CURT: A Real-Time Scheduling Algorithm for Coexistence of LTE and Wi-Fi in Unlicensed Spectrum

Yan Huang Yongce Chen Y. Thomas Hou Wenjing Lou
Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

***Abstract*—Carrier-Sensing Adaptive Transmission (CSAT) is a major approach from industry to address coexistence between LTE and Wi-Fi in unlicensed bands. Under CSAT, a key problem is the design of a scheduling algorithm to allocate radio resources across multiple channels and a large number of sub-channels. This paper investigates this scheduling problem through an optimization formulation with the objective of minimizing LTE's adverse impact on Wi-Fi users. This is achieved by optimal allocation of radio resources at channel and sub-channel levels to meet each LTE user's uplink and downlink rate requirements. Special considerations of channel conditions are given during LTE scheduling. A major challenge here is to obtain an optimal (or near-optimal) scheduling solution on ∼1 ms time scale — a stringent timing requirement for the algorithm to be useful in the field. Our main contribution is the development of CURT, a scheduling algorithm that can obtain near-optimal solution in ∼1 ms. CURT exploits the unique structure of the underlying optimization problem and decomposes it into a large number of independent sub-problems. These sub-problems can be solved efficiently and in parallel by GPU multi-processors. By implementing CURT on Nvidia GPU/CUDA platform, we demonstrate that CURT can indeed deliver near-optimal scheduling solution in ∼1 ms and meet all our design objectives.**

***Index terms*—** Optimization, real-time, unlicensed LTE, coexistence, spectrum sharing, resource scheduling

I. Introduction

There is a strong interest from cellular carriers to use existing unlicensed spectrum (e.g., the 5 GHz UNII bands) to boost cellular services. This approach is appealing for a number of reasons: (i) unlicensed spectrum is free (no need of auction and a license fee), (ii) the underlying bandwidth is substantial (e.g., 775 MHz available bandwidth in 5 GHz UNII bands), (iii) coexisting with other unlicensed wireless technologies (e.g., Wi-Fi) bears significantly fewer operational risk concerns when compared to sharing spectrum on the military bands (e.g., with radar systems). As a result, there have been significant activities on coexistence of cellular (LTE) and Wi-Fi in unlicensed bands from both industry [1], [2], [3], [4], [5] and academia [6], [7], [8], [9], [10], [11], [12].

A key consideration in the design and operation of Unlicensed LTE (U-LTE) systems is to ensure fairness when they coexist with Wi-Fi. LTE was originally designed to work exclusively in operator-owned licensed bands. Its transmissions are centrally controlled and have no consideration for cross-technology coexistence [4]. In contrast, Wi-Fi employs CSMA/CA and is based on distributed contention. It can only transmit when the operating channel is clear and after its backoff period. Such incompatibility makes Wi-Fi highly vulnerable to the presence of LTE in the same band.

To address this issue, a number of mechanisms have been proposed for U-LTE, such as Listen-Before-Talk (LBT) [3], [5], and Carrier-Sensing Adaptive Transmission (CSAT) [1], [2]. LBT is a random access approach similar to Wi-Fi's CSMA/CA. On the other hand, CSAT is based on centralized scheduling, which is native to LTE's operation. With proper design, both CSAT and LBT could achieve fair spectrum sharing between LTE and Wi-Fi. Although CSAT may cause collisions to Wi-Fi's on-going packets, such impact can be mitigated by configuring longer duration for each LTE transmission burst [8]. CSAT is fully compatible with 3GPP Release 10/11 and does not require any change of LTE specifications [7]. It can be quickly launched in countries that do not mandate implementing LBT (e.g., the US and China). The US operator T-Mobile has already supported CSAT-based U-LTE in six cities in the US since early 2017 [13].

In this paper, we employ the CSAT coexistence mechanism and study a scheduling problem for U-LTE and Wi-Fi in 5 GHz unlicensed bands. In the 5 GHz spectrum, there are multiple bands that can be used by LTE simultaneously. Under CSAT, the air time of each channel is divided into periodic LTE "on/off" cycles, where the "on" and "off" periods are used by LTE and Wi-Fi for channel access, respectively. Optimal division of "on" and "off" periods is determined by the LTE eNodeB (eNB) based on Wi-Fi's traffic load measured from carrier sensing. Within LTE's "on" period of a channel, the bandwidth of the channel is expanded into a group of sub-channels and it is at this level that the so-called Resource Blocks (RBs) are allocated to LTE users. Suppose we have a different set of Wi-Fi users on each channel. To support a set of LTE users on these channels, where each user may have its own uplink (UL) and downlink (DL) rate requirements, the problem becomes how to perform radio resource allocation to minimize U-LTE's impact on Wi-Fi while meeting various constraints and requirements.

The scheduling problem as outlined above can be formulated into an optimization problem. But getting an optimal (or near optimal) solution is not trivial. The main challenge we are facing is to find a scheduling solution in real-time, with a computational time of ∼1 ms. This timing requirement comes from the fact that channel coherence time in 5 GHz bands is at most 10s of ms, meaning that a channel-dependent scheduling solution is only effective for 10s of ms. If the computation

This research was supported in part by NSF under grants 1800650, 1642873, 1617634, 1343222 and ONR under grant N00014-15-1-2926.

978-1-5386-5191-9/18/$31.00 ©2018 IEEE

time is beyond this time limit, the solution would not be good since channel conditions may have already changed. We aim to solve the scheduling problem in $\sim$1 ms so that U-LTE can operate with the optimized scheduling solution for the bulk portion of the coherence time period. We call our proposed solution CURT, which can be considered as an abbreviation of CSAT based U-LTE Real-Time resource scheduling (from coexistence scheme's perspective) or CUDA-based Real-Time resource scheduling (from implementation's perspective). We summarize the main contributions of this paper as follows:

- In the design of CURT, we take a wide range of parameters into considerations so as to best resemble what one would encounter in the field. These include (i) multiple channels available for U-LTE/Wi-Fi coexistence; (ii) both UL and DL rate requirements from LTE users; (iii) variation of channel conditions across sub-channels. We formulate our scheduling problem into an optimization problem with the objective of minimizing the adverse impact on Wi-Fi while meeting U-LTE users' rate requirements.
- For such highly complex optimization problem, traditional approaches based on iterative computation cannot meet the real-time requirement in practice. Instead, by exploiting the unique problem structure, we decompose the original scheduling problem into a large number ($\sim 10^3$) of independent sub-problems that enumerate all possible cases of U-LTE's impact on Wi-Fi. Then by evaluating all sub-problems in parallel by massive GPU processing cores, we are able to determine a near-optimal solution for each scheduling frame.
- By implementing CURT on Nvidia GPU/CUDA platform, we show that CURT can find a near-optimal scheduling solution in $\sim$1 ms under realistic network settings. This represents the first known CSAT-based scheduler design that can achieve real-time and near-optimal scheduling on a per-frame level for the coexistence between U-LTE and Wi-Fi.

II. Related Work

Our literature review covers existing works that are most relevant to this paper. In the research community, there have been a number of studies on LTE/Wi-Fi coexistence, such as modeling and analysis of U-LTE's impact on Wi-Fi [9], [10], optimizations of coexistence mechanisms [6], [12], and radio resource management of U-LTE [11].

In [9], the authors developed an analytical model for Wi-Fi's collision probability and throughput when coexisting with CSAT-based LTE. In [10], the authors proposed a general framework to evaluate the performance of multiple technologies operating in the same unlicensed bands. Both [9] and [10] focused on modeling of LTE/Wi-Fi coexistence and did not address allocation of radio resources.

In [12], the authors derived optimal division of LTE "on" and "off" periods under CSAT for a given number of LTE and Wi-Fi users. In [6], the authors considered joint optimization of channel selection and CSAT parameters. A fairness criterion was derived for LTE and Wi-Fi sharing multiple unlicensed channels. The criterion requires LTE not to impact Wi-Fi more than another Wi-Fi network with the same traffic load. The efforts in [12] and [6] addressed optimizations of CSAT parameters, but fell short on addressing resource management for U-LTE at the RB level.

In [11], the authors aimed to maximize energy efficiency for CSAT-based U-LTE by studying RB allocation over licensed and unlicensed bands. The analysis in this work, however, did not consider channel fading effect on each individual RB, which is what would be encountered in practice. Further, it is unclear whether or not the proposed iterative algorithm can find a solution within the channel coherence time.

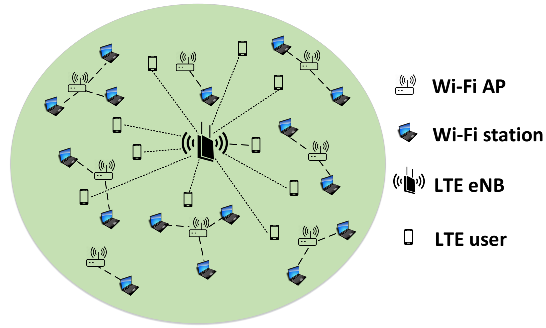


Fig. 1: Coexistence of LTE and Wi-Fi in an area.

TABLE I: Notation

Symbol	Definition
$\mathcal{F}$	A set of channels shared between LTE and Wi-Fi
$\mathcal{S}_i$	A set of sub-channels on channel i
(i,j)	The jth sub-channel in $\mathcal{S}_i$
$\mathcal{K}$	A set of LTE users offloaded to unlicensed bands
T_0	Duration of a TTI
T_{SF}	Duration of a scheduling frame
N_{SF}	The number of TTIs in a scheduling frame
M	The number of radio frames in a scheduling frame
I_i^{UL}	Binary variable indicating whether or not channel i is selected for UL transmission
I_i^{DL}	Binary variable indicating whether or not channel i is selected for DL transmission
$n_{(i,j)}^{k,\text{UL}}$	Integer variable denoting the amount of TRBs allocated to user k for UL transmission on sub-channel (i,j)
$n_{(i,j)}^{k,\text{DL}}$	Integer variable denoting the amount of TRBs allocated to user k for DL transmission on sub-channel (i,j)
$n_{i,\max}^{\text{UL}}$	The amount of TRBs reserved for LTE's UL transmission on channel i
$n_{i,\max}^{\text{DL}}$	The amount of TRBs reserved for LTE's DL transmission on channel i
$C_{(i,j)}^{k,\text{UL}}$	UL achievable rate of user k on sub-channel (i,j)
$C_{(i,j)}^{k,\text{DL}}$	DL achievable rate of user k on sub-channel (i,j)
$R^{k,\text{UL}}$	UL rate requirement of user k on unlicensed bands
$R^{k,\text{DL}}$	DL rate requirement of user k on unlicensed bands
Q_i	The maximum number of TTIs that can be used for LTE scheduling on channel i
U_i	The number of Wi-Fi nodes on channel i
w_i	The weight reflecting the Wi-Fi traffic load on channel i
z	Objective value in Problem OPT-R

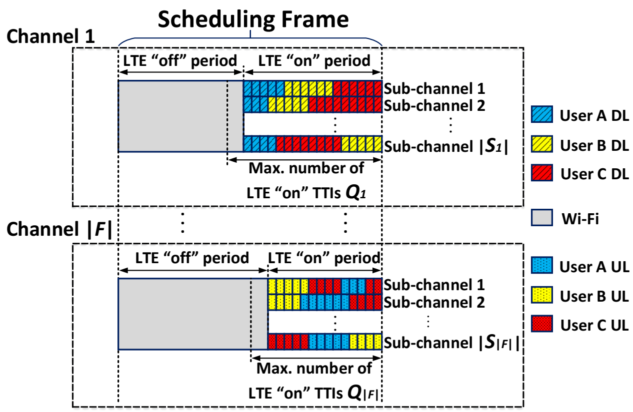


Fig. 2: An illustration of a scheduling frame.

III. System Architecture

We consider a deployment scenario where a U-LTE small cell overlaps with multiple Wi-Fi APs, as shown in Fig. 1. Table I lists notation in this paper. A set of LTE users $\mathcal{K}$ are served by a single U-LTE eNB while each Wi-Fi user is served by a nearby Wi-Fi AP.[1] Each LTE user $k \in \mathcal{K}$ has both UL and DL rate requirements in the unlicensed spectrum, which are denoted by $R^{k,\text{UL}}$ and $R^{k,\text{DL}}$, respectively. The rate requirements should be configured dynamically by a traffic management mechanism on unlicensed bands as described in Section V-E.

Suppose there is a number of channels in the unlicensed band that can be used by both LTE and Wi-Fi networks. Due to dense Wi-Fi deployment, there may not be enough clear channels for LTE and thus LTE has to coexist with Wi-Fi on some of these channels. In this paper, we focus on this subset of channels where LTE and Wi-Fi coexist, which is denoted by $\mathcal{F}$. Denote U_i as the number of Wi-Fi nodes on channel $i \in \mathcal{F}$. For LTE, its transmission scheduling is centrally controlled by the eNB, and it can combine multiple channels for UL and DL transmissions via FDD carrier aggregation. Every channel $i \in \mathcal{F}$ (used for either UL or DL) is further divided into a set of sub-channels $\mathcal{S}_i$. Thus the frequency granularity of LTE is on sub-channel level. For Wi-Fi, an AP or station typically occupies the entire bandwidth of a channel and employs CSMA for DL or UL data transmission.

CSAT-based U-LTE Scheduling We employ CSAT for U-LTE scheduling [2]. As shown in Fig. 2, CSAT is a TDM-like channel access mechanism where each CSAT cycle (a.k.a a scheduling frame) consists of an LTE "off" and "on" period. During the "off" period, LTE transmission is suspended so that co-channel Wi-Fi nodes can access the medium. Once the "off" period is over, LTE network starts transmission regardless of channel occupancy status. Since Wi-Fi senses the channel before transmission, it will cease transmission during LTE's "on" period. In a CSAT cycle, the division of "off" and "on" periods is determined by the centralized LTE scheduler. Through carrier sensing, the eNB measures Wi-Fi traffic load on each channel, and uses it as input to determine "off" and "on" periods.

Radio Resource Arrangement in LTE In LTE, the radio resource on a channel is organized as a two-dimensional resource grid [4]. In frequency domain, each channel is divided into a set of sub-channels, each with a bandwidth of 180 KHz. In time domain, we have consecutive *radio frames*, each with a duration of 10 ms. A radio frame consists of 10 sub-frames. The duration of a sub-frame is 1 ms, which is termed a *Transmission Time Interval (TTI)*. A TTI is further divided into two time slots, each having a duration of 0.5 ms. A resource block (RB) is defined as a time-frequency resource unit with 180 kHz in frequency (a sub-channel) and 0.5 ms in time (a time slot). LTE's scheduling resolution is two consecutive RBs in a sub-frame, which we call a *Twin RBs* (TRB). Thus on each sub-channel, a radio frame contains 10 TRBs.

Scheduling Frame and Coherence Time We define a *scheduling frame* (SF) as a consecutive M radio frames. Since a radio frame is 10 ms, the duration of a SF, denoted by T_{SF}, is equal to $10M$ ms.

The maximum number of radio frames that can be packed into a scheduling frame, M, is upper limited by the coherence time of the underlying channel. That is, M should be small enough so that there is no significant change of LTE users' channel conditions (as well as their achievable data rates) over a period of T_{SF}. As an example, consider the 5 GHz spectrum for an indoor deployment scenario. The channel coherence time T_{C} can be calculated by $T_{\text{C}} = \sqrt{\frac{9}{16\pi f_{\text{M}}^2}}$ [17], where $f_{\text{M}} = v/\lambda$ denotes the maximum Doppler shift, v is the user speed, and λ is the carrier wavelength. In an indoor small cell, assuming a user speed of 3 km/h [3], the coherence time on 5 GHz spectrum is $T_{\text{C}} = 30.58$ ms. Therefore, the maximum value M can take is 3 ($\leq 30.58/10$). That is, $T_{\text{SF}} = 30$ ms.

Denote the number of TTIs in a SF by N_{SF}. By definition, we have $N_{\text{SF}} = T_{\text{SF}}/T_0 = 10M$, where T_0 ($= 1$ ms) denotes the duration of a TTI. Under a CSAT "on/off" cycle, we will have integral numbers of TTIs for both "on" and "off" periods. Further, we assume perfect time synchronization so that the boundaries of TTIs and SFs across all channels occupied by LTE are perfectly aligned.

Problem Statement We are interested in addressing the following problem: *Given that a set $\mathcal{K}$ of LTE users are to coexist with Wi-Fi, how do we minimize LTE traffic's adverse impact on Wi-Fi users while meeting each LTE user's UL and DL rate requirements?* To answer this question, we must address the following sub-problems:

- For LTE, since each user has both UL and DL data traffic, we must decide how to use each channel in $\mathcal{F}$. That is, should a channel $i \in \mathcal{F}$ be used for UL or DL transmission?
- For Wi-Fi/LTE coexistence on each channel $i \in \mathcal{F}$, we must decide the durations of "off" and "on" periods

[1] The set $\mathcal{K}$ of LTE users that are offloaded to unlicensed bands is determined by the carrier load balancing function defined in the LTE specification [14]. How this function works to determine $\mathcal{K}$ is out of the scope of this paper.

within each SF. The "on" period directly translates into adverse impact on Wi-Fi users. Our objective is to divide "off" and "on" periods on each channel optimally to minimize such adverse impact across all channels.

- To meet each LTE user's UL and DL rate requirements, we need to allocate TRBs on each sub-channel to users. A user's rate requirements can be fulfilled by allocating TRBs from multiple channels. This is not trivial because the achievable data rate of a user varies on different sub-channels, due to frequency-selective channel fading.
- Last but perhaps most significant is that we are interested in a *real-time* scheduling algorithm. By real-time, we mean that the LTE scheduling solution must be found within the "off" periods of the SFs (more precisely, the smallest "off" period across all channels in $\mathcal{F}$). This will ensure that the centralized LTE scheduler can convey the optimal scheduling result to all users via control signaling so that LTE users can follow the pre-computed, optimal traffic pattern on all channels. Given that T_{SF} is typically several 10s of ms and optimal "off" periods may be less than 10 ms, the eNB must determine the scheduling solution within a few ms. In this paper, we aim for 1 ms for our scheduling solution.

IV. Mathematical Modeling

In this section, we develop a mathematical model for the U-LTE resource scheduling problem.

UL/DL Channel Assignment Referring to Fig. 2, consider the set of channels $\mathcal{F}$ where each channel is shared among LTE and Wi-Fi nodes. For LTE, denote I_i^{UL} and I_i^{DL} as binary variables to indicate whether channel $i \in \mathcal{F}$ is used for UL and DL transmissions, respectively, i.e.,

$$I_i^{\text{UL}} = \begin{cases} 1, & \text{if channel } i \in \mathcal{F} \text{ is selected for UL;} \\ 0, & \text{otherwise.} \end{cases}$$

$$I_i^{\text{DL}} = \begin{cases} 1, & \text{if channel } i \in \mathcal{F} \text{ is selected for DL;} \\ 0, & \text{otherwise.} \end{cases}$$

Since each channel can only be used by LTE for either UL or DL transmission, but not both, we have:

$$I_i^{\text{UL}} + I_i^{\text{DL}} \le 1 \quad (i \in \mathcal{F}). \tag{1}$$

Effective Occupancy by LTE on A Channel Referring to Fig. 2, for each channel $i \in \mathcal{F}$, there is a set of sub-channels $\mathcal{S}_i$. Scheduling for LTE is performed on sub-channel level. On each sub-channel of channel i, LTE's transmission time (either UL or DL) may not terminate at the same time. Denote $(i,j) \in \mathcal{S}_i$ as sub-channel j on channel i. Then as far as Wi-Fi is concerned, channel i is available only if LTE ceases transmissions on *all* sub-channels.

To model this effective channel occupancy by LTE, denote $n_{(i,j)}^{k,\text{UL}}$ and $n_{(i,j)}^{k,\text{DL}}$ as the number of TRBs on sub-channel $(i,j) \in \mathcal{S}_i$ within a SF that are allocated to user $k \in \mathcal{K}$ for UL and DL transmissions, respectively. If channel i is selected for UL transmission (i.e., $I_i^{\text{UL}} = 1$), then LTE's usage of TTIs on sub-channel (i,j) across all users in $\mathcal{K}$ is $\sum_{k\in\mathcal{K}} n_{(i,j)}^{k,\text{UL}}$. Denote $n_{i,\max}^{\text{UL}}$ as the effective channel occupancy by LTE on channel i across all $|\mathcal{S}_i|$ sub-channels. Then we have:

$$n_{i,\max}^{\text{UL}} = \max_{j\in\mathcal{S}_i} \sum_{k\in\mathcal{K}} n_{(i,j)}^{k,\text{UL}} \quad (i \in \mathcal{F}). \tag{2}$$

Likewise, if channel i is selected for DL transmission (i.e., $I_i^{\text{DL}} = 1$), then LTE's usage of TTIs on sub-channel (i,j) across all users in $\mathcal{K}$ is $\sum_{k\in\mathcal{K}} n_{(i,j)}^{k,\text{DL}}$. Denote $n_{i,\max}^{\text{DL}}$ as the effective channel occupancy by LTE on channel i. We have:

$$n_{i,\max}^{\text{DL}} = \max_{j\in\mathcal{S}_i} \sum_{k\in\mathcal{K}} n_{(i,j)}^{k,\text{DL}} \quad (i \in \mathcal{F}). \tag{3}$$

Within a SF on channel i, the usable time duration (in number of TTIs) for LTE is determined by $n_{i,\max}^{\text{UL}}$ (if $I_i^{\text{UL}} = 1$) or $n_{i,\max}^{\text{DL}}$ (if $I_i^{\text{DL}} = 1$). While the time duration left for Wi-Fi is $N_{\text{SF}} - n_{i,\max}^{\text{UL}}$ (for UL) or $N_{\text{SF}} - n_{i,\max}^{\text{DL}}$ (for DL) TTIs.

Upper Bound on LTE Usage To ensure that LTE does not monopolize radio resource of each channel $i \in \mathcal{F}$, it is important to set up an upper bound on LTE's transmission time on each channel [3]. Let Q_i ($Q_i < N_{\text{SF}}$) denote the upper bound on the number of TTIs that LTE can use for UL or DL transmission on channel i within a SF. Then

$$n_{i,\max}^{\text{UL}} \le I_i^{\text{UL}} Q_i \quad (i \in \mathcal{F}), \tag{4}$$

$$n_{i,\max}^{\text{DL}} \le I_i^{\text{DL}} Q_i \quad (i \in \mathcal{F}). \tag{5}$$

The setting of Q_i typically depends on some fairness criteria.[2] In this paper, we assume that Q_i is given *a priori* by the network operator. The discussion of specific criterion to set Q_i is beyond the scope of this paper. Note that how Q_i is determined will not affect our scheduling solution in Section V.

Meeting LTE User Rate Requirement For each user $k \in \mathcal{K}$, to ensure both of its UL and DL rate requirements are met, we have the following constraints:

$$R^{k,\text{UL}} \le \frac{\sum_{i\in\mathcal{F}} \sum_{j\in\mathcal{S}_i} n_{(i,j)}^{k,\text{UL}} C_{(i,j)}^{k,\text{UL}} T_0}{T_{\text{SF}}} \quad (k \in \mathcal{K}), \tag{6}$$

$$R^{k,\text{DL}} \le \frac{\sum_{i\in\mathcal{F}} \sum_{j\in\mathcal{S}_i} n_{(i,j)}^{k,\text{DL}} C_{(i,j)}^{k,\text{DL}} T_0}{T_{\text{SF}}} \quad (k \in \mathcal{K}), \tag{7}$$

where $C_{(i,j)}^{k,\text{UL}}$ and $C_{(i,j)}^{k,\text{DL}}$ are UL and DL achievable data rates for user k on sub-channel (i,j), respectively. The achievable data rates can be modeled by the Shannon's equation, i.e., $C_{(i,j)}^{k,\text{UL}} = B\log_2\left(1 + \frac{\rho^{\text{UL}} l_i^k |h_{(i,j)}^k|^2}{\sigma_0^2}\right)$ and $C_{(i,j)}^{k,\text{DL}} = B\log_2\left(1 + \frac{\rho^{\text{DL}} l_i^k |h_{(i,j)}^k|^2}{\sigma_0^2}\right)$, where B is the bandwidth of a sub-channel, ρ^{UL} and ρ^{DL} are an LTE user's UL transmit power and eNB's DL transmit power on each sub-channel, respectively, l_i^k is the pathloss between the eNB and user $k \in \mathcal{K}$ on channel $i \in \mathcal{F}$, $h_{(i,j)}^k$ is the Rayleigh fading coefficient between the eNB and user $k \in \mathcal{K}$ on sub-channel

[2] An example criterion for setting Q_i is to ensure that the impact of U-LTE on Wi-Fi on a channel does not exceed that of another Wi-Fi network with the same traffic load under LTE [3], [6].

(i, j), and σ_0^2 denotes the noise power. The pathloss is dependent on the carrier frequency and is modeled by the Friis transmission equation $l_i^k = G_t G_r \left(\frac{\lambda_i}{4\pi D_k}\right)^2$, where G_t and G_r are respectively the transmit and receive antenna gains, λ_i is the wavelength on channel i, and D_k denotes the distance between the eNB and user k.

In practice, $C_{(i,j)}^{k,\text{UL}}$'s and $C_{(i,j)}^{k,\text{DL}}$'s can be estimated from LTE's channel state information (CSI) report [5] and remain constant during a SF (by the definition of SF in Section III).

Objective Function and Problem Formulation In a SF, the transmission time of LTE on channel $i \in \mathcal{F}$ is determined by $I_i^{\text{UL}} \cdot n_{i,\max}^{\text{UL}} + I_i^{\text{DL}} \cdot n_{i,\max}^{\text{DL}}$. For the same transmission time duration, LTE's impact on Wi-Fi depends on the traffic load of Wi-Fi. The heavier traffic that is being served by Wi-Fi on the same channel, the greater the impact of LTE on Wi-Fi. To take this into consideration, we introduce a weight parameter w_i to reflect Wi-Fi's traffic load on channel $i \in \mathcal{F}$. A simple example is to set w_i to the number of Wi-Fi nodes on channel i, i.e., $w_i = U_i$. To find U_i on each channel, the eNB can monitor Wi-Fi's channel usage during "off" periods. For example, with the methods proposed in [15] and [16], U_i can be determined online based on the proportion of observed busy time slots. This is feasible since an U-LTE eNB is expected to be able to perform carrier sensing [1]. For a given w_i, the impact of LTE on Wi-Fi on channel i can be quantitatively measured by $w_i(I_i^{\text{UL}} \cdot n_{i,\max}^{\text{UL}} + I_i^{\text{DL}} \cdot n_{i,\max}^{\text{DL}})$. Since LTE's impact on Wi-Fi varies from channel to channel, a plausible objective for the network operator is to minimize the maximum of LTE's impact across all channels. This is the objective that we use in this paper. That is, $\min \max_{i\in\mathcal{F}} w_i(I_i^{\text{UL}} \cdot n_{i,\max}^{\text{UL}} + I_i^{\text{DL}} \cdot n_{i,\max}^{\text{DL}})$.

Our optimization problem is formally stated as follows:

OPT

minimize $\max_{i\in\mathcal{F}} w_i(I_i^{\text{UL}} \cdot n_{i,\max}^{\text{UL}} + I_i^{\text{DL}} \cdot n_{i,\max}^{\text{DL}})$

subject to UL/DL channel assignment: (1),
Effective channel occupancy by LTE: (2), (3),
Upper bound on LTE usage: (4), (5),
Per-user rate requirement: (6), (7),
$n_{i,\max}^{\text{UL}}, n_{i,\max}^{\text{DL}}, n_{(i,j)}^{k,\text{UL}}, n_{(i,j)}^{k,\text{DL}} \in \mathbb{N}$,
$I_i^{\text{UL}}, I_i^{\text{DL}} \in \{0, 1\} \quad (i \in \mathcal{F}, j \in \mathcal{S}_i, k \in \mathcal{K})$.

Reformulation In Problem OPT, since the objective function involves integer variables and two levels of max functions (due to (2) and (3)), it needs to be reformulated. In particular, in the presence of constraints (4) and (5), the objective function can be simplified to $\max_{i\in\mathcal{F}} w_i(n_{i,\max}^{\text{UL}} + n_{i,\max}^{\text{DL}})$. To remove these two levels of max functions, we define $z = \max_{i\in\mathcal{F}} w_i \left(n_{i,\max}^{\text{UL}} + n_{i,\max}^{\text{DL}}\right)$ as the new objective function. Then we have the following constraint:

$$z \geq w_i \left(n_{i,\max}^{\text{UL}} + n_{i,\max}^{\text{DL}}\right) \quad (i \in \mathcal{F}). \tag{8}$$

By constraint (1), at most one of the two terms, $n_{i,\max}^{\text{UL}}$ and $n_{i,\max}^{\text{DL}}$, can be nonzero. Then (8) can be reformulated to $z \geq w_i \cdot n_{i,\max}^{\text{UL}}$ and $z \geq w_i \cdot n_{i,\max}^{\text{DL}}$ for $i \in \mathcal{F}$. By definitions of $n_{i,\max}^{\text{UL}}$ and $n_{i,\max}^{\text{DL}}$ in (2) and (3), we have:

$$n_{i,\max}^{\text{UL}} \geq \sum_{k\in\mathcal{K}} n_{(i,j)}^{k,\text{UL}} \quad (i \in \mathcal{F}, j \in \mathcal{S}_i),$$

$$n_{i,\max}^{\text{DL}} \geq \sum_{k\in\mathcal{K}} n_{(i,j)}^{k,\text{DL}} \quad (i \in \mathcal{F}, j \in \mathcal{S}_i).$$

Therefore, we have the following constraints on z:

$$z \geq w_i \sum_{k\in\mathcal{K}} n_{(i,j)}^{k,\text{UL}} \quad (i \in \mathcal{F}, j \in \mathcal{S}_i), \tag{9}$$

$$z \geq w_i \sum_{k\in\mathcal{K}} n_{(i,j)}^{k,\text{DL}} \quad (i \in \mathcal{F}, j \in \mathcal{S}_i). \tag{10}$$

The constraints in (2), (3), (4) and (5) can be simplified by eliminating $n_{i,\max}^{\text{UL}}$ and $n_{i,\max}^{\text{DL}}$ and removing the max functions. We have:

$$\sum_{k\in\mathcal{K}} n_{(i,j)}^{k,\text{UL}} \leq I_i^{\text{UL}} Q_i \quad (i \in \mathcal{F}, j \in \mathcal{S}_i), \tag{11}$$

$$\sum_{k\in\mathcal{K}} n_{(i,j)}^{k,\text{DL}} \leq I_i^{\text{DL}} Q_i \quad (i \in \mathcal{F}, j \in \mathcal{S}_i). \tag{12}$$

Finally we have the reformulated optimization problem:

OPT-R

minimize z

subject to Adverse impact of LTE on Wi-Fi: (9), (10),
UL/DL channel assignment: (1),
Upper bound on LTE usage: (11), (12),
Per-user rate requirement: (6), (7),
$z \geq 0,\ n_{(i,j)}^{k,\text{UL}}, n_{(i,j)}^{k,\text{DL}} \in \mathbb{N}$,
$I_i^{\text{UL}}, I_i^{\text{DL}} \in \{0, 1\} \quad (i \in \mathcal{F}, j \in \mathcal{S}_i, k \in \mathcal{K})$.

Problem OPT-R is a mixed integer linear program (MILP). It can be shown that this problem is NP-complete based on a reduction from the partition problem. The proof is omitted due to space limit. The numbers of variables and constraints are both on the order of $O(|\mathcal{F}||\mathcal{S}_i||\mathcal{K}|)$. For such a problem size, a solver based on branch-and-bound (BB) [18] (and perhaps augmented with cutting planes, a.k.a., BB-CP) cannot offer a solution in ~ 1 ms time scale (refer to Section VI).

V. CURT – A Real-Time Scheduling Algorithm

Before presenting the design of CURT, we first explain why conventional methods fail to meet our target of providing near-optimal solution to OPT-R in real-time. We discuss the following approaches: (i) reusing LTE schedulers designed for licensed bands, (ii) solving linear programming (LP) relaxation of OPT-R and rounding up the solution to integers, and (iii) using an exact algorithm such as BB to solve OPT-R directly.

Existing LTE schedulers used on licensed bands, although meeting the real-time requirement, cannot be readily extended to solve the scheduling optimization problem on unlicensed bands. These schedulers are simple metric-based algorithms that allocate TRBs on a per-TTI basis [19]. In every TTI, each

TRB (or group of TRBs) is allocated to the user that has the highest metric (e.g., achievable rate, rate requirement, delay, or fairness index) on it. On the other hand, on unlicensed band, a U-LTE scheduler needs to decide the air time division of each SF for LTE and Wi-Fi on different channels. Such time division captures how severely U-LTE impacts Wi-Fi on each channel. Existing LTE schedulers cannot offer optimal or near-optimal time division solution across multiple channels since they are designed for per-TTI resource allocation.

Standard optimization techniques such as LP relaxation and BB cannot meet the real-time requirement of $\sim$1 ms. Although an LP relaxation of OPT-R can be solved efficiently by Simplex or interior-point methods, the computation time is much larger than 1 ms (as will be shown in Section VI). BB can be used to solve an MILP such as OPT-R. The basic idea of a BB-based approach is to find upper and lower bounds of each sub-problem as well as the global upper and lower bounds across all sub-problems at each iteration. The gap between upper and lower bounds is expected to shrink after each iteration until it is within the desired optimality gap. The main problem with such an approach is that the overall running time to close the gap is usually very long.

A. Our Proposed Approach

Different from existing LTE schedulers used on licensed bands, CURT jointly addresses time division between U-LTE and Wi-Fi across multiple channels and TRB allocation within LTE "on" periods, with the target of getting optimal or near-optimal solution to OPT-R in real-time ($\sim$1 ms). The design of CURT is based on problem decomposition and parallel execution of sub-problems on a massive number of GPU cores. To pursue near-optimality, CURT first decomposes OPT-R into a large number of sub-problems, each with a fixed assignment of UL/DL channels and "off/on" time division pattern across all channels, and then solves all sub-problems in parallel by GPU cores. In particular, our proposed decomposition ensures that there is no inter-dependency among sub-problems, so that all sub-problems can be executed concurrently. Thus the time complexity of solving OPT-R is reduced to that of solving one sub-problem. Further, the sub-problems are of the same structure, which allows them to be solved within a close-to-same amount of time. The parallel design of CURT ensures that all possible cases of U-LTE's impact on Wi-Fi can be evaluated (for feasibility) in parallel. By comparing among the objectives of all feasible solutions, CURT has a high probability to find a near-optimal solution.

B. Problem Decomposition

Figure 3 shows how we decompose OPT-R into smaller independent sub-problems. There are three levels of decomposition: 1) fixing UL and DL channel assignments, 2) fixing achievable objective values, and 3) separating UL and DL sub-problems. Details of these steps are as follows.

Fixing UL and DL Channel Assignments Given a set of channels $\mathcal{F}$ for LTE/Wi-Fi coexistence, there is a total of $\binom{|\mathcal{F}|}{1} + \binom{|\mathcal{F}|}{2} + \cdots + \binom{|\mathcal{F}|}{|\mathcal{F}|-1} = 2^{|\mathcal{F}|} - 2$ different ways

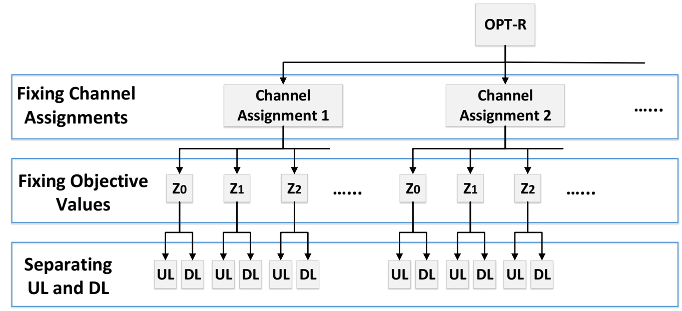


Fig. 3: Decomposition of OPT-R.

for assigning the channels for UL and DL transmissions. In practice, $|\mathcal{F}|$ is not a large number due to limitations on signal processing capability at U-LTE eNBs (small cell access points) and users. For example, when $|\mathcal{F}| = 5$, there is a total of 30 different channel assignments. Problem OPT-R can thus be decomposed into a set of $(2^{|\mathcal{F}|} - 2)$ sub-problems, each with a given UL and DL channel assignment.

Fixing Achievable Objective Values From constraints (9) and (10), it is clear that the range of the optimal objective value of OPT-R under a given UL/DL channel assignment contains only a finite number of values. Specifically, since both $w_i \sum_{k \in \mathcal{K}} n_{(i,j)}^{k,\text{UL}}$ and $w_i \sum_{k \in \mathcal{K}} n_{(i,j)}^{k,\text{DL}}$ (for all $j \in \mathcal{S}_i$ on channel $i \in \mathcal{F}$) can only take values from $\mathcal{Z}_i = \{0, w_i, 2w_i, \cdots, Q_i w_i\}$, the optimal objective value z^* must be within the set

$$\mathcal{Z} = \bigcup_{i \in \mathcal{F}} \mathcal{Z}_i. \quad (13)$$

With fixed objective values, each sub-problem under a specific channel assignment can be further split into $|\mathcal{Z}|$ sub-problems. Since each set $\mathcal{Z}_i$ ($i \in \mathcal{F}$) consists of 0 and Q_i nonzero elements (if $Q_i > 0$), we have

$$|\mathcal{Z}| \leq 1 + \sum_{i \in \mathcal{F}} Q_i \leq |\mathcal{F}| \cdot N_{\text{SF}}, \quad (14)$$

where the second inequality follows from the definition $Q_i < N_{\text{SF}}$ for all $i \in \mathcal{F}$.

After this decomposition, the $|\mathcal{Z}|$ sub-problems under a given UL/DL channel assignment include all possible "off/on" time division patterns across channels in $\mathcal{F}$.

Separating UL/DL Sub-Problems For each sub-problem under a given channel assignment and objective value, we need to check whether or not it is feasible to meet all users' rate requirements. This feasibility check can again be decomposed into two parallel problems, one for UL and the other for DL.

Now the original problem OPT-R is decomposed into a total of $2(2^{|\mathcal{F}|}-2) \cdot |\mathcal{Z}|$ UL/DL feasibility check sub-problems (each with a given channel assignment and objective value). These sub-problems are independent and can be solved in parallel by massive GPU cores. Once the feasibility checks for all UL/DL sub-problems are completed, we pick the smallest feasible objective value under all channel assignments that has both its corresponding UL and DL sub-problems pass the feasibility checks. This is our output solution.

Next, we present the design of a feasibility check algorithm for each sub-problem.

C. Feasibility Check of Sub-Problems

In the feasibility check of a UL/DL sub-problem, we aim to determine whether or not each user's UL/DL rate requirement can be met under the given UL/DL channel assignment. This problem can be formulated into an integer linear program (ILP), which is NP-hard in general and cannot be solved exactly under our tight time constraint ($\sim$1 ms). So a fast and efficient heuristic algorithm is needed.

We now present our proposed feasibility check algorithm. Since the objective value is pre-determined in each UL/DL sub-problem, the maximum duration of "on" period on each channel as well as the number of usable TRBs on each sub-channel are fixed. Our TRB allocation scheme is to iteratively schedule usable TRBs on each sub-channel to users in $\mathcal{K}$ based on certain user selection criterion. The criterion we choose is a user's UL/DL achievable rate on a sub-channel weighted (multiplied) by its remaining data volume (bits) that needs to be filled for meeting its UL/DL rate requirement. Specifically, on each sub-channel, we select the user with the maximum weighted rate and allocate TRBs to it until its remaining data volume is zero. A user will not be considered again for TRB allocation once its rate requirement is met. If there is still TRBs left on this sub-channel after the allocation, we move on to the next user. After allocating all TRBs, we move on to the next sub-channel. If all users' rate requirements are met during or after the TRB allocation, we conclude that the given UL/DL sub-problem is feasible (infeasible otherwise).

D. Computational Complexity

The time complexity of CURT is determined by the feasibility check of a DL/UL sub-problem, while its space complexity is determined by the number of DL/UL sub-problems. For each DL/UL feasibility check we need to go through at most $|\mathcal{F}||\mathcal{S}_i|$ sub-channels. On each sub-channel, user selection is on the order of $O(|K|)$. So the time complexity of a feasibility check is $O(|\mathcal{F}||\mathcal{S}_i||\mathcal{K}|)$. For space complexity, we need to determine how many processors are needed for parallel feasibility checks. From our analysis in Section V-B, we know that the total number of parallel UL/DL sub-problems is $2(2^{|\mathcal{F}|}-2)|\mathcal{Z}|$. Based on (14), it is upper bounded by $2(2^{|\mathcal{F}|}-2)|\mathcal{F}|N_{\text{SF}}$, which is independent of the number of LTE and Wi-Fi users. That is, the number of GPU cores needed by CURT does not increase with the number of LTE/Wi-Fi nodes in the network.

E. Guaranteeing Feasibility via Traffic Management

Problem OPT (and OPT-R) may have no feasible solution to meet the constraints on rate requirements $R^{k,\text{UL}}$'s and $R^{k,\text{DL}}$'s for all $k \in \mathcal{K}$ and U-LTE's channel usage Q_i's for all $i \in \mathcal{F}$. On the other hand, CURT may not be able to find a feasible solution when OPT-R is indeed feasible since CURT does not traverse the entire search space of OPT-R. In fact, it is impossible for any algorithm to go through all possible solutions of an NP-hard problem such as OPT-R while meeting the real-time requirement of $\sim$1 ms. To guarantee feasibility, CURT should work in concert with a traffic management mechanism. Specifically, if a user's UL/DL rate requirements cannot be fully met after scheduling a SF, the traffic management module may negotiate with this user to decrease its rate requirements or switch it to licensed band. Detailed design of the traffic management module is beyond the scope of this paper.

VI. Experimental Results

In this section, we investigate the performance of CURT.

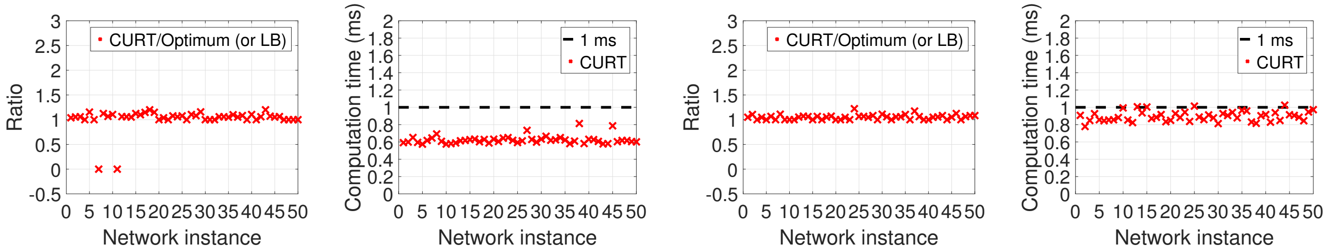
Experiment Platform As a proof of concept, we implement CURT with Nvidia's CUDA architecture [20]. CUDA is a GPU-based general-purpose parallel computing platform. Logically, CUDA performs a computation task (termed a kernel) through a threads hierarchy. This hierarchy has a two-layer structure. The upper layer is a grid of thread blocks. Given that the original task has been decomposed into a large set of sub-tasks, each sub-task can be assigned to a thread block to solve. At the lower layer, each thread block expands into a grid of threads. A thread is the finest computing granularity under CUDA. Threads within a thread block can be used to process parallel computations involved in a sub-task.

Under the CUDA architecture, all $|\mathcal{Z}|$ UL or DL feasibility checks under a specific channel assignment are solved by a thread block. Thus the total number of thread blocks is $2(2^{|\mathcal{F}|}-2)$. Within a thread block, each feasibility check is done independently by a thread.

Our implementation is done on a Dell desktop computer with an Intel Xeon E5-2687W v4 CPU (3.0 GHz) and dual Nvidia Quadro P6000 GPUs (each with 30 SMs and 3840 CUDA cores). We use IBM CPLEX Optimizer (version 12.7.1) on the same computer to compute optimal or lower-bound solution to OPT-R and solve LP relaxation of OPT-R. To address potentially prohibitively long computation time by CPLEX, we set a time limit of 1 hour. The lower-bound solution is taken as benchmark when CPLEX cannot find optimal solution by 1 hour.

Network Setting Assume that all LTE and Wi-Fi nodes in the small cell are within each other's transmission and interference ranges and there is no hidden-node. Suppose that $|\mathcal{F}|=5$ channels in the 5 GHz unlicensed spectrum are chosen for U-LTE/Wi-Fi coexistence, with carrier frequencies being 5.20, 5.22, 5.24, 5.26, and 5.28 GHz, respectively. Each channel has 20 MHz bandwidth and is divided into $|\mathcal{S}_i|=100$ sub-channels (each with 180 kHz bandwidth). The time duration of a SF $T_{\text{SF}}=30$ ms. The fast fading coefficients are randomly generated by Rayleigh fading model with mean 0 and variance 1. Antenna gains are set as $G_t=G_r=1$. Considering that the transmit power of eNB is typically higher than that of user terminals, we set $\rho^{\text{DL}}/\sigma_0^2=120$ dB and $\rho^{\text{UL}}/\sigma_0^2=115$ dB. For Q_i, we let $Q_i=\left\lfloor N_{\text{SF}}\frac{|\mathcal{K}|/|\mathcal{F}|}{|\mathcal{K}|/|\mathcal{F}|+U_i}\right\rfloor$.

Results Following 3GPP's evaluation methodology [3], we consider a maximum of $|\mathcal{K}|=20$ users in a U-LTE small cell. Results under $|\mathcal{K}|=10$ and $|\mathcal{K}|=20$ are shown in Fig.



(a) Ratios of objectives. $|\mathcal{K}| = 10$. (b) Computation time. $|\mathcal{K}| = 10$. (c) Ratios of objectives. $|\mathcal{K}| = 20$. (d) Computation time. $|\mathcal{K}| = 20$.

Fig. 4: Performance of CURT over 50 network instances. Objective value of 0 indicates infeasibility.

4. In both cases, distances between eNB and LTE users are randomly and uniformly generated from $[1, 30]$ m. As shown in Section V-D, the computational complexity of CURT is independent of the number of Wi-Fi users. Without loss of generality, we assume that U_i on each channel $i \in \mathcal{F}$ is randomly chosen from $\{1, 2, 3\}$ so that on average there are around 10 Wi-Fi users sharing the spectrum with U-LTE [3].

In Fig. 4(a) and 4(b), UL and DL rate requirements of the 10 users are randomly generated from $[10, 40]$ Mb/s. Fig. 4(a) shows ratios between objective values achieved by CURT and optimal (or lower-bound) solutions found by CPLEX over 50 network instances. For each instance, if CURT finds the optimum, the ratio of objective values equals to one. When CURT fails to find a feasible solution, we set the ratio to zero. Among 50 network instances, CURT finds optimal solutions for 14 instances. We observe that in 2 instances CURT cannot find feasible solution.[3] That is, the percentage that CURT can find a feasible solution is 96%. This is reasonable since CURT does not traverse the entire search space of OPT-R. Among the instances where CURT can find feasible solutions, the average of CURT's ratios is 1.04, with a variance of 0.0021. Fig. 4(b) shows computation time of CURT. Mean computation time of CURT is 0.62 ms, with a maximum of 0.81 ms and a variance of 0.0022. In contrast, the mean of CPLEX's computation time for finding the optimal (or lower-bound) solution is 1246.35 s. As another comparison, we employ CPLEX to solve LP relaxation of OPT-R (refer to Section V). The mean of CPLEX's computation time for solving LP relaxation is 89.81 s.

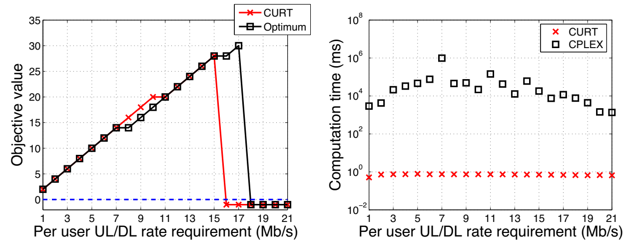
In Fig. 4(c) and 4(d), the rate requirements of the 20 users are randomly generated from $[5, 20]$ Mb/s. Fig. 4(c) shows that in 18 out of 50 instances, CURT achieves optimum. Also, for all 50 instances, CURT is able to find feasible solutions. The average of ratios by CURT (over optimum) is 1.04, with a variance of 0.0014. Mean computation time of CURT is 0.90 ms, with a maximum of 1.02 ms and a variance of 0.0034. In contrast, the mean of CPLEX's computation time for finding optimal or lower-bound solution is 987.86 s. The mean of CPLEX's computation time for solving LP relaxation is 48.62 s. Numerical results in Fig. 4 demonstrate that CURT can indeed deliver near-optimal solution while meeting the real-time requirement of $\sim$1 ms.

As shown in Section V-D, CURT's complexity and computation time depend on the number of LTE users $|\mathcal{K}|$. For a worst-case study, consider an extremely dense network setting with $|\mathcal{K}| = 50$, $U_i = 10$ for all channels, and UL/DL rate requirement of 5 Mb/s for all users. Under such setting, there is a total of 100 LTE and Wi-Fi users sharing the same spectrum. Results of 1000 experiments for CURT show that the mean computation time of CURT is 3.06 ms, with a maximum of 3.22 ms and a variance of 0.0016. That is, CURT can find a solution in about $\sim 10\%$ duration of a SF (with the "off" period), which still meets the real-time requirement based on our discussion in Section III.

Next, we conduct experiments using representative LTE schedulers designed for licensed bands for comparison with CURT. The *maximum throughput (MT)* and *throughput to average (TTA)* schedulers [19] are considered. MT allocates each TRB to the user that has the highest achievable rate on it. The allocation metric of TTA is a user's per-RB achievable rate divided by its average rate across all UL (DL) channels. For both schedulers, the ratio between numbers of UL and DL channels is chosen to be the closest to $\sum_k R^{k,\text{UL}} / \sum_k R^{k,\text{DL}}$, with UL and DL channels randomly assigned. In each SF, TRBs from all channels are allocated on a per-TTI basis (subject to the restrictions Q_i's). A user is no longer considered for TRB allocation after its rate requirement is met. We run experiments under $|\mathcal{K}|$ =10 and 20, with scenario settings being the same as those for Fig. 4. Results show that under $|\mathcal{K}| = 10$, the means of ratios between the achieved objective and optimum for MT and TTA are 1.70 and 1.73, respectively; Under $|\mathcal{K}| = 20$, the means of ratios for MT and TTA are 1.91 and 1.84, respectively. Clearly, the performance of both MT and TTA is far from optimum and that of CURT (with a mean of 1.04 for both $|\mathcal{K}|$ =10 and 20), which indicates that LTE schedulers designed for licensed bands cannot be reused for coexistence with Wi-Fi on unlicensed bands.

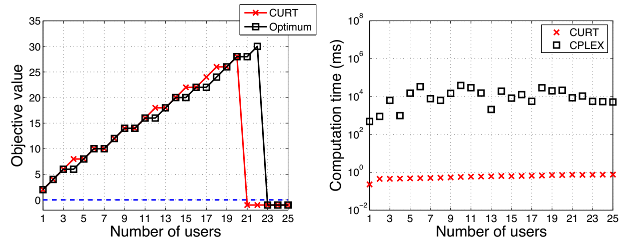
We now evaluate the behavior of CURT under varying LTE traffic load. Suppose the cell has 30 users. To identify each user distinctly, we name them as user 1 to 30, with their distances to the eNB randomly generated as follows (in meter): 24.58, 1.29, 5.03, 6.88, 6.76, 18.51, 8.89, 6.77, 1.44, 22.66, 13.91, 28.02, 14.51, 13.14, 25.54, 16.23, 6.88, 20.49, 25.31, 1.57,

[3]In this case, the traffic management module may be invoked to ensure feasibility.



(a) Objective values. (b) Computation time.

Fig. 5: Performance of CURT under increasing per-user rate requirement. Objective value of -1 indicates infeasibility.



(a) Objective values. (b) Computation time.

Fig. 6: Performance of CURT under increasing number of users. Objective value of -1 indicates infeasibility.

20.76, 12.01, 25.12, 15.58, 21.57, 13.44, 9.83, 6.50, 6.61, 20.78. Note that the U-LTE cell may not be able to meet rate requirements of all these users. It may only serve a subset of users and transfer the remaining users to licensed band (via the traffic management module) as described in Section V-E. We set $Q_i = N_{\text{SF}}/2$ and $U_i = 2$ for all channels in $\mathcal{F}$.

In Fig. 5(a), we choose the first 20 users and increase their UL/DL rate requirements simultaneously from 1 Mb/s. We see that CURT can support a maximum per-user UL/DL rate requirement of 15 Mb/s, while the optimal solution can support up to 17 Mb/s. On the other hand, in Fig. 5(b), we see that CURT's computation time is consistently less than 1 ms while CPLEX's computation time varies from 1.36 s to 984.33 s.

In Fig. 6(a), we fix the per-user UL/DL rate requirements to 15 Mb/s and increase the number of LTE users (starting from user 1). It shows that CURT can satisfy the first 20 users, while the optimal solution can support the first 22 users. In Fig. 6(b) we see that computation time of CURT is no greater than 1 ms while CPLEX's computation time varies from 484 ms to 38.47 s.

VII. Conclusions

In this paper, we studied scheduling problem for coexistence between LTE and Wi-Fi under CSAT in unlicensed bands. We formulated the scheduling problem into an optimization problem, which involves assigning channels for UL and DL, dividing each channel's air time into "off" and "on" periods, scheduling TRBs on a large number of sub-channels based on LTE users' channel conditions and UL/DL rate requirements. The objective is to minimize LTE's adverse impact on Wi-Fi users. We presented CURT, a scheduling algorithm based on GPU platform that is able to obtain optimal (or near-optimal) solution in real-time ($\sim$1 ms). We implemented CURT on Nvidia P6000 GPU/CUDA platform and demonstrated that CURT can deliver optimal (or near-optimal) scheduling solution on $\sim$ 1 ms time scale and meet all of our design objectives.

References

[1] Qualcomm Research, "LTE in unlicensed spectrum: Harmonious coexistence with Wi-Fi." Available: https://www.qualcomm.com/documents/lte-unlicensed-coexistence-whitepaper

[2] LTE-U Forum, "LTE-U CSAT procedure TS version 1.0." Available: http://www.lteuforum.org/uploads/3/5/6/8/3568127/lte-u_forum_lte-u_sdl_csat_procedure_ts_v1.0.pdf

[3] 3GPP TR 36.889 version 13.0.0, "Study on licensed-assisted access to unlicensed spectrum." Available: http://www.3gpp.org/DynaReport/36-series.htm

[4] 3GPP TS 36.211 version 14.4.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation." Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2425

[5] 3GPP TS 36.213 version 14.4.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures." Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2427

[6] Z. Guan and T. Melodia, "CU-LTE: Spectrally-efficient and fair coexistence between LTE and Wi-Fi in unlicensed bands," in *Proc. IEEE INFOCOM*, pp. 1-9, San Francisco, CA, USA, Apr. 2016.

[7] Y. Huang, Y. Chen, Y.T. Hou, W. Lou, and J.H. Reed, "Recent advances of LTE/Wi-Fi coexistence in unlicensed spectrum," *IEEE Network*, vol. 32, no. 2, pp. 107-113, Mar.-Apr. 2018.

[8] C. Cano, D.J. Leith, A. Garcia-Saavedra, and P. Serrano, "Fair coexistence of scheduled and random access wireless networks: Unlicensed LTE/WiFi," *IEEE Trans. Networking,* vol. 25, no. 6, pp. 3267-3281, Dec. 2017.

[9] A. Abdelfattah and N. Malouch, "Modeling and performance analysis of Wi-Fi networks coexisting with LTE-U," in *Proc. IEEE INFOCOM*, pp. 1-9, Atlanta, GA, USA, May. 2017.

[10] A. M. Voicu, L. Simic, and M. Petrova, "Inter-technology coexistence in a spectrum commons: A case study of Wi-Fi and LTE in the 5-GHz unlicensed band," *IEEE J. Sel. Areas Commun.,* vol. 34, no. 11, pp. 3062-3077, Nov. 2016.

[11] Q. Chen, G. Yu, R. Yin, A. Maaref, G. Y. Li, and A. Huang, "Energy efficiency optimization in licensed-assisted access," *IEEE J. Sel. Areas Commun.,* vol. 34, no. 4, pp. 723-734, Apr. 2016.

[12] Q. Chen, G. Yu, and Z. Ding, "Optimizing unlicensed spectrum sharing for LTE-U and WiFi network coexistence," *IEEE J. Sel. Areas Commun.,* vol. 34, no. 10, pp. 2562-2574, Oct. 2016.

[13] "T-Mobile completes nation's first live commercial network test of license assisted access (LAA)." Available: https://newsroom.t-mobile.com/news-and-blogs/lte-u.htm

[14] 3GPP TS 36.300 version 14.4.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (EUTRAN); Overall description; Stage 2." Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2430

[15] G. Bianchi and I. Tinnirello, "Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network," in *Proc. IEEE INFOCOM,* pp. 844-852, San Francisco, CA, USA, Mar. 2003.

[16] A.L. Toledo, T. Vercauteren, and X. Wang, "Adaptive optimization of IEEE 802.11 DCF based on Bayesian estimation of the number of competing terminals," *IEEE Trans. Mobile Computing,* vol. 5, no. 9, pp. 1283-1296, Sep. 2006.

[17] T. S. Rappaport, *Wireless Communications: Princlples and Practice.* Upper Saddle River, NJ: Prentice-Hall, 1996.

[18] Y.T. Hou, Y. Shi, and H.D. Sherali, *Applied optimization methods for wireless networks,* Chapter 5, Cambridge University Press, 2014.

[19] F. Capozzi, G. Piro, L.A. Grieco, G. Bogga, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE Commun. Surveys & Tutorials,* vol. 15, no. 2, pp. 678-700, 2013.

[20] Nvidia, "CUDA Toolkit Documentation v8.0." Available: http://docs.nvidia.com/cuda/index.html#axzz4Qrv7Umfs