Downlink Decoding Based Accurate Measurement of LTE Spectrum Tenancy

Rui Zou, *Member, IEEE,* and Wenye Wang, *Fellow, IEEE*

Abstract—Mobile networks are embracing Dynamic Spectrum Access (DSA) to unleash data capacities of spectrum holes caused by tidal traffic. Being the largest mobile system, LTE has been standardized to operate in the DSA mode where the knowledge on the spectrum tenancy of LTE systems are required. Although there exists rich literature on spectrum sensing, measurement and modeling, they cannot satisfy the needs of accurately acquiring the spectrum tenancy of LTE systems. This is because most traditional measurements only provide inaccurate tenancy in coarse granularities, and therefore models built upon them are defective. To enable the precise discovery of spectrum assignments of an LTE cell from an outsider perspective, we build U-CIMAN to *Un*C*over* spectrum occupancy and user *I*nformation in *M*obile *A*ccess *N*etworks. The LTE protocol fields parsed by U-CIMAN not only accurately reveal the spectrum occupancy at the same granularity with LTE scheduling, but also provide important details associated with spectrum usage, i.e., rough user locations and traffic types. Besides insightful observations based on measurements enabled by U-CIMAN, we propose to characterize LTE spectrum occupancy using Vector Autoregression that captures the statistical distributions of spectrum tenancy intervals in multiple channels and the correlations among them.

Index Terms—LTE, dynamic spectrum access, measurements, software defined radio, test bed, modeling

1 Introduction

Due to the low utilization of spectrum resources caused by exclusive spectrum assignment [1], Dynamic Spectrum Access (DSA) has been proposed as the new paradigm for 5G/6G systems and beyond [2]. It is foreseeable that a potpourri of various radio access technologies (RATs) which are either originally developed for exclusive or shared spectrum usage, are expected to collocate on the same spectrum bands with equal usage rights. This trend in standardization activities has been pioneered by efforts, such as MulteFire, Licensed Assisted Access (LAA), and LTE-U [3], [4], [5], to enable the most advance mobile system in production deployment, namely the 4G LTE system, to work in shared spectrum bands, for instance the ISM 2.4 GHz and 5 GHz bands. These standards have already been supported by leading commercial cell phone modems, such as Snapdragon X55. In this setting, LTE systems need to sense the spectrum usage of other systems which in turn desire the spectrum tenancy knowledge of LTE cells, so the collocated wireless systems with various RATs are able to function properly without interfering with one another.

To avoid interference and improve spectrum utilization, the ideal spectrum sensing accuracy should reach the time and frequency granularity of the radio resource unit of the RATs' scheduling schemes. This requirement on spectrum sensing accuracy is a direct outcome of the scheduling schemes. For example, if two Frequency Division Duplex (FDD) LTE cells are sharing the spectrum resources, then they should sense the spectrum usage of the other party in the basic unit of LTE scheduling, also known as Resource Blocks (RBs) that are spectrum slices of 1 millisecond (ms) by 180 kHz. In this way, the two cognizant LTE cells are able to learn the spectrum holes left by each other, and make use of them without hampering other's transmissions. Because LTE systems reassign the occupancy of RBs every 1 ms, the tenancy and the vacancy of spectrum slices also changes at the same pace, entailing the spectrum sensing at the same time frequency granularity if all the spectrum holes are to be identified correctly.

Although the importance of sensing spectrum occupancy has long been identified in the existing DSA literature [6], [7], [8], [9], prior spectrum sensing works cannot achieve the desired accuracy. According to Table VI in [10], the results of existing measurement campaigns cannot achieve the time and frequency granularities of LTE scheduling since the sweep time is mostly tens of seconds. The low measurement resolutions are determined by their spectrum tenancy detection methods. One of the most widely methods is the energy detection which is achieved by deploying a spectrum analyzer to measure the power levels within the spectrum bins of certain sizes. This method is only appropriate if the spectrum tenancy remains the same within the sweep time, which is not true for LTE bands. For a 10 MHz LTE system, the sweep time is the total bandwidth divided by the square of the bin size, usually chosen between 10 kHz and 30 kHz according to application manuals of spectrum analyzer manufacturers and detailed measurement reports on LTE systems [11], resulting in 11.1 to 100 ms, much longer than the 1 ms scheduling interval after which the spectrum usage changes. Despite the existence of other spectrum tenancy measurement methods such as matched filter detection and cyclostationary feature detection [12], they are based on similar principles with the energy detection method, i.e., scanning through the entire spectrum bands or searching signal patterns in one spectrum slice after another, so their measurement results cannot reach the fine time resolution required for the studies on LTE spectrum tenancy.

Since the existing measurement methods and datasets cannot show the true characteristics of LTE spectrum usage, the models developed based on those measurement cannot be reflective of the actual spectrum tenancy patterns. For

example, the time domain data samples are collected every 1 μs, or 1 MHz, in [13], which is much lower than the minimum sampling rate of 15.36 MHz required for processing signals in an LTE cell with 10 MHz bandwidth. Thus, the data collected in this campaign is useful for examining the existence of low bandwidth signals, but is not able to reveal the spectrum usage of LTE systems. Another well known dataset collected by the energy detection method with 1.8 seconds sweep time is adopted in [14]. Since LTE spectrum tenancy may have changed as many as 1800 times in this time interval, this dataset cannot serve the purpose of building spectrum tenancy models of LTE systems. As we will show in the measurement results, most of the occupancy times of LTE channels do not last longer than 1 second, so the 1.8 seconds sweep time will hugely mislead the results, if such dataset is adopted for LTE spectrum tenancy studies. Since there are no accurate spectrum usage measurement method, data, and models to facilitate the participation of LTE systems in the DSA regime, we identify and answer two research questions in this paper, 1) *how to accurately measure spectrum tenancy of an LTE cell in fine granularity*, and 2) *what model characterizes the spectrum tenancy of an LTE cell*.

To answer the first research question, we propose to measure the spectrum tenancy of an LTE cell by parsing downlink control messages and decoding raw bytes of user data. Inspired by the LTE sniffing technologies and equipped with the emerging Software Defined Radio (SDR) hardware and software libraries [15], [16], [17], [18], [19], we develop U-CIMAN to *U*n*C*over spectrum and user *I*nformation in *M*obile *A*ccess *N*etworks, which decodes the downlink control messages and raw bytes of user data transmitted from an LTE base station, or an eNB (Evolved Node B). In this way, the cell-wide spectrum occupancy can be parsed from the messages carrying spectrum resource assignments aired by the eNB. Thus, U-CIMAN achieves accurate spectrum tenancy at the same granularities with LTE scheduling, i.e., the frequency resolution of 180 kHz and the time granularity of 1 ms. The protocol fields decoded by U-CIMAN also facilitates the inference of important facts related to spectrum tenancy, i.e., the rough user locations and the traffic types. Compared with pioneering LTE deciphering works which target downlink control messages only [15], [16], U-CIMAN further decodes the raw user data bytes, and exploits the Time Advance (TA) and packet size headers for spectrum tenancy studies. TA values and packet sizes are highly indicative of the locations and traffic types of LTE devices, respectively, which have appeared as key assumptions in many DSA proposals [20], [21]. How the TA and packet sizes in bytes relate to user locations and traffic types are explained and validated in Section 2. It is worth noting that the payload of user data is protected by LTE encryption, so obtaining raw bytes in enciphered form causes no security or privacy issues.

With the accurate LTE spectrum tenancy obtained by U-CIMAN, we characterize the LTE spectrum occupancy with both on/off model and Vector Autoregression (VAR). VAR outperforms on/off model according to our analysis where their performance is evaluated from three aspects, goodness of fit to the distributions of measured on-time, off-time, and interval lengths, correlations among adjacent channels, and correlations between adjacent idle and busy periods.

As mobile access networks are embracing the DSA paradigm, this paper timely enables the understanding of LTE spectrum tenancy from an outsider point of view. Equipped with U-CIMAN, we measure the spectrum occupancy of a commercial LTE cell for four months, and make insightful observations. For example, spectrum tenancy is upper bounded to around 10^4 ms, which is consistent with practical systems, but in contrast to analytic results of heavy-tailed distributions [22], [23]; the Modulation and Coding Scheme (MCS) of the spectrum slices are highly indicative of the occupancy status in the next time slot. We compare the performance of VAR and on/off model in characterizing LTE spectrum occupancy, and the proposed VAR model outperforms the widely used on/off model. The main contributions are summarized as below.

1) We design and implement U-CIMAN to decode LTE downlink control messages and user data bytes. The performance of U-CIMAN is validated in two setups. One is our lab environment where a working LTE system is realized with Amarisoft with accessible logs [24], while in the other scenario U-CIMAN is applied to decode data fields in a commercial LTE cell with realistic user mobility.
2) Utilizing protocol fields decoded by U-CIMAN, we conduct accurate and detailed spectrum occupancy studies on a commercial LTE cell. We show that the observed LTE spectrum usage characteristics are substantially impacted by the measurement granularity, and the accurate results enabled by U-CIMAN are key to analyzing LTE spectrum tenancy.
3) We find that VAR outperforms the widely used on/off model in characterizing LTE spectrum occupancy, due to its superior capabilities to capture the distributions of busy and idle time lengths, and occupancy correlations among adjacent channels.

The rest of this paper is organized as follows. The design and implementation details of U-CIMAN are described in Section 2. In Section 3, thorough performance validations of U-CIMAN are presented, as well as its potential applications. Then, we explain the measurement results that are characterized by the on/off and VAR models, and the observations in Section 4. Related work is discussed in Section 5. Lastly, the paper is concluded in Section 6.

2 U-CIMAN Design and Implementation

This section describes the LTE preliminaries, as well as the design and the implementation of U-CIMAN, the accurate measurement tool for LTE spectrum tenancy.

2.1 LTE Preliminaries

Since U-CIMAN measures LTE spectrum tenancy by decoding downlink information, we introduce relevant LTE domain knowledge before the system design. We explain why some messages aired by eNBs are susceptible to eavesdropping without causing privacy or security concerns, and the three chosen LTE data fields for spectrum tenancy measurement and the reasons behind the choice.

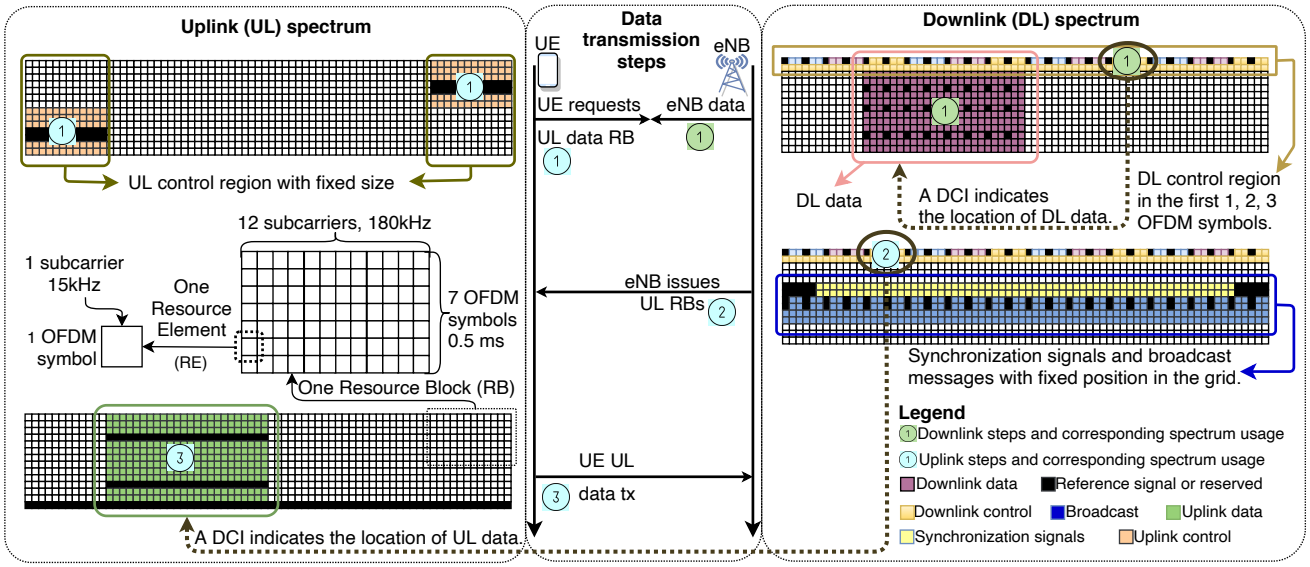

Fig. 1. The structures of LTE spectrum resources and the steps of the data transmissions.

2.1.1 Two types of unencrypted fields

Though LTE provides an integral encryption mechanism for both the control and the data planes, two types of LTE data cannot be encrypted. *Type I* unencrypted data is left in clear text because they are transmitted before the encryption setup. There are several steps for an LTE User Equipment (UE) to undergo before setting up ciphered data exchanges with the network, so the downlink information transmitted before the completion of encryption setup has to be in clear text. *Type II* unencrypted data is found in the messages or headers generated in the protocol layers under the Packet Data Convergence Protocol (PDCP) layer that is responsible for encryption [25], so the information in those data fields generated below the PDCP layer cannot be protected.

Due to the existence of the two types of unprotected data fields, the messages in those fields can be understood by *any devices* following LTE specifications once the corresponding raw data are decoded. Since both types of unprotected messages either carry information for initial access of an LTE cell or pertain only to operations below PDCP layer, their leakage can hardly be related to specific users whose identifiers in the two types of unencrypted information are represented as rapidly changing Radio Network Temporary Identifiers (RNTIs). Though the unencrypted messages cannot be linked to specific users, the information contained therein can still be of vital importance. For example, other DSA systems only need to know the spectrum usage of collocated LTE devices, but not the identities of LTE users.

The three types of data fields decoded by U-CIMAN to reveal LTE spectrum tenancy are the Downlink Control Information (DCI), TA, and packet sizes, all of which are *Type II* unencrypted data. How they are related to spectrum tenancy and how to decode them will be explained in the rest of this section.

2.1.2 Structure of LTE spectrum tenancy

Since the goal of U-CIMAN is to measure spectrum tenancy, the structure of LTE spectrum usage is briefly explained. In the time domain, an LTE *subframe* is 1 ms, which is the time interval for an eNB to schedule the spectrum resources. Thus, a subframe is also known as a Transmission Time Interval (TTI). In each subframe, the spectrum resources in two dimensional frequency-time grids are divided into RBs that are the smallest unit of eNB resource assignment [26]. As shown in the left part of Fig. 1, an RB is 180 kHz by 0.5 ms, and it comprises 12 subcarriers each of which typically carries 7 symbols. The smallest spectrum resource is called Resource Element (RE) that carries one symbol on one subcarrier [27].

The right part of Fig. 1 illustrates the resource structures of two typical LTE downlink subframes. For all downlink subframes, the first one to three REs are for various control channels in the downlink, and the rest carry user data. Downlink synchronization signals and physical broadcast channel are transmitted by REs in fixed positions in the center of a subframe at regular intervals. Different from downlink subframes, the data and the control regions in the uplink are split by the frequency, as shown in the left of Fig. 1. The RBs in the middle of the uplink frequency range carry data, while the RBs on the two ends bear control messages. Since the control messages and broadcasts happen at fixed positions in the time-frequency grid, the key to measure LTE spectrum tenancy is locating the RBs dynamically scheduled every TTI for user data transmission.

2.1.3 Transmission steps of LTE user data

Central to the dynamic scheduling of LTE data RBs is the DCI carried in the downlink control region, or Physical Downlink Control CHannel (PDCCH), in the first one to three REs in a subframe. The roles played by unencrypted DCIs in LTE spectrum resource allocations are illustrated in Fig. 1 where the time grows vertically downwards and the frequency increases to the right horizontally. The uplink data transmission steps for an actively connected UE are numbered by light blue circles. First, a UE that intends to transmit uplink data sends its request for spectrum resources. The request is carried by uplink control channels that do not require dynamic allocations per TTI. After receiving the request, the eNB schedules the uplink spectrum resources and puts the decision in a DCI to inform the UE.

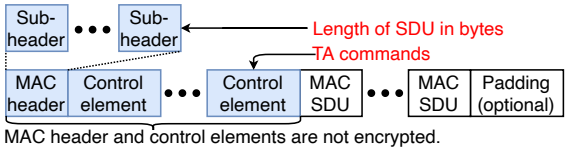

Fig. 2. Locations of the unencrypted packet sizes and the TA commands.

DCIs indicate the spectrum resources and MCS for transmissions. Finally, the UE finds its DCI, and then transmits in accordance with the information in the DCI.

For downlink transmissions, only one step is needed. As shown by the circled green number in Fig. 1, the eNB puts the DCI in control channel region and the corresponding data in the data RBs. Similar to DCIs for uplink transmission, DCIs for downlink data inform UEs where data is and the MCSs for demodulation. To receive downlink data, a UE blindly searches for its DCIs in the small downlink control region. If DCIs pointing to downlink data are found, UEs locate and decode the RBs according to the resource assignment and MCS values in the DCIs.

2.1.4 The packet size and the TA commands

In addition to the DCIs that specify which LTE RBs are occupied, there are another two data fields that show essential details associated with the used spectrum resources, how much data is carried by the spectrum slices and TA of the UEs. As shown in Fig. 2, the length of Media Access Control (MAC) Service Data Unit (SDU) in bytes is contained in the MAC header, and the TA commands are in the control element field. Though they are part of the user data, they are generated below the PDCP layer for LTE encryption, so they are in clear texts. As we will show later in the paper, packet sizes are indicative of user applications, the knowledge of which can serve as the enabling function for many DSA proposals that are traffic pattern or application dependent [21], [28], [29]. On the other hand, TA values are related to user locations that are able to satisfy the needs of many location based DSA algorithms [20], [30].

2.2 The design of U-CIMAN

From the previous subsection, we have identified the three data fields that accurately specify the LTE spectrum usage with associated details, DCIs, TA values, and packet sizes. Achieving these data fields as a user inside the LTE cell is straightforward, since each UE has all the input data required for decoding its own control and user data. However, decoding the DCIs, TA values, and packet sizes of all the UEs as an outsider faces the challenge of missing important inputs in the decoding procedures. In this subsection, we first introduce the overall steps of U-CIMAN decoding, and highlight the key inputs that are unknown to outsiders. Then, the designs to uncover those inputs are explained.

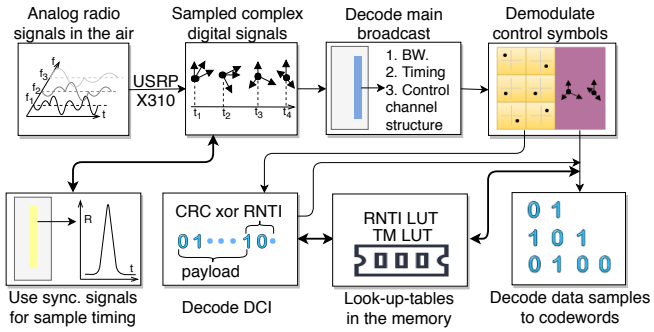

Fig. 3. Overall data processing steps of U-CIMAN.

2.2.1 Overall work flow

Fig. 3 depicts the high level work flow of U-CIMAN, from analog radio signals to decoded raw bytes of user data. First, the SDR front end Universal Software Radio Peripheral (USRP) X310 converts analog radio signals to complex samples, and sends them to the host computer. U-CIMAN in the host computer utilizes LTE downlink synchronization signals to update sampling time and frequency range of the SDR. After getting time and frequency synchronized with the eNB, U-CIMAN decodes main broadcast messages to discover system time, bandwidth, and the structure of downlink control channel which is then employed to locate REs carrying downlink control data. Because the Transmission Modes (TMs), i.e., the multiple antenna schemes of control channels are known from decoding main broadcast messages, and the modulation scheme of control channels is fixed as Quadrature Phase Shift Keying (QPSK), U-CIMAN is able to decode the complex samples of DCIs into raw bytes after obtaining the positions of the control channel REs. So far, these steps pose no challenges to outsiders to the cell, because synchronization and main broadcast messages are designed to be decodable for any devices executing the corresponding LTE routines.

Challenges show up in later steps which require two user specific configurations, RNTI and TM, to validate DCI messages and decode raw bytes of downlink user data. For a normal UE, the eNB assigns an RNTI to the UE during the random access, and the RNTI is used for the validation of DCI decoding and the generation of scrambling sequences for user data protection against burst errors. Different from downlink control messages transmitted in narrow range of spots in the RB grid, RNTI assignments are irregular and not always adopted by UEs, making the direct decoding inefficient. Another challenge is obtaining TMs, because they are configured by the network side and transmitted to UEs through encrypted messages. With unknown multiple antenna configurations, i.e., the TMs, user data bytes cannot be decoded even if the corresponding DCI is attained. The last three steps in the second row of Fig. 3 and Fig. 4 demonstrate how U-CIMAN obtains RNTIs and TMs to decode control messages and user data bytes, which is explained in later paragraphs in this subsection.

2.2.2 Reliable decoding of RNTI and DCI

Though there are some prior works on RNTI and DCI decoding [15], [16], U-CIMAN has two major distinctions from them. First, U-CIMAN ensures the validity of the decoded RNTIs. Moreover, U-CIMAN further decodes the raw bytes of user data in addition to RNTI, which is proposed for the first time. In [15], the DCI-based RNTI-derivation method is proposed, where the trailing bits after DCI payload are exploited. Because the last two bytes of DCIs are the XOR of the RNTI and the CRC checksum of DCI payload, as shown in Fig. 3, RNTI can be obtained by computing the checksum

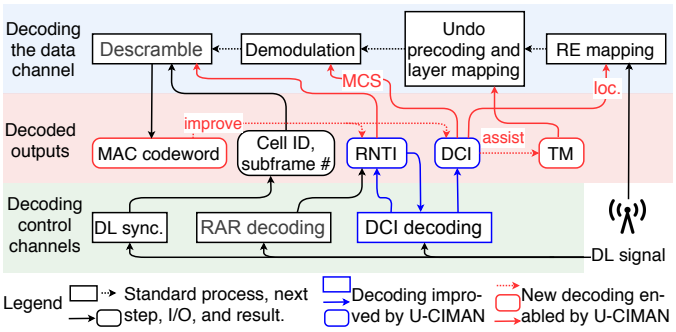

Fig. 4. U-CIMAN design to decode raw bytes of user data.

and then XORing it with the last two bytes, assuming that the entire DCI is correctly decoded. This method has been verified to suffer from low reliability, because the internal LTE error detection mechanism for DCI is forfeited [31]. Another decoding method is proposed in [16], where the RNTIs are decoded in the random access stage when they are initially assigned and transmitted to users in Random Access Response (RAR) messages. However, the RNTIs contained in RARs are temporary, and may not necessarily be adopted by UEs [32].

To improve the existing RNTI decoding schemes, U-CIMAN first collects a pool of RNTIs using both the methods, decoding DCIs from RARs and reverse engineering the checksum fields of DCIs. Since these results contain invalid RNTIs due to the reasons discussed in the previous paragraph, we further filter the collected pool of RNTIs by applying the them to the decoding of corresponding downlink user data. For an RNTI decoded from RAR messages or DCIs, if there is no decodable user data that corresponds to the RNTI in the next 10 ms, the RNTI is considered invalid. This is because RNTI is used as an input for descrambling user data as shown in Fig. 4, which cannot output correct codewords if the RNTIs are erroneous. This screening process validates RNTIs derived from RAR messages or downlink DCIs, eliminating the invalid ones. As shown in Fig. 3, the RNTI Look-Up Table (LUT) stores initial results of decoded RNTIs and the ones that have been validated. When U-CIMAN decodes DCIs and user data bytes, it first tries the stored RNTIs before deducing them. If the RNTIs stored in the LUT do not yield successful decoding of DCIs or user data for 10 consecutive TTIs, the RNTIs are removed from the LUT; otherwise, the 10 ms timer is refreshed.

2.2.3 *Decoding raw bytes of user data*

The design to decode raw bytes of user data is illustrated in Fig. 4. According to the figure, the decoding of LTE downlink data channel can be achieved by standard procedures shown in black, but U-CIMAN needs to obtain the inputs to data decoding in ways different from normal UEs. First, U-CIMAN locates REs for a codeword in MAC layer according to decoded DCIs. Then, the corresponding TM is required to undo precoding and layer mapping. U-CIMAN utilizes MCS parsed from DCI to conduct demodulation whose output is then descrambled with a sequence based on cell ID, subframe number, and RNTI. Among the inputs to the processing chain, how the decoding of RNTIs and DCIs are improved by U-CIMAN over existing methods has been explained, so the ones awaiting expositions are cell ID, subframe number, and TM.

As shown in Fig. 4, cell ID and subframe number are achieved by decoding unencrypted downlink synchronization signals, so the key is how to achieve TM. Unlike RNTIs or DCIs, TMs are configured in an enciphered downlink control message by eNB, decoding it over the air interface is impossible without breaking LTE encryption. To uncover the TMs, we utilize two LTE mechanisms to deduce TMs more efficiently than the brute-force method of trying all possible TM values every time. One is the mapping between TMs and the formats of DCI as summarized in Table 9.2 in [26]. Since many DCI formats map to a very limited set of possible TMs, we first use DCI type to reduce the size of TM search space. The other LTE mechanism which helps TM inference is that TMs are reconfigured at a much lower rate than that of RNTIs, so we store the TMs corresponding to RNTIs in the LUT as well for later lookup. In this way, U-CIMAN obtains the TMs efficiently.

Having achieved the RNTIs, DCIs, and TMs, U-CIMAN decodes user data bytes from the complex samples in the same way as a normal UE. Though most decoded user data bytes are encrypted, headers added below the PDCP layer are in clear texts and can be parsed by U-CIMAN. Thus, this design realizes our goal of accurately detecting the spectrum usage of all users in an LTE cell.

2.2.4 *Timeliness of U-CIMAN decoding*

Since U-CIMAN decodes all the DCIs and user data bytes in an LTE cell in real time, the timeliness of the decoding is worth discussions. Based on the estimation in section 9.3.5.5 of the LTE canon [26], the data rate requirements for decoding all the DCIs is proportional to the system bandwidth. For a an LTE system with K MHz bandwidth, the processing time for blindly searching and decoding all possible DCIs is equivalent to receiving a data stream at the rate of about $0.4K$ Mbps, which is a small overhead compared to the maximum LTE data rate of hundreds of megahertz. In terms of data decoding, the U-CIMAN design adds no extra requirements to the existing UE processing chain, because a normal UE may also use up to the total amount of RBs. Thus, the U-CIMAN design does not require a data processing chain capable of receiving a much higher rate than a normal UE, in order to accurately measure LTE spectrum tenancy. As long as the design is implemented on hardware that provides the similar processing capability of the receiver chain of an ordinary UE, U-CIMAN is able to measure in real time.

2.3 The implementation of U-CIMAN

We implement the overall data flow as described in the previous subsection, as well as the solutions for efficiently finding the two key inputs, RNTI and TM. The implementation of U-CIMAN is facilitated by the open source LTE library srsLTE [19], and the functions for DCI decoding from OWL [16], because standard processing routines can be safely reused, as shown in Fig. 4 by the black arrows and boxes. Besides implementing the main U-CIMAN design, we parse data fields relevant to LTE spectrum occupancy, and record them in files. Three types of data fields are

recorded, resource assignments, TA values, and packet sizes. Resource assignment fields provide the fine granularity of spectrum measurement. TA fields indicate rough locations of users, and the size of the physical layer packets, or codewords, reflects user traffic types.

TABLE 1
Descriptions for recorded data fields.

SFN	System frame number, in $\{0, \cdots, 1023\}$
Subframe	Index of LTE subframe, in $\{0, \cdots, 9\}$
RNTI	User identifier, in $\{0, \cdots, 65535\}$
Direction	Resource assignment direction, in $\{0, 1\}$
MCS	Modulation, coding scheme, in $\{0, \cdots, 31\}$
Total	Total number of RBs, in $\{1, \cdots, 100\}$
RA type	Resource assignment types, in $\{0, 1, 2\}$
RA1	The first field indicating RB assignment
RA2	The second field indicating RB assignment
RA3	The third field indicating RB assignment
CFI	Size of PDCCH, in $\{1, 2, 3\}$
RAR TA	TA values in RAR, in $\{0, \cdots, 1282\}$
TA	TA updates, in $\{0, \cdots, 63\}$
Length	Packet size in bytes, $\{0, \cdots, 65535\}$

2.3.1 Data fields in U-CIMAN records

U-CIMAN generates one data record per DCI, and all data fields in a record are listed in Table 1. The first 11 fields for spectrum tenancy are decoded from DCIs or other physical layer control channels. Most of them have been explained, except the four whose names include 'RA'. These fields describe the assigned RBs in the same way as DCIs do. Because there are three types of spectrum resource allocation in LTE, and each type adopts different data structures to indicate the assigned RBs, we use four 'RA' fields to record RB assignments. The field for resource allocation type, 'RA type', shows the type of LTE resource allocation, taking values from 0, 1, or 2. For RA type 0, 'RA 1' field is a bitmap indicating the allocated RBs, and 'RA 2' and 'RA 3' fields are left unused. For RA type 1, 'RA 1' field is a different type of bitmap that requires additional information called 'subset' and 'shift' that are stored in fields 'RA 2' and 'RA 3' to describe RB assignments. For RA Type 2, U-CIMAN stores the starting RB position in field 'RA 1' and the number of assigned RBs in 'RA 2'. Since these three types of spectrum resource assignments are standardized by LTE specifications, interested readers can find them in [33], to get the details on how to determine the exact index of occupied RBs in each allocation type.

Within the TA category, there are two types of data fields, the 'RAR TA' and the 'TA'. 'RAR TA' is the initial TA value obtained from RAR messages. 'TA' field is the TA update value decoded from the unencrypted headers in downlink user data. The payload size is the number of bytes in decoded downlink user packets at the physical layer. The top 11 fields in Table 1 are present for every record, while the others may not be. 'RAR TA' is only for RAR messages. TA updates are conducted by eNB at regular time intervals, so they are in the headers of downlink packets near the TA update time. The 'length' field is nonempty when the DCI points to a downlink data packet whose raw bytes are decoded successfully.

2.3.2 Hardware and software environment

U-CIMAN is implemented in the user space of a Linux computer, as depicted in Fig. 5(a). The four main function

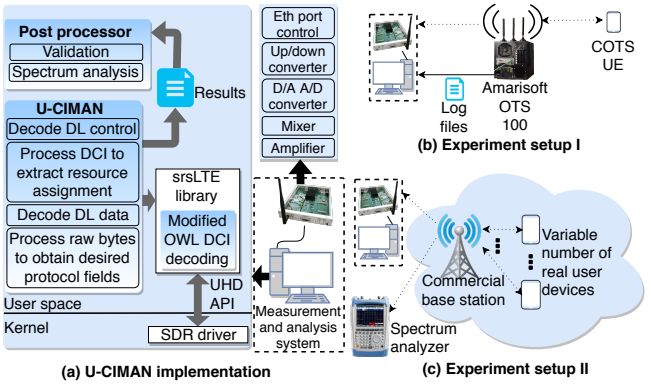

Fig. 5. Implementation and experiment setups.

blocks of U-CIMAN are decoding the raw bytes of DCI, parsing DCI messages, decoding raw bytes of user data, and parsing the unencrypted headers of user data. U-CIMAN calls the srsLTE library for existing LTE functions, including DCI decoding routines provided by OWL. The measurement results of U-CIMAN are written into files that are fed to post-processing scripts for performance validation and result analysis. Through the SDR driver API, U-CIMAN calls USRP Hardware Driver (UHD) [18] version 3.9.7 to communicate with the SDR front end. The SDR system includes a USRP X310 mother board [17] and two SBX-120 wide-band daughter-boards [34]. The SDR boards contain function blocks to convert analog signals to complex samples. The host computer has a quad-core CPU and 16 GB memory, running Ubuntu 16.04. The host computer is connected to the USRP with a Gigabit Ethernet cable.

3 Validations and Applications

To validate the performance of U-CIMAN, comprehensive experiments are conducted to evaluate the decoding accuracy and the potential applications. Based on decoded data, the application of U-CIMAN to spectrum tenancy measurement achieves accurate spectrum measurement at the frequency-time granularity of 180 kHz by 1 ms. Besides spectrum tenancy, packet sizes and TA values associated with occupied RBs are also revealed through decoding headers of user data. We also show that the distributions of packet sizes are highly indicative of traffic types and TA values reflect user mobility.

Two setups, shown in Fig. 5(b) and (c), are used in the experiments. The first one involves a Commercial Off-The-Shelf (COTS) UE, and a commercial grade LTE network realized by Amarisoft OTS 100 that functions as the eNB and the core network [24]. This setup allows access to log files of the LTE network, so the U-CIMAN decoding accuracy can be validated. Though Amarisoft system is close to real LTE networks, it comes with one COTS UE and has limited radio coverage, which cannot reflect statistical characteristics of spectrum tenancy in an actual LTE cell with many real users. To overcome these limitations, U-CIMAN is also validated by decoding the downlink of a nearby commercial eNB, which is experiment setup II. The following performance validations are conducted in setup I or II based on which one better supports the experiment goals.

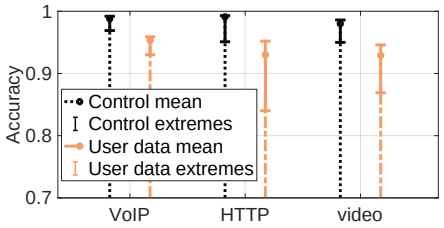

Fig. 6. Accuracy of U-CIMAN decoding.

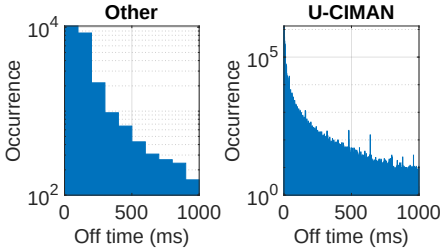

Fig. 7. Time granularity comparison.

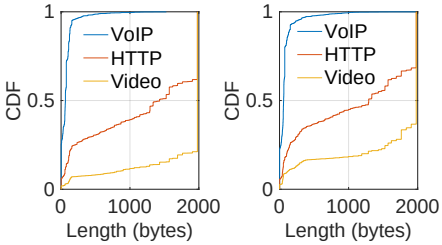

Fig. 8. Distributions of packet lengths.

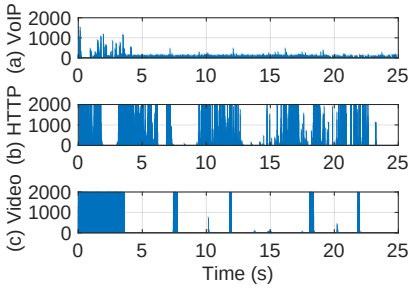

Fig. 9. Packet length versus time under low traffic.

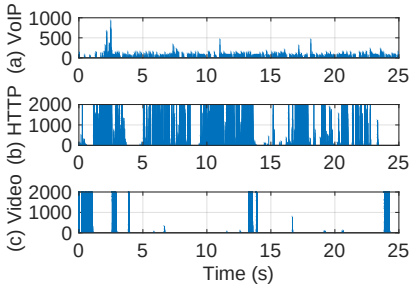

Fig. 10. Packet length versus time under high traffic.

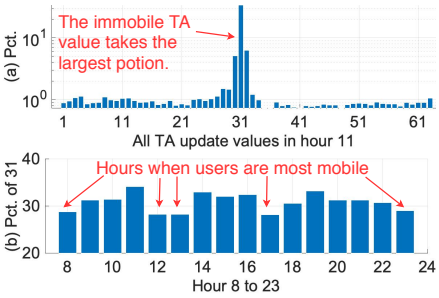

Fig. 11. Distributions of TA upates.

3.1 Decoding and measurement accuracy

Setup I is used to validate decoding accuracy. We run three different applications on the commercial UE multiple times, and compare the downlink control messages and user data bytes decoded by U-CIMAN with those recorded in system log files of Amarisoft OTS. The three applications each generate VoIP, HTTP, and video streaming traffic. Every time the three applications run individually around 25 seconds, and we ensure that no applications other than the one under test generate wireless traffic. Ten rounds of experiments are conducted in total. The accuracy for control messages is the ratio between the number of correctly decoded DCIs and the total number of DCIs. The accuracy for user data bytes is the ratio between the number of correctly decoded codewords and the total number of codewords. The results are shown in Fig. 6 where the dots show the arithmetic mean and the caps are the extremes. The black plot shows the accuracy for control messages, while the yellow plot depicts the performance for user data decoding. U-CIMAN achieves over 95% accuracy for control messages, and over 90% accuracy for user data. The results show that U-CIMAN is capable of decoding downlink control messages and user data bytes with high accuracy for different traffic.

To demonstrate the benefits of fine measurement granularity, we compare the distributions of time length when an RB is not occupied, the off-time, since it is the only LTE channel usage statistics at RB frequency granularity reported in other studies to our best knowledge. In [9], LTE spectrum occupancy is measured with good accuracy since the frequency resolution is the same with that of an LTE RB and time resolution is 100 ms. In comparison, U-CIMAN measures spectrum tenancy at time resolution of an LTE subframe, or 1 ms. Shown in Fig. 7 are off-time distributions of RB 5 according to the measurement in [9], and that obtained by U-CIMAN in setup II. The system bandwidth in the two measurements are both 10 MHz. off-times over one second are omitted due to their small percentage. Though 100 ms time resolution is good compared to most measurements in [10], the majority of the off-time falls in only ten bins of 100ms. In comparison, the off-time distribution achieved by our measurements is much more fine grained. Our off-time distribution shows that off-times are mostly under 30 ms, which cannot be observed with 100 ms time resolution. Thus, we claim that U-CIMAN provides an accessible way to achieve the highly accurate spectrum tenancy measurement which has long been desired [35].

3.2 Traffic types indicated by codeword size

Based on the accurate decoding of both control and user data, the spectrum tenancy measurements achieved by U-CIMAN provide insight on mobile applications. Since codeword sizes are obtained by decoding user data, traffic types can be inferred from the distribution of the codeword sizes. We use setup I, and run three applications generating VoIP, HTTP, and video streaming traffic one at a time for around 25 seconds on a COTS UE. In the first scenario, there is no background traffic, so the only mobile traffic in the cell is generated by the COTS UE. We plot length of codewords in bytes versus time for a single run of the three applications in Fig. 9. For the VoIP traffic, there are a few large packets in the beginning, and the packets are short afterwards. The codeword sizes of HTTP traffic have a wide range. For video streaming, the packet sizes are mostly very large. We redo the same experiment in another scenario with heavy background traffic realized by adding one set of USRP and PC that emulates large amount of user traffic by running Amarisoft UE 100 [36]. According to Fig. 10, the codeword size versus time plots show similar trends to those in Fig. 9, so cell traffic load has little impacts on codeword size characteristics of different applications.

In addition to distinctive trends of codeword sizes along the time horizon, distributions of packet sizes of different traffic demonstrate clear separations as shown in Fig. 8. The plot on the left shows the distribution of codeword sizes under light traffic, and the one on the right shows the distribution in heavy traffic. Though the increased traffic load slightly shifts packet sizes of HTTP and video traffic to the low end, packet size distributions of various applications

still show clear differences in both scenarios with diverse traffic loads. Hence, the decoded packet sizes at physical layer are highly indicative of user application types, regardless of varying cell traffic volume. Detecting spectrum occupancy and the traffic type at the same time can serve as the enabling function for many DSA proposals that are traffic pattern or application dependent [21], [28], [29].

3.3 TA updates indicate UE mobility

The correctness of TA data fields decoded by U-CIMAN is validated in setup II where mobility is generated by real users. Due to the lack of actual UE coordinates, our experiments focus on the validation of TA updates which correspond to user mobility. In mobile networks, downlink synchronizations can be achieved by mobile devices individually; on the uplink, however, a base station needs to adjust the timing of user transmissions such that they arrive at the same time, making it possible for the base station to synchronize with all the uplink transmitters.

In LTE, TA is realized by commanding UEs far from eNB to transmit with larger time advance than the nearby UEs, so TA values reveal the distance of a UE to the eNB. The maximum LTE cell radius is 100 km and the corresponding largest TA is 1282, so the UE-eNB distance of around 78 meters maps to one in TA value. When a UE attempts random access, the eNB determines the TA value and puts it in the RAR. When the eNB finds that the TA needs to be adjusted due to UE mobility, TA update commands are placed in MAC headers. The TA update is a six-bit value, and the updated TA is the sum of the original value, TA update, and -31. Thus, a TA update of 31 means that the UE remains its previous distance from the eNB. The farther the TA update values are from 31, the quicker the UE moves towards or away from the eNB.

To demonstrate TA update fields decoded by U-CIMAN, we apply U-CIMAN to the decoding of TA update values of a commercial LTE cell where the mobility is generated by real users. Fig. 11 shows two distributions of TA updates. The upper figure shows how TA update values in the 11th hour of a day distribute across all 64 values, 0 to 63. The observation is that most TA updates are static or near static. Since TA update value 31 takes the largest portion, we plot how the percentage of TA update value 31 varies across different hours. The result illustrated in the lower part of Fig. 11 agrees with life experience, since the portion of static users are small during commute hours 8, 12, 13, 17, and 23. Thus, U-CIMAN single-handedly succeeds in obtaining the spectrum tenancy of all users in a cell together with their rough location and mobility as shown by TA values and TA updates, which requires less equipment than triangulation [15] while serving the needs of location based DSA algorithms [20], [30].

Based on the above performance validations and applications of U-CIMAN to spectrum tenancy measurement, application inference based on packet sizes, and user mobility detection enabled by TA values, the capabilities of U-CIMAN to measure LTE spectrum tenancy with packet sizes and user mobility have been proved.

4 Measurement Result Analysis

To achieve accurate measurement of LTE spectrum occupancy in fine granularity, we apply U-CIMAN to the spectrum tenancy measurement of a commercial LTE cell. We first search the LTE bands in commercial operations in our area. Then, a spectrum analyzer is employed to verify their existence, and to find the nearby cell with the best signal to noise ratio. The downlink system bandwidth of the cell is 10 MHz which accommodates 50 LTE RBs. We collect LTE spectrum tenancy data of the nearby commercial LTE cell in band 17 with U-CIMAN for four months, and conduct the post-processing to present the measurement results. Due to the space limitation, only the downlink measurement results are presented. As far as we know, this is the first long time LTE spectrum occupancy measurement at RB granularity with packet sizes and TA values. Due to the fine granularity, the tenancy data in a single day is over 1 GB when zipped. Thus, we randomly choose the tenancy data from five days, based on which we present the following results.

4.1 Tenancy characteristics of a single channel

In this subsection, we first introduce the on/off model, and then fit our measurement data to this model. To highlight the importance of time granularity, we study how coarse measurement data affects the tenancy characteristics.

To study the spectrum tenancy of a single LTE channel, we choose the on/off model among the many single-channel occupancy models surveyed in [37], due to its wide usage. Our occupancy data comprises 0 and 1 to indicate whether an LTE channel is idle or occupied. Assume the time lengths of idle or busy periods of one channel to be independent and identically distributed (i.i.d.). Define the vectors $(Y_n, Z_n), n \in \mathbb{N}^+$ where Y_n and Z_n are i.i.d. random variables representing time lengths of the nth idle and busy periods, respectively. We fit five widely used distributions, exponential, Weibull, Lognormal, Generalized Pareto, and Gamma distributions to the observed samples, and the parameters are estimated using Maximum Likelihood Estimation (MLE). The goodness-of-fit is obtained by conducting K-S test, also used in [38] [39] for the same purpose. K-S test is a tool for comparing the closeness of two distributions. We employ K-S test to compare the empirical distributions of the time statistics of measurement data with the distributions obtained by the on/off model fitting. The empirical distribution $G(x)$ of a random variable X that has n observed samples x_i is

$$G(x) = \mathbb{P}(X < x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i<x\}}. \quad (1)$$

Denote the upper bound of the difference between the empirical distribution and the fitting model distribution as D,

$$D = sup|G(x) - F_0(x)|. \quad (2)$$

If the empirical distribution G and fitting model distribution F_0 are identical, the distribution of the random variable D, denoted as D^*, is independent of the fitting distribution. Let G be the cumulative distribution function of D^*. The p value is defined as $p = 1 - G(D)$, so the larger the p value, the more likely D obeys the distribution of D^*, meaning that

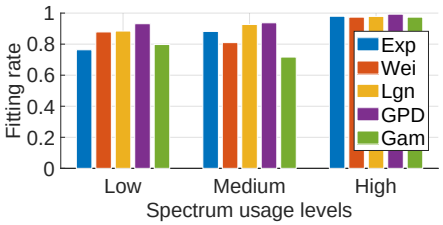

Fig. 12. On-time fitting rates achieved by on/off models with different distributions.

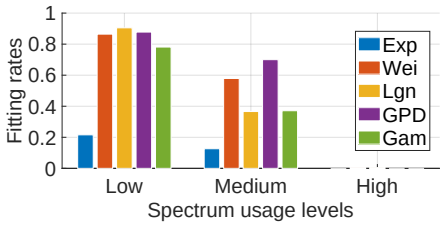

Fig. 13. Off-time fitting rates achieved by on/off models with different distributions.

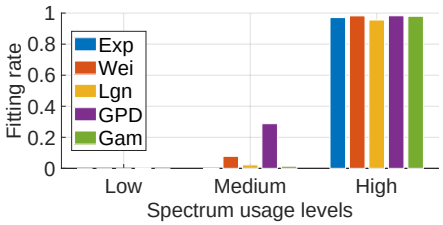

Fig. 14. Interval length fitting rates achieved by on/off models with different distributions.

$G(x)$ and $F_0(x)$ are more likely to be the same. A threshold value $p = 0.05$ is chosen, so the null hypothesis that the samples follow the distribution F_0 is accepted if $p \geq 0.05$.

4.1.1 Fitting results based on U-CIMAN measurement

To identify the distributions for the time lengths of on-times and off-times, we fit the five distributions to 100 groups of randomly chosen off-times and on-times. Each group contains 100 samples of off-times and on-times. We calculate the fitting rate as the number of groups for which a certain distribution fits the samples according to K-S test at the significance level of $p = 0.05$ over 100, the total number of groups. We investigate the fitting rates under three spectrum usage levels where the average numbers of occupied LTE channels per scheduling interval are around 5, 25, and 45 out of the total number of 50. The three spectrum usage scenarios are referred to as low, medium, and high spectrum usage levels, respectively.

As shown in Fig. 12, the fitting rates of on-time obtained by the five distributions are mostly over 80% in all the three spectrum tenancy levels, and generalized Pareto distribution (GPD) achieves the best performance with its fitting rate over 90% under all spectrum tenancy conditions. The off-times cannot be well fitted by any of the five distributions during high spectrum usage as shown in Fig. 13, so we further study the fitting of the interval length which is the summation of adjacent off-time and on-time. Fitting results for interval lengths are illustrated in Fig. 14, showing that good fitting rates are achieved in high spectrum tenancy, which complements the poor fitting rate of off-time when the spectrum usage is high.

Based on the fitting rate study, the spectrum usage model for a single LTE channel is summarized in Table 2, where the parameters are the medians of the values obtained from fitting to the 100 groups of data samples. The location parameters of all the GPDs are zero, so they are not shown in the table. The time unit of the values in the table is one millisecond, the scheduling interval of the LTE system. The best fitting scheme is to fit the off-time and the following on-time with GPDs during low and medium spectrum usage, and apply GPD to the on-time and the interval lengths during high spectrum usage, as suggested by the fitting rates study in Figs. 12 to 14.

TABLE 2
A summary of the fitting models in different traffic conditions.

Usage	Schemes	Shape	Scale	Mean	Variance
Low	On, GPD	0.4166	1.4385	2.8840	45.8264
	Off, GPD	0.4969	13.3262	23.1905	913.8778
Med.	On, GPD	0.5620	2.3151	4.8323	79.2201
	Off, GPD	0.4809	4.3246	7.6043	189.7997
High	On, GPD	0.0209	21.6998	22.0619	507.8245
	Int., GPD	−0.0238	23.7513	23.1243	507.6929

4.1.2 Fitting results based on coarse data

To highlight the impacts of measurement granularity on the spectrum tenancy models built on measurement results, we fit the on/off model with the same five distributions to LTE spectrum tenancy data with coarse time resolution. The coarse time resolution is chosen to be 1.8 seconds, the same as the well-known data set in [14], [40].

To obtain the coarse spectrum usage with time resolution of 1.8 seconds, we evenly pick one data point from every 1800 samples in U-CIMAN measurement results. The fitting results are summarized in Table 3, where the data are all in the unit of 1.8 seconds. For low and medium spectrum usage, the time lengths of the coarse spectrum usage data cannot be fitted by any of the five distributions according to K-S test. However, the closest GPD parameters, the mean, and the variance are still presented in the table. The GPD fittings of on-times and interval lengths during high spectrum usage are able to pass K-S test with significance $p = 0.05$. As we can see, the average time lengths are hugely different, as well as the shape of the distributions. For example, the average on-time according to U-CIMAN data in low spectrum usage condition is 2.884 ms, but it is $1.2683 \times 1800 = 2282.94$ ms according to the coarse version. The shape parameters of the GPDs in Table 2 and 3 are clearly disparate, as they are either several times of their counterpart in the other table or take different signs.

Thus, the measurement granularity is of utmost importance in the study of spectrum tenancy, because it fundamentally affects the tenancy characteristics. U-CIMAN and its measurement results are an essential base for an accurate study of LTE spectrum usage.

TABLE 3
A summary of fitting models to measurement data with the time resolution of 1.8 seconds.

Usage	Schemes	Shape	Scale	Mean	Variance
Low	On, N/A	−0.2894	1.5575	1.2683	0.4037
	Off, N/A	0.1311	2.8375	3.3073	21.9053
Med.	On, N/A	0.2007	2.7366	3.4755	26.4225
	Off, N/A	0.0594	2.4375	2.5982	9.1273
High	On, GPD	0.1922	21.8622	26.8592	962.5513
	Int., GPD	0.1295	24.3510	27.9014	962.6330

4.2 Tenancy characteristics of multiple channels

For the modeling of spectrum tenancy in multiple channels, we propose to adopt the VAR model, which is the first time as far as we know. We regard the occupancy of multiple channels at each time slot as a sample of a multivariate normal random variable which is the sum of a constant, white noise, and multivariate normal random variables

representing the tenancy in previous time slots. The channel usage at time instant n is a random vector, denoted as $\mathbf{y}_n$,

$$\mathbf{y}_n = \mathbf{a} + \sum_{i=1}^{k} \boldsymbol{\phi}_i \mathbf{y}_{n-i} + \boldsymbol{\varepsilon}_n. \quad (3)$$

The constant vector is $\mathbf{a}$, and $\boldsymbol{\varepsilon}_n$ is the noise term. $\mathbf{y}_{n-i}$, where $1 \leq i \leq k$, is the channel usage in a previous time slot no earlier than the time lag k, and its linear relations with $\mathbf{y}_n$ are described by the matrix $\boldsymbol{\phi}_i$. We fit VAR models to the measurement data with different time lags, and the parameters are estimated using MLE.

To decide on the time lag k, we compare the performance of VAR models with different time lags. The performance is compared by employing the Akaike Information Criterion (AIC) which is defined as

$$AIC = 2n - 2\log(L), \quad (4)$$

where L is the optimized scalar value of log-likelihood objective function, and n is the number of parameters that need to be estimated in the model. AIC measures the relative qualities of statistical models fitted to a data set, and models with small AIC values are preferred because they capture statistical features of data better using fewer parameters.

We fit VAR models with different k values, 1, 2, 3, 4, 8, and 12, to measurement data, and the AIC values are calculated for the models with different time lags for comparison. Because the differences among the six AIC values of different models are negligible, the time lag is chosen to be 1, which has the fewest parameters to estimate.

Now that we have obtained the segment length, the on/off model fitting strategy, and the time lag of VAR, we compare the performance of on/off and VAR models from the aspects of D values of K-S tests, correlations among adjacent channels, and correlations between adjacent off-time and on-time. Specifically, we extract the spectrum occupancy in 10^5 time slots of 10 LTE channels during different spectrum usage levels, and then fit both the single-channel and the multi-channel models to the data. For the VAR model, fitting multiple channels requires only adjusting the number of elements in the vectors $\mathbf{y}_n$, $\mathbf{c}$, $\boldsymbol{\varepsilon}$, and the matrices $\boldsymbol{\phi}$ in (3). For the on/off model, we fit 10 on/off models to each of the 10 channels independently, and obtain 10 sets of parameters. Using the two types of models, we produce synthetic spectrum occupancy of the same size with the measurement data for the three performance comparisons.

4.2.1 Compare the similarity of tenancy time distributions

To study the extent to which the two sets of synthetic data resemble the measurement results, we calculate the D values between the measurement and the two sets of synthetic data using (1) and (2), where $F_0(x)$ is the empirical distribution of the data generated by the two models.

The D values between the distributions of measurement and synthetic spectrum occupancy are presented in Table 4. In all the three spectrum usage levels, synthetic data generated by VAR model achieves on-time distributions with smaller D values. In terms of the similarity comparison of off-time distributions, VAR model has significantly lower D values during low and high spectrum usage, meaning that the on-times of synthetic occupancy generated by VAR models resemble those in measurements much closer than the on/off model in those cases. Though the on/off model outperforms VAR in terms of the similarity of interval length distributions in low and medium spectrum usage, the advantages in these cases are not pronounced.

TABLE 4
D value comparisons of the two models in different traffic.

Time	Model	Low	Medium	High
On	on/off	0.4891	0.3150	0.8654
	VAR	0.0723	0.1621	0.0803
Off	on/off	0.4088	0.2561	0.9226
	VAR	0.1811	0.2714	0.0154
Interval	on/off	0.1157	0.1012	0.0524
	VAR	0.2991	0.1160	0.0421

Overall, VAR achieves better resemblance of the distributions of channel usage times to those of the measurement data than the on/off model.

4.2.2 Compare tenancy correlations in adjacent channels

As indicated in previous studies, spectrum occupancy of the same radio access technology are correlated [41], we study how the different channels in the same cell are correlated and whether the correlations can be captured by our models. The correlation coefficient is Pearson correlation coefficient.

We compare how closely the two sets of synthetic data resemble the measurement results in terms of correlations among adjacent channels. Fig. 15 presents the pairwise correlation coefficients between the spectrum tenancy of the first channel and all the ten channels. The black line shows the channel tenancy correlations between the first channel and the other channel whose index is shown in the x-axis. The measured spectrum tenancy in the first channel shows very high correlations with those of the three nearest channels, and the correlations decrease gradually as the frequency distance grows. This trend is captured very well by the data generated by the VAR model, though the correlations in blue are lower than those of measurements. Since on/off model is a single-channel model, its synthetic tenancy has zero correlations among adjacent channels, as shown by the red line.

Thus, the tenancy correlations of adjacent channels are better captured by VAR than the on/off model.

4.2.3 Compare correlations between on/off times

Since the negative correlations between adjacent off-time and on-time are suggested in previous studies [9], we investigate this correlation reflected by measurement and synthetic data for LTE channels.

It has been suggested in previous studies that off-times and the following on-times are negatively correlated [9]. However, LTE spectrum tenancy does not show this phenomenon as illustrated in Fig. 16. The correlations between idle periods and the following busy periods are studied for measurements and synthetic data. The correlations among adjacent idle and busy periods in the three groups of data are close to zero, as they are bounded within $[-0.2, 0.1]$ in all three spectrum usage levels, meaning that off-times and the following on-times are weakly correlated. The phenomenon is due to the fact that LTE systems schedule spectrum

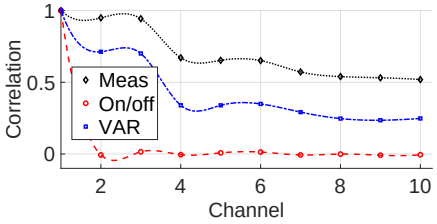

Fig. 15. Spectrum tenancy correlations among adjacent channels.

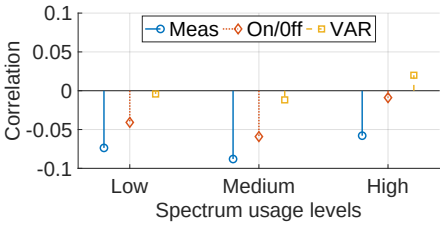

Fig. 16. Correlations between on-time and the following off-time.

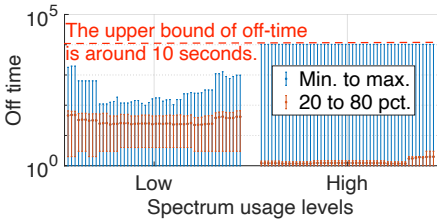

Fig. 17. MCS distributions conditioned on next subframe occupancy.

resources every 1 ms, so the off-times are independent of the next on-times many scheduling intervals away.

Among the three aspects considered in the performance comparisons, VAR outperforms on/off model in LTE spectrum tenancy modeling due to superior capabilities to capture the statistical features of busy and idle time lengths, and occupancy correlations among adjacent channels.

4.3 Other observations on spectrum tenancy

Besides characterizing the LTE spectrum tenancy with on/off and VAR models, the spectrum usage data obtained by U-CIMAN enables other observations. In this subsection, we present the observations on the off-time characteristics, and the factors affecting the spectrum tenancy.

4.3.1 Characteristics of off-times

The off-time of an RB, defined as the time period from the subframe when the RB becomes idle to the subframe right before it turns to be occupied, is a key parameter in spectrum usage studies. Denote the off-time of all the 50 RBs as a group of random variables, $X_{Off}(c)$, where $c \in \{1, \cdots, 50\}$. $X_{Off}(c)$ is a random variable on the sample space Ω_c^T where each element ω is an off-time of RB c. The pmf of $X_{Off}(c)$ is

$$\mathbb{P}(X_{Off}(c) = x) = \frac{|\{\omega \in \Omega_c^T | X_{Off}(c)(\omega) = x\}|}{|\Omega_c^T|}, \quad (5)$$

where $x \in \mathbb{N}^+$. The distribution function of $X_{Off}(c)$ is

$$F_{X_{Off}(c)}(x) = \sum_{y \le x} \mathbb{P}(X_{Off}(c) = y), \quad (6)$$

where $x \ge 0$. Fig. 17 presents the rough off-time distributions of 50 RBs during low and high spectrum tenancy. The red dot in the middle is the mean value, and the red caps show the 20th and the 80th percentiles of off-times. The blue caps are the extreme values.

Remark 1: We make two observations on off-time from Fig. 17. One observation is that the off-time distributions of different RBs under the same spectrum usage level are similar. According to the figure, the key percentiles and mean values of the off-times in different channels are close to one anther under the same spectrum usage level, which is especially true for the channels in the center when the usage is low and all the channels when usage is high. The similar off-time among different RBs stems from the scheduling that treats the spectrum resources in different frequency equally. The other observation is that the off-time exhibits a common upper bound regardless of the spectrum usage levels, and RBs. Fig. 17 shows the off-time is capped around 10^4 ms.

4.3.2 Factors affecting channel tenancy

As we have observed that the distributions of off-times are similar across different channels, next we study whether the usage of different RBs depend on user locations or traffic types. We consider the sample space Ω^{RP} of RBs allocated for 2000 randomly chosen packets each of which may require multiple RBs. Define the random variable X_{TA} to be the TA value of a UE receiving packets carried by the RBs in the sample space. The pmf of X_{TA} is

$$\mathbb{P}(X_{TA} = x) = \frac{|\{\omega \in \Omega^{RP} | X_{TA}(\omega) = x\}|}{|\Omega^{RP}|}, \quad (7)$$

where $x \in \{0, \cdots, 1282\}$. The bars in the bottom right of Fig. 19 show the distribution of X_{TA} before normalized by the size of sample space Ω^{TA}. The heights of the bars are

$$|\{\omega \in \Omega^{RP} | X_{TA}(\omega) = x\}| = |\Omega^{RP}| \mathbb{P}(X_{TA} = x). \quad (8)$$

The other three plots depict the distributions of X_{TA} conditioned on another random variable, the index of the RBs, X_{iRB}. The probability of TA values conditioned on RB indexes $\mathbb{P}(X_{TA} = x | X_{iRB} = y)$ is

$$\mathbb{P}(x|y) = \frac{|\{\omega \in \Omega^{RP} | X_{TA}(\omega) = x, X_{iRB}(\omega) = y\}|}{|\{\omega \in \Omega^{RP} | X_{iRB}(\omega) = y\}|}, \quad (9)$$

where $y \in \{1, \cdots, 50\}$. The conditional probability mass, $\mathbb{P}(X_{TA} | X_{iRB} = y_0)$ is similar to $\mathbb{P}(X_{TA})$, and three examples are given in Fig. 19 where $y_0 \in \{20, 30, 40\}$, showing that TA values are independent of RB indexes, i.e., $X_{TA} \perp\!\!\!\perp X_{iRB}$.

After studying the relationship between RB usage and UE-eNB distance, we investigate the impact of packet size on the frequency of assigned spectrum resources. Define the random variable X_{PS} for packet sizes of RBs in the sample space Ω^{RP}. The packet size of an RB is the number of codeword bytes carried by the RB. For packets carried by multiple RBs, we regard all the RBs correspond to the same size. The pmf of X_{PS} is

$$\mathbb{P}(X_{PS} = x) = \frac{|\{\omega \in \Omega^{RP} | X_{PS}(\omega) = x\}|}{|\Omega^{RP}|}, \quad (10)$$

where $x \in \mathbb{N}^+$. Bars in the bottom right of Fig. 18 show the distribution of X_{PS} before normalized by the size of sample space Ω^{RP}, which is

$$|\{\omega \in \Omega^{RP} | X_{PS}(\omega) = x\}| = |\Omega^{RP}| \mathbb{P}(X_{PS} = x). \quad (11)$$

The other three plots show distributions of X_{PS} conditioned on RB indexes of packets, X_{iRB}. The probability of packet sizes conditioned on RB indexes $\mathbb{P}(X_{PS} = x | X_{iRB} = y)$ is

$$\mathbb{P}(x|y) = \frac{|\{\omega \in \Omega^{RP} | X_{PS}(\omega) = x, X_{iRB}(\omega) = y\}|}{|\{\omega \in \Omega^{RP} | X_{iRB}(\omega) = y\}|}, \quad (12)$$

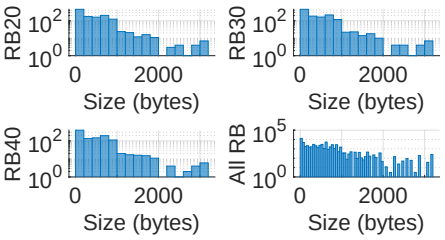

Fig. 18. Packet size distributions conditioned on RB index.

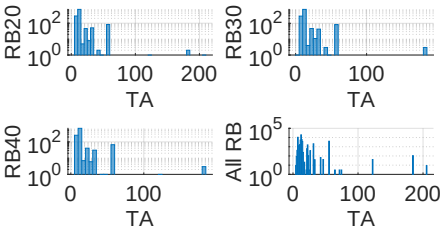

Fig. 19. TA distributions conditioned on RB index.

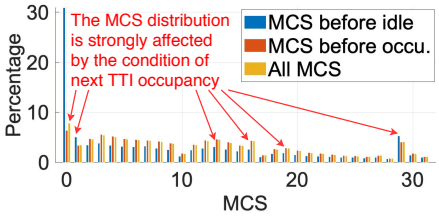

Fig. 20. MCS distributions conditioned on next subframe occupancy.

where $y \in \{1, \cdots, 50\}$. The conditional probability mass, $\mathbb{P}(X_{PS}|X_{iRB} = y_0)$ is similar to that of X_{PS}, and three examples are given in Fig. 18 where $y_0 \in \{20, 30, 40\}$, showing that packet sizes are independent of RB indexes, $X_{PS} \perp\!\!\!\perp X_{iRB}$.

Thus, the RB occupancy does not depend on its frequency, UE-eNB distance, and packet size. However, MCS values strongly affect RB occupancy. According to former introduction, MCS is specified along with spectrum resource assignments in DCIs. MCS is five-bit long, taking values from 0 to 31. Define X_M as the random variable for MCS values of RBs. The sample space is the set of all occupied RBs in a day, Ω^M. Define another random variable X_{NO} for RB occupancy in the next TTI on the same sample space. The pmf for X_M,

$$\mathbb{P}(X_M = x) = \frac{|\{\omega \in \Omega^M | X_M(\omega) = x\}|}{|\Omega^M|}, \tag{13}$$

where $x \in \{0, \cdots, 31\}$, shown by yellow bars in Fig. 20 in percentage. The blue and the orange bars show the probability of X_M conditioned on X_{NO} being 0 for unoccupied, or 1 for occupied in percentage. The conditional pmf, $\mathbb{P}(X_M = x|X_{NO} = y)$, is

$$\mathbb{P}(x|y) = \frac{|\{\omega \in \Omega^M | X_M(\omega) = x, X_{NO}(\omega) = y\}|}{|\{\omega \in \Omega^M | X_{NO}(\omega) = y\}|}. \tag{14}$$

Fig. 20 illustrates that MCS values of an occupied RB have different distributions conditioning on whether the RBs are occupied next TTI, and the differences are quite obvious at some MCS values. For example, the MCS values of 1 and 29 are much more likely to appear when RBs will be idle next TTI than in RBs continuing to be occupied. Hence, MCS values of currently occupied RBs strongly correlates with their tenancy in the next TTI, i.e. $X_M \not\perp\!\!\!\perp X_{NO}$.

Remark 2: It is observed that spectrum resources on different frequency are treated similarly, irrespective of UE-eNB distance and packet sizes. Hence, they provide little information on the frequency of occupied spectrum. However, MCS values are highly indicative of channel occupancy in the next time slot.

5 Related Work

In the existing literature, there are two categories of researches closely related to the measurement enabled by U-CIMAN, the spectrum usage measurement based on traditional methods, and LTE signal decoding by outsiders.

The fundamental importance of spectrum tenancy studies has long been recognized and there exist many spectrum tenancy measurement campaigns. A recent survey study [10] on those campaigns provides in depth summaries from various aspects. According to the survey, these previous spectrum measurements typically adopt the energy detection method, because they measure the spectrum occupancy spanning a range of several gigahertz where the signals are too diverse for other measurement methods, such as the matched filter detection. This causes the results to be of coarse time and frequency resolutions, and the detection threshold has to be chosen empirically, introducing another source of errors [42]. Spectrum tenancy data in coarse granularity is useful for studies on radio activities that remain steady for long periods of time, such as television broadcast. However, scheduling of spectrum resources in modern cellular systems happen at millisecond time scale, far exceeding the granularity provided by existing measurement data. It is worth mentioning that sampling at one millisecond time scale in the LTE spectrum bands using USRPs is not enough to detect LTE spectrum usage, as in [13], since the minimum sampling rate of 15.36 MHz is required to study the tenancy of an LTE system with 10 MHz bandwidth with out aliasing.

Existing works that decode LTE protocol fields as an outsider have targeted only the control plane in the physical layer, since DCIs contain many useful data fields that are able to satisfy the needs of diverse applications as pointed out in [15]. While decoding DCIs, the main challenge is how to obtain the RNTIs. Both of the two existing solutions proposed in [15], [16] suffer from the shortcoming of being unable to validate the decoded RNTIs [31]. U-CIMAN overcomes this issue by applying the RNTIs to decoding user data bytes, so the RNTIs can be validated if the user data bytes are correctly decoded. In this way, U-CIMAN not only checks the correctness of RNTIs decoded from DCIs or RARs, but also obtains the raw bytes of user data which contains more information on spectrum tenancy, such as the TA values and packet sizes.

6 Conclusion

To accurately measure LTE spectrum tenancy, we design and implement a new sniffing tool U-CIMAN that decodes both downlink control messages and raw data bytes without breaking LTE encryption. We apply U-CIMAN to the four-month measurement of spectrum tenancy of a commercial LTE cell. Compared with existing measurements, our results are more accurate in terms of time-frequency granularities, and provide important details of spectrum users, such as the inferred traffic types and rough locations. The accurate measurement enables new observations, such as the 10 seconds upper bound of idle time and the predictive power of MCS on spectrum tenancy. Based on the fine measurement, we characterize LTE spectrum tenancy measurements with

both on/off and the proposed VAR models. The accurate spectrum tenancy data provided by U-CIMAN enables analysis and new understanding of LTE spectrum tenancy that used to be shadowed by coarse measurement data.

REFERENCES

[1] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in *Proceedings of the first international workshop on Technology and policy for accessing spectrum*. ACM, 2006, p. 1.

[2] I. F. Akyildiz, A. Kak, and S. Nie, "6g and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, 2020.

[3] C. Rosa, M. Kuusela, F. Frederiksen, and K. I. Pedersen, "Standalone lte in unlicensed spectrum: Radio challenges, solutions, and performance of multefire," *IEEE Communications Magazine*, vol. 56, no. 10, pp. 170–177, 2018.

[4] H. Lee, H. Kim, H. J. Yang, J. T. Kim, and S. Baek, "Performance analysis of license assisted access lte with asymmetric hidden terminals," *IEEE Transactions on Mobile Computing*, vol. 17, no. 9, pp. 2141–2154, 2018.

[5] H. He, H. Shan, A. Huang, Q. Ye, and W. Zhuang, "Edge-aided computing and transmission scheduling for lte-u-enabled iot," *IEEE Transactions on Wireless Communications*, 2020.

[6] T. Harrold, R. Cepeda, and M. Beach, "Long-term measurements of spectrum occupancy characteristics," in *2011 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2011, pp. 83–89.

[7] B. Li, M. Sun, X. Li, A. Nallanathan, and C. Zhao, "Energy detection based spectrum sensing for cognitive radios over time-frequency doubly selective fading channels," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 402–417, 2015.

[8] Y. Chen and H.-S. Oh, "A survey of measurement-based spectrum occupancy modeling for cognitive radios," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 848–859, 2016.

[9] T. A. Hall, A. Sahoo, C. Hagwood, and S. Streett, "Exploiting lte white space using dynamic spectrum access algorithms based on survival analysis," in *Communications (ICC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–7.

[10] M. Höyhtyä, A. Mämmelä, M. Eskola, M. Matinmikko, J. Kalliovaara, J. Ojaniemi, J. Suutala, R. Ekman, R. Bacchus, and D. Roberson, "Spectrum occupancy measurements: A survey and use of interference maps," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2386–2414, 2016.

[11] G. A. Sander and J. E. Carroll, "Emission spectrum measurements of a 3.5 ghz lte hotspot," in *NTIA report series*. U.S. DEPARTMENT OF COMMERCE, National Telecommunications and Information Administration, 2015.

[12] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[13] M. López-Benítez and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE transactions on vehicular technology*, vol. 62, no. 5, pp. 2091–2104, 2013.

[14] J. Sun, J. Wang, J. Chen, G. Ding, and F. Lin, "Clustering analysis for internet of spectrum devices: Real-world data analytics and applications," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4485–4496, 2020.

[15] S. Kumar, E. Hamed, D. Katabi, and L. Erran Li, "Lte radio analytics made easy and accessible," in *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4. ACM, 2014, pp. 211–222.

[16] N. Bui and J. Widmer, "OWL: a Reliable Online Watcher for LTE Control Channel Measurements," in *ACM All Things Cellular (MobiCom Workshop)*, Nov. 2016.

[17] E. Research, "x310-kit," February 2019. [Online]. Available: https://www.ettus.com/all-products/x310-kit/

[18] ——, "Uhd," November 2017. [Online]. Available: https://kb.ettus.com/UHD

[19] I. Gomez-Miguelez, A. Garcia-Saavedra, P. D. Sutton, P. Serrano, C. Cano, and D. J. Leith, "srslte: an open-source platform for lte evolution and experimentation," in *Proceedings of the Tenth ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation, and Characterization*. ACM, 2016, pp. 25–32.

[20] H. Hu, H. Zhang, and N. Li, "Location-information-assisted joint spectrum sensing and power allocation for cognitive radio networks with primary-user outage constraint," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 2, pp. 658–672, 2015.

[21] M. E. Ahmed, J. B. Song, and Z. Han, "Traffic pattern-based reward maximization for secondary user in dynamic spectrum access," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2013, pp. 291–296.

[22] T. A. Hall, A. Sahoo, C. Hagwood, and S. Streett, "Dynamic spectrum access algorithms based on survival analysis," *IEEE transactions on cognitive communications and networking*, vol. 3, no. 4, pp. 740–751, 2017.

[23] H. Eltom, S. Kandeepan, R. J. Evans, Y. C. Liang, and B. Ristic, "Statistical spectrum occupancy prediction for dynamic spectrum access: a classification," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 29, 2018.

[24] Amarisoft, "Amarisoft ots 100," December 2017. [Online]. Available: https://www.amarisoft.com/2016/12/28/amarisoft-presents-amari-ots-100/

[25] *Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) specification (3GPP TS 36.323 version 14.3.0 Release 14)*, ETSI 3GPP, Jul. 2017, version 14.3.0.

[26] S. Sesia, M. Baker, and I. Toufik, *LTE-the UMTS long term evolution: from theory to practice*. John Wiley & Sons, 2011.

[27] E. U. T. R. Access, "Physical channels and modulation, 3gpp ts 36.211," *V10*, vol. 2, 2009.

[28] J. Huang, H. Wang, Y. Qian, and C. Wang, "Priority-based traffic scheduling and utility optimization for cognitive radio communication infrastructure-based smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 78–86, 2013.

[29] M. Hoyhtya, S. Pollin, and A. Mammela, "Classification-based predictive channel selection for cognitive radios," in *2010 IEEE International Conference on Communications*. IEEE, 2010, pp. 1–6.

[30] X. Xu, L. Li, Y. Cai, X. Chen, and M. Zhao, "Transmission rate optimization in cooperative location-aware cognitive radio networks," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–6.

[31] X. Xie, X. Zhang, S. Kumar, and L. E. Li, "pistream: Physical layer informed adaptive video streaming over lte," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 413–425.

[32] *Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification (3GPP TS 36.321 version 13.5.0 Release 13)*, ETSI 3GPP, Apr. 2017, version 13.5.0.

[33] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (3GPP TS 36.213 version 13.0.0 Release 13)*, ETSI 3GPP, May 2016, version 13.0.0.

[34] E. Research, "Sbx120," February 2019. [Online]. Available: https://www.ettus.com/all-products/SBX120/

[35] R. Chen, J.-M. Park, and K. Bian, "Robust distributed spectrum sensing in cognitive radio networks," in *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*. IEEE, 2008, pp. 1876–1884.

[36] Amarisoft, "Amarisoft ue 100," December 2017. [Online]. Available: https://www.amarisoft.com/products/test-measurements/amari-ue-simbox/

[37] Y. Saleem and M. H. Rehmani, "Primary radio user activity models for cognitive radio networks: A survey," *Journal of Network and Computer Applications*, vol. 43, pp. 1–16, 2014.

[38] S. Geirhofer, L. Tong, and B. M. Sadler, "A measurement-based model for dynamic spectrum access in wlan channels," in *Military Communications Conference, 2006. MILCOM 2006. IEEE*. IEEE, 2006, pp. 1–7.

[39] L. Stabellini, "Quantifying and modeling spectrum opportunities in a real wireless environment," in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*. IEEE, 2010, pp. 1–6.

[40] M. Wellens, "Empirical modelling of spectrum use and evaluation of adaptive spectrum sensing in dynamic spectrum access networks," *Unpublished doctoral dissertation). RWTH University of Aachen, Germany. http://darwin. bth. rwth-aachen. de/opus3/volltexte/2010/3248*, 2010.

[41] S. Yin, D. Chen, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: a large-scale spectrum measurement study," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 1033–1046, 2012.

[42] S. Saleem and K. Shahzad, "Performance evaluation of energy detection based spectrum sensing technique for wireless channel,"

International journal of multidisciplinary sciences and engineering, vol. 3, no. 5, pp. 31–34, 2012.

Rui Zou (S'19) received the B.S. degrees in electrical engineering from the joint program provided by Beijing University of Posts and Telecommunications (BUPT) and Queen Mary University of London in 2012. He then obtained the M.S. degree in electrical engineering from BUPT in 2015.

He is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA. His current research interests include measurement and modeling of spectrum tenancy, and security issues in dynamic spectrum access systems.

Wenye Wang (F'17) received the M.S.E.E. and Ph.D. degrees in computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1999 and 2002, respectively.

She is a Professor with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA. Her current research interests include mobile and secure computing, modeling and analysis of wireless networks, network topology, and architecture design.

Dr. Wang was a recipient of the NSF CAREER Award 2006. She was a co-recipient of the 2006 IEEE GLOBECOM Best Student Paper Award Communication Networks and the 2004 IEEE Conference on Computer Communications and Networks Best Student Paper Award. She has been a member of the Association for Computing Machinery since 1998 and a member of the Eta Kappa Nu and Gamma Beta Phi honorary societies since 2001.