BLU: 𝔹lue-printing Interference for Robust 𝕃TE Access in 𝕌nlicensed Spectrum

Ramanujan K Sheshadri[1 *], Karthikeyan Sundaresan[2], Eugene Chai[2], Amir Khojastepour[2], Sampath Rangarajan[2], and Dimitrios Koutsonikolas[1]
[1]University at Buffalo, SUNY. [2]NEC Labs America, Inc.
(ramanuja,dimitrio)@buffalo.edu
(karthiks,eugene,amir,sampath)@nec-labs.com

ABSTRACT

Deploying LTE networks in unlicensed spectrum requires us to move beyond coexistence mechanisms and understand the suitability of LTE's synchronous operation in a spectrum that is governed by asynchronous access principles. Our study reveals a fundamental conflict in LTE uplink access that arises between the scheduled nature of LTE's multi-user transmissions – critical for leveraging the diversity (OFDMA) and multiplexing (multi-user MIMO) gains – and the asynchronous nature of interference on the clients. The result is a significant loss in spectrum utilization and throughput that scales with the number of interfering terminals.

To tackle this critical challenge on the LTE uplink, we propose Blu. Blu transforms today's LTE schedulers into speculative schedulers that leverage *interference diversity* across clients to intelligently *over-schedule* clients on the same spectral resources to prevent this utilization loss. Blu's challenges lie in how to over-schedule appropriate clients on the same resources without paying the penalty of collisions, while containing the exponential overhead incurred in measuring the required interference dependencies between clients. The under-pinning of Blu's design includes a novel mechanism to blue-print the very source of interference on LTE clients along with their dependencies, which allows for a constant, significantly reduced overhead. Blu can be realized in today's LTE base stations. Its realization in an enterprise environment with SDRs (hosting LTE release 10) reveals appreciable gains of 1.5-2x in both utilization and throughput over existing schemes for SISO and MU-MIMO transmissions in unlicensed spectrum.

CCS CONCEPTS

• **Networks** → *Network protocol design*;

KEYWORDS

LTE-LAA, LTE-Unlicensed, WiFi, Interference, Hidden-terminals

CoNEXT '17, December 12–15, 2017, Incheon, Republic of Korea

ACM ISBN 978-1-4503-5422-6/17/12…$15.00
https://doi.org/10.1145/3143361.3143373

1 INTRODUCTION

LTE in Unlicensed Spectrum. Equipping cellular networks with the ability to operate in unlicensed spectrum is a critical first step towards opening them up for innovation. Several efforts are ongoing in the industry to bring LTE to unlicensed spectrum [4, 17, 23, 25]. These can be broadly classified into two categories: license-assisted (LTE-U [17] and LTE-LAA/eLAA [4, 23]), where unlicensed carriers are aggregated with existing licensed carriers, and stand-alone (e.g. MulteFire [25]), where LTE is deployed completely on unlicensed carriers. As expected, while the license-assisted mode is favored by traditional operators, the stand-alone mode is already opening LTE for innovation from green-field providers (Alphabet, Federated Wireless, Comcast, etc.) in newer bands like Citizen Broadband Radio Service (CBRS [9]).

The initial focus in all these efforts is to enable co-existence with other incumbents (e.g. WiFi and other LTE providers) in the unlicensed spectrum. This requires an otherwise always-on, synchronous LTE node to adopt asynchronous access principles of energy sensing and back-off to access the medium. Understandably, such co-existence mechanisms form the first step, where substantial progress has been made in terms of both the standards [4, 9] and research [5, 10, 11, 20, 21, 27, 31]. The next critical step is to understand whether LTE is equipped to operate efficiently in unlicensed spectrum that is governed by asynchronous access principles. This work takes an important step in that direction.

Conflict between Concurrency and Asynchronous Interference. One of the key differences between LTE and WiFi is the synchronous and scheduled nature of LTE transmissions as shown in Fig. 1 (compared to the asynchronous WiFi transmissions). Synchronous transmissions in LTE contribute to increased capacity through multi-user diversity (OFDMA) and spatial multiplexing (multi-user MIMO) gains, especially on the uplink, where it is otherwise challenging to synchronize clients (UEs). However, as we show in Section 2.2, these very same features make it particularly challenging for realizing gains in unlicensed spectrum, where the impact of asynchronous interference (through hidden terminals from WiFi or other LTE nodes) on concurrent transmissions is exacerbated. This reveals a fundamental conflict between the scheduled, multi-user transmissions in LTE and asynchronous access in unlicensed spectrum.

*Ramanujan conducted this work during his internship at NEC Labs America, Inc., Princeton, NJ

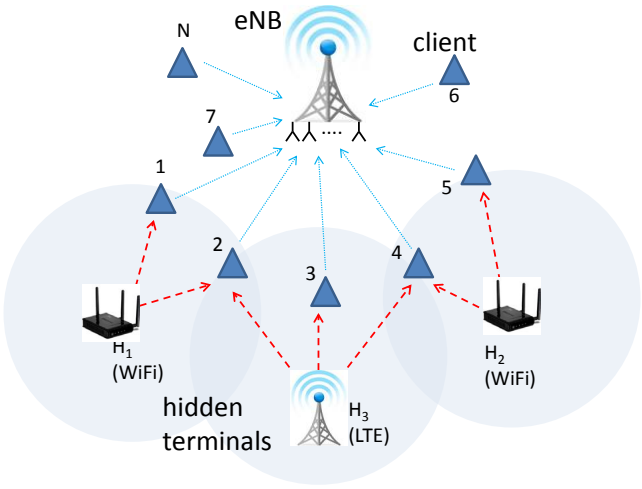
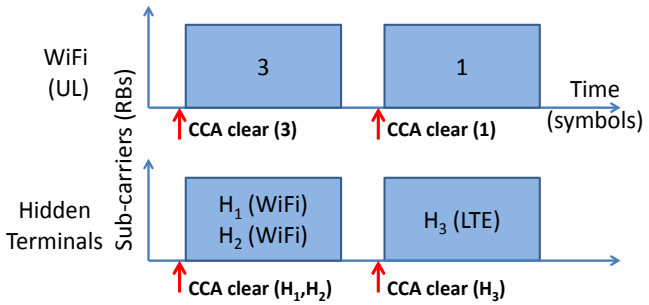
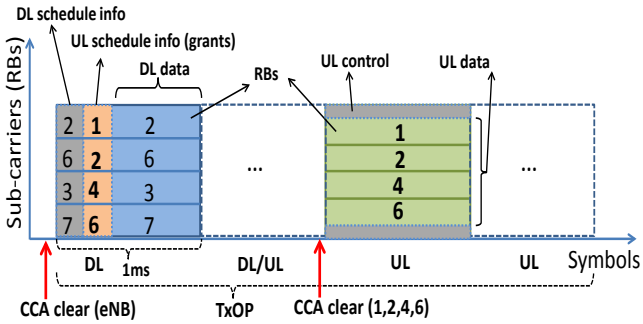


Figure 1: LTE uplink in unlicensed spectrum affected by hidden terminals (Topology used for ensuing examples).



(a) WiFi (if both eNB & clients are WiFi): Individual WiFi clients 3 and 1, access UL channels independently and counter the effects of hidden terminals H1, H2, H3.



(b) LTE: eNB uses CCA/backoff; schedules DL+UL; clients do CCA; TxOP (2-10 ms with flexible DL/UL split).

Figure 2: WiFi vs. LTE in unlicensed spectrum.

Pronounced Impact on Uplink Access: One of the manifestations of this interference is collisions (similar to WiFi) in both LTE's downlink (DL) and uplink (UL). Our prior work, Ultron [10] alleviated such collisions by homogenizing channel sensing between LTE and WiFi. However, LTE's UL access faces an additional, unique layer of impact, which remains un-resolved. Contrary to WiFi, where individual clients asynchronously access the channel on the uplink based on their channel availability (Fig. 2a), it is the eNB that schedules clients on the UL through grants (sent on DL, Fig. 2b). When multiple clients are jointly scheduled (on different frequency resources) to leverage OFDMA in the same UL transmission, or on the same frequency resource to leverage multiple antennas at the eNB (through MU-MIMO), interfering sources (hidden terminals to eNB) in the vicinity of the clients will prevent clients from utilizing their allocated grants, resulting in wastage of spectrum resources. The impact of this wastage (un-utilized resources) is exacerbated when multiple clients are jointly scheduled, where some are able to utilize the grants, while others are not (e.g. Fig. 3). This problem, unique to leveraging higher efficiencies from scheduled, multi-user transmissions in LTE, amplifies the impact of hidden terminals on UL access. As we show in Section 2, the resulting under-utilization grows with the number of hidden terminals and can be well over 50% in several cases. Given the growing importance of uplink traffic from mobile services such as live streaming on social media (Facebook, periscope), and interactive applications (AR/VR), we aim to tackle this important and timely problem in UL access so that LTE can be efficiently deployed in unlicensed spectrum.

Our Proposal: Blu. Tackling this problem requires us to address the fundamental conflict between *scheduled*, synchronous transmissions (for increased gains from concurrency) and *distributed*, asynchronous access (for coexistence). While we know how to leverage one in the absence of the other (e.g., LTE and WiFi in isolation), the key conundrum facing us is - *Can we have both?* We take an important step in answering this question through our LTE-compliant proposal: Blu. Blu adopts the notion of *speculative scheduling* for UL access in LTE, whereby it leverages *interference diversity* across clients in a novel manner to jointly (over-)schedule multiple (fM, $f > 1$) single-antenna clients on the same UL transmission resources in an M antenna eNB to compensate a priori for the potential under-utilization that results during access (see Fig. 5). While intuitive in principle, the real challenge and hence our contributions lie in how to leverage interference diversity effectively in practice: Specifically, (i) how to determine which set of clients can be over-scheduled jointly for both SISO and MU-MIMO transmissions on UL to increase utilization? Wrong decisions can lead to collisions (when $> M$ transmissions received) and to a much worse performance than the under-utilization itself (e.g. compare UL sub-frame 2 in Figs. 5 and Fig. 3), especially with increased concurrency of transmissions in MU-MIMO, (ii) how does the scheduler adapt its mechanisms to incorporate this paradigm to increase its efficiency while still adhering to its fairness principles?

Contributions. At the heart of its design, Blu aims to speculatively (over-)schedule its clients to cope with un-predictable (stochastic) interference from hidden terminals. In this regard, Blu makes two important contributions:

(1) To avoid wrong decisions and determine the correct set of clients that need to be jointly over-scheduled, one needs the joint (stochastic) distribution of access from these clients. This poses a scalability problem in collecting appropriate measurements needed

to capture this joint distribution of all possible $[1, fM]$-tuple clients (scales exponentially in M as $O(N^{fM})$), and is even infeasible in some cases with large M. BLU addresses this key challenge through a scalable, novel framework that blue-prints the source interference experienced by the clients (from hidden terminals) through a graphical structure. The latter estimates the set of hidden terminals in the network, along with their access distributions and impact on specific clients to a high degree of accuracy from only *pair-wise* access distributions of clients. This in turn results in a fixed (w.r.t. M), significantly reduced measurement overhead that scales as $O(N^2)$) and does not depend on the concurrency of transmissions (M). More importantly, the blue-printed interference structure captures all the interference dependencies between different hidden terminals and the clients and hence makes it possible to infer the higher-order joint access distributions of all clients.

(2) Using the joint access information of its clients, BLU transforms the proportional-fair (PF) scheduling algorithms employed in eNBs today into speculative scheduling algorithms that leverage interference diversity to increase spectrum utilization for both SISO and MU-MIMO transmissions on the uplink, while still adhering to the PF principle.

(3) BLU is readily compatible with LTE specifications. We build a version of BLU using WARP SDRs as eNB and clients (running LTE release 10), and evaluate its performance in an enterprise environment. Our test-bed experiments are supplemented with a larger scale emulation on larger traces collected from the test-bed as well NS3 simulations. Our evaluations reveal that BLU is able to infer the interference topology with a high median accuracy of over 90% with just pair-wise client access measurements. This allows BLU's speculative scheduler to yield a substantial gain of 1.5-2× in both utilization and throughput over today's schedulers for SISO and MU-MIMO, thereby helping retain the concurrency benefits of LTE in unlicensed spectrum.

Broader Impact: In addition to providing the information needed for intelligent speculative scheduling, BLU's approach of blue-printing stochastic wireless interference has wider applications for LTE in indoor environments, namely: channel selection for unlicensed LTE operation based on assessment of hidden terminal impact on candidate channels; coarse localization of clients in indoor environments by using inferred hidden terminals as landmarks, etc.

2 MOTIVATION

2.1 Background

LTE Overview: LTE is a synchronous, scheduled access system designed for operation in the licensed spectrum. The eNB is responsible for scheduling both the downlink (DL) and uplink (UL) clients in its sub-frames (1 ms long, see DL sub-frame in Fig. 2b), which consists of two-dimensional resource elements spanning both time (symbols) and frequency (sub-carriers), called resource blocks (RBs). LTE employs OFDMA (orthogonal frequency division multiple access), whereby multiple clients are scheduled in each sub-frame on *different* RBs - in the case of multi-user MIMO, multiple clients are scheduled on the *same* RB. The schedule for both DL and UL transmissions is conveyed to the clients through the control part of the DL sub-frames.

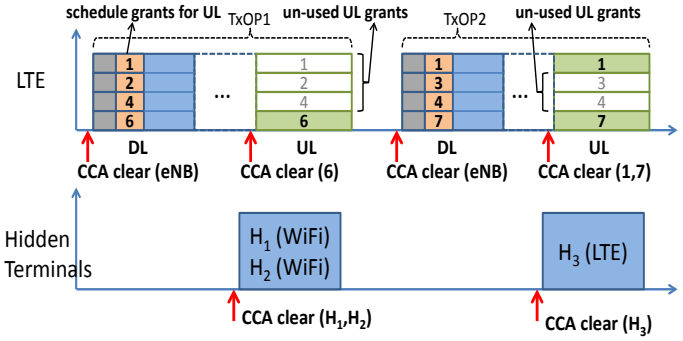


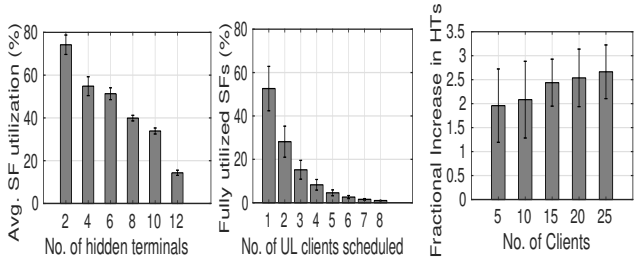
Figure 3: eNB schedules multiple clients on UL resources (SISO): clients 1,2,4 (3,4) in UL sub-frame 1 (2) are unable to use their grants due to hidden terminals.

LTE in Un-licensed Spectrum: Unlike traditional LTE that operates in an always-on mode in licensed spectrum, operating in un-licensed spectrum requires LTE to adopt asynchronous access principles of clear-channel assessment (CCA through energy sensing) and back-off for co-existence with the incumbents. There are two categories of solutions: license-assisted (LTE-U [17] and LTE-LAA/eLAA [3, 4, 23]), where unlicensed carriers are aggregated with existing licensed carriers (latter serving as anchors for control signaling); and stand-alone (e.g. MulteFire [25, 26]), where LTE is deployed and anchored completely on unlicensed carriers. Unlike the initial version (LAA), its predecessor (enhanced LAA) allowed for the challenging, UL multi-user access to also be executed in unlicensed spectrum. This has paved the way for the stand-alone mode, where both DL and UL access operate in the unlicensed carrier in a time-divisioned mode (Fig. 2b). The latter has opened LTE up for innovation from green-field operators in newer 3.5 GHz bands like CBRS [9]. Our work is applicable to both eLAA and stand-alone modes, where UL access operates in unlicensed spectrum.

2.2 The Conflict

Increased Interference in LTE-WiFi Environments: While WiFi employs carrier sensing to detect (and avoid) other WiFi nodes (interference) with higher sensitivity (-85 dBm), a heterogeneous set-up of LTE and WiFi nodes have to settle for interference detection through energy sensing that has a lower sensitivity ([-70,-65] dBm). This increases the number of interfering (hidden to transmitter) terminals in LTE-WiFi environments (compared to all WiFi) by an appreciable margin. Fig. 4c indicates this increase can be well over two times when a single WiFi cell (AP and clients) is replaced by an LTE cell in a network of otherwise WiFi nodes.

Pronounced Impact on UL Access: Increased hidden terminals contribute to more collisions in both DL and UL of LTE and WiFi, which can be alleviated by homogenizing the asymmetric channel sensing between LTE and WiFi (e.g. Ultron [10]). However, unique to LTE, they also create an additional layer of substantial impact on UL access that remains un-resolved. It is the eNB in LTE (instead of clients in WiFi) that accesses the channel for a transmission opportunity (TxOP); then schedules the synchronous access of multiple clients on the UL sub-frames that immediately follow its



(a) Impact of hidden terminals. (b) Impact of multi-user access. (c) Increase in HTs for LTE.

Figure 4: Interference impact on UL transmissions.

DL sub-frames in the TxOP (Fig. 2b). The scheduled clients simply perform a CCA before transmitting on the UL grants [3, 26]. This synchronous feature makes it possible to realize the gains from OFDMA and MU-MIMO, which are otherwise not possible on the UL (e.g. WiFi). However, since the instantaneous channel (interference) state of the clients cannot be known a priori at the eNB in unlicensed spectrum, a scheduled LTE client that is inhibited by an interfering transmission (hidden to the eNB) during its CCA will not be able to utilize its allocated UL grant (e.g. Fig. 3). This leads to an under-utilization of spectral resources – a problem that is not encountered in WiFi (Fig. 2a) and exacerbated by multi-user access in LTE (Fig. 3).

Impact Scales with Interference: To understand the magnitude of this problem, we collect access traces of 8 clients from our test-bed (described in Section 4), where the UL access of LTE clients in the cell are impacted by WiFi hidden terminals as shown in Fig. 1. The result in Fig. 4a presents the loss in spectral (sub-frame) utilization. It can be seen that the loss in utilization scales with the number of hidden terminals, which increases the probability of scheduled grants going un-used in a subframe (SF), and can be over 50% even for a small number of hidden terminals in the network. With both OFDMA and MU-MIMO relying on scheduled, multi-user access on the uplink, such under-utilization is un-avoidable as seen in Fig. 4b (fraction of completely occupied sub-frames). This reveals a fundamental conflict between leveraging the concurrency gains (diversity from OFDMA and multiplexing from MU-MIMO) from LTE's scheduled, synchronous transmissions on the UL, and coexistence (asynchronous interference) in un-licensed spectrum.

2.3 The Problem

Moving to single-user transmissions on the UL (similar to WiFi) can side-step the conflict and alleviate the issue of under-utilization, albeit at the expense of the gains from concurrent transmissions. Hence, the problem we aim to address is the conflict itself: *Can we retain the benefits of concurrent UL transmissions in an asynchronous access environment?*

A natural approach to retain the gains from concurrency on the UL, is to over-schedule clients on RBs to compensate a priori for the potential under-utilization that may result from un-predictable interference as shown in Fig. 5. Although simple in idea at the outset, digging deeper into the approach raises several important questions and associated challenges:

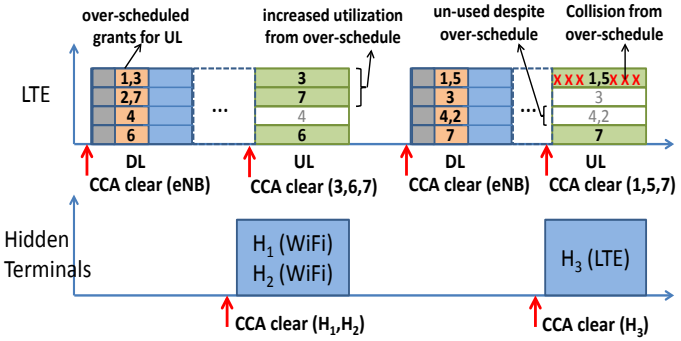


Figure 5: eNB over-schedules clients 3,7 (TxOP 1) & 5,2 (TxOP 2) on SISO UL; Increased utilization in TxOP 1, but decreased utilization (1 & 5 collide, 2 & 4 blocked by H_3) in TxOP 2.

(i) Is it even feasible to over-schedule clients on UL sub-frames in today's LTE? While it is not possible to schedule more than M clients on an RB (in a M antenna eNB) in the DL, it is indeed possible (LTE compliant) to schedule grants for more than M clients on an UL RB as we show in Section 4. However, eNB will not be able to resolve (decode) when more than M transmissions are received on M antennas, resulting in collisions (relevance to future non-orthogonal multiple access schemes is discussed in Section 5).

(ii) Given feasibility, how to over-schedule clients on the same RB to increase utilization without paying the penalty of collisions? Depending on the interference dependencies between clients scheduled on the same RB, over-scheduling can either increase utilization (e.g. UL of TxOP 1, Fig. 5) or decrease it further (from collisions, e.g. UL of TxOP 2).

(iii) How to determine the interference dependencies across clients in the cell? LTE's ability to schedule multi-user transmissions allows us infer interference dependencies indirectly by measuring the joint access patterns (probability distributions) of the scheduled clients on the UL over time (sub-frames). However, orchestrating these measurements is a challenge in itself, as the overhead (spectrum resources) incurred in measuring all the desired information, scales exponentially with the concurrency of transmissions, M.

3 BLU: DESIGN COMPONENTS

3.1 Overview

Can we design an efficient, speculative scheduler that leverages the joint access distributions of all clients to increase utilization of concurrent transmissions on LTE uplink (with stochastic interference), but do so with a small, fixed measurement overhead that does not scale with the concurrency of transmissions?

For an easier exposition, we present BLU's design components (overall design described later in Fig. 9) by working backwards. First, we present the scheduling algorithms traditionally employed at eNBs. Then, we show how these algorithms can be transformed to speculative scheduling algorithms to leverage interference diversity and hence increase the spectrum utilization. We highlight the scalability issues that arise in estimating the access distribution information needed to execute these speculative scheduling algorithms. Then, we present the under-pinning of BLU's design – a

scalable, novel graphical framework to *blue-print* the interference experienced by the clients (from hidden terminals). This framework allows Blu to require only pair-wise access distributions of clients to be measured (a fixed overhead that does not scale with the concurrency of transmissions). The resulting interference structure and dependencies then helps infer all the higher-order joint access distributions of clients. This further allows for estimating all the information required for executing speculative scheduling algorithm in Blu to deliver increased utilization. While Blu targets UL access, we discuss how Blu can deliver benefits to the DL access in §3.7.

3.2 Speculative Scheduling in Blu

3.2.1 Current LTE Scheduler. LTE schedulers employ orthogonal frequency division multiple access (OFDMA) to leverage multi-user diversity. The spectrum (e.g. 20 MHz channel) is partitioned into resource blocks (groups of OFDM sub-carriers) and a user (users in case of MU-MIMO) with a higher rate on a RB is assigned to it, while accounting for fairness across clients. Proportional fair (PF) scheduling is the most popular scheduling model adopted in eNBs today as it strikes a good balance between throughput efficiency and fairness, allowing for clients with better channels to achieve a proportionally higher throughput.

The optimal scheduling policy can be obtained through a utility optimization framework that maximizes the aggregate utility of all the clients ($\sum_i U_i$). For PF, the utility function is the logarithm of the client's average throughput $U_i = \log(R_i)$. Being a convex optimization problem, picking a schedule that maximizes the gradient of the utility (i.e. marginal utility, $\frac{dU_i}{dt} = \frac{r_i(t)}{R_i(t-1)}$) at each sub-frame t, achieves proportional fairness over a longer time period [29]. The scheduling problem for each sub-frame with B sub-frames and N clients (with single antenna) now reduces to,

$$S^*(t) \underset{\text{SISO}}{=} \arg\max_{x \in S}\{\sum_{b=1}^{B}\sum_{i=1}^{N}\frac{x_{i,b}r_{i,b}(t)}{R_i(t-1)}\},\ \text{s.t.}\ \sum_{i=1}^{N} x_{i,b} \leq 1, \forall b$$

$$\underset{\text{MU–MIMO}}{=} \arg\max_{y \in S}\{\sum_{b=1}^{B}\sum_{i=1}^{N}\frac{y_{i,b}r_{i,b,g}(t)}{R_i(t-1)}\},\ \text{s.t.}\ \sum_{i=1}^{N} y_{i,b} \leq M, \forall b \quad (1)$$

where $r_{i,b}(t)$ and $r_{i,b,g}(t)$ are the instantaneous rates of client i (measured at eNB) on RB b in SISO and MU-MIMO respectively, while x and y are binary variables capturing the schedule. $r_{i,b,g}(t)$ depends on the group of clients g selected for MU-MIMO and their respective channels. The above scheduling problem can be decoupled into multiple (individual) RB-level scheduling[1] problems, $S^*_b(t)$. After each schedule, the average throughput of a client i is updated as,

$$R_i(t) \underset{\text{SISO}}{=} \frac{1}{\alpha}\sum_{b=1}^{B} x^*_{i,b}r_{i,b}(t) + (1-\frac{1}{\alpha})R_i(t-1)$$

$$\underset{\text{MU–MIMO}}{=} \frac{1}{\alpha}\sum_{b=1}^{B} y^*_{i,b,g}r_{i,b,g}(t) + (1-\frac{1}{\alpha})R_i(t-1)$$

[1]Coupling constraints across RBs (e.g. finite buffer data for clients) as well as multi-antenna clients can be accommodated through simple extensions to proposed scheduler.

where α is an exponential weighting constant. We will now focus on a sub-frame and hence drop the subscript of t for an easier exposition.

3.2.2 Scheduler Leveraging Interference Diversity. Since the clients are scheduled by the eNB on the UL, interferers (WiFi or other LTE nodes) to the clients that are hidden from the eNB will prevent the clients from utilizing the allocated resource grants.

If $p(i)$ is the probability that client i is able to utilize its allocated grant, then the expected utility value of the schedule S^* (for SISO) reduces to,

$$E(S^*) = \sum_{b=1}^{B}\sum_{i \in S^*_b}\frac{p(i)\cdot r_{i,b}}{R_i} \quad (2)$$

Depending on the impact of hidden terminals (reduced $p(i)$), existing schedulers, albeit efficient for licensed spectrum, can lead to significant under-utilization in unlicensed spectrum (as shown in Section 2).

Blu transforms the very challenge posed by scheduled, multi-user LTE transmissions into an opportunity as follows. Different clients in the same cell could be interfered by different hidden terminals (e.g. clients 1 and 3 in Fig. 1) and hence may not be silenced at the same time. Blu leverages this *interference diversity* across clients, coupled with LTE's ability to simultaneously schedule multiple users in an UL sub-frame, to *over-schedule* multiple users ($> M$) on the same UL resource block to increase utilization (Fig. 5 UL sub-frame 1). However, executing this intelligently by identifying which clients need to be over-scheduled on the same RB is paramount, as multiple client transmissions (> 1 for SISO and $> M$ for MU-MIMO) on the same RB will lead to collisions (e.g. clients 1 and 5 in UL sub-frame 2) and hence a much worse performance than the under-utilized schedule.

Blu makes its decisions based on the expected utility of a schedule that accounts for the *joint* (dependent) stochastic access patterns of the clients. At each step, it determines the schedule for one of the remaining (un-scheduled) RBs, which maximizes the incremental utility. In computing the schedule for a given RB b, it adds users one at a time – for an existing set of clients (G_b) scheduled on RB b, Blu selects and adds another client ℓ^* that provides the maximum incremental utility to the current schedule on that RB.

$$\ell^* = \arg\max_{\ell \notin G_b}\{E(G'_b) - E(G_b)\};\ \text{where},\ G'_b \leftarrow G_b \cup \ell \quad (3)$$

In its most generic form, the expected utility of a schedule on a RB depends on the total number of its scheduled clients, who can use the grant, being less than or equal to the total number of antennas (M) at the eNB, their joint access distribution, and the utility of those specific clients in the group.

$$E(G'_b) = \sum_{g: g \subseteq G'_b\ \&|g|<=M}\left(\mathbb{P}(g, \overline{G'_b \setminus g})\sum_{i \in g}\frac{r_{i,b,g}}{R_i}\right) \quad (4)$$

where, $\mathbb{P}(g, \overline{G'_b \backslash g})$[2] represents the joint access distribution of the group – the probability that all the clients in g (e.g. clients 1,2 in $\mathbb{P}(1, 2, \bar{3}, \bar{4})$) are able to utilize the grants, while all the remaining clients ($j \in G'_b \backslash g$; clients 3 and 4 in our e.g.) are not able to (i.e. $\bar{j}$). The size of g represents the eventual transmissions on the RB and hence can be upto M (number of antennas) - otherwise, this would lead to collisions on all the transmissions on the RB. The addition of clients to the RB's schedule stops, when no remaining client can further increase the schedule's utility. As the number of clients carefully scheduled in an RB continues to increase beyond M, it increases the potential for utilization but it also increases the risk of collisions from over-scheduling, thereby resulting in diminishing returns. BLU's *speculative scheduler* strikes a fine balance and typically over-schedules between $[M, 2M]$ clients (i.e. $f = 2$) on an RB as determined by Eqns. 3, 4.

Importance of Joint Access Distribution: Joint access distribution of clients is critical for over-scheduling. In its absence, one can devise a weighted proportional fair schedule that accounts for the individual access probabilities of clients, but will not have the interference dependency information needed to intelligently over-schedule (over-scheduling clients sharing common hidden terminals can lead to collisions or under-utilization as in Fig. 5, UL sub-frame 2). We refer to this as the *access-aware scheduler*, where

$$E(G'_b) = \sum_{i \in G'_b \ \& |G'_b| <= M} \frac{\mathbb{P}(i) \cdot r_{i,b,G'_b}}{R_i} \tag{5}$$

Example: As an example, consider a SISO speculative schedule on an RB. The first client is chosen as $s_1 = \arg\max_i \{\mathbb{P}(i) \cdot \frac{r_{i,b}}{R_i}\}$. The next client to be (over-)scheduled on the same RB is chosen as,

$$s_2 = \arg\max_{i \neq s_1} \left\{ \mathbb{P}(i, \bar{s_1}) \cdot \frac{r_{i,b}}{R_i} + \mathbb{P}(\bar{i}, s_1) \cdot \frac{r_{s_1,b}}{R_{s_1}} \right\}$$

where $\mathbb{P}(i, \bar{s_1})$ indicates the probability that i is able to transmit, while s_1 is not, and vice versa. Note that for SISO, $\mathbb{P}(i, s_1)$ and $\mathbb{P}(\bar{i}, \bar{s_1})$ don't contribute to useful transmissions, leading to collision and no-transmission respectively. s_2 is then over-scheduled, only if the access distributions (interference diversity) of the two clients s_1 and s_2 are such that they allow for a better utilization than the current schedule,

$$\left\{ \mathbb{P}(s_2, \bar{s_1}) \cdot \frac{r_{s_2,b}}{R_{s_2}} + \mathbb{P}(\bar{s_2}, s_1) \cdot \frac{r_{s_1,b}}{R_{s_1}} \right\} > \left\{ \mathbb{P}(i) \cdot \frac{r_{s_1,b}}{R_{s_1}} \right\}$$

Subsequent clients to be over-scheduled on the same RB are iteratively evaluated in a similar procedure using Eqns. 3, 4.

3.3 Scalable Measurement Overhead in BLU

The challenge in executing the proposed scheduler in BLU is the need to estimate the joint access distribution of clients ($\mathbb{P}(g, \overline{G'_b \backslash g})$). For example, to over-schedule 4 clients in a $M = 2$ MU-MIMO speculative schedule, one would need to estimate $\mathbb{P}(1, 2, \bar{3}, \bar{4})$, $\mathbb{P}(1, 3, \bar{2}, \bar{5})$,

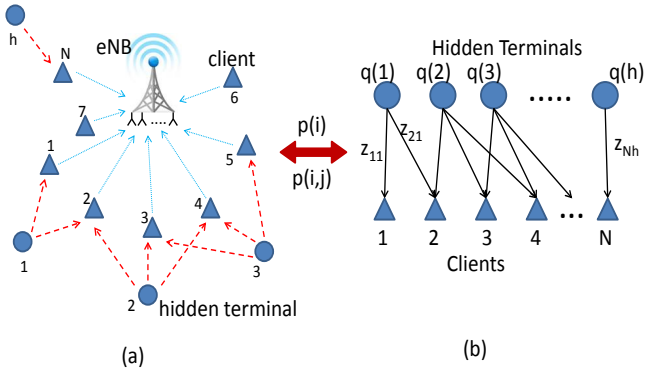


Figure 6: Topology Inference Objective

etc. LTE's ability to leverage OFDMA on the UL, allows BLU to estimate the joint access (probability) distributions of clients directly from their transmissions – schedule a desired set of clients *jointly* in UL sub-frames and measure the fraction of those sub-frames the clients were able to use (i) or not use ($\bar{i}$) those scheduled grants *jointly*. Although data is transferred during these measurement sub-frames, the client schedule is optimized for obtaining the desired access information rather than for performance. Hence, it is imperative to keep the overhead of this measurement phase as small as possible.

The number of *distinct* clients (K) that can be scheduled together in each sub-frame[3] is typically much smaller (less than 10) than the number of clients in a cell (N). This raises two issues: (i) for larger MU-MIMO systems, it is not feasible to get any k-client ($k \in [1, 2M]$) joint distribution when $k > K$ - e.g. estimating $\mathbb{P}(1, 2, 3, \bar{4}, \bar{5})$ (i.e. $k = 5$) is not possible when at most $K = 4$ distinct clients can be scheduled in a sub-frame; and (ii) even when $k \leq K$, if T samples (sub-frames) are needed to measure the joint distribution of each k-client tuple, then the associated overhead (minimum number of sub-frames) for estimating all such k-tuples is $\lceil \frac{\binom{N}{k}}{\binom{K}{k}} T \rceil$ sub-frames, which scales exponentially with k (and hence M) as $O(\frac{N^{\min\{k, N-k\}}}{K^{\min\{k, K-k\}}})$. For example, measuring all 6-client joint distributions (for $M = 3$ MU-MIMO) in a cell of 20 clients with $K = 8$ requires a minimum of $\lceil \frac{\binom{20}{6}}{\binom{8}{6}} T \rceil \approx 1384T$ sub-frames.

In contrast, BLU proposes to work with just pair-wise client distributions ($\mathbb{P}(i, j)$), which results in a *constant* (w.r.t. M), significantly reduced overhead of $F_{\min} = \lceil \frac{\binom{N}{2}}{\binom{K}{2}} T \rceil$ sub-frames (only $< 7T$ sub-frames for the above example) that is $O(\left(\frac{N}{K}\right)^2)$ and completely independent of M. While ensuing subsections will demonstrate how this is sufficient, first we need to determine the schedule of clients for successive measurement sub-frames that will estimate all the pair-wise access distributions needed in $F_{\min}$ sub-frames (lower bound). This being a hard problem in itself, BLU employs the following scheduling algorithm 1 (in the measurement period) to

[2] To avoid confusion, we use $\mathbb{P}(\cdot)$ to indicate probability, especially for joint distributions, where $\mathbb{P}(\cdot) = p(\cdot)$.

[3] K is limited by control signaling requirements. However, each client can be assigned multiple RBs in a sub-frame.

estimate these distributions with as small a number of sub-frames as possible (close to $F_{\min}$).

Algorithm 1 Scheduling Measurements

Output: $S(t)$, $t \in [1, t_{\max}]$ % schedule for measurements
$\mathcal{N} = \{1, 2, \ldots, N\}$; $C(\mathcal{N}) = \{(1,2), (1,3), \ldots, (N-1, N)\}$
$t = 0$; $c_j = 0, \forall j \in C(\mathcal{N})$
while $c_j \neq T, \ \forall j \in C(\mathcal{N})$ **do**
  $t \leftarrow t + 1$; $S(t) = \{\}$
  for $i = 1 : K$ **do**
    $\ell^* = \arg\max_{\ell \notin S(t)} \{\sum_{j \in C(S(t) \cup \ell)} \log(\frac{1+c_j}{1+T})$
      $- \sum_{j \in C(S(t))} \log(\frac{1+c_j}{1+T})\}$
    $S(t) \leftarrow S(t) \cup \ell^*$
  end for
  $c_j \leftarrow c_j + 1, \forall j \in S(t)$
end while
$t_{\max} = t$

In each sub-frame during the measurement period, Blu schedules K clients that will contribute the most value towards measuring pair-wise distributions; i.e. K clients are chosen, whose resulting pair-wise distributions have the least number of measurements thus far. A logarithmic function of the measurement count is employed to ensure that each pair is sampled for approximately the same number of times at any point during the measurement period. This provides for flexibility in using the measurements even before the end of the period, if desired.

Differentiating between Fading and Hidden Terminal Loss: This is achieved with the help of UL reference signals (a.k.a. pilots) in LTE, which are sent at the lowest modulation, and are much more resistant to channel fading compared to data signals. Hence, when the eNB fails to receive any UL signal (including pilots) from a scheduled UE, this is due to the client backing-off due to a hidden terminal transmission with high probability. On the other hand, when the eNB receives the pilot(s) but is unable to decode the data signals, there are two cases in Blu. Note that even when clients are over-scheduled on the same RB, their pilots are still kept orthogonal (non-overlapping). This allows Blu to detect if the decoding failure is due to collision from over-scheduled clients (both pilots are received) or due to fading (only one pilot is received).

3.4 Blue-printing Interference

Instead of spending the measurement overhead to estimate all the joint access distributions, Blu aims to leverage just the pair-wise access distribution measurements (Section 3.3) to "blue-print" the source of the interference itself, which in turn is responsible for *all* the joint client access distributions.

The challenge lies in how to blue-print the hidden terminal interference on the clients (Fig. 6(a))? In other words, given the individual ($\mathbb{P}(i) = p(i)$) and pairwise ($\mathbb{P}(i,j) = p(i,j)$) client access distributions, can we determine the topology (Fig. 6(b)) characterized by (i) the number of hidden terminals (h), (ii) their access distributions ($q(k), \ k \in [1, h]$), as well as (iii) their impact on specific clients (edges, $z_{ik}, \ i \in \mathcal{N}, k \in [1, h]$), that will contribute to these observed distributions? Here, an edge from a hidden terminal to a client indicates that the latter can sense the former's transmission, when it exists and will defer its own.

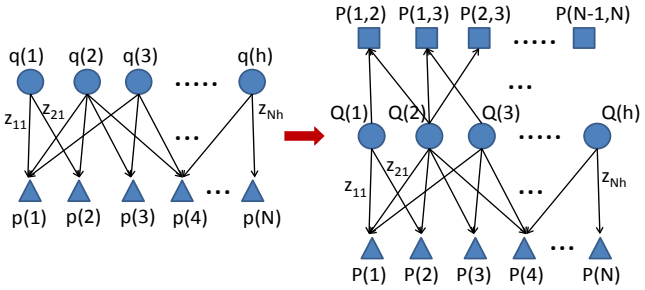


Figure 7: Graphical Representation of Topology Inference

Similar to wired network topology inference problems [8, 12, 22], one could employ Bayesian learning to estimate our wireless interference topology. Specifically, we have applied Monte Carlo Markov Chain (MCMC [14]) based techniques, where the interference topology is adapted based on likelihood estimates such that the topology distribution converges to a stationary distribution that maximizes the posterior probability of the observed data (client access distributions). However, in addition to the time for convergence, note that the topology only converges in distribution in such an approach. Hence, when the topology information needs to be used for real-time scheduling of clients, one needs to sample this distribution to pick an actual topology – mis-matches from the ground-truth topology could lead to sub-optimality.

While such Bayesian approaches are better suited for large scale networks with multiple-hops, the wireless topology that we are interested in has a single layer of nodes (hidden terminals) and their interference edges (to clients) and distributions that need to be estimated. Hence, Blu aims to design an alternate deterministic solution that can leverage this inherent structure to infer the topology with high accuracy. Blu accomplishes this in two steps.

3.4.1 Step 1: Graph Transformation. Blu's goal is to infer topology and access patterns of hidden terminals that contribute to the observed $p(i)$ and $p(i,j)$ of the clients in the cell. Let $q(k)$ be access probability of hidden terminal k. Blu applies a transformation to access probabilities as follows.

$$P(i) = -\log(p(i)); \quad Q(k) = -\log(1 - q(k))$$
$$P(i,j) = -\log\left(\frac{p(i) \cdot p(j)}{p(i,j)}\right)$$

The transformation allows us to operate with sum of the transformed variables as opposed to the product of the original variables (probabilities). This allows us to now formulate the topology inference problem as a graphical constraint satisfiability problem as shown in Fig. 7. The first and third layer of nodes correspond to each of the input constraints (transformed access distributions, $P(i)$ and $P(i,j)$) that we want to satisfy, while the second layer of nodes represents an "un-known" number (h) of hidden terminals, whose access distributions ($Q(k)$) and interference impact (edges, z_{ik})

we want to infer. Specifically, we need to determine the topology (h, Q, Z) that satisfies the following constraints.

$$\begin{aligned} P(i) &= \sum_{k=1}^{h} z_{ik} Q(k), \quad \forall i \in \mathcal{N} \\ P(i,j) &= \sum_{k=1}^{h} z_{ik} z_{jk} Q(k), \quad \forall i,j \in \mathcal{N} \end{aligned} \tag{6}$$

where Z is a matrix, whose entries, $Z(i,k) = \{z_{ik}\},\ \forall i,k$ are binary variables capturing the impact of hidden terminal k on client i. The first set of constraints captures the access probability of a client i as the product [4] of the idle probabilities $(1 - q_k)$ of all hidden terminals k impacting it (i.e. $z_{ik} = 1$). The second set of constraints indicates that the point mass mutual information $(P(i,j))$ between two clients (i,j) is given by the product of the idle probabilities of all hidden terminals that impact *both* clients. Using more variables (hidden terminals, h) than the constraints, can result in an under-determined system with potentially many solutions. Blu aims to limit the solutions to those that satisfy the above constraints while minimizing the number of HTs (h).

3.4.2 Step 2: Topology Inference. Blu infers the topology by starting with an initialized topology (initialization discussed shortly) and then adapts the topology in each iteration through a gradient approach to improve the satisfiability of the constraints. At each iteration, it determines the constraint that is maximally violated. Then, it selects a hidden terminal $\hat{k}$, along with its appropriate topology adaptation $(\hat{h}, \hat{Q}, \hat{Z})$ that will resolve this violation, while minimizing the violation caused to the other constraints in the process. It terminates when all the constraints are satisfied (zero violation), or the maximum number of iterations is reached, in which case the configuration with the least aggregate violation is chosen.

Topology Adaptation: There are multiple cases to consider during the adaptation process in each iteration.
Case 1: If the constraint chosen for restoring violation is an individual access constraint, $P(i)$, two sub-cases arise based on the type of violation. Let $c_i = \sum_{k=1}^{h} z_{ik} Q(k) - P(i)$.
(i) Over-contribution ($c_i > 0$): Blu reduces the contribution by determining whether to decrease the appropriate contribution ($\hat{Q}(k) \leftarrow Q(k) - c_i$); (or) remove an edge completely ($\hat{z}_{ik} = 0$) from one of the existing hidden terminals k (impacting client i), where $k : z_{ik} = 1$.
(ii) Under-contribution ($c_i < 0$): Blu determines whether to increase the appropriate contribution ($\hat{Q}(k) \leftarrow Q(k) + |c_i|$) from one of its hidden terminals k; (or) add an edge to one of the existing hidden terminals k (where $z_{ik} = 0$) to avail its contribution ($Q(k)$) to $P(i)$; (or) add a new hidden terminal k' with an edge to it ($\hat{z}_{ik'} = 1$) that provides the missing contribution ($\hat{Q}(k') = |c_i|$).
Case 2: Similarly, if the constraint chosen is a joint access constraint, $P(i,j)$, the corresponding scenarios are slightly more involved. Let $c_{i,j} = \sum_{k=1}^{h} z_{ik} z_{jk} Q(k) - P(i,j)$.
(i) Over-contribution ($c_{i,j} > 0$): Blu determines whether to reduce the appropriate contribution ($\hat{Q}(k) \leftarrow Q(k) - c_{i,j}$) from one of the contributing hidden terminals, $k : z_{ik} z_{jk} = 1$; (or) remove an edge from one or both of the clients ($\hat{z}_{ik} = 0$ and/or $\hat{z}_{jk} = 0$) impacted by that hidden terminal.
(ii) Under-contribution ($c_{i,j} < 0$): Blu determines whether to increase the appropriate contribution ($\hat{Q}(k) \leftarrow Q(k) + |c_{i,j}|$) from one of its contributing hidden terminals $k : z_{ik} z_{jk} = 1$; (or) add edge(s) to a hidden terminal k to avail its contribution ($Q(k)$), where an edge to only one or neither clients (i and j) exists, i.e. $k : z_{jk} + z_{ik} \leq 1$; (or) add a new hidden terminal (k') with two edges, one each to i and j ($\hat{z}_{ik'} = 1, \hat{z}_{jk'} = 1$) that provides the missing contribution ($\hat{Q}(k') = |c_{i,j}|$) to $P(i,j)$.

At the end of the adaptation, hidden terminals left with no edges to clients are removed and the resulting topology $(\hat{h}, \hat{Q}, \hat{Z})$ serves as input (h, Q, Z) to the next iteration.

Topology Initialization: Given the non-linear nature of the problem, a gradient based approach is not guaranteed to converge to an optimal solution and could end up in a locally optimal topology. To alleviate the resulting sub-optimality and also to minimize the number of hidden terminals employed, Blu runs the inference algorithm by initializing with different starting topologies and picking the inferred topology with least number of hidden terminals that yields the smallest violation. In addition to starting with random topologies with varied number of hidden terminals, it also picks from those that satisfy only one set of constraints as starting topologies. Given the single layer of variables that need to be inferred, such a multi-point initialization is able to overcome local optima in most cases, enabling Blu's deterministic algorithm to yield high accuracies in topology inference.

3.5 Discussions

Skewed Topologies: Occasionally, when the number of hidden terminals is much larger than clients, multiple topologies (solutions) may satisfy the observed pair-wise client access distributions, making it in-feasible to pin-point the ground-truth topology. However, even in such cases, there is a large similarity between the topology inferred by Blu and ground-truth, which leads to a minimal degradation in Blu's scheduler performance. Further, in such scenarios, additional joint access distribution of clients (beyond pair-wise, say triplets) that maybe available (obtained) from existing (new) measurements, can provide additional constraints, which will significantly reduce the number of feasible topologies.

Interference Impact: Blu's topology inference currently assumes that the interference impact of a hidden terminal on different clients has a binary {0,1} effect. While this will capture scenarios where clients are either strongly or weakly interfered by the hidden terminal, it may not accurately capture the fractional [0,1] impact resulting from fading related interference variations. However, the sub-optimality resulting from this assumption is restricted to the specific clients in question. Hence, this does not appreciably affect the benefits to speculative scheduling, especially in the presence of a reasonable number of clients in the cell.

Stationarity and Mobility: Blu's measurement of interference statistics and its application to speculative scheduling, operate at a finer time granularity compared to the time-scales of topology (e.g. client and hidden terminal mobility) and traffic dynamics (Sec. 3.7). This allows Blu to infer and leverage interference dependencies within their stationary regime.

[4] Product becomes sum in the transformed domain.

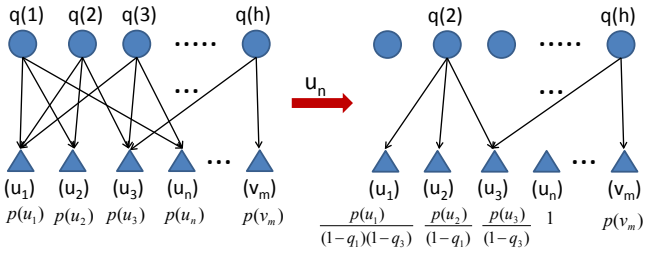


Figure 8: Topology Conditioning

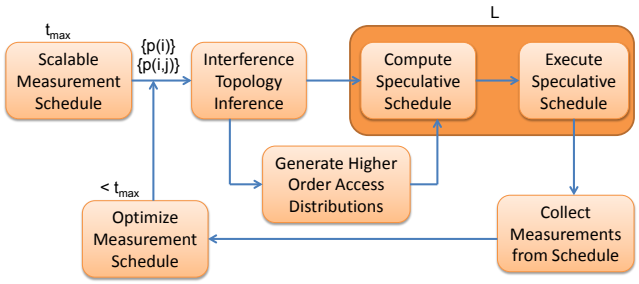


Figure 9: Blu Overview

3.6 Generating Higher-order Distributions

Having inferred the blue-print of the interference topology $\mathcal{T} = \{h^*, Q^*, Z^*\}$, we now demonstrate how Blu can compute the higher order access distributions from just the individual client access distributions, $\mathbb{P}(u_i)$.

Recall from Equation 4, we need to compute $\mathbb{P}(g, \overline{G'_b \backslash g})$, i.e. the probability that all the clients in g are able to utilize the grants, while all the remaining clients (in $G'_b \backslash g$) are not able to. Without loss of generality, let us assume,

$$U_n = \{u_1, u_2, \ldots, u_n\}; \qquad V_m = \{v_1, v_2, \ldots, v_m\}$$
$$g = U_n; \qquad G'_b = U_n \cup V_m$$

Hence, we are interested in computing $\mathbb{P}(U_n, \overline{V_m})$. Applying Bayes' theorem, we have,

$$\mathbb{P}(U_n, \overline{V_m}) = \mathbb{P}(\overline{V_m}|U_n) \cdot \mathbb{P}(U_n) \tag{7}$$

With the help of the inferred topology $\mathcal{T}$, we can now compute $\mathbb{P}(U_n)$ and $\mathbb{P}(\overline{V_m}|U_n)$ easily. $P(U_n)$ can be further simplified as,

$$\begin{aligned}\mathbb{P}(u_1, \ldots, u_n) &= \mathbb{P}((u_1, \ldots, u_{n-1})|u_n) \cdot \mathbb{P}(u_n) \\ &= \mathbb{P}(U_{n-1}|u_n) \cdot \mathbb{P}(u_n)\end{aligned}$$

Computing $\mathbb{P}(U_{n-1}|u_n)$ on $\mathcal{T}$ is equivalent to computing just $\mathbb{P}(U_{n-1})$ but on a modified topology that is conditioned given the occurrence of u_n as shown in Fig. 8. Given u_n's occurrence, the topology gets updated (conditioned) by removing the hidden terminals $\hat{k}$ that have an edge to u_n (i.e. $z_{u_n,\hat{k}} = 1$) and the access probabilities on this conditioned topology ($\mathcal{T}|u_n$) are updated using Eqns. 6 and represented as $\mathbb{P}_{u_n}(\cdot)$, where

$$\mathbb{P}_{u_n}(u_i) = \frac{\mathbb{P}(u_i)}{\prod_{\hat{k}:(z_{u_i\hat{k}}=1)}(1 - q(\hat{k}))}$$

Thus, Equation 8 can be computed by recursively conditioning the topology ($\mathcal{T}|u_n, u_{n-1}, \ldots$) on the occurrence of each client in U_n till it consists of just the individual client access probabilities.

$$\begin{aligned}\mathbb{P}(U_n) &= \mathbb{P}(u_n) \cdot \mathbb{P}_{u_n}(u_{n-1}) \cdot \mathbb{P}_{u_n,u_{n-1}}(u_{n-2}) \ldots \\ &= \mathbb{P}(u_n) \cdot \left(\prod_{a=1}^{n-1} \mathbb{P}_{u_n,\ldots,u_{n-a+1}}(u_{n-a})\right)\end{aligned} \tag{8}$$

Focusing on the other term, $\mathbb{P}(\overline{V_m}|U_n)$ in Eqn. 7, this is essentially $\mathbb{P}(\overline{V_m})$ on the topology conditioned by the occurrence of all the clients in U_n, i.e. $\mathbb{P}(\overline{V_m}|U_n) = \mathbb{P}_{U_n}(\overline{V_m})$. Applying Bayes' theorem, we have

$$\begin{aligned}\mathbb{P}_{U_n}(\overline{V_m}) &= \left(1 - \frac{\mathbb{P}_{U_n}(\overline{V_{m-1}}|v_m) \cdot \mathbb{P}_{U_n}(v_m)}{\mathbb{P}_{U_n}(\overline{V_{m-1}})}\right) \cdot \mathbb{P}_{U_n}(\overline{V_{m-1}}) \\ &= \left(1 - \frac{\mathbb{P}_{U_n,v_m}(\overline{V_{m-1}}) \cdot \mathbb{P}_{U_n}(v_m)}{\mathbb{P}_{U_n}(\overline{V_{m-1}})}\right) \cdot \mathbb{P}_{U_n}(\overline{V_{m-1}})\end{aligned} \tag{9}$$

As before, the above equation can be simplified by recursively conditioning the topology on the various clients in V_m till it consists of just the individual access probabilities of clients in V_m at various stages of the topology conditioning. Using Equations 8 and 9 in Equation 7, we are now able to compute the required higher order access distributions from just the individual client access distributions on the various conditioned topologies.

Example: In a four client (2 user MU-MIMO) schedule grant, the joint access distribution of clients 3 and 4 being able to transmit, while 1 and 2 not being able to, can be computed using the source interference topology and its conditioned versions as,

$$\begin{aligned}\mathbb{P}(\overline{1}, \overline{2}, 3, 4) &= \mathbb{P}((\overline{1}, \overline{2})|(3, 4)) \cdot \mathbb{P}(3, 4) \\ \text{where, } \mathbb{P}(3, 4) &= \mathbb{P}(3|4) \cdot \mathbb{P}(4) = \mathbb{P}_4(3) \cdot \mathbb{P}(4) \\ \mathbb{P}((\overline{1}, \overline{2})|(3, 4)) &= \mathbb{P}_{3,4}(\overline{1}, \overline{2}) = \left(1 - \frac{\mathbb{P}_{3,4,2}(\overline{1}) \cdot \mathbb{P}_{3,4}(2)}{\mathbb{P}_{3,4}(\overline{1})}\right) \cdot \mathbb{P}_{3,4}(\overline{1})\end{aligned}$$

3.7 Putting It All Together

Blu orchestrates its various design components to execute its speculative scheduler at eNBs as shown in Fig. 9. Blu operates the uplink eNB schedule in two phases repeatedly: a measurement schedule phase for $t_{\max}$ sub-frames, and a speculative schedule phase for L sub-frames ($L >> t_{\max}$). In the measurement phase (Section 3.3), clients are scheduled and transfer data, albeit with the objective of obtaining the desired client access distributions ($p(i), p(i, j)$) with minimal overhead. In the second phase, Blu first blue-prints the source interference topology (Section 3.4) from the measured distributions and uses it to determine the higher-order joint client access distributions (Section 3.6), needed especially for MU-MIMO transmissions. Blu then uses this information to speculatively schedule clients (Section 3.2) for higher utilization and efficiency.

Tracking Dynamics: L is chosen to infer (and leverage) interference topology reliably in the face of topology and traffic dynamics (from both clients and interference), which happens at the granularity of tens of seconds to minutes. Hence, L is chosen to be several thousand sub-frames (1000 sub-frames = 1 s), while $t_{\max}$ is in the order of few hundred sub-frames – for a 20 (N) client cell with a 50 (T) measurement sample requirement per client-pair, and maximum of 8 (K) distinct clients per sub-frame schedule, $t_{\max} \approx 340$ sub-frames. Note that, other than the first time that the eNB is operated, the measurement phase is run for less than $t_{\max}$ sub-frames, as the outcome of the schedule during the speculative phase will implicitly contribute to measurements as well. Thus, the measurement phase constitutes a very small part of the whole UL schedule that is predominantly used to maximize UL utilization through speculative scheduling.

Applicability to DL Access: In contrast to under-utilized scheduled grants on UL, the conflict between concurrent transmissions and asynchronous interference manifests in the form of increased collisions in the DL. While over-scheduling transmissions themselves (instead of grants) on the RBs is not feasible on the DL, Blu's methodology in inferring interference topology and dependencies between clients can be leveraged to enable *access-aware scheduling* for OFDMA and MU-MIMO transmissions on the DL (Eqn. 5). This would enable scheduling decisions that minimize collisions and increase overall efficiency on the DL.

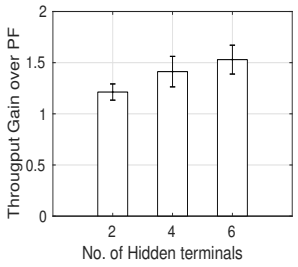


Figure 10: Blu SISO throuthput gains

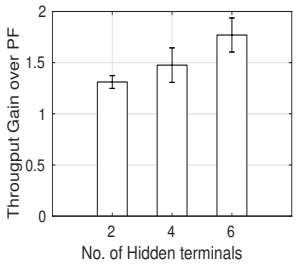


Figure 11: Blu MU-MIMO throughput gains

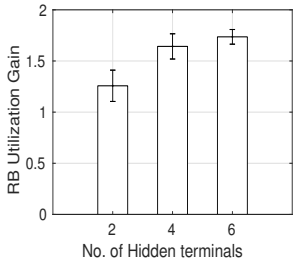


Figure 12: Blu SISO RB utilization gains

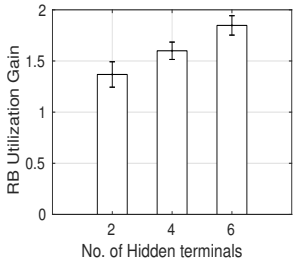


Figure 13: Blu MU-MIMO RB utilization

4 IMPLEMENTATION AND EVALUATION

4.1 Testbed Evaluation

Implementation:

In order to show the practicality of the design, and to evaluate the performance of Blu on a real testbed, we implement Blu on a WARPv3 platform. The testbed contains four single antenna LTE clients (UEs), and one LTE base station (eNB). We use six laptops equipped with `ath9k` wireless cards, supporting 802.11 a/b/g/n standards as hidden terminals. The eNB and the UEs run a *Release 10 standard compliant LTE stack* built using the MATLAB LTE Toolbox [1]. Blu's topology inference and speculative scheduling for both SISO and MU-MIMO are incorporated into eNB's LTE scheduler code. We use a 10 MHz LTE signal (sampling rate = 15.36 MHz), which is then up-sampled to the 40MHz sampling rate of the WARP board before transmission. The received I/Q data is down-sampled back to 15.36 MHz before being decoded by the LTE stack. To implement the real-time energy sensing on the UEs, we modified the WARPLab firmware (v7.7.1) to incorporate the LAA channel access mechanism (CCA and backoff). The eNB schedules grants to each UE in bursts of three subframes. The UEs on receiving the grant generates three subframes to occupy the appropriate RBs that are allocated by the eNB. However, before transmission the UE senses the channel for clearance. If the energy on the channel is below the stated threshold the UE transmits, else it backs off from transmission. The eNB receives the uplink UE transmissions using the WARPv3 platform. The received LTE I/Q data is then decoded using the MATLAB LTE toolbox to recover the transmitted UE data.

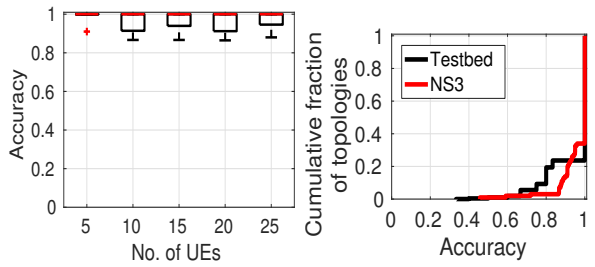
Performance: To generate hidden terminal traffic, we make laptops send UDP data to each other using the `iperf` application. The laptops use dynamic rate selection to ensure that the best bitrate is used at the sender. Each UE is affected by the hidden terminal traffic differently, based on their spatial distance between the hidden terminals and the UE. Each UE transmits 500 frames, with each frame containing three subframes. Each UE performs the LAA channel access (CCA – energy detection and backoff) before transmitting each frame. We operate the eNB in both SISO and 2-user (2 antenna) MU-MIMO configurations. To evaluate the efficacy of Blu, we conduct this experiment over multiple topologies by varying both the locations of the UEs and the hidden terminals.

Figs. 10,12 and Figs. 11,13 show the aggregate throughput and RB utilization gains of Blu over the PF scheduler (Eqn. 1), for both SISO and 2-user MU-MIMO. Increasing the hidden terminal interference (number of hidden terminals per UE), leads to increased asynchronicity, channel un-availability at the UEs and hence under-utilization in the native scheme. However, this provides more room for Blu to leverage interference diversity and *intelligently over-schedule clients to boost RB utilization by 80% and deliver throughput gains as high as 50-80% for both SISO and MU-MIMO.*

4.2 Trace-Based Evaluation

4.2.1 Methodology: To evaluate Blu in large topologies, we run a trace-based emulation of an LTE/WiFi interference network. Since real world LTE traces/datasets are unavailable for public use, we collect data traces from our testbed. We collect two different traces from our test-bed with each trace lasting for 5 minutes. LTE channel traces between each UE and the eNB, and traces of the WiFi interference between the hidden terminals and the UEs.

LTE Channel Traces: We collect channel traces from each UE with data being sent to the eNB that uses four receive antennas. We use the afore-mentioned, modified (with energy-sensing) WARPLab



(a) NS3: Variable no. of UEs **(b) NS-3 and Testbed**

Figure 14: Accuracy of topology inferred by BLU.

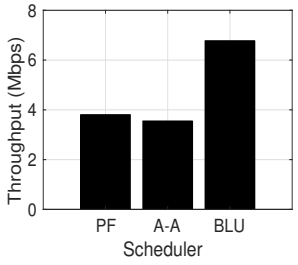


Figure 15: LTE throughput; speculative scheduling.

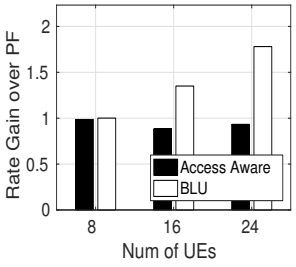


Figure 16: LTE SISO throughput with varying UEs.

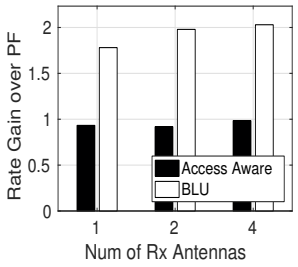


Figure 17: LTE throughput gain with 24 UEs.

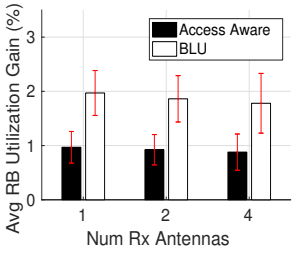


Figure 18: Average RB utilization per subframe.

firmware in our UEs. We configure the WARPLab to continuously transmit a stream of standards-compliant LTE subframes from each UE antenna to the eNB. The eNB decodes these subframes to obtain the corresponding per-subframe CSI for the client during data collection duration.

WiFi Interference Traces: In each of these testbed topologies, we also obtain PHY-layer packet traces of WiFi activity. In each topology, the WARP UEs use the WARP 802.11 Reference design v1.6.2 to overhear the transmitted WiFi packets from the hidden terminals placed at different locations. We time-synchronize the WARP and WiFi devices so that all collected packets in each topology can be globally ordered in time. In total we collect traces for 150 different topologies of UEs and hidden terminals.

We emulate larger topologies by combining the traces collected from different testbed topologies. E.g., for a given UE set-up, we collect time synchronized data traces by moving hidden terminals to different locations. Later, we combine the data traces collected from different hidden terminal locations to emulate a larger spatially separated hidden terminal topology for a given UE set-up. Similarly, we emulate large UE topologies by combining traces from different smaller UE topologies for a given hidden terminal set-up. We thus build large network topologies consisting up to 24 UEs and 36 WiFi hidden terminals mimicking real-world scenarios where UEs and hidden terminals are not close to each other, and the hidden terminal impact on UEs vary from one-another.

Experiment Setup: We generate Release 10 standards compliant 10 MHz LTE subframes, with 3 subframes per burst (identical to that used in the testbed). The interference power seen at each UE at an emulated time instance is determined from the WiFi interference traces collected before. If the channel access is successful (i.e. interference energy below a threshold), the LTE transmission is modulated according to the CSI obtained from the earlier collected LTE traces, and transmitted over the emulated channel to the eNB. Using these data traces we evaluate the various components in BLU.

4.2.2 Topology Inference. We first evaluate BLU's topology inference accuracy for multiple topology-traces collected from both our testbed and NS3 simulator.

NS3 traces: We generate 300 large topology traces from NS3 to stress test BLU's topology inference algorithm. We use the existing NS3-LAA implementation [2] and modify the UE's implementation by enabling it to capture the WiFi traffic in the promiscuous mode. The topologies are generated by varying the number of UEs and WiFi nodes from 5 to 25, in steps of 5. For each topology, we randomly distribute the locations of eNB, UEs and WiFi nodes. The WiFi nodes are transfer UDP traffic to random neighbors at a bitrate chosen by the rate adaptation algorithm. The data from WiFi nodes hidden to eNB, but captured by UEs is used to determine the ground truth topology between HTs and UEs.

Results: WiFi activity traces collected from both testbed and NS3 are used to calculate the channel-access probabilities - *P(i), and P(i,j)*. We use a stringent accuracy metric, calculated as the fraction of the hidden terminals that are inferred with the exact same interference edges to specific UEs, when compared to the ground truth (even a single missing edge will prevent the match). Fig. 14 shows the CDFs of the accuracy of inferred topologies. We see for both testbed and NS3 topologies, *BLU's accuracy of topology inference is 100% for nearly 70% of the cases and above 90% for 90% of the cases, with little room for error in practical network topologies.* Also, Fig. 14(a) shows that increasing the number of UEs (larger topologies) does not affect the accuracy (median ≈ 100%).

4.2.3 Speculative Scheduling. We now compare the joint performance of topology inference and speculative scheduling in BLU with access-aware (AA, Eqn. 5) and native PF schedulers (Eqn. 1), first for SISO, followed by MU-MIMO.

Isolating Performance of Speculative Scheduler: To understand scheduler's performance in isolation, we first consider a SISO set-up with perfect knowledge of the interference from hidden terminals. We consider SISO transmissions between the 24 UEs and the eNB (single antenna UE and eNB) and compute access probabilities $p(i)$ and $p(i,j)$ directly from the traces (instead of computing $p(i,j)$ from the topology). This optimal $p(i,j)$ is used by both AA and BLU schedulers. The schedulers allocate resources amongst

24 UEs, and are limited to scheduling up to 10 UEs per subframe. Fig. 15 shows the throughput achieved by each of the schedulers in the emulated network environment. *Observe that while PF and AA schedulers achieve an aggregate throughput of 3.8 and 3.5Mbps, Blu achieves 6.8Mbps – a 1.8 and 1.9× gain in performance over PF and AA respectively.* This captures the gains possible from speculative scheduling alone, when not impacted by the topology inference component.

Impact of Topology Inference on Scheduler Performance: To understand the impact of inaccuracies in Blu's topology inference on speculative scheduling, we compare Blu's SISO performance, when joint access distributions are estimated from the inferred topology (§3.6) instead of the traces. Fig. 15. Fig. 16 shows that the throughput gains achieved by Blu over PF for SISO is close to that (1.8×) in Fig. 15 (the 24 UE scenario), indicating that *(low) inaccuracies in Blu's topology inference contribute to a minimal impact on its scheduling performance.* The gains are more with larger number of UEs, as it provides more room for leveraging interference diversity to over-schedule appropriate clients.
Note: *Computing joint access distribution of clients directly from the traces in* real-time *is impractical even for a 2-user MU-MIMO (complexity scales exponentially with M and number of UEs)*, and is observed in our set-up. This provides additional advantage, in addition to measurement overhead, for Blu's topology (inference)-driven approach to speculative scheduling.

MU-MIMO Scheduler Performance: Uplink MU-MIMO transmissions increase LTE's spectrum utilization. Here, the eNB leverages its multiple antennas to schedule a larger number of concurrent uplink UEs on the same RB.

Fig. 17 shows the throughput gains achieved by Blu and the AA scheduler over PF. Observe that *as MIMO degrees-of-freedom or concurrency (M) increases, Blu achieves a larger gain of 2x (for 4 antenna MU-MIMO) over PF and AA schedulers.* This is because with greater DoFs, the potential for a large number of scheduled UEs to not use their grants (due to channel access conflicts), also increases. Blu ensures that the set of UEs chosen for over-scheduling maximizes the probability of utilizing the DoFs in each RB, whose benefits increase for larger *M*.

Spectrum Efficiency: Fig 18 shows the gain in RB utilization (spectrum efficiency), achieved by Blu over PF and AA. All RBs are allocated to UEs in every subframe to fully utilize the available resources. With conventional UL transmissions, un-coordinated and unknown channel availability at UEs results in approximately half the assigned RBs un-utilized in each subframe. On the other hand, *due to its accurate estimation of interference topology, Blu almost doubles RB utilization over PF* for both SISO and MU-MIMO configurations. The AA, being unable to compensate for under-utilization during access, cannot improve spectrum utilization.

5 RELATED WORK

LTE in Unlicensed Spectrum: Existing works focus on the fair co-existence between LTE and WiFi in un-licensed spectrum [5, 10, 11, 20, 21, 27, 31], where the goal is to alleviate the asymmetries in channel sensing and access between the two technologies. With research and standards [4, 9] making reasonable progress in this direction, Blu focuses on the next critical problem, namely the operational efficiency of LTE in unlicensed spectrum and tackle the fundamental conflict that arises between concurrent LTE transmissions (increased gains) and asynchronous access (coexistence).

LTE Scheduling: LTE scheduling decisions that deal with the allocation of time-frequency-antenna resources at sub-frame granularity, are deterministic and combinatorial in nature. They have received a lot of attention in the last several years [6, 7, 15, 16, 24, 29]. Blu leverages stochastic tools and applies them intelligently to enable speculative (yet deterministic) over-scheduling for increased utilization, while retaining the underlying principles of LTE's PF scheduling.

Non-orthogonal Multiple Access: NOMA [13, 18, 28] is a future access technique that leverages successive interference cancellation and power control to schedule multiple clients on the same UL transmission resource. Being designed for licensed spectrum, the benefits from Blu's speculative scheduler in counteracting the effects of asynchronous interference in unlicensed spectrum, will apply to NOMA too.

Topology Inference: Topology inference has been studied primarily in the context of wired networks [8, 12, 22]. The goal is largely to identify lossy links [8, 22] or high level structure of the topology (especially multicast) [12] using stochastic sampling techniques (e.g. Markov Chain Monte Carlo [14]). These have been applied to infer limited information such as connectivity (given node information) in sensor networks [19] and communication patterns (given topology) in WLANs [30]. While the underlying stochastic techniques could be instrumented for our problem (which we have), Blu leverages the structure of LTE's cell and mechanisms to design an end-end, deterministic solution that blue-prints the complete interference topology with scalable overhead.

6 CONCLUSION

We tackled a fundamental challenge in realizing the concurrency gains offered by LTE's synchronous transmissions in the uplink, when deployed in unlicensed spectrum, characterized by un-predictable, asynchronous interference. We proposed Blu, a novel scheduling system that delivers the concurrency gains through intelligent, over-scheduling of clients in LTE base stations. Blu's intelligence lies in leveraging the interference diversity across clients – information that is measured in a scalable manner by blue-printing the source of the interference itself and its dependencies on clients. Blu's realization in LTE base stations today reveals substantials gains to spectrum utilization and throughput for LTE uplink access in unlicensed spectrum.

7 ACKNOWLEDGEMENT

We would like to thank all the anonymous reviewers for their insightful comments and feedback.

REFERENCES

[1] [n. d.]. Matlab - LTE Toolbox. ([n. d.]).
[2] [n. d.]. NS3 - LAA LAA-WiFi-Coexistence. ([n. d.]).
[3] 3GPP Release 14, 36.213. 2016. Channel Access Procedures for LAA. (2016).
[4] 3GPP Technical Report 36.889. 2015. Feasibility Study on Licensed-Assisted Access to Unlicensed Spectrum. (2015).
[5] E. Almeida, A. M. Cavalcante, R. C. D. Paiva, F. S. Chaves, F. M. Abinader, R. D. Vieira, S. Choudhury, E. Tuomaala, and K. Doppler. 2013. Enabling LTE/WiFi coexistence by LTE blank subframe allocation. In *Communications (ICC), 2013*

IEEE International Conference on. 5083–5088. https://doi.org/10.1109/ICC.2013.6655388

[6] M. Andrews and L. Zhang. 2007. Scheduling algorithms for multi-carrier wireless data systems. In *ACM MOBICOM*.

[7] B. Bai, W. Chen, Z. Cao, and K. B. Letaief. 2010. Uplink Cross-Layer Scheduling with Differential QoS Requirements in OFDMA Systems. *EURASIP Journal on Wireless Communications and Networking* 2010, 168357 (2010).

[8] R. Caceres, N. G. Duffield, J. Horowitz, D. Towsley, and T. Bu. 1999. Multicastbased inference of network internal characteristics. In *IEEE INFOCOM*.

[9] CBRS Alliance. 2016. Citizen Broadband Radio Service. (2016).

[10] E. Chai, K. Sundaresan, M. Khojastepour, and S. Raagarajan. 2016. LTE in Unlicensed Spectrum: Are We There Yet?. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom '16)*.

[11] F. S. Chaves, E. P. L. Almeida, R. D. Vieira, A. M. Cavalcante, F. M. Abinader, S. Choudhury, and K. Doppler. 2013. LTE UL Power Control for the Improvement of LTE/Wi-Fi Coexistence. In *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th*. 1–6. https://doi.org/10.1109/VTCFall.2013.6692275

[12] Mark Coates, Rui Castro, Robert Nowak, Manik Gadhiok, Ryan King, and Yolanda Tsang. 2002. Maximum Likelihood Network Topology Identification from Edge-based Unicast Measurements. In *Proceedings of the 2002 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '02)*. 11–20.

[13] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C-L. I, and H. V. Poor. 2016. Application of Non-orthogonal Multiple Access in LTE and 5G Networks. In *Arxiv*.

[14] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. 1996. Markov Chain Monte Carlo in Practice. *Chapman and Hall* (1996).

[15] S.B. Lee, S. Choudhury, A. Khoshnevis, S. Xu, and S. Lu. 2009. Downlink MIMO with Frequency-Domain Packet Scheduling for 3GPP LTE. In *IEEE INFOCOM*.

[16] S.B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu. 2009. Proportional Fair Frequency-Domain Packet Scheduling for 3GPP LTE Uplink. In *IEEE INFOCOM Mini-conference*.

[17] LTE-U Forum White Paper. 2014. LTE-U. (2014).

[18] F-L. Luo and C. Zhang. 2016. Non-Orthogonal Multiple Access (NOMA): Concept and Design. In *Signal Processing for 5G: Algorithms and Implementations*.

[19] D. Marinakis and G. Dudek. 2006. A Practical Algorithm for Network Topology Inference. In *IEEE International Conference on Robotics and Automation*.

[20] A. Mukherjee, Jung-Fu Cheng, S. Falahati, L. Falconetti, A. Furuskar, B. Godana, Du Ho Kang, H. Koorapaty, D. Larsson, and Yu Yang. 2015. System architecture and coexistence evaluation of licensed-assisted access LTE with IEEE 802.11. In *Communication Workshop (ICCW), 2015 IEEE International Conference on*. 2350–2355. https://doi.org/10.1109/ICCW.2015.7247532

[21] T. NihtilÃď, V. Tykhomyrov, O. Alanen, M. A. Uusitalo, A. Sorri, M. Moisio, S. Iraji, R. Ratasuk, and N. Mangalvedhe. 2013. System performance of LTE and IEEE 802.11 coexisting on a shared frequency band. In *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*. 1038–1043. https://doi.org/10.1109/WCNC.2013.6554707

[22] V. Padmanabhan, L. Qui, and H. Wang. 2002. Passive network tomography using Bayesian inference. In *2nd ACM Workshop on Internet Measurements*.

[23] Qualcomm White Paper. [n. d.]. License Assisted Access (LAA). https://www.qualcomm.com/invention/technologies/lte/laa. ([n. d.]).

[24] A. Pokhariyal, G. Monghal, K. I. Pedersen, P. E. Mogensen, I. Z. Kovacs, C. Rosa, and T. E. Kolding. 2007. Frequency domain packet scheduling under fractional load for the UTRAN LTE downlink. In *IEEE VTC*.

[25] Qualcomm White Paper. [n. d.]. MulteFire. ([n. d.]).

[26] Qualcomm White Paper. 2016. MulteFire Specification. (2016).

[27] Shweta Sagari, Samuel Baysting, Dola Saha, Ivan Seskar, Wade Trappe, and Dipankar Raychaudhuri. 2015. Coordinated dynamic spectrum management of LTE-U and Wi-Fi networks. In *Dynamic Spectrum Access Networks (DySPAN), 2015 IEEE International Symposium on*. IEEE, 209–220.

[28] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi. 2013. Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access. In *IEEE Vehicular Technology Conference (VTC Spring)*.

[29] G. Song and Y. Li. 2005. Cross-layer optimization for OFDM wireless networks - Part I: Theoretical Framework. *IEEE Transactions on Wireless Communications* 4, 2 (Mar 2005).

[30] P. Tilghman and D. Rosenbluth. 2013. Inferring Wireless Communications Links and Network Topology from Externals using Granger Causality. In *IEEE MILCOM*.

[31] Sangki Yun and Lili Qiu. 2015. Supporting WiFi and LTE co-existence. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*. 810–818. https://doi.org/10.1109/INFOCOM.2015.7218451