" communications. This design results in very poor throughput performance due to limited frequency channel bandwidth. In this study, we propose a "*hybrid mode*" communication paradigm to enhance the performance of M$^3$C$^2$Ns by taking advantage of local available channels and link rate adaptivity.

### A. Hybrid Mode Communications

In hybrid mode communications, we only let a fraction of nodes close to a base station communicate with the base

[1]Note that in this study we only consider the minimum length scheduling in one cell to focus on the optimization problem and make it easier to understand. The interference from other cells can be addressed by frequency planning and our interference model that will be introduced later. Many related works on cross-layer optimization for cognitive cellular networks also focus on one cell only [11]–[15]. Besides, the presented study here can be easily extended to multi-cell scenarios with minor changes.

[2]Note that the interference model we adopt in this study is the Protocol Model introduced in [20], which considers one interfering link at a time. [20] also introduces the Physical Model, according to which a transmission is successful if its signal-to-interference plus noise ratio (SINR) is above a threshold. It has been shown in [20] that these two interference models can be equivalent in terms of network capacity by setting the interference range in Protocol Model appropriately. [21] also studies how to set the optimal interference range in Protocol Model to bridge the gap between it and Physical Model in analyzing throughput of multi-hop wireless networks. Protocol Model has been widely adopted in works on cross-layer wireless network optimization and design [3], [22], [23].

[3]Note that this link capacity is the same no matter which radios the transmitter and the receiver use.

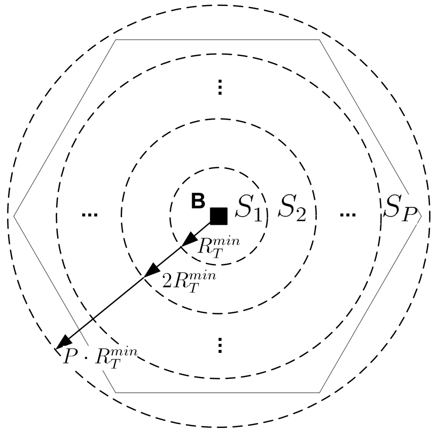


Fig. 2. Hybrid mode communications.

station directly in one hop. The other nodes farther away need communicate via multiple hops, i.e., in ad hoc mode, with some of the above nodes in order to communicate with the base station.

We further illustrate the hybrid mode communication paradigm in Fig. 2. Consider a regular cell $\mathcal{C}$, at the center of which there is a base station denoted by $B$. We denote the minimum transmission range of all the users on all the channels by $R_T^{min}$, i.e., $R_T^{min} = \min_{i\in\mathcal{N},b\in\mathcal{M}}\{R_T^{i,b}\}$, and define an area $S_p$ $(1 \le p \le P)$ as follows:

$$S_p = \{\rho|(p-1)\cdot R_T^{min} \le d(\rho, B) < p \cdot R_T^{min}\}$$

where $\rho$ denotes a point in the network and its position as well, and $B$ denotes the position of the base station. Letting the largest distance between the BS and a point in the cell be $D$, we can have $P = \lceil D/R_T^{min} \rceil$. Then, we choose one of the above areas, say $S_g$ $(1 \le g \le P)$, as *the proxy region*. In particular, define by $\delta(p)$ $(1 \le p \le P)$ a binary function, which is equal to 1 when area $S_p$ is selected as the proxy region and 0 otherwise. Consequently, we have

$$\sum_{p=1}^{P} \delta(p) = 1, \text{ and } g = \delta^{-1}(1).$$

Let $A_2 = S_g$. Then, the nodes in $A_1 = \bigcup_{p=1}^{g} S_p$, which we call *the infrastructure region*, communicate with the base station directly in one hop. The nodes in $A_3 = \mathcal{C}\backslash A_1$, which we call *the ad hoc region*, communicate via multiple hops with the nodes in $A_2$ in order to communicate with the base station.

Note that the hybrid mode communication paradigm changes into the traditional infrastructure mode when $g = P$, and becomes the pure ad hoc mode when $g = 1$. Although ad hoc mode transmissions lead to higher frequency reuse, more ad hoc mode transmissions in the network, i.e., lower $g$, do not necessarily lead to higher performance. For instance, when $g = 1$, all the traffic burden will be put on the nodes in $S_1$, which may not be the optimal strategy resulting in the best performance when all the network users have the same available channels, not to mention the fact that the users in $S_1$ may even have fewer available channels with lower bandwidths. Therefore, we need to find out an optimal proxy region in the network.

In addition, although only the nodes in the infrastructure region communicate with the base station in one hop, the base station still needs to maintain the transmission power level to cover the whole cell in order to guarantee voice services, and controlling and signalling. So we assume the base station's transmission range is the same on all the channels, i.e., to cover the whole cell.

### *B. Construction of Conflict Graph and Independent Sets*

Taking into account the local channel availability and the existence of a powerful base station, we construct a two-layer conflict graph as follows to characterize the interference among the communication links in a cell.

In particular, we denote the conflict graph by $G(V, E)$, where $V$ is the vertex set and $E$ is the edge set. Each vertex corresponds to a link-radio-channel (LRC) tuple defined as $((i,j),(m,n),b)$, where $i, j \in \mathcal{N}$, $j \in \mathcal{T}_i^b$, $m \in \mathcal{R}_m$, $n \in \mathcal{R}_n$, and $b \in \mathcal{M}_i \bigcap \mathcal{M}_j$. Here, $\mathcal{T}_i^b$ is the set of nodes within node $i$'s transmission range on channel $b$. The LRC tuple indicates that node $i$ transmits to node $j$ on channel $b$ with $i$ and $j$ using radio $m$ and radio $n$, respectively. The conflict graph has two layers: the higher layer contains all the vertices involving the base station, while the lower layer consists of all the other vertices, i.e., the LRC tuples with respect to the users only. Besides, we say that two LRC tuples interfere with each other if 1) the receiving node in one tuple is within the interference range of the transmitting node in the other tuple given that the two tuples are using the same channel, or 2) the two tuples use the same radio at one or two nodes. We connect two vertices in $V$ with an undirected edge if the corresponding LRC tuples interfere with each other.

In the two-layer conflict graph $G(V, E)$, we define a variable $w_{xy}$, where $x, y \in V$, as follows:

$$w_{xy} = \begin{cases} 1, & \text{if there is an edge between vertex } x \text{ and } y \\ 0, & \text{otherwise} \end{cases}.$$

Thus, if there is a vertex (i.e., LRC tuple) set $\mathcal{I} \subseteq V$ and a vertex $x \in \mathcal{I}$ satisfying $\sum_{y\in\mathcal{I},x\ne y} w_{xy} < 1$, the transmission on the LRC tuple $x$ can be carried out successfully even if all the other LRC tuples belonging to the set $\mathcal{I}$ transmit at the same time. If every $x \in \mathcal{I}$ satisfies the above condition, we can schedule the transmissions over all these LRC tuples in $\mathcal{I}$ to be active simultaneously. Such a vertex set $\mathcal{I}$ is called *an independent set*. If adding any more LRC tuples into an independent set $\mathcal{I}$ results in a non-independent one, $\mathcal{I}$ is defined as *a maximal independent set*.

### *C. Link Scheduling and Routing Constraints*

*1) Link Scheduling Constraints:* Given the constructed conflict graph $G = (V, E)$, suppose we can list all the maximal independent sets[4] as $\mathcal{K} = \{\mathcal{I}_1, \mathcal{I}_2, ..., \mathcal{I}_Q\}$, where $Q = |\mathcal{K}|$, and $\mathcal{I}_q \subseteq V$ for $1 \le q \le Q$. Then, in the conflict graph $G$,

[4]We will show in the next section that we do not really need to find all the maximal independent sets.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMC.2014.2314107



at any time instance, there should be only one active maximal independent set to ensure the success of all the transmissions. We denote the maximal independent set $\mathcal{I}_q$'s time share (out of unit time 1) to be active by $w_q$. Therefore, we have

$$\sum_{1\leq q\leq Q} w_q \leq 1, \quad w_q \geq 0 \ (1 \leq q \leq Q).$$

Let $c^b_{ij,mn}(\mathcal{I}_q)$ be the data rate on the LRC tuple $((i,j),(m,n),b)$ when the maximal independent set $\mathcal{I}_q$ is active. Then, $c^b_{ij,mn}(\mathcal{I}_q)$ is equal to 0 if the LRC tuple $((i,j),(m,n),b)$ is not in $\mathcal{I}_q$, and equal to the link capacity calculated according to (1) otherwise. Thus, letting $f_{ij}(l)$ denote the flow rate of session $l$ over link $(i,j)$ (over all the channels), where $i \in \mathcal{N}$, $l \in \mathcal{L}$, and $j \in \bigcup_{b\in\mathcal{M}_i} \mathcal{T}^b_i$, the traffic rate on link $(i,j)$, i.e., $\sum_{l\in\mathcal{L}} f_{ij}(l)$, should not exceed the capacity of the link. Consequently, the schedule of the maximal independent sets should satisfy the following constraint:

$$\sum_{l\in\mathcal{L}} f_{ij}(l) \leq \sum_{q=1}^{Q} w_q \sum_{b\in\mathcal{M}_i\cap\mathcal{M}_j} \sum_{m\in\mathcal{R}_i} \sum_{n\in\mathcal{R}_j} c^b_{ij,mn}(\mathcal{I}_q). \quad (2)$$

*2) Routing Constraints:* Recall that we consider downlinks in this study, which means that the base station is the source node for all the flows. In our hybrid mode communication paradigm presented in Section III-A, we have defined the infrastructure region $A_1$, the proxy region $A_2$, and the ad hoc region $A_3$. The destinations in the infrastructure region receive packets from the base station in one hop and hence on a single path, while packets intended for the destinations in the ad hoc region reach the proxy region first, which may go through multiple paths.

Thus, at the base station, we have the following constraints.

$$\begin{aligned} \sum_{j\in\underline{\mathcal{N}}} f_{jB}(l) &= 0, & (3)\\ f_{Bd(l)}(l) &= r(l) \quad \forall d(l) \in A_1, & (4)\\ \sum_{j\in A_2} f_{Bj}(l) &= r(l) \quad \forall d(l) \in A_3. & (5) \end{aligned}$$

The first constraint means that the incoming data rate at the base station is 0. The second constraint indicates that the traffic intended for any destination in the infrastructure region is delivered in one hop on a single path. The third constraint means that the traffic for any destination in the ad hoc region goes through the proxy region and may be delivered on multiple paths.

Remember that we define by $d(l)$ $(l \in \mathcal{L})$ the destination node of session $l$. We then define by $d^{-1}(j)$ the session whose destination is $j$. We further let $\mathcal{D}(\mathcal{L})$ be the set of all the destination nodes in the network, i.e., $\mathcal{D}(\mathcal{L}) = \{d(l)|l \in \mathcal{L}\}$, and $\mathcal{T}_i = \bigcup_{b\in\mathcal{M}_i} \mathcal{T}^b_i$. Then, we can have the constraints below.

- For any $j \in A_1$ and $j \in \mathcal{D}(\mathcal{L})$, we have

$$f_{Bj}(l) = r(l), \quad (6)$$

  i.e., the destination nodes in the infrastructure region $A_1$ receive their packets from the base station directly in one hop on a single path. Note that this constraint holds whenever the constraint (4) holds.
- For any $j \in A_2 \bigcup A_3$, $d(l) \in A_3$, and $j \neq d(l)$, we have

$$\sum_{\{p|j\in\mathcal{T}_p\}} f_{pj}(l) = \sum_{i\in\mathcal{T}_j} f_{ji}(l), \quad (7)$$

  which indicates that each node in the proxy region $A_2$ and in the ad hoc region $A_3$ can act as a relay node for the destination nodes in the ad hoc region, and hence its total incoming data rate is equal to its total outgoing data rate.
- For any $j \in A_3$ and $j \in \mathcal{D}(\mathcal{L})$, we have $\sum_{\{i|j\in\mathcal{T}_i\}} f_{ij}(l) = r(l)$, which means that the total incoming data rate to each destination node in the ad hoc region is equal to its throughput demand. Note that this constraint holds whenever the constraints (5) and (7) hold.

### *D. Scheduling Length Optimization for M$^3$C$^2$Ns*

The objective of this study is to exploit both the cellular and the local available channels to minimize the scheduling length, i.e., $\Psi = \sum_{1\leq q\leq Q} w_q$, required to support certain traffic demands in M$^3$C$^2$Ns. Gathering information about channel availability in the network, the service provider can achieve this goal by optimally selecting proxy region, determining end-to-end paths, and scheduling the transmissions. Note that a minimum value of $\Psi$ greater than 1 indicates that the current traffic demands exceed the system capacity and cannot be satisfied.

Considering the hybrid communication paradigm, the scheduling length optimization problem under the aforementioned link scheduling and routing constraints can be formulated as follows:

**OOP:** **Minimize** $\Psi = \sum_{1\leq q\leq Q} w_q$

**s.t.** Equations (2) and (3)

$$\sum_{p=1}^{P} \delta(p) = 1, \delta(p) \in \{0,1\} (1 \leq p \leq P), g = \delta^{-1}(1) \quad (8)$$

$$A_1(g) = \bigcup_{p=1}^{g} S_p, A_2(g) = S_g, A_3(g) = \mathcal{C}\backslash A_1(g) \quad (9)$$

$$f_{Bd(l)}(l)\mathbf{1}\{d(l) \in A_1(g)\} = r(l) \quad (l \in \mathcal{L}) \quad (10)$$

$$\sum_{j\in A_2} f_{Bj}(l)\mathbf{1}\{d(l) \in A_3(g)\} = r(l) \quad (l \in \mathcal{L}) \quad (11)$$

$$\sum_{\{p|j\in\mathcal{T}_p\}} f_{pj}(l)\mathbf{1}\{d(l) \in A_3(g), j \in A_2(g) \cup A_3(g), j \neq d(l)\} =$$
$$\sum_{i\in\mathcal{T}_j} f_{ji}(l)\mathbf{1}\{d(l) \in A_3, j \in A_2(g) \cup A_3(g), j \neq d(l)\} (l \in \mathcal{L}) \quad (12)$$

$$w_q \geq 0 \ (1 \leq q \leq Q) \quad (13)$$

$$f_{ij}(l) \geq 0 \ (l \in \mathcal{L}, i \in \mathcal{N}, j \in \mathcal{T}_i) \quad (14)$$

where (2) indicate that the flow rate over link $(i,j)$ cannot exceed the link capacity, (8)–(9) characterize the hybrid communication paradigm and define the infrastructure region, the proxy region, and the ad hoc region, and (3), (10)–(12) are the routing constraints. Here, we define an indicator function $\mathbf{1}\{A\}$ which is equal to 1 if the event $A$ is true, and 0 otherwise[5].

Given all the maximal independent sets in the network, we can find that the formulated optimization problem above is a Mixed Integer Non-Linear Programming problem, which is in general NP-hard to solve [24], [25]. In the rest of this paper, we call this optimization problem the original optimization problem (OOP), and denote the minimum scheduling length by $\Psi^*$.

Note that the solution to the OOP consists of the following four parts: $\delta(p)$'s, $f_{ij}(l)$'s, $s^b_{ij,mn}$'s and $w_q$'s, which determine the proxy region, routing, frequency-domain scheduling, and time-domain scheduling, respectively. Particularly, $w_q$'s represent the time share (out of unit time 1) for the maximal independent set $I_q$ to be active. The service provider can preassign an index number to each of the maximal independent sets, and schedule all the independent sets (i.e., the link-radio-channel tuples) to be active following a certain order in each time slot, e.g., from the lowest index number to the highest. In so doing, a detailed schedule can be obtained and each LRC tuple knows when and how long it needs to be active in each time slot. Besides, if the traffic demand changes or some users join/leave the network, the OOP problem will be computed again to find a new solution. Otherwise, the same solution will be adopted.

## IV. A Column Generation Based Efficient $\epsilon$-Bounded Approximation Algorithm

In this section, we propose a column generation based $\epsilon$-bounded approximation algorithm, which can efficiently find the $\epsilon$-bounded approximate results and the optimal result as well, i.e., when $\epsilon = 0$ in the algorithm, without finding all the maximal independent sets. The definition of $\epsilon$-bounded approximate solution will be given later.

### A. Decomposition of the Original Optimization Problem

The OOP is a mixed integer non-linear programming problem because of $\delta(p)$'s in constraint (8) and the non-linear constraints (10)–(12). Notice that when the proxy region $A_2$, i.e., $\delta(p)$'s, is fixed, the infrastructure region $A_1$ and the ad hoc region $A_3$ will both be determined, which can turn the OOP into a linear programming problem. Therefore, we can solve the OOP as follows: we first solve the $P$ scheduling length optimization problems separately considering that one of the $P$ subareas (as shown in Fig. 2) is selected as the proxy region, and then compare the $P$ results and find the minimum scheduling length among them as the global optimization result for the OOP. Notice that when the user density is sparse, it is possible that there are no users in one or some (but obviously not all) of the $P$ subareas. Thus, when an empty subarea is selected as the proxy region, the traffic for destinations in the ad hoc region cannot be supported, and hence we set the minimum scheduling length in that case to be infinity.

[5]Note that each maximal independent set's time share to be active is a real number. OOP is formulated based on the maximal independent sets as in [3]–[5].

Specifically, we decompose the OOP into $P$ linear programming problems, each of which we call a Master Problem (MP). Notice that the optimal result of OOP remains the same when we consider all the independent sets $\overline{\mathcal{K}}$ which include all the maximal independent sets $\mathcal{K}$. Thus, when the proxy region is $S_g$ $(1 \le g \le P)$, the MP is formulated as follows:

$$\textbf{MP:} \qquad \textbf{Minimize} \quad \psi_g = \sum_{1 \le q \le |\overline{\mathcal{K}}|} w_q$$

$$\textbf{s.t.} \quad \text{Equations } (2), (3), (13), \text{ and } (14)$$

$$f_{Bd(l)}(l) = r(l) \quad (l \in \mathcal{L}, d(l) \in A_1(g)) \tag{15}$$

$$\sum_{j \in A_2(g)} f_{Bj}(l) = r(l) \quad (l \in \mathcal{L}, d(l) \in A_3(g)) \tag{16}$$

$$\sum_{\{p|j \in \mathcal{T}_p\}} f_{pj}(l) = \sum_{i \in \mathcal{T}_j} f_{ji}(l)$$
$$(l \in \mathcal{L}, j \in A_2(g) \cup A_3(g), d(l) \in A_3(g), j \neq d(l)) \tag{17}$$

However, after the decomposition, there are still two difficulties in solving this linear programming problem. First, each MP is a linear programming problem if we can find all the independent sets, which is nonetheless an NP-complete problem itself [26], [27]. Second, even if we can find all the independent sets, the number of such sets increases exponentially as the number of LRC tuples and hence can be huge. In the following, we propose a column generation based approach to circumvent these difficulties and efficiently solve each MP.

### B. Column Generation

Column generation (CG) is an iterative approach for solving huge linear or nonlinear programming problems, in which the number of variables (columns) is too large to be considered completely [6]. Generally, only a small subset of these variables are positive values in an optimization solution, while the rest of the variables (called *nonbasis*) are zeros. Therefore, CG leverages this idea by generating only those critical variables that have the potential to improve the objective function. In our case, each MP is further decomposed into a Restricted Master Problem (RMP) and a Pricing Problem (PP). The strategy of this further decomposition procedure is to operate iteratively on two separate, but easier, problems. During each iteration, PP tries to determine whether any columns (i.e., independent sets) uninvolved in RMP exist that have a negative reduced cost[6], and adds the column with the most negative reduced cost to the corresponding RMP, until the algorithm terminates at, or satisfyingly close to, the optimal solution.

Notice that the formulation of MP considers the entire set of independent sets $\overline{\mathcal{K}}$, while RMP only starts with a set of initial feasible independent sets, say $\mathcal{K}'$, which can be easily formed by placing just one LRC tuple in each of them. Thus,

[6]Reduced cost [6] refers to the amount by which the objective function would have to improve before the corresponding column is assumed to be part of optimal solution. In the case of a minimization problem like in this paper, improvement in the objective function means a decrease of its value, i.e., a negative reduced cost. In finding the column with the most negative reduced cost, the objective is to find the column that has the best chance to improve the objective function.

an RMP can be formulated as follows:

$$\textbf{RMP:} \qquad \textbf{Minimize} \quad \overline{\psi_g} = \sum_{1 \le q \le |\mathcal{K}'|} w_q$$

$$\textbf{s.t.} \quad \text{Equations } (3), (13), (14), \text{ and } (15) - (17)$$

$$\sum_{l \in \mathcal{L}} f_{ij}(l) \le \sum_{q=1}^{|\mathcal{K}'|} w_q \sum_{b \in \mathcal{M}_i \cap \mathcal{M}_j} \sum_{m \in \mathcal{R}_i} \sum_{n \in \mathcal{R}_j} c_{ij,mn}^b(\mathcal{I}_q)$$
$$(i \in \mathcal{N}, j \in \mathcal{T}_i, \text{ and } \mathcal{I}_q \in \mathcal{K}') \tag{18}$$

RMP is a small-scale linear programming problem that can be easily solved in polynomial time by the simplex algorithm [28]. We can thus obtain its primal optimal solution and a Lagrangian dual optimal solution. Since RMP uses only a subset of all the independent sets (i.e., columns) used by MP, i.e., $\mathcal{K}' \subseteq \overline{\mathcal{K}}$, the optimal result for RMP serves as an upper bound on the optimal result for MP. By introducing more independent sets to the RMP, column generation may be able to decrease the upper bound. Therefore, we need to determine which column can potentially improve the optimization result the most and when the optimal result of RMP is exactly the same as or satisfyingly close to the optimal result of MP.

### C. Introducing More Columns to RMP

During every iteration, when RMP is solved, we need to verify whether any new independent set can improve the current solution. In particular, for each independent set $\mathcal{I}_q \in \overline{\mathcal{K}} \setminus \mathcal{K}'$, we need to examine if any of them has a negative reduced cost. The reduced cost $u_g(\mathcal{I}_q)$ for a column $\mathcal{I}_q \in \overline{\mathcal{K}} \setminus \mathcal{K}'$ can be calculated as [28]:

$$u_g(\mathcal{I}_q) = 1 - \sum_{i \in \mathcal{N}, j \in \mathcal{T}_i} \lambda_{ij} \sum_{b \in \mathcal{M}_i \cap \mathcal{M}_j} \sum_{m \in \mathcal{R}_i} \sum_{n \in \mathcal{R}_j} c_{ij,mn}^b(\mathcal{I}_q), \tag{19}$$

where $\lambda_{ij}$'s are the Lagrangian dual optimal solution corresponding to (18). Since there are totally $|\mathcal{N}| \times (|\mathcal{N}| - 1)$ constraints generated from (18), the total number of $\lambda_{ij}$'s is also $|\mathcal{N}| \times (|\mathcal{N}| - 1)$.

Notice that we need to find the column which can produce the most negative reduced cost. Consequently, this column to be added to RMP can be obtained by solving

$$\underset{\mathcal{I}_q \in \overline{\mathcal{K}}/\mathcal{K}'}{\textbf{Minimize}} \qquad u_g = u_g(\mathcal{I}_q), \tag{20}$$

or equivalently

$$\underset{\mathcal{I}_q \in \overline{\mathcal{K}} \setminus \mathcal{K}'}{\textbf{Maximize}} \; v_g = \sum_{i \in \mathcal{N}, j \in \mathcal{T}_i} \lambda_{ij} \sum_{b \in \mathcal{M}_i \cap \mathcal{M}_j} \sum_{m \in \mathcal{R}_i} \sum_{n \in \mathcal{R}_j} c_{ij,mn}^b(\mathcal{I}_q), \tag{21}$$

which is called a Pricing Problem (PP). Denote by $u_g^*$ and $v_g^*$ the optimal solutions to the above two problems, respectively. Then, if $u_g^* \ge 0$ or $v_g^* \le 1$, it means that there is no negative reduced cost and hence the current solution to RMP optimally solves MP as well. Otherwise, we add to RMP the column derived from (21), and repeat re-optimizing RMP. We leave how to solve PP in Section IV-D.

### D. Solving PP

Next, we study how to solve PP, i.e., the optimization problem formulated in (21). Our objective is to find out the independent set, i.e., all the LRC tuples that can be active at the same time, which can maximize $v$.

We define a variable $s_{ij,mn}^b$ as follows: $s_{ij,mn}^b$ is equal to 1 if node $i$, using radio $m$, transmits to node $j$, using radio $n$, on channel $b$, and equal to 0 otherwise. Then, the independent set we need to find out is $\{((i,j),(m,n),b) | s_{ij,mn}^b = 1\}$ that can maximize $v_g$ in (21).

Recall that we let $\mathcal{T}_i^b$ denote the set of nodes that can access channel $b$ and are within the transmission range of node $i$, and $\mathcal{R}_i$ the set of radios at node $i$. We can prove that a node cannot transmit to or receive from multiple nodes on the same channel due to interference, even if it has multiple radios with different transmission power. Thus, we have

$$\sum_{j \in \mathcal{T}_i^b} \sum_{m \in \mathcal{R}_i} \sum_{n \in \mathcal{R}_j} s_{ij,mn}^b \le 1, \; \sum_{\{i | j \in \mathcal{T}_i^b\}} \sum_{m \in \mathcal{R}_i} \sum_{n \in \mathcal{R}_j} s_{ij,mn}^b \le 1. \tag{22}$$

Besides, a node $j$ cannot use the same channel or radio for transmission and reception at the same time. Therefore, we get

$$\sum_{\{i | j \in \mathcal{T}_i^b\}} \sum_{m \in \mathcal{R}_i} \sum_{n \in \mathcal{R}_j} s_{ij,mn}^b + \sum_{q \in \mathcal{T}_j^b} \sum_{y \in \mathcal{R}_j} \sum_{z \in \mathcal{R}_q} s_{jq,yz}^b \le 1. \tag{23}$$

$$\sum_{b \in \mathcal{M}_j} \sum_{\{i | j \in \mathcal{T}_i^b\}} \sum_{m \in \mathcal{R}_i} s_{ij,mn}^b + \sum_{p \in \mathcal{M}_j} \sum_{q \in \mathcal{T}_j^b} \sum_{z \in \mathcal{R}_q} s_{jq,nz}^p \le 1. \tag{24}$$

Moreover, the total number of communication links, transmitting or receiving, at any node $j$ should be no larger than the number of radios node $j$ has, which means

$$\sum_{b \in \mathcal{M}_j} \sum_{\{i | j \in \mathcal{T}_i^b\}} \sum_{m \in \mathcal{R}_i} \sum_{n \in \mathcal{R}_j} s_{ij,mn}^b + \sum_{b \in \mathcal{M}_j} \sum_{q \in \mathcal{T}_j^b} \sum_{y \in \mathcal{R}_j} \sum_{z \in \mathcal{R}_q} s_{jq,yz}^b$$
$$\le |\mathcal{R}_j| = R_j. \tag{25}$$

In addition to the above constraints at the same node, there are also scheduling constraints due to potential interference among the nodes in the network. In particular, if node $i$ uses channel $b$ to transmit data to node $j \in \mathcal{T}_i^b$, then any other node that may interfere with the reception at node $j$ should not use this channel. To model this constraint, we let $\underline{\mathcal{I}}_j^b$ represent the set of nodes that can produce interference at node $j$ on channel $b$, i.e., $\underline{\mathcal{I}}_j^b = \{p \mid d_{pj} \le R_I^{p,b}, p \ne j, T_p^b \ne \emptyset\}$. The interpretation of $\mathcal{T}_p^b \ne \emptyset$ in the above definition is that node $p$ may interference with the reception at node $j$ only if there are some nodes within $p$'s transmission range on channel $b$ which $p$ can transmit to. Based on the definition of $\underline{\mathcal{I}}_j^b$, we have

$$\sum_{\{i | j \in \mathcal{T}_i^b\}} \sum_{m \in \mathcal{R}_i} \sum_{n \in \mathcal{R}_j} s_{ij,mn}^b + \sum_{q \in \mathcal{T}_p^b} \sum_{y \in \mathcal{R}_p} \sum_{z \in \mathcal{R}_q} s_{pq,yz}^b \le 1. \tag{26}$$

Consequently, considering the above constraints, the pricing problem (PP) (21) of finding the optimal column can be formulated as follows:

$$\textbf{PP: Maximize} \sum_{i\in\mathcal{N},j\in\mathcal{T}_i}\lambda_{ij}\sum_{b\in\mathcal{M}_i\cap\mathcal{M}_j}\sum_{m\in\mathcal{R}_i}\sum_{n\in\mathcal{R}_j}c^b_{ij,mn}s^b_{ij,mn}$$

$$\textbf{s.t.} \quad \text{Equations } (22)-(26)$$

$$s^b_{ij,mn}=0 \quad \text{or} \quad 1$$

where $s^b_{ij,mn}$ are the optimization variables. Recall that $\lambda_{ij}$'s are the Lagrangian dual optimal solutions to RMP, and $c^b_{ij,mn}$'s are calculated according to (1). Since $s^b_{ij,mn}$ can only take value of 0 or 1, PP is a Binary Integer Programming (BIP) problem and thus NP-complete [25]. Instead of using the traditional *branch-and-bound* or *branch-and-cut* [24] approach, we follow a similar idea to that in [5], [29] to develop a greedy algorithm to find a suboptimal solution to PP, which is called the sequential-fix (SF) algorithm.

The main idea of SF is to fix the values of $s^b_{ij,mn}$'s sequentially through a series of relaxed linear programming problems. Specifically, in each iteration, we first relax all the 0-1 integer constraints on $s^b_{ij,mn}$'s to $0 \le s^b_{ij,mn} \le 1$ to transform the problem to a linear programming (LP) problem. Then, we solve this LP to obtain an optimal solution with each $s^b_{ij,mn}$ being between 0 and 1. Among all the values, we set the largest $s^b_{ij,mn}$ to 1. After that, by (25), among all the $s^c_{pj,hk}$ and $s^d_{jq,yz}$ for any $c,d\in\mathcal{M}_j$, $\{p|j\in\mathcal{T}^c_p, p\neq i\}$, $q\in\mathcal{T}^d_j$, $h\in\mathcal{R}_p$, $k,y\in\mathcal{R}_j$, $z\in\mathcal{R}_q$, we randomly choose $R_j-1$ of them and set them to 1, while having the rest set to 0. Then, by (26), we can fix $s^b_{pj,hk}=0$ and $s^b_{tq,yz}=0$ for any $\{p|j\in\mathcal{T}^b_p, p\neq i\}$, $t\in\mathcal{P}^b_j$, $q\in\mathcal{T}^b_t$, $h\in\mathcal{R}_p$, $k\in\mathcal{R}_j$, $y\in\mathcal{R}_t$, $z\in\mathcal{R}_q$.

Having fixed some $s^b_{ij,mn}$'s in the first iteration, we remove all the terms associated with those already fixed $s^b_{ij,mn}$'s, eliminate the related constraints in (25) and (26), and update the problem to a new one for the second iteration. Similarly, in the second iteration, we solve an LP with reduced number of variables, and then determine the values of some other unfixed $s^b_{ij,mn}$'s based on the same process. The iteration continues until we fix all $s^b_{ij,mn}$'s to be either 0 and 1.

Recall that when the optimal result of PP is less than 1, i.e., $v^*_g \le 1$, it means that there is no negative reduced cost and RMP can be optimally solved. Unfortunately, the SF algorithm developed above does not find the optimal solution to PP. Nevertheless, when the optimal result of the relaxed PP (formulated by relaxing binary variables in PP to variables between 0 and 1), denoted by $\overline{v^*_g}$, is less than 1, we have $v^*_g \le \overline{v^*_g} \le 1$ and hence RMP can still be optimally solved.

### *E. $\epsilon$-Bounded Approximate Solutions*

Since the number of independent sets in $\overline{\mathcal{K}}$ increases exponentially as the number of links in the network, the number of iterations (of PP) needed to find all the independent sets producing negative reduced cost might be very large, especially in large-size networks. However, it has been observed in the context of column generation algorithms [6], [7] that one can usually determine solutions that are at least $95\%-99\%$ of the global optimality fairly quickly, although the tail-end convergence rate in obtaining the optimal solution can be slow in some classes of problems. Here, we propose an $\epsilon$-bounded approximation algorithm to find $\epsilon$-bounded approximate solutions more efficiently.

We first give the definition of $\epsilon$-bounded solutions as follows.

***Definition** 1:* Let $0\le\epsilon<1$ be a predefined parameter and $\psi^*_g$ be the optimal result of the MP when the proxy region is $S_g$ $(1\le g\le P)$. Then, a solution is called an $\epsilon$-bounded approximate solution if its corresponding result $\psi_g$ satisfies

$$(1-\epsilon)\psi^*_g \le \psi_g \le (1+\epsilon)\psi^*_g.$$

Then, we can have the following lemma.

***Lemma** 1:* Denote by $\psi^u_g$ and $\psi^l_g$ the upper bound and lower bound on the optimal result $\psi^*_g$ of the MP when the proxy region is $S_g$ $(1\le g\le P)$. Then, $\epsilon$-bounded approximate solutions $(0\le\epsilon<1)$ can be obtained when the iteration stops at $\overline{v^*_g}\le 1$ or

$$\frac{\psi^l_g}{\psi^u_g} \ge \frac{1}{1+\epsilon}. \tag{27}$$

*Proof:* When $\frac{\psi^l_g}{\psi^u_g} \ge \frac{1}{1+\epsilon}$, we can get that $\psi^u_g \le (1+\epsilon)\psi^l_g \le (1+\epsilon)\psi^*_g$ and $\psi^l_g \ge \psi^u_g/(1+\epsilon) \ge (1-\epsilon)\psi^u_g \ge (1-\epsilon)\psi^*_g$. Thus, any obtained result between the upper and lower bounds, i.e., $\psi^l_g \le \psi_g \le \psi^u_g$, satisfies $\psi_g \le \psi^u_g \le (1+\epsilon)\psi^*_g$ and $\psi_g \ge \psi^l_g \ge (1-\epsilon)\psi^*_g$, and hence is an $\epsilon$-bounded approximate solution by definition. Besides, when $\overline{v^*_g}\le 1$, as mentioned before, the obtained solution is the optimal solution and hence an $\epsilon$-bounded approximate solution as well. ∎

Notice that in (27), $\epsilon$ is predetermined, e.g., $3\%$. As mentioned before, the optimal result of RMP in each iteration is an upper bound on the optimal result of MP, i.e., $\psi^u_g$. A lower bound can be obtained by [28]

$$\psi^l_g = \psi^u_g + \mathcal{Y}u^*_g \le \psi^*_g, \tag{28}$$

where $u^*_g$ is obtained by solving (20) optimally, and $\mathcal{Y} \ge \sum_{1\le q\le|\overline{\mathcal{K}}|} w_q$ holds for the optimal solution to RMP [28]. We set $\mathcal{Y}=1$. Then, if a traffic demand can be supported, the optimal solution must satisfy $\sum_{1\le q\le|\overline{\mathcal{K}}|} w_q \le \mathcal{Y} = 1$. Thus, if an optimal solution leads to $\sum_{1\le q\le|\overline{\mathcal{K}}|} w_q > 1$, then the corresponding traffic demand cannot be supported. However, since we actually do not obtain $u^*_g$ with the SF algorithm, the lower bound can be set to $\psi^l_g = \psi^u_g + \mathcal{Y}\underline{u^*_g}$, which is less than $\psi^u_g + \mathcal{Y}u^*_g$ and hence $\psi^*_g$. Here, $\underline{u^*_g}$ is obtained by the optimal result of the relaxed PP, i.e., $\underline{u^*_g} = 1-\overline{v^*_g}$. In addition, since $u^*_g$ is negative, $\psi^l_g$ may be negative as well. Therefore, we finally calculate $\psi^l_g$ by

$$\psi^l_g = \max\{\psi^u_g + \mathcal{Y}\underline{u^*_g}, 0\}. \tag{29}$$

We finally detail an $\epsilon$-bounded approximation algorithm for the scheduling length optimization problem in Algorithm 1. Note that in the algorithm we choose the solution of an RMP, whose result, $\psi^u_g$, serves as an upper bound on the optimal result of the corresponding MP, as the $\epsilon$-bounded approximate solution of the MP since we have found the corresponding scheduling and routing solutions. It is easy to prove that $\min_{1\leq g\leq P}\{\psi_g^u\}$ among all the MPs is an $\epsilon$-bounded approximate solution to the OOP[7].

**Algorithm 1** **$\epsilon$-Bounded Approximation Algorithm**

**Input:** Approximation factor $\epsilon$, traffic demand $\mathbf{r}(\mathbf{l})$, initial independent set $\mathcal{K}'$, $P = \left\lceil \frac{D}{R_T^{min}} \right\rceil$, $\psi_g^u = \infty$, $\psi_g^l = 0$, $u_g = -\infty$ and $\overline{v_g^*} = +\infty$ for any $g \in [1, P]$.

**Output:** $g^*$, $\psi_{g^*}^u$, $\psi_{g^*}^l$, $w_q^*$, $f_{ij}^*(l)$

Decompose OOP into $P$ RMPs by considering $A_2$ is chosen as one of the $P$ subareas in the cell;
**for** $g = 1$ to $P$ **do**
  **while** $\frac{\psi_g^l}{\psi_g^u} < \frac{1}{1+\epsilon}$ and $\overline{v_g^*} > 1$ **do**
    Solve RMP with current region plan $g$ and $\mathcal{K}'$, and obtain its optimal result $\psi_g^u$ and dual optimal solution $\lambda_{ij}$'s;
    **while** $\{s_{ij,mn}^b | i,j \in \mathcal{N}, m \in \mathcal{R}_i, n \in \mathcal{R}_j, b \in \mathcal{M}_i \bigcap \mathcal{M}_j\} \neq \emptyset$ **do**
      Solve PP by relaxing $0 \leq s_{ij,mn}^b \leq 1$;
      Search for the $s_{ij,mn}^b$ with the largest value and set it to be 1;
      Find $s_{pj,hk}^c$ and $s_{jq,yz}^d$ for any $c, d \in \mathcal{M}_j$, $\{p | j \in \mathcal{T}_p^c, p \neq i\}$, $q \in \mathcal{T}_j^d$, $h \in \mathcal{R}_p$, $k, y \in \mathcal{R}_j$, $z \in \mathcal{R}_q$, and set $R_j - 1$ of them to 1 and the others to 0;
      Fix $s_{pj,hk}^b = 0$ and $s_{tq,yz}^b = 0$ for any $\{p | j \in \mathcal{T}_p^b, p \neq i\}$, $t \in \mathcal{P}_j^b$, $q \in \mathcal{T}_t^b$, $h \in \mathcal{R}_p$, $k \in \mathcal{R}_j$, $y \in \mathcal{R}_t$, $z \in \mathcal{R}_q$
      Remove the fixed $s_{ij,mn}^b$'s from $\{s_{ij,mn}^b\}$;
    **end while**
    Update $\mathcal{K}' = \mathcal{K}' \cup \mathcal{I}_q$, where $\mathcal{I}_q = \{((i,j),(m,n),b) | s_{ij,mn}^b = 1\}$;
    Obtain the current optimal result of the relaxed PP $\overline{v_g^*}$;
    $\underline{u_g^*} = 1 - \overline{v_g^*}$;
    $\overline{\psi_g^l} = \psi_g^u + \mathcal{Y}\underline{u_g^*}$;
  **end while**
**end for**
Sort the scheduling time under each proxy region plan in increasing order, $\psi_1^u \leq \cdots \leq \psi_P^u$;
**if** $\psi_1^u \leq 1$ **then**
  Current traffic demand can be supported, and $g^* = \arg \psi_1^u$, $\psi_{g^*}^u = \psi_1^u$, $\psi_{g^*}^l = \psi_1^l$;
**else if** $\psi_1^u > 1 + \epsilon$ or the corresponding lower bound $> 1$ **then**
  Current traffic demand cannot be supported;
**else**
  Set $\epsilon = 0$ to see if the traffic demand can be supported.
**end if**

[7]Denote the optimal result of the OOP by $\Psi^*$. Define $\psi_{g^*}^u = \min_{1\leq g\leq P}\{\psi_g^u\}$. Then we have $\psi_{g^*}^u \leq \psi_g^u \leq \psi_g^*(1+\epsilon)$ for any $g \neq g^*$, where $\psi_g^*$ is the optimal result of the MP corresponding to $\psi_g^u$. Since $\Psi^*$ is equal to either one of the $\psi_g^*$'s or $\psi_{g^*}^*$, we have $\psi_{g^*}^u \leq (1+\epsilon)\Psi^*$. Besides, since $\psi_{g^*}^u \geq \psi_{g^*}^l \geq \psi_{g^*}^*(1-\epsilon)$ and $\psi_{g^*}^* \geq \Psi^*$, we can get $\psi_{g^*}^u \geq (1-\epsilon)\Psi^*$. In fact, we also have $\psi_{g^*}^u \geq \Psi^*$ since $\psi_{g^*}^u \geq \psi_{g^*}^* \geq \Psi^*$.

### F. Computational Complexity Analysis

As we mentioned before, although each MP is an LP problem, solving it directly still requires a high computational complexity since finding all the independent sets is an NP-complete problem. Note that each node needs to have $\Theta(\log n)$ neighbors on average in order to achieve asymptotic connectivity in wireless networks as proved in [30]. Considering a connected network, the number of LRC tuples in it, denoted by $G$, will be $O(|\mathcal{N}|(\log|\mathcal{N}|)|\mathcal{R}|^2|\mathcal{M}|)$ where $|\mathcal{R}|$ denotes the maximum number of radios a node can have. Thus, the number of independent sets is at most $2^G$, i.e., $O(2^{|\mathcal{N}|(\log|\mathcal{N}|)})$. Since usually only a small number of independent sets would be useful in a scheduling problem, the developed column generation based algorithm finds the useful ones one-by-one iteratively. We analyze the computation complexity of our algorithm as follows.

***Theorem** 1:* The computational complexity of our proposed column generation based algorithm for MP is $O(K^4 + |\mathcal{N}|^8)$ when there are $K$ iterations in the algorithm, and $O(2^{4|\mathcal{N}|\log|\mathcal{N}|})$ in the worst case.

*Proof:* In our proposed column generation based algorithm, one RMP and one PP are solved in each iteration. In RMP, the variables include $w_q$'s and $f_{ij}(l)$'s. Note that the initial independent sets are formed by placing one LRC in each of them. Thus, in the $k$th iteration, the numbers of $w_q$'s and $f_{ij}(l)$'s are $G+k$ and $|\mathcal{N}|(\log|\mathcal{N}|)L$, respectively. Since RMP is an LP problem, it can be solved by the polynomial interior algorithm introduced in [31], whose computation complexity is $O(n^3)$ where $n$ is the number of the variables in a problem. Therefore, the computation complexity of RMP in the $k$th iteration is $O((G+k+|\mathcal{N}|(\log|\mathcal{N}|)L)^3)$, and hence $O((G+k+|\mathcal{N}|^2L)^3)$. In PP, we develop an SF algorithm that consists of multiple rounds of computation for relaxed LP problems with a decreasing number of variables, i.e., $s_{ij,mn}^b$'s, in each round. Note that the number of variables is clearly upper bounded by $G$. Thus, the computation complexity in each round is no larger than $O(G^3)$. Besides, notice that in each round, SF fixes one of $s_{ij,mn}^b$'s to 1 and other interfering variables to 0 according to constraints (22)-(26). Particularly, from the first inequality in (22), we can know that if $s_{ij,mn}^b = 1$, then $s_{ik,yz}^b = 0$ ($k \neq j$ or $y \neq m$ or $z \neq n$). Therefore, all the variables $s_{ij,mn}^b$'s in PP can be determined in at most $|\mathcal{N}||\mathcal{M}|$ rounds. Consequently, the computation complexity of PP in the $k$th iteration is upper bounded by $O(G^3|\mathcal{N}||\mathcal{M}|)$. We can see that the computational complexity in the $k$the iteration of our algorithm is $O\big((G+k+|\mathcal{N}|^2L)^3 + G^3|\mathcal{N}||\mathcal{M}|\big)$.

Thus, the computational complexity of our column generation algorithm when there are $K$ iterations is

$$\begin{aligned}
& O(\sum_{k=1}^{K}\left[(G+k+|\mathcal{N}|^2L)^3 + G^3|\mathcal{N}||\mathcal{M}|\right] \\
= \; & O\big((G+K+|\mathcal{N}|^2L)^4 + KG^3|\mathcal{N}||\mathcal{M}|\big) \\
= \; & O(K^4 + |\mathcal{N}|^8).
\end{aligned}$$

The first step is due to $\sum_{k=1}^{K} k^3 = K^2(K+1)^2/2$. In the worst case that all the independent sets need to be found, our algorithm needs to have at most $2^G - G$ iterations and hence

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMC.2014.2314107



its computational complexity is $O((2^{|\mathcal{N}|\log|\mathcal{N}|})^4 + |\mathcal{N}|^8)$, i.e., $O(2^{4|\mathcal{N}|\log|\mathcal{N}|})$. ■

Note that the computational complexity of our algorithm is polynomial if the number of iterations needed in the algorithm is at most a polynomial function of $|\mathcal{N}|$. Our later simulations show that usually only a small number of iterations are needed. In contrast, if we solve MP directly, the computational complexity is $O\big((2^G + |\mathcal{N}|(\log|\mathcal{N}|)L)^3\big)$, i.e., $O(2^{3|\mathcal{N}|\log|\mathcal{N}|})$, and hence always exponential. Thus, our algorithm's computational complexity can be much lower.

## V. Uncertain Channel Availability

So far we have assumed that frequency channels in M$^3$C$^2$Ns have constant bandwidths. However, in practice, the vacancy/occupancy of the secondary channels (or licensed channels) is uncertain and dynamic at different times, due to the unpredictable activities of the primary users. To model this unique feature of M$^3$C$^2$Ns, we consider that the bandwidths of secondary channels, i.e., $W^b$'s $(1 \le b \le M)$, are independent random variables, which is inspired by the statistical results of frequency channels obtained by experiments in [8]–[10]. Thus, from (1), we can know that link capacities $c^b_{ij}$'s are random variables as well.

Taking uncertain frequency channel availability into consideration, we can reformulate constraint (2) in OOP as follows:

$$\mathbf{Pr}\left(\sum_{q=1}^{Q} w_q \sum_{b\in\mathcal{M}_i\cap\mathcal{M}_j}\sum_{m\in\mathcal{R}_i}\sum_{n\in\mathcal{R}_j} c^b_{ij,mn}(\mathcal{I}_q) \ge \sum_{l\in\mathcal{L}} f_{ij}(l)\right) \ge \beta \tag{30}$$

where $\beta$ is a control parameter describing network operator's requirements on link quality. In so doing, the original MINLP becomes a Stochastic Optimization Problem (SOP), with random variables involved in its constraints. Obviously, we cannot directly apply our previously proposed method to solve this problem.

On the other hand, according to Markov inequality, we have

$$\begin{aligned}
&\mathbf{Pr}\left(\sum_{q=1}^{Q} w_q \sum_{b\in\mathcal{M}_i\cap\mathcal{M}_j}\sum_{m\in\mathcal{R}_i}\sum_{n\in\mathcal{R}_j} c^b_{ij,mn}(\mathcal{I}_q) \ge \sum_{l\in\mathcal{L}} f_{ij}(l)\right)\\
&\le \frac{E\left[\sum_{q=1}^{Q} w_q \sum_{b\in\mathcal{M}_i\cap\mathcal{M}_j}\sum_{m\in\mathcal{R}_i}\sum_{n\in\mathcal{R}_j} c^b_{ij,mn}(\mathcal{I}_q)\right]}{\sum_{l\in\mathcal{L}} f_{ij}(l)}\\
&= \frac{\sum_{q=1}^{Q} w_q \sum_{b\in\mathcal{M}_i\cap\mathcal{M}_j} E\left[\sum_{m\in\mathcal{R}_i}\sum_{n\in\mathcal{R}_j} c^b_{ij,mn}(\mathcal{I}_q)\right]}{\sum_{l\in\mathcal{L}} f_{ij}(l)},
\end{aligned}$$

which can give us a relaxed linear constraint as follows:

$$\beta\sum_{l\in\mathcal{L}} f_{ij}(l) \le \sum_{q=1}^{Q} w_q \sum_{b\in\mathcal{M}_i\cap\mathcal{M}_j}\sum_{m\in\mathcal{R}_i}\sum_{n\in\mathcal{R}_j} E\left[c^b_{ij,mn}(\mathcal{I}_q)\right]. \tag{31}$$

As a result, the SOP can be transformed back to an MINLP, which can be efficiently solved using our proposed $\epsilon$-bounded approximation algorithm. Notice that since (31) is a relaxed constraint compared to (30), the obtained optimal results serve as lower bounds on the optimal results of SOP.

## VI. Simulation Results

In this section, we carry out extensive simulations to evaluate the performance of the proposed algorithms. Simulations are conducted under CPLEX 12.4 and C++ on a computer with a 2.8 GHz CPU and 24 GB RAM. Besides, notice that previous works obtain suboptimal results that are either unbounded or far away from the optimal results, and many works based on conflict graphs assume all the maximal independent sets are given. Since our developed $\epsilon$-bounded approximation algorithm relaxes this assumption and is able to find tight $\epsilon$-bounded approximate solutions and the optimal solution as well, it is not very fair to compare it with other algorithms. Moreover, the new M$^3$C$^2$N architecture and the new hybrid communication paradigm make the problem studied in this paper a very unique problem, and hence it is not very appropriate to compare with other algorithms. In fact, due to a similar reason, most previous papers, e.g., [3], [11], [18], [19], [29], focus on the performance evaluation of their proposed algorithms and do not compare with other ones, either.

Specifically, we consider a square network of area $1000m \times 1000m$. A base station (BS) is located at the center, while 30 nodes are uniformly and randomly distributed in the area. Assume that each node has a downlink session from the BS and has a traffic demand of 100Kbps. The number of radio interfaces at the BS and at each user are 5 and 2, respectively. Some important simulation parameters are listed as follows. The path loss exponent is 4 and $C = 62.5$. The transmission power spectral density of nodes is $8.1\times10^7\eta$, and the reception threshold and interference threshold are both $10\eta$. Thus, the transmission range and the interference range on all channels are all equal to 150m. Besides, we set the reception power density of nodes to be $4.5 \times 10^7\eta$ based on the fact that the ratio of transmission power to reception power of wireless adaptors is usually 1:1 to 2:1 [32]. The transmission power spectral density of the BS is $5.06\times10^{10}\eta$, and hence the BS's transmission range is 750m on all channels, i.e., covering the whole network area. Note that since the location of the BS is (500m, 500m) and $R_T = 150$m, there are $P = \lceil\frac{500\sqrt{2}}{150}\rceil = 5$ proxy region candidates. Moreover, we assume the basic channel, i.e., the cellular channel available at the nodes and the BS, has bandwidth of 1MHz. The available secondary channels at each user are 4 randomly chosen channels from the channel set $[10, 20, ..., 100]$KHz, all of which are available at the BS. In addition, in the case of uncertain channel availability, we consider all the secondary channels' bandwidths follow the same normal distribution $\mathcal{N}$(50KHz, (5KHz)$^2$).

In the following, we evaluate the cost of solving RMP in Section VI-A. We evaluate the cost of solving PP, and compare the performance of the proposed SF algorithm with that of one traditional algorithm in Section VI-B. Then, we show the minimum scheduling length and the maximum network throughput, and compare the performance of our proposed architecture with that of pure ad hoc mode and that of traditional cellular networks in Section VI-C. We finally demonstrate the results under uncertain channel availability in Section VI-E.

TABLE II
SOLVING RMP WITH DIFFERENT $\epsilon$'S.

| $\epsilon$ | Iteration Number | Running Time (s) |
|---|---|---|
| 5% | 102 | 16.51 |
| 3% | 113 | 18.28 |
| 1% | 131 | 21.20 |
| 0% | 151 | 24.41 |

TABLE III
SOLVING RMP WITH DIFFERENT NETWORK SIZES.

| Network Size | Iteration Number | Running Time (s) |
|---|---|---|
| $N = 20$ | 55 | 6.38 |
| $N = 25$ | 79 | 11.34 |
| $N = 30$ | 151 | 24.41 |

### A. Cost of Solving RMP

We first study the cost of solving RMP under different network settings. Note that in order to well investigate the cost of solving RMP, we apply a traditional algorithm (provided by CPLEX), which can solve BIPs, to solve PP in inner iterations. Table II shows the iteration numbers and running time needed to solve RMP in order to obtain $\epsilon$-bounded approximate solutions when $g = 3$. We can see that when $N = 30$, it takes 102 iterations and 16.51 seconds to obtain 5%-bounded results, and 151 iterations and 24.41 seconds to get the optimal result, i.e., when $\epsilon = 0\%$. Table III gives the cost of solving RMP when $\epsilon = 0\%$ in networks of different sizes. We can observe that it only takes a few iterations and very little time to obtain the results.

Besides, Fig. 3 illustrates the convergence property of upper and lower bounds on optimal scheduling length when we use traditional algorithms to solve PP in inner iterations, given that $\epsilon = 0\%$ and there are 30 nodes in the network. Specifically, we assume the proxy region is at $g = \delta^{-1}(1) = 3$. At each iteration, we compute the lower and upper bounds on the minimum scheduling length of MP and track their progresses. Recall that in each iteration the upper bound, denoted by $\psi_u$ for simplicity, is the optimal result for RMP, while the lower bound, denoted by $\psi_l$ for simplicity, is calculated according to (29). We can find that, even though the gap between the lower and upper bounds is initially high, the gap narrows down quickly in the first 100 iterations. In addition, note that there is a sharp decrease of $\psi_u$ at the beginning. This is because the initial set of independent sets $\mathcal{K}'$ used for solving RMP is very small, and can be easily well improved. Thus, it demonstrates that the choice of initial independent sets does not have much impact on the convergence performance. Moreover, we find that the minimum scheduling length achieved is 0.78, which is smaller than 1. This means that with current proxy region plan, we can find a valid optimal solution to support users' traffic demands.

### B. Cost of Solving PP

We then evaluate the cost of solving PP by the proposed SF algorithm under different network settings. Particularly, Table IV compares the iteration numbers (for solving RMP) and the total running time of SF with those of a traditional algorithm provided by CPLEX when $g = 3$. We notice that although using SF leads to a few more iterations than using

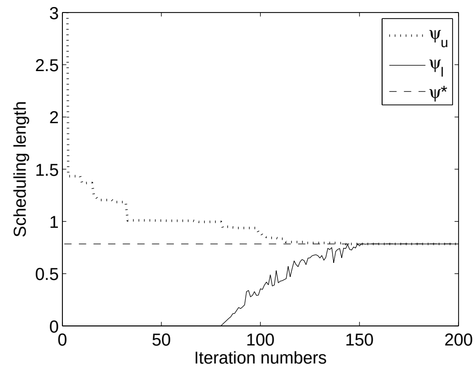


Fig. 3. Convergence property when using the traditional algorithm to solve PP.

TABLE IV
SOLVING PP USING SF WITH DIFFERENT $\epsilon$'S.

| $\epsilon$ | Iter. No. (Trad. Alg.) | Run. Time (S) (Trad. Alg.) | Iter. No. (SF) | Run. Time (s) (SF) |
|---|---|---|---|---|
| 5% | 102 | 8.48 | 115 | 4.80 |
| 3% | 113 | 9.32 | 129 | 5.35 |
| 1% | 131 | 10.95 | 151 | 6.37 |
| 0% | 151 | 12.67 | 173 | 7.34 |

the traditional algorithm, the running time when using SF is much less due to much lower complexity. We then evaluate in Table V the cost of solving PP in networks of different sizes when $\epsilon = 0\%$ and $g = 3$. Obviously, the proposed SF outperforms (in terms of running time) traditional algorithms more as network size gets larger.

Moreover, from Table II and Table IV, we can see that when $N = 30$ and $\epsilon = 0\%$, the running time for RMP and PP are 24.41 seconds and 7.34 seconds, respectively, resulting in total running time of 31.75 seconds. When $N = 30$ and $\epsilon = 5\%$, the running time for RMP and PP are 16.51 seconds and 4.81 seconds, respectively, resulting in total running time of 21.32 seconds. Note that the number of iterations for solving RMP is 151 when $\epsilon = 0\%$ and 102 when $\epsilon = 5\%$. Thus, we can see that given the same network size, the total running time is approximately proportional to the number of iterations for solving RMP ($151 : 102 \approx 31.75 : 21.32 \approx 1.48$).

In order to further illustrate the iteration number (and hence running time as well) with regard to $\epsilon$ of our proposed $\epsilon$-bounded approximation approach, we show in Fig. 4 the iteration ratio under different $\epsilon$'s. Here, the iteration ratio is defined as the ratio of the iteration number under certain $\epsilon$ to that to achieve optimality, i.e., $\epsilon = 0$. We can see that the iteration ratio decreases very fast as $\epsilon$ increases. In particular, even when $\epsilon$ is very tight, say 0.01, the iteration ratio is only about 0.87. When $\epsilon$ is larger than 0.1, the iteration ratio drops below 0.62. This reveals similar results observed by [6], [7]

TABLE V
SOLVING PP USING SF WITH DIFFERENT NETWORK SIZES.

| Network Size | Iter. No. (Trad. Alg.) | Run. Time (S) (Trad. Alg.) | Iter. No. (SF) | Run. Time (s) (SF) |
|---|---|---|---|---|
| $N = 20$ | 55 | 2.66 | 70 | 2.15 |
| $N = 25$ | 79 | 4.08 | 99 | 3.82 |
| $N = 30$ | 151 | 12.67 | 173 | 7.34 |

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMC.2014.2314107



that column generation approaches can determine solutions that are at least $95\% - 99\%$ of the global optimality fairly quickly.

In addition, Fig. 5 shows the convergence property of upper and lower bounds on optimal scheduling length when we use SF to solve PP in inner iterations, given that $\epsilon = 0\%$, $g = 3$, and there are 30 nodes in the network. We can see that the minimum scheduling length is also 0.78, which is the same result as that in Fig. 3.

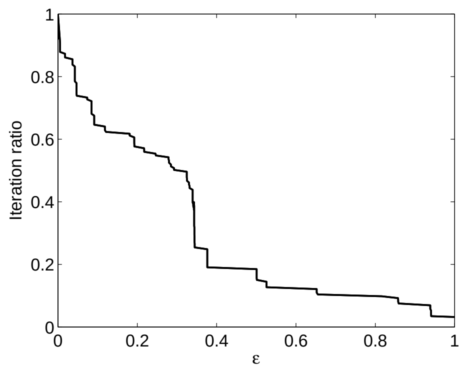


Fig. 4. The relationship between $\epsilon$ and iteration ratio.

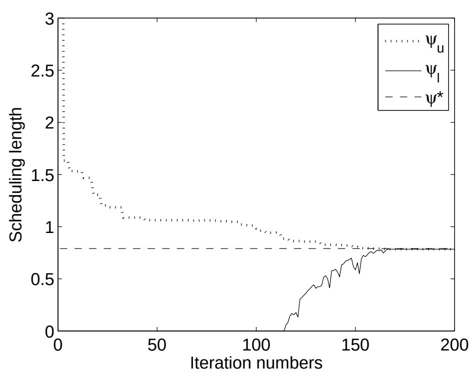


Fig. 5. Convergence property when using SF to solve PP.

### C. Minimum Scheduling Length and End-to-End Throughput

Next, we study the minimum scheduling length and the maximum end-to-end throughput with different proxy region allocation plans.

Fig. 6(a) depicts the minimum scheduling length $\psi_g^*$ when $g$ ranges from 1 to 5. We can see that the minimum scheduling length $\psi^*$ is equal to 0.61 when $g = 2$, which means that the second proxy region candidate $S_2$ is the optimal proxy region. Besides, when $g = 1$, i.e., *in the case of pure ad hoc mode transmissions*, $\psi_g^*$ is 0.97. When $g = 5$, the M$^3$C$^2$N becomes *a one-hop cognitive cellular network* (one-hop mode communications with both cellular and secondary channels available). Obviously, this structure cannot exploit spectrum spatial reuse among nodes and results in $\psi_g^* = 1.05$, which is larger than 1. It indicates that we cannot support users' traffic demands with this proxy region plan. Accordingly, we can derive the maximum end-to-end throughput $r_g^*(l)$, i.e., saturated throughput when the minimum scheduling length is 1, for each node under each region allocation plan $g$, assuming that all nodes still have equal traffic demands, which is $r_g^*(l) = r(l)/\psi_g^*$.

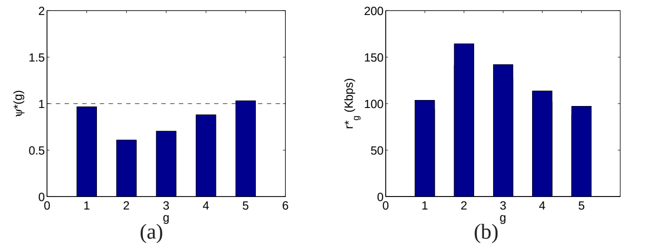


Fig. 6. Performance results with different proxy region plans. (a) Minimum scheduling length. (b) Maximum end-to-end throughput.

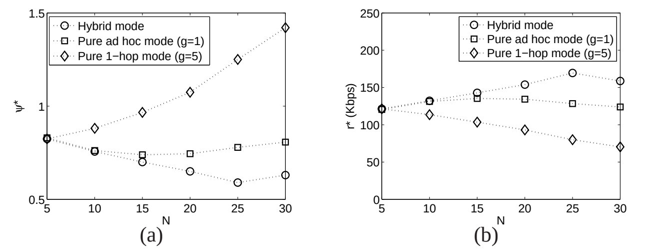


Fig. 7. Performance results in different transmission modes. (a) Minimum scheduling length. (b) Maximum end-to-end throughput.

To more clearly compare the performance of our proposed hybrid mode communications with that of pure ad hoc mode transmissions ($g = 1$) and that of pure one-hop mode transmissions ($g = 5$), we show the scheduling length and maximum throughput of these three architectures with different number of users in Fig. 7. We can easily see that the hybrid mode architecture has the best performance.

We further demonstrate the minimum scheduling length as well as the maximum end-to-end throughput in Fig. 8 under four different network architectures, i.e., single-radio single-channel single-hop traditional cellular network (Type A), single-radio single-channel multi-hop cellular network (Type B), single-radio single-channel multi-hop cognitive cellular network, (Type C) and multi-radio multi-channel multi-hop cognitive cellular network (Type D). Note that we still consider that each node has a downlink traffic demand of 100Kbps. In particular, Type A architecture represents the current typical cellular network. Its scheduling lengths are all larger than 1 when $N$ ranges from 5 to 30, indicating that it cannot fully support all the data traffic demands. Besides, the

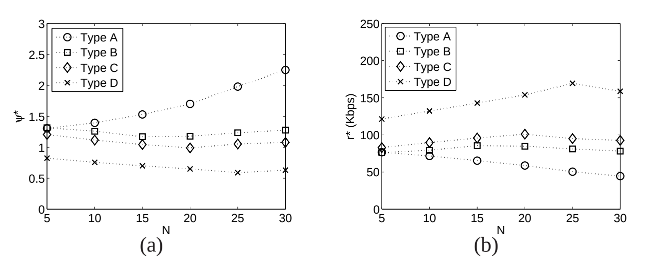


Fig. 8. Performance Results under different architectures. (a) Minimum scheduling length. (b) Maximum end-to-end throughput.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMC.2014.2314107



scheduling length keeps on increasing as $N$ grows. For Type B architecture, it allows multi-hop transmissions in cellular networks and hence can support more traffic than Type A due to increased spectrum spatial reuse and link rate adaptivity. Type C architecture can further enhance network throughput because of more available channel resources brought by cognitive radio. Type D architecture, i.e., the proposed one in this paper, has the best performance.

### D. Energy Consumption

We also compare the energy consumption, including transmission energy and reception energy, per bit of traffic in our proposed hybrid mode with that in pure ad hoc mode and that in pure one-hop mode. Specifically, it is calculated by dividing the total power consumption by the maximum end-to-end throughput achieved. Since energy consumption at mobile users is more critical in wireless networks, Fig. 9 shows the energy consumption per bit of traffic at all mobile users, when $N$ ranges from 5 to 30. The energy per bit consumed in pure 1-hop mode is the lowest since the mobile users' transmission energy is 0. Besides, we find that given the same $N$, the mobile users consume much more energy per bit in pure ad hoc mode than in hybrid mode. This is because first, more nodes may be involved in the data transmissions in pure ad hoc mode, and second, hybrid mode can support higher end-to-end throughput as illustrated in Fig. 7(b). Note that the results are obtained when downlink traffic is considered. Obviously, the energy saving in the hybrid mode compared with that in the pure one-hop mode will be more significant for uplink traffic.

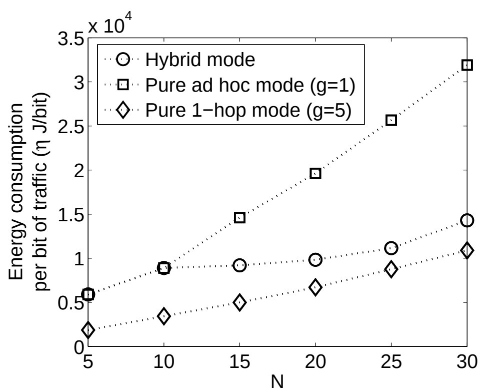


Fig. 9. Energy consumption at mobile users under different transmission mode.

### E. Minimum Scheduling Length Under Uncertain Channel Availability

We first illustrate the results of minimum scheduling length under uncertain channel availability with different $\beta$'s in Fig. 10. Note that we consider a network of 30 nodes, and set $g = 3$ as the proxy region. We can see that the minimum scheduling length (when the results are stable) when $\beta = 0.9$ is shorter than that when $\beta = 0.95$. This is intuitively true because a smaller $\beta$ indicates a lower requirement on service quality, and hence the minimum scheduling length can be lower. Interestingly, we also notice that it takes more iterations for the

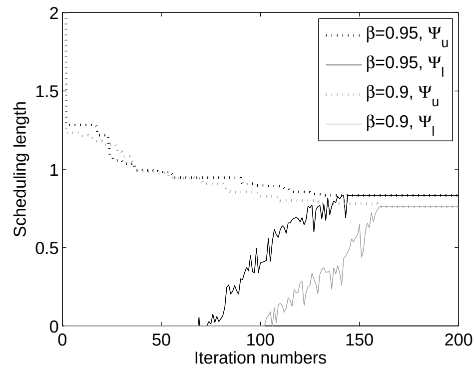


Fig. 10. Performance comparison under different $\beta$'s for uncertain channel availability.

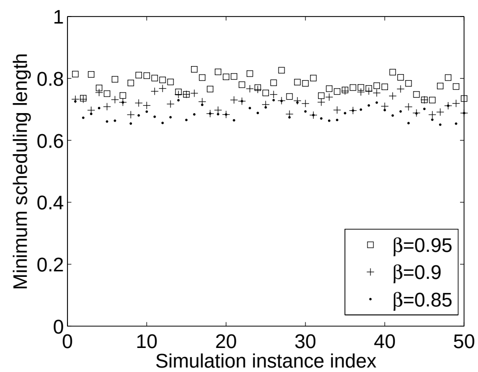


Fig. 11. Minimum scheduling length under different $\beta$'s for uncertain channel availability.

result to become stable when $\beta$ is smaller. We also show in Fig. 11 the minimum scheduling length under different $\beta$'s. Note that we have assumed that all secondary channels have the same expected bandwidths of 50 KHz. For each $\beta$, we conduct 50 simulations, each with a newly generated topology and bandwidths for all the secondary channels according to their distributions. We find that the minimum scheduling length achieved when $\beta = 0.85$ is the lowest among the three, due to the same reason as that for Fig. 10. In addition, we show in Fig. 11 the minimum scheduling length under secondary channels of different expected bandwidths $E[W^b]$'s, which indicates different levels of secondary channel availability. We assume that all secondary channels have the same standard variance of 5 KHz. For each $E[W^b]$, we conduct 50 simulations, each of which uses a newly generated topology. We set $\beta = 0.9$. As shown in Fig. 12, we find that the minimum scheduling length achieved when $E[W^b] = 60$KHz is the lowest since it has the highest level of channel availability.

## VII. Related Work

In this section, we discuss related work on multihop cellular networks and cross-layer throughput optimization for wireless networks.

### A. Cellular Network Architecture

In traditional cellular networks [1], [33]–[35], ad hoc communications are introduced to deliver information between

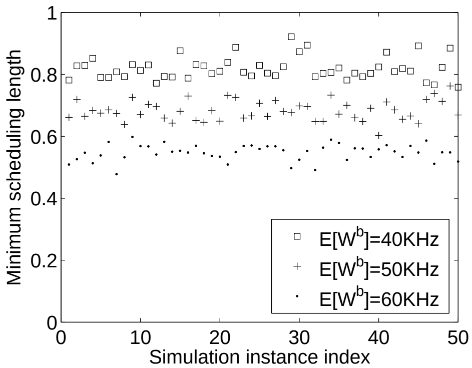


Fig. 12. Minimum scheduling length under different $E[W^b]$'s for uncertain channel availability.

users, but every user still communicates with base stations directly in one hop, which leads to low frequency spatial reuse and hence low throughput. Considering multi-hop communications between nodes and base stations, some works such as [1], [2] investigate the capacity of multihop cellular networks, which has been shown to be much higher than that of traditional cellular networks. However, these works only consider the case where nodes share the cellular channels and have not exploited the local available channels or multi-radio as we propose in this study. Besides, although asymptotic capacity bounds have been studied, the exact optimal throughput value remains unknown.

### B. Cross-Layer Optimization for Wireless Networks

There has been some work on joint link scheduling and routing for throughput optimization in traditional ad hoc networks [4], [36]–[38], multi-radio multi-channel mesh networks [26], [39]–[43], and cognitive ad hoc networks [3], [18], [19], [29], [44].

*Traditional ad hoc networks:* Jain et al. [36] study the throughput bounds of one source-destination pair considering both routing and interference. Zhai and Fang [4] study the impact of routing metrics on path capacity by formulating a joint link scheduling and routing optimization problem. A single traffic flow is considered in the paper. Lin and Shroff [37] employ a joint rate control and scheduling approach to maximize the utility of the network subject to the constraint that the network is stable. The same authors [38] then design a distributed algorithm to solve the problem, which is shown to achieve a constant factor of the capacity region. Routing is not considered in their optimization problem. In these works, they study single-radio single-channel networks.

*Multi-radio multi-channel networks:* Alicherry et al. [39], [40] investigate the throughput optimization problem by joint channel assignment and routing, assuming a CSMA/CA medium access control protocol, and find a suboptimal result that is within $K \cdot c(q)/I$ of the optimal result, where $K$ is the number of channels, $c(q)$ is a constant greater than 4, and $I$ is the number of radios at each node. Kodialam and Nandagopal [41] study the throughput bounds of multi-radio multi-channel mesh networks following a similar approach. Li et al. [26] explore the capacity region by formulating a joint scheduling, channel assignment, and routing optimization problem, and develop a heuristic algorithm to obtain capacity upper bounds. Zhou et al. [42] study the video streaming problem to minimize the video distortion by jointly considering channel assignment, rate allocation, and routing. Link scheduling is not considered. Lin and Rasool [43] design distributed algorithms that jointly solve the channel assignment, scheduling, and routing problem, and show that the algorithms can achieve an efficiency ratio of $1/(K+2)$ or $1/K$, where $K$ is the interference degree of the network.

*Cognitive ad hoc networks:* Tang et al. [3] try to maximize network throughput considering frequency channel allocation and scheduling but not routing, and find suboptimal results. Feng and Yang [18] study a joint transport, routing, and frequency channel allocation optimization problem to maximize network utility. Ding et al. [44] attempt to maximize network utility by joint relay selection, routing, and frequency channel allocation, and propose both centralized and distributed algorithms to obtain suboptimal results. Hou et al. [19], [29] try to minimize space-bandwidth product by joint frequency channel sharing and routing. Heuristic algorithms are proposed to find lower bounds on the optimal results. In such works, time-domain link scheduling is not considered.

TABLE VI
COMPARISON OF OUR WORK AND EXISTING WORKS ON CROSS-LAYER OPTIMIZATION FOR WIRELESS NETWORKS.

| | Multi-radio multi-channel | Frequency-domain scheduling | Time-domain scheduling | Optimal solution |
|---|---|---|---|---|
| [4] | × | √ | √ | × |
| [45] | × | × | √ | × |
| [37] | × | × | √ | √ |
| [38] | × | × | √ | √ |
| [39] | √ | √ | × | × |
| [40] | √ | √ | × | × |
| [41] | √ | √ | √ | × |
| [26] | √ | √ | √ | × |
| [42] | √ | √ | × | × |
| [43] | √ | √ | √ | × |
| [3] | × | √ | √ | × |
| [18] | × | √ | √ | × |
| [44] | × | √ | √ | × |
| [29] | √ | √ | × | × |
| [19] | √ | √ | × | × |
| Our work | √ | √ | √ | √ |

We summarize the main differences between the above works and ours in Table VI. In general, only a few perform a complete study on joint link scheduling (both in time domain and in frequency domain) and routing. Besides, previous works only obtain suboptimal results that are either unbounded or still far from the optimal results. In this paper, we are able to find $\epsilon$-bounded approximate results, which are less than $(1+\epsilon)$ and larger than $(1-\epsilon)$ of the optimal result and hence much tighter. Here, $0 \leq \epsilon < 1$ is a system control parameter. Besides, some conflict graph based works like [3], [4] assume all the maximal independent sets are given, while we do not make this assumption in this study.

Moreover, a few works [11]–[13] study cognitive cellular networks, but they focus on resource management rather than join link scheduling and routing optimization, and still enforce one-hop direct communications between nodes within the transmission range of base stations and the base stations. In addition, since each node has only one radio, such networks can be considered as a special case in the proposed $M^3C^2$Ns.

Furthermore, most previous research on the minimum length scheduling has been conducted on traditional ad hoc networks [46]–[48] and only considers link scheduling. In contrast, we study this problem in cognitive networks by taking into consideration both scheduling and routing. Besides, most previous works only derive an upper bound on the minimum scheduling length, while we obtain both upper and lower bounds in $\epsilon$-bounded approximate solutions, which converge quickly, and can also efficiently find the optimal result when $\epsilon = 0\%$.

In addition, most previous works study homogeneous networks, where each node has a fixed set of available channels and all nodes have the same transmission range on all the channels. While in $M^3C^2$Ns, we consider heterogeneous networks and take uncertain channel availability into account, which is an intrinsic feature of cognitive radio networks but has rarely been studied before. Besides, although column generation has been adopted in the study of wireless networks [49]–[54], our problem formulation and the algorithm design are completely different. We propose an $\epsilon$-bounded approximation algorithm for the first time in the literature.

## VIII. Conclusions

In this paper, we have proposed a novel Multi-radio Multi-channel Multi-hop Cognitive Cellular Network ($M^3C^2$N) architecture to enhance the throughput of current cellular networks. We study the minimum length scheduling problem in $M^3C^2$Ns, and formulate it as a maximal independent set based joint scheduling and routing optimization problem, which we call OOP and is a Mixed Integer Linear Programming (MINLP) and hence generally NP-hard problem. Then, we decompose OOP into a sequence of MPs, each of which is further decomposed into an RMP and a PP. Noticing that RMP can be solved in polynomial time but PP is still an NP-complete problem, we then design a sequential-fix (SF) algorithm which can find a suboptimal solution to PP in polynomial time. After that, an $\epsilon$-bounded approximation algorithm is developed which can find the $\epsilon$-bounded approximate result and the optimal result (when $\epsilon = 0\%$) quickly. Consequently, we are able to solve OOP very efficiently, in terms of both $\epsilon$-bounded approximate solutions and optimal result as demonstrated by simulations, without having to find all the maximal independent sets. Furthermore, although most previous research only assumes constant channel bandwidth, we take uncertain channel availability into consideration to account for practical issues, e.g., the unpredictable activities of primary users.

## References


[1] Y. Lin and Y. Hsu, "Multihop cellular: a new architecture for wireless communications," in *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM'00)*, Tel Aviv, Israel, March 2000.

[2] P. Li, X. Huang, and Y. Fang, "Capacity scaling of multihop cellular networks," in *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM'11)*, Shanghai, China, April 2011.

[3] J. Tang, S. Misra, and G. Xue, "Joint spectrum allocation and scheduling for fair spectrum sharing in cognitive radio wireless networks," *Computer Networks (Elsevier) Journal*, vol. 52, no. 11, pp. 2148–2158, August 2008.

[4] H. Zhai and Y. Fang, "Impact of routing metrics on path capacity in multirate and multihop wireless ad hoc networks," in *Proc. of the IEEE International Conference on Network Protocols, ICNP 2006*, Santa Barbara, CA, November 2006.

[5] M. Pan, C. Zhang, P. Li, and Y. Fang, "Joint routing and link scheduling for cognitive radio networks under uncertain spectrum supply," in *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM'11)*, Shanghai, China, April 2011.

[6] L. S. Lasdon, *Optimization Theory for Large Systems*. NY: Dover, 2002.

[7] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows, 3rd ed.* NY: Wiley, 2005.

[8] M. McHenry, "Spectrum white space measurements," New America Foundation Broadband Forum, June 20, 2003.

[9] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements and analysis and a long-term studies proposal," in *Proc. of TAPAS 2006*, Boston, MA, August 2006.

[10] D. Chen, S. Yin, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: a large-scale spectrum measurement study," in *Proc. of international conference on Mobile computing and networking (Mobicom'09)*, Beijing, China, September 2009.

[11] H. Xu and B. Li, "Efficient resource allocation with flexible channel cooperation in ofdma cognitive radio networks," in *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM'10)*, San Diego, California, USA, March 2010.

[12] J. Jin and B. Li, "Cooperative resource management in cognitive wimax with femto cells," in *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM'10)*, San Diego, California, USA, March 2010.

[13] J. Jin, , H. Xu, and B. Li, "Multicast scheduling with cooperation and network coding in cognitive radio networks," in *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM'10)*, San Diego, California, USA, March 2010.

[14] R. Wang, V. K. N. Lau, L. Lv, and B. Chen, "Joint cross-layer scheduling and spectrum sensing for ofdma cognitive radio systems," *IEEE Transactions on Wireless Communication*, vol. 8, no. 5, pp. 2410–2416, May 2009.

[15] T. C. yam Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks,," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 328–339, February 2007.

[16] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surv. Tutorials*, vol. 11, no. 1, p. 116130, First quarter, 2009.

[17] Y. Zeng, Y.-C. Liang, A. T. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: challenges and solutions," *EURASIP Journal on Advances in Signal Processing - Special Issue on Advanced Signal Processing for Cognitive Radio Networks*, January 2010.

[18] Z. Feng and Y. Yang, "Joint transport, routing and spectrum sharing optimization for wireless networks with frequency-agile radios," in *Proc. of IEEE Conference on Computer Communications (INFOCOM'09)*, Rio de Janeiro, Brazil, April 2009.

[19] Y. T. Hou, Y. Shi, and H. D. Sherali, "Spectrum sharing for multi-hop networking with cognitive radios," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 146–155, January 2008.

[20] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[21] Y. Shi, Y. T. Hou, J. Liu, and S. Kompella, "Bridging the gap between protocol and physical models for wireless networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 7, pp. 239–258, 2013.

[22] X. Che, X. Liu, X. Ju, and H. Zhang, "Adaptive instantiation of the protocol interference model in mission-critical wireless networks," in *IEEE Proceedings of SECON*, Boston, MA, June 2010.

[23] M. Pan, C. Zhang, P. Li, and Y. Fang, "Spectrum harvesting and sharing in multi-hop crns under uncertain spectrum supply," *IEEE Journal on Selected Areas in Communications - Special Issue on Cooperative Networking Challenges and Applications*, vol. 30, no. 2, pp. 369–378, February 2012.

[24] Y. Pochet and L. Wolsey, *Production Planning by Mixed Integer Programming*. Secaucus, 2006.

[25] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[26] H. Li, Y. Cheng, C. Zhou, and P. Wan, "Multi-dimensional conflict graph based computing for optimal capacity in MR-MC wireless networks," in *Proc. of International Conference on Distributed Computing Systems (ICDCS'10)*, Genoa, Italy, June 2010.

[27] R. Diestel, *Graph Theory*. Springer, 2005.

[28] D. Bertsimas and J. Tsitsiklis, *Introduction to Linear Optimization*. Athena Scientific, 1997.

[29] Y. T. Hou, Y. Shi, and H. D. Sherali, "Optimal spectrum sharing for multi-hop software defined radio networks," in *Proc. of IEEE Conference on Computer Communications (INFOCOM'07)*, Anchorage, AL, May 2007.

[30] P. Gupta and P. Kumar, "Critical power for asymptotic connectivity in wireless networks," *Stochastic Analysis, Control, Optimization and Applications*, vol. A Volume in Honor of W.H. Fleming, pp. 547–566, 1998.

[31] Y. Ye, "An $o(n^3 l)$ potential reduction algorithm for linear programming," *Math. Programming*, vol. 50, 1991.

[32] B. Chen, K. Jamieson, H. Balakrishnamurthy, and S. Dao, "Span: an energy-efficient coordination algorithm for topology maintainance in ad hoc wireless networks," *ACM Transactions on Wireless Networks*, vol. 8, no. 5, pp. 481–494, 2002.

[33] H. Luo, R. Ramjee, P. Sinha, L. Li, and S. Lu, "Ucan: A unified cellular and ad-hoc network architecture," in *Proceeding of ACM MobiCom (MobiCom'03)*, San Diego, CA, USA, June 2003.

[34] P. Li, C. Zhang, and Y. Fang, "Capacity and delay of hybrid wireless broadband access networks," *IEEE Journal on Selected Areas in Communications (JSAC) - Special Issue on Broadband Access Networks*, vol. 27, no. 2, pp. 117–125, February 2009.

[35] A. Zemlianov and G. Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 3, March 2005.

[36] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proc. of Mobile Computing and Networking (Mobicom'03)*, San Diego, CA, September 2003.

[37] X. Lin and N. B. Shroff, "Joint rate control and scheduling in multihop wireless networks," in *Proc. of IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 2004.

[38] ——, "The impact of imperfect scheduling on cross-layer congestion control in wireless networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 2, pp. 302–315, April 2006.

[39] M. Alicherry, R. Bhatia, and L. E. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proc. of international conference on Mobile computing and networking (Mobicom'05)*, Cologne, Germany, August 2005.

[40] M. Alicherry, R. Bhatia, and L. Li, "Joint channel assignment and routing for throughput optimization in multiradio wireless mesh networks," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 24, no. 11, pp. 1960–1971, November 2006.

[41] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. of international conference on Mobile computing and networking (Mobicom'05)*, Cologne, Germany, August 2005.

[42] L. Zhou, "Distributed scheduling scheme for video streaming over multi-channel multi-radio multi-hop wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 409–419, 2010.

[43] X. Lin and S. B. Rasool, "Distributed and provably efficient algorithms for joint channel-assignment, scheduling, and routing in multichannel ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 6, pp. 1874–1887, 2009.

[44] L. Ding, T. Melodia, S. N. Batalama, and J. D. Matyjas, "Distributed routing, relay selection, and spectrum allocation in cognitive and cooperative ad hoc networks," in *Proceedings of IEEE SECON'10*, Boston, Massachusetts, June 2010.

[45] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wireless Networks*, vol. 11, July 2005.

[46] P. Djukic and S. Valaee, "Delay aware link scheduling for multi-hop tdma wireless networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 870–883, 2009.

[47] D. M. Blough, G. Resta, and P. Santi, "Approximation algorithms for wireless link scheduling with sinr-based interference," *IEEE/ACM Transactions on Networking*, vol. 18, no. 6, pp. 1701–1712, 2010.

[48] O. Goussevskaia, Y. V. Oswald, and R. Wattenhofer, "Complexity in geometric sinr," in *Proceeding of ACM MobiHoc*, Montreal, QC, Canada, September 2007.

[49] P. V. Patrik Björklund and D. Yuan, "Resource optimization of spatial tdma in ad hoc radio networks: A column generation approach," in *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM'03)*, San Francisco, CA, USA, March 2003.

[50] L. Fu, S. C. Liew, and J. Huang, "Fast algorithms for joint power control and scheduling in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 3, pp. 1186–1197, 2010.

[51] X. Zheng, F. Chen, Y. Xia, and Y. Fang, "A class of cross-layer optimization algorithms for performance and complexity trade-offs in wireless networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 10, pp. 1393–1407, 2009.

[52] M. Johansson and L. Xiao, "Cross-layer optimization of wireless networks using nonlinear column generation," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 435–445, 2006.

[53] M. Cao, X. Wang, S.-J. Kim, and M. Madihian, "Multi-hop wireless backhaul networks: A cross-layer design paradigm," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 25, no. 4, pp. 738–748, 2007.

[54] J. Luo, C. Rosenberg, and A. Girard, "Engineering wireless mesh networks: Joint scheduling, routing, power control, and rate adaptation," *IEEE Transactions on Networking*, vol. 18, no. 5, pp. 1387–1400, 2010.