FLuMe: Understanding Differential Spectrum Mobility Features in High Resolution

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

***Abstract*—Existing measurements and modeling of radio spectrum usage have shown that exclusive access leads to low efficiency. Thus, the next generation of wireless networks is adopting new paradigms of spectrum sharing and coexistence among heterogeneous networks. However, two significant limitations in current spectrum tenancy models hinder the development of essential functions in nonexclusive spectrum access. First, these models rely on data with much coarser resolutions than those required for wireless scheduling, rendering them ineffective for spectrum prediction or characterizing spectrum access behavior in a wireless coexistence setting. Second, due to a lack of detailed data, current models cannot describe the access dynamics of individual users, leading to unjustified adoption of simplistic traffic models, such as the on/off model and the M/G/1 queue, in spectrum access algorithm research. To address these limitations, we propose the Frame-Level spectrum Model (FLuMe), a data-driven model that characterizes individual spectrum usage based on high-resolution data. This lightweight model tracks the spectrum tenancy movements of individual users using four variables. The proposed model is applied to high-resolution LTE spectrum tenancy data, from which model parameters are extracted. Comprehensive validations demonstrate the goodness-of-fit of the model and its applicability to spectrum prediction.**

***Index Terms*—Data driven model, data resolution, LTE, spectrum tenancy.**

I. Introduction

RADIO spectrum is the fundamental resource in wireless networks, enabling the transmission of signals via electromagnetic waves that must occupy a portion of the spectrum. Consequently, the management of wireless spectrum has been extensively studied to monitor usage and improve efficiency [1], [2]. A key finding from these studies is the significant underutilization of many exclusively assigned spectrum resources. For example, it has been observed that radio channels auctioned for exclusive access can remain idle for up to 50% of the time [3]. This low efficiency in exclusive spectrum allocations has spurred the development of various technologies and standards aimed at promoting spectrum sharing and wireless coexistence. Technologies such as MulteFire [4] and Licensed Assisted Access (LAA) [5] facilitate the deployment of Long-Term Evolution (LTE) and 5G New Radio (NR) in TV White Space (TVWS), as well as the 2.4 GHz and 5 GHz Industrial, Scientific, and Medical (ISM) spectrum bands. These bands coexist with other wireless networks like WiFi and Low-Power Wide Area Networks (LPWANs) [6], [7]. Furthermore, the coexistence of various Radio Access Technologies (RATs) that typically occupy exclusive frequency channels has attracted significant research interest. This is evident from the numerous proposals addressing the coexistence of LTE and 5G cells on overlapping channels [8], [9].

Spectrum sharing and wireless coexistence critically depend on spectrum usage models. Spectrum prediction is essential for avoiding collisions in open spectrum access settings, and prediction algorithms are either directly based on spectrum usage models or reliant on training data synthesized by these models [10]. However, spectrum usage models based on coarse data fail to guide spectrum access decisions effectively. For instance, two spectrum predictions rely on models based on tenancy data with resolutions much coarser than the scheduling of spectrum resources [11], [12]. The LTE spectrum prediction in [12] is governed by a model based on tenancy data with a time resolution of 100 ms. As LTE spectrum dynamically changes every ms, accessing it in chunks of 100 ms either wastes spectrum holes or incurs collisions, since the probability of having the same tenancy over 100 consecutive scheduling intervals is very small. Similarly, the prediction of Global System for Mobile Communications (GSM) spectrum tenancy based on an on/off model with data resolution lower than scheduling [11] faces issues. GSM spectrum is dynamically rescheduled per 4.61538 ms (the GSM frame time length), while the data resolution in [11] ranges from 0.1 to 0.9 seconds. Exploiting GSM spectrum holes in such coarse resolution either leads to serious waste or inevitable collisions, rendering this GSM spectrum prediction impractical for spectrum access. In summary, wireless coexistence urgently requires spectrum tenancy models based on high-resolution data with the same granularity as the scheduling of spectrum resources.

Besides spectrum prediction, the design of spectrum access algorithms in a wireless coexistence environment heavily relies on spectrum tenancy models, especially the spectrum usage behavior of individual users. However, intuitive classical models have been adopted to describe individual users' spectrum usage

Manuscript received 13 December 2023; revised 12 June 2024; accepted 30 July 2024. Date of publication 13 August 2024; date of current version 5 November 2024. This work was supported in part by NSF under Grant CNS-1527696, in part by NSF under Grant CNS-1824518, and in part by ARO under Grant W911NF-15-2-0102. Recommended for acceptance by D. Wang. *(Corresponding author: Rui Zou.)*

Rui Zou is with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27606 USA, and also with Business and Computer Technology Division, Central Arizona College, Coolidge, AZ 85128 USA (e-mail: rzou@ncsu.edu).

Wenye Wang is with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27606 USA (e-mail: wwang@ncsu.edu).

Digital Object Identifier 10.1109/TMC.2024.3442151

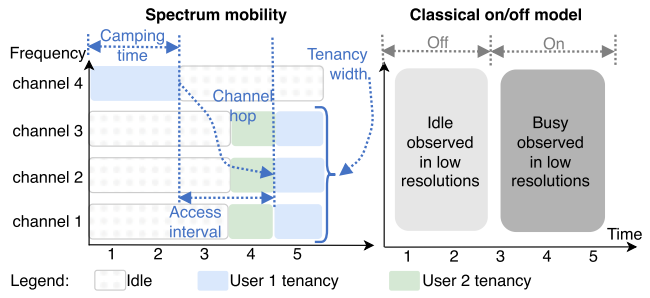


Fig. 1. Spectrum mobility versus the classical on/off model.

behavior without serious consideration of real-world data [13], [14], [15]. For instance, the spectrum usage behavior of individual users is modeled as a preemptive M/G/1 queue in [13], [16]. A non-preemptive M/G/1 queue is adopted in [13], [14], [16]. Another design [15] characterizes the spectrum tenancy of individual devices using an on/off model. However, it has not been verified by measurement data that the spectrum usage of individual devices follows any of these models, compromising the validity of research based on these unverified spectrum models. The lack of models for individual devices is due to existing data-driven spectrum models relying on measurements that cannot differentiate the spectrum activities of individual wireless devices.

In essence, existing spectrum models suffer from two major limitations: the lack of fine data resolution matching the granularity of spectrum resource scheduling and the inability to characterize the spectrum usage of individual users, hindering their applicability for spectrum sharing and wireless coexistence. To overcome these limitations, we address the question of what spectrum usage model characterizes both aggregate and individual spectrum tenancy based on high-resolution data in the same time-frequency granularity as scheduling. We adopt a data-driven approach to develop a *spectrum mobility model*. This model refers to the new spectrum tenancy model presented in this paper and should not be confused with other concepts, such as spectrum handoff in cognitive radio networks, where users move from one frequency channel to another [17].

Our model leverages a two-dimensional time-frequency grid to represent spectrum usage, visualizing occupied spectrum slices as they dynamically move across frequencies over time. We characterize individual mobile device behavior using four key parameters: camping times, access intervals, channel hops, and tenancy widths. Camping time and access interval represent the busy and idle periods of spectrum usage. Channel hops describe the movement of the carrier frequency across different channels, and tenancy width indicates the number of unit spectrum slices allocated in a scheduling interval. The concepts of these parameters are illustrated in the left part of Fig. 1, with definitions and specific values explained in Fig. 2 in Section II.

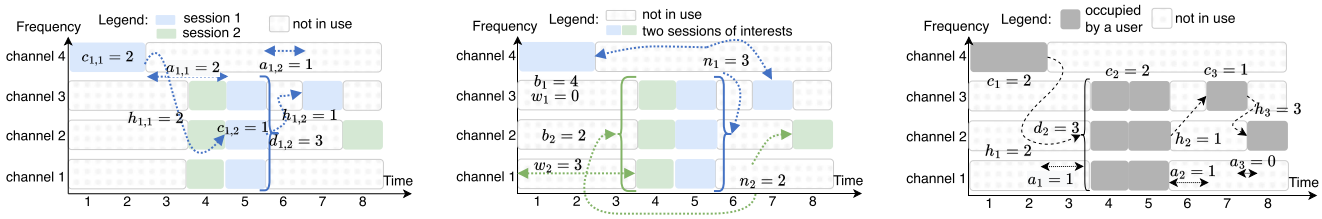
Compared to classical spectrum tenancy models, spectrum mobility captures the dynamic usage of individual users across different channels over time. For instance, Fig. 1 illustrates the comparison between spectrum mobility and the classical on/off model. The on/off model merely approximates the duration of busy and idle periods across frequency channels based on a majority vote of the tenancy at its true resolutions, without considering which specific devices are utilizing the spectrum resources or assuming any interconnections among the tenancy across various channels. In contrast, the spectrum mobility model incorporates factors such as camping times, access intervals, channel hops, and tenancy widths to accurately characterize the spectrum usage of individual devices. This feature is particularly pertinent for modeling the spectrum usage of Orthogonal Frequency-Division Multiplexing (OFDM) systems, which allocate spectrum resources across different frequency channels to various users, as seen in LTE and 5G NR systems.

Given the current dominance of LTE systems, their seamless transition to future 5G, and their emerging coexistence with other wireless systems supported by technologies like LAA and MulteFire, we focus on the spectrum usage of a commercial LTE cell under diverse wireless traffic conditions. To this end, we set up a measurement system to capture raw spectrum tenancy data. High resolution is achieved by decoding downlink control messages that contain the spectrum assignment decisions from the LTE base station. Consequently, the spectrum tenancy resolutions match the LTE scheduling granularity, specifically 1 ms by 180 kHz. To extract samples of the variables defined in the spectrum mobility model, the measurement data undergo further processing. This includes identifying spectrum users and excluding tenancy not related to user data transmissions, such as spectrum slices used for cell broadcast. Although we apply the proposed spectrum mobility model to the spectrum tenancy of a specific LTE cell in this work, it is important to note that our model can characterize the spectrum usage of various wireless systems, provided spectrum tenancy data are available for the four parameters. Even if high-resolution spectrum tenancy data for individual users are not available, the proposed spectrum mobility model can still characterize aggregated spectrum tenancy as explained in Section II and depicted in Fig. 2.

After conducting measurements and post-processing, we analyze the distributions of variables defined in the spectrum mobility model. These distribution studies yield valuable insights into the high-resolution spectrum usage of individual users within the model. Notably, some findings are counter-intuitive and contradict existing studies. For example, we observe decreased busy times during periods of high wireless traffic and surprisingly short idle times even during low traffic conditions.

Our contributions are summarized as follows. First, we develop the Frame-Level spectrum Model (*FLuMe*), which includes the *camping time*, the *inter-access time*, the *hopping distance*, and the *tenancy width*, by monitoring spectrum tenancy in the same time-frequency resolutions as scheduling. The model parameters are studied under low, medium, and high wireless traffic conditions,[1] and the details of the model definitions and the data collection process are elaborated in Sections II and III. The spectrum mobility model addresses the two challenging features required by emerging spectrum sharing and coexistence: high-resolution data and individual spectrum models. Second, we study the distributions of the model parameters based on large amounts of spectrum data obtained through real-world measurement and post-processing. Contrary to previous studies,

[1]The dataset will be made available to the public on *GitHub* upon acceptance of the manuscript with an agreement of use.



(a) Access variables. (b) Session variables. (c) *Aggregated* spectrum strides.

Fig. 2. An example of *FLuMe* parameters.

our findings reveal that busy time lengths are skewed towards the lower end when wireless traffic is high. The idle times, or inter-access times, concentrate on small values below 10 ms, even in low traffic scenarios. Third, the proposed spectrum mobility model can generate synthetic spectrum tenancy traces with distributions identical to our large-scale measurement data. By tuning the parameters, the spectrum mobility model can also synthesize spectrum usage patterns reported in other studies [12], [18], [19].

II. The *FLuMe* Framework

Since modeling the spectrum usage of individual wireless devices is missing yet crucial for improving spectrum efficiency, we propose the *FLuMe* framework. It models spectrum usage behavior of individual devices in two levels, the *access variables*, and the *session variables*, which is inspired by wireless traffic models that categorize user data arrivals into sessions to capture both intra-session and inter-session features [20], [21].

A. Access Variables

Access variables describe how the spectrum usage of individual users move along the two dimension time-frequency grid when wireless transmissions are actively taking place in a session. Each move is characterized by a camping time, a hopping distance, an inter-access time, and a tenancy width.

Specifically, we assume that spectrum resources are assigned in orthogonal frequency slices and discrete time intervals, which is reasonable in most wireless systems that use OFDM or OFDMA techniques, e.g. LTE/5G. In the example illustrated in Fig. 2(a), the number of orthogonal frequency channels $m = 4$ is shown by the vertical axis, and the horizontal axis records spectrum mobility in discrete time intervals. The differential spectrum mobility strides are denoted by $\{s_{j,i_j} = (c_{j,i_j}, h_{j,i_j}, a_{j,i_j}, d_{j,i_j}), j = 1, 2, \ldots, J, i_j = 1, 2, \ldots, I_j\}$, where the four elements, $c_{j,i_j}, h_{j,i_j}, a_{j,i_j}, d_{j,i_j}$, are instances of the random variables C, H, A, D for camping time, hopping distance, inter-access time, and tenancy width. The session index j ranges from 1 to J, and i_j is the index for individual-access variable samples in the session j, from 1 to I_j. Specifically, the camping time C is the amount of time when a piece of spectrum resources is continuously occupied, starting from the beginning of the first time slot to the end of the last one. For example, the first camping time in session 1 shown in blue in Fig. 2(a) is $c_{1,1} = 2$, and the second sample is $c_{1,2} = 1$. The hopping distance is the difference between the index of the starting frequency channel and that of the ending channel. When multiple frequency channels are utilized, the channel index is considered as the arithmetic mean of the index all the occupied channels rounded up to the nearest integer. For instance, the second hopping distance is $h_{1,2} = 3 - 2 = 1$, as the destination channel index is 3 and the source channel index of the center frequency is 2; the first hopping distance of the blue user is $h_{1,1} = 2 - 4 + m = 2$, where the number of channels $m = 4$ is added to bring the negative hopping distance to a positive value. The inter-access time is the number of time slots between two spectrum accesses, and it is measured from the end of the previous channel usage to the beginning of the next, e.g., the first inter-access time of the blue user is $a_{1,1} = 2$, and its second one is $a_{1,2} = 1$. The tenancy width D is the number of occupied frequency channels in one spectrum access in a session. In Fig. 2(a), the second access in session one has a tenancy width of three, i.e., $d_{1,2} = 3$.

B. Session Variables

Now that single spectrum accesses within sessions have been captured by the access variables, we continue to propose the session variables, the starting channel, the inter-session times, and the number of accesses within a session.

We characterize the spectrum mobility sessions by a sequence of session variables, $\{\bar{s}_j = (b_j, w_j, n_j), j = 1, 2, \cdots\}$. The b_j is a sample of the random variable B, the index of the frequency channel on which a session starts. The number of time slots between the starting time of consecutive sessions $j - 1$ and j is denoted by w_j, an instance of the random variable W. The number of discontinuous spectrum accesses in session j is n_j, an instance of the random variable N. In Fig. 2(b), we illustrate two sessions for different users, shown in blue and green. Since the blue session starts from the first time slot, its inter-session time is deemed to be zero, i.e., $w_1 = 0$, and $w_2 = 3$. The index of starting frequency channels and the numbers of discontinuous spectrum accesses of the two sessions are $b_1 = 4$, $n_1 = 3$, and $b_2 = 1, n_2 = 2$.

The session variables are a necessary complement to the access variables to specify the numbers of user strides, the starting frequency channel, and the time between sessions. By grouping spectrum accesses into sessions, there are fewer outliers in the samples of the spectrum mobility variables, so their distributions can be well captured by functions with a small number of parameters.

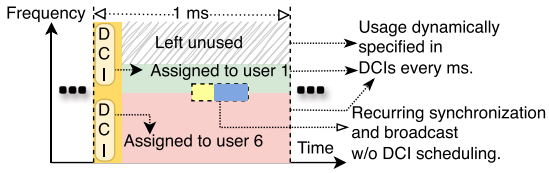


Fig. 3. Scheduling of downlink LTE spectrum usage.

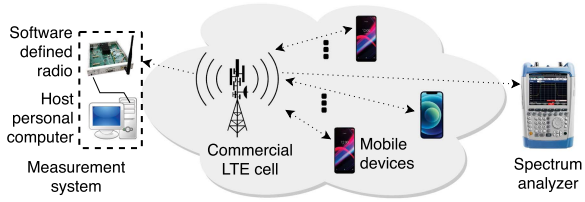


Fig. 4. The setup of the measurement for *FLuMe* variables.

The *FLuMe* framework is a versatile spectrum model, since it not only characterizes the spectrum activities of individual devices, but also models the aggregate usage, which is the only function of classical spectrum usage models that do not differentiate spectrum users [22]. When the *FLuMe* framework is applied to the aggregate spectrum tenancy, we assume that the inter-session times are zero, because the chance that all individual users in a wireless network remain silent over extended time is rare. Due to the assumption of zero inter-access time, all the aggregate spectrum usage is regarded to be in a single session, so the access variables alone are enough to describe the aggregate spectrum tenancy. In the example in Fig. 2(c), we apply the access variables to describe the aggregate spectrum mobility without differentiating individual users. There are three complete strides. The first aggregate tenancy camps on a single channel for two slots, and then hop two channels after one time slot, i.e., $c_1 = 2, h_1 = 2, a_1 = 1, d_1 = 1$. Similarly, the second and the third movement of the aggregate spectrum usage in this example can also be fully captured by the access variables as illustrated.

III. Data Collection and Processing

To apply the *FLuMe* framework to high resolution data, here we present the details of how we collect spectrum usage in *fine granularity* and how they are processed for extracting features for the proposed *FLuMe* framework.

A. LTE Spectrum Usage

Since *FLuMe* reflects spectrum tenancy with the same resolution as scheduling, we first provide a brief primer on spectrum usage in LTE downlink to ensure self-containment.

In the downlink of an LTE cell with frequency-division duplexing, the assignment of spectrum resources for user traffic is centrally scheduled by the base station. It schedules the spectrum resources in the basic time unit of 1 ms. In the frequency domain, the size of the basic unit is, 180 kHz, the same with that of an LTE Resource Block (RB) which is 0.5 ms in time. The assignment of RBs is contained in unencrypted Downlink Control Information (DCI) that reside in the beginning of every scheduling interval. Fig. 3 depicts an example of spectrum usage in one scheduling interval. During a scheduling interval of 1 ms, the symbols in the beginning of the interval contains two DCIs that dynamically allocate some RBs to user 1 and user 6, and some RBs are left unused. The center of the spectrum resources are for synchronization and system information broadcast, which occur at predefined intervals and do not require dynamic scheduling.

B. Data Collection

To collect the desired features of spectrum usage, measurements must address three challenges. First, spectrum usage must be captured with the same time-frequency resolution as users, specifically 1 ms by 180 kHz in LTE systems. Second, spectrum usage data must be associated with individual devices and their traffic sessions to establish a mapping between traffic sessions, spectrum usage, and individual spectrum access. Lastly, the dataset must be sufficiently large to incorporate spectrum tenancy patterns under various wireless traffic conditions, ensuring the general applicability of the *FLuMe* framework to different traffic levels. Although several large-scale measurement campaigns exist [23], [24], the data resolutions in these studies are far coarser than the LTE scheduling interval of 1 ms, with time granularities ranging from 50 to 3750 seconds. Moreover, these datasets do not allow for tracing the spectrum usage of individual users. Given the absence of spectrum usage datasets that address all these challenges, we must establish a spectrum measurement system to collect the necessary data for *FLuMe* modeling.

The measurement setup is illustrated in Fig. 4. Our measurement tool is composed of a Software Defined Radio (SDR) system and a host personal computer (PC). The SDR includes a Universal Software Radio Peripheral (USRP) X310 motherboard [25], and two SBX-120 wide-band daughter-boards [26]. The host computer has a quad-core CPU and 16GB memory, running Ubuntu 16.04. The radio front end realized by the SDR communicates with the host PC using USRP Hardware Driver (UHD) [27] version 3.9.7. To scan the spectrum usage data of a commercial LTE cell, we first search LTE bands in commercial operations near our lab. The existence of LTE control channels is then verified by a spectrum analyzer. We choose the nearby cell with the best signal-to-noise ratio (SNR).

The key steps to achieve the decoding based high resolution measurement for user spectrum activities are depicted in Fig. 5. First, the SDR radio front tunes onto the LTE band verified by the spectrum analyzer, and turns the analog carrier frequency signals to baseband complex samples. Through an Ethernet cable, the complex samples are passed to the PC for further processing.

Within the host PC, the first step in decoding spectrum usage is to achieve time and frequency synchronization based on the periodical LTE main broadcast and synchronization signals in the center of the time-frequency spectrum resources in LTE subframes as shown in Fig. 5. Next, we need to decode the DCIs and Radio Network Temporary Identifier (RNTI) to find out the assignment of spectrum resources and the associated user IDs. Because DCIs are not encrypted, and their locations and

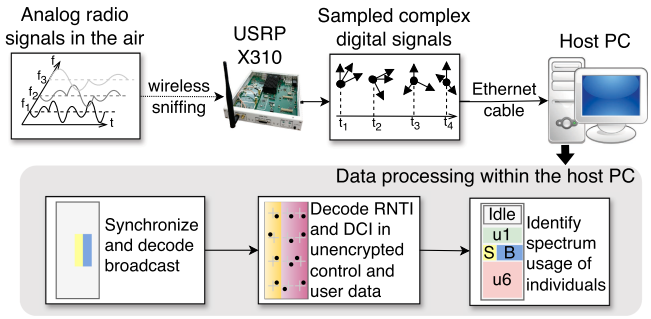


Fig. 5. Key steps in the measurement.

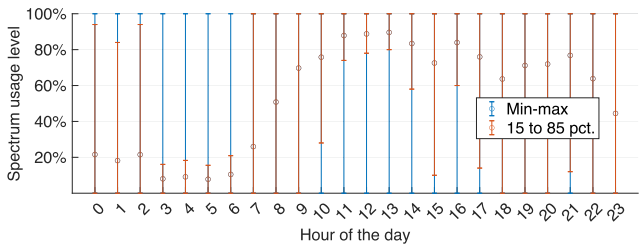


Fig. 6. Various spectrum usage levels.

modulations are fixed, we can do a blind decoding of DCIs. However, inferring RNTIs and validating decoded DCIs are more challenging. Existing works have proposed an effective way to infer RNTIs by computing the DCI checksum and then XORing it with the last two bytes, because they are the XOR of the RNTI and its CRC checksum [28], [29]. This method assumes that the entire DCI is correctly decoded, and forfeits the error detection power of DCI checksum. To reliably decode DCIs and RNTIs, we adopt the method that validates RNTIs by further applying them to user data decoding [2]. Since the adopted decoding method keeps the original checksum mechanism in LTE DCI, the decoded DCIs are error free, which yields the accurate high-resolution measurement.

After the DCI and RNTI decoding, the spectrum tenancy of the LTE cell is parsed from DCIs and measured exactly as the scheduling. However, part of the spectrum usage in an LTE cell is not the results of user data transmissions. For example, some LTE RBs are occupied regularly to carry LTE System Information Blocks (SIBs) to announce physical layer information related to cell access and selection, even if there is no user traffic at all. Such transmissions consume spectrum resources, but they do not reflect spectrum activities due to user data transmissions. As the goal of the spectrum mobility model is to link spectrum tenancy to data transmission activities rather than study the static spectrum usage of Media Access Control (MAC) and physical (PHY) layer signaling, we need to find a way to identify the spectrum tenancy independent of user data transmissions. According to Table 7.1-1 in [30], the range of RNTIs that correspond to user data transmissions are 2401 to 65523. Hence, we further identify the spectrum tenancy for user data transmissions based on RNTI values in the last step of the measurement, by marking RBs as occupied by RNTIs in the range of 2401 to 65523 or unused by 0. Compared with classical spectrum modeling data sets [24], our data achieves high resolutions and links directly to spectrum tenancy due to user data transmissions with the spectrum usage of MAC and PHY layers filtered out.

The downlink system bandwidth of the cell is 10 MHz that accommodates 50 LTE RB, which is equivalent to the number of frequency channels in *FLuMe*, i.e., $m = 50$. We collect spectrum usage data of the nearby LTE cell in different time periods of the day, obtaining 4320 million binary tenancy in total. Each spectrum usage is either marked as unused by zero, or occupied by the user RNTI, i.e., the temporary user ID in the LTE physical layer. To ensure that the spectrum usage covers all possible conditions, we choose a day in which the LTE cell under observation exhibits the most diverse distributions of spectrum usage levels over the 24 hours as illustrated in Fig. 6. Each bar shows the distribution of the 3.6 million spectrum usage level in one of the 24 hours, by marking the 0th, 15th, 50th, 85th, and 100th percentiles by bars and circles. As we can see, these 24 hours cover diverse spectrum tenancy levels, as some hours have the majority of the scheduling intervals occupied by over 90%, while some hours witness low spectrum usage below 10% most of the time. Such diverse distributions of spectrum usage levels in the dataset ensure that the *FLuMe* model is applicable to all spectrum usage conditions.

C. Data Processing

Though we have achieved the spectrum tenancy of user data in the same resolution as scheduling, further data processing is still required to extract data samples for the *FLuMe* variables. Two important practical issues addressed in post-processing are elaborated as the following.

Trace the user IDs: The first challenge is how to trace individual spectrum usage given that the recorded tenancy is marked by RNTIs, a temporary ID, rather than a user ID with traffic information. This challenge stems from the RNTI design, which is supposed to hide user identities in the physical layer by frequent updates of RNTI values. According to LTE specifications, RNTIs need to be updated whenever a user goes from an idle to a connected state [31], which occurs around tens of seconds [32]. In practical systems, RNTI update intervals can be as long as several hours in the US [33]. Recall that the *FLuMe* variables described in Section II, camping times C, inter-access times A, and inter-session times W, are much shorter than the RNTI update intervals in practical systems. This is because these variables reflect the dynamic changes in spectrum usage. In fact, they are very *unlikely* to be larger than 200 ms, as $\mathbb{P}(C > 200) = 0$, $\mathbb{P}(A > 100) = 0.0537$, $\mathbb{P}(W > 100) = 0.0631$ according to our analysis, where $\mathbb{P}()$ measures the probability. Hence, the RNTI update intervals are at least 50 times larger than the differential camping times, inter-access times and inter-session times. Thus, the RNTI updates only affect a small portion of the samples of C, A and W by splitting them into different spectrum mobility sessions.

Find the session end: The second challenge is how to determine the end of a spectrum mobility session. Because RNTIs

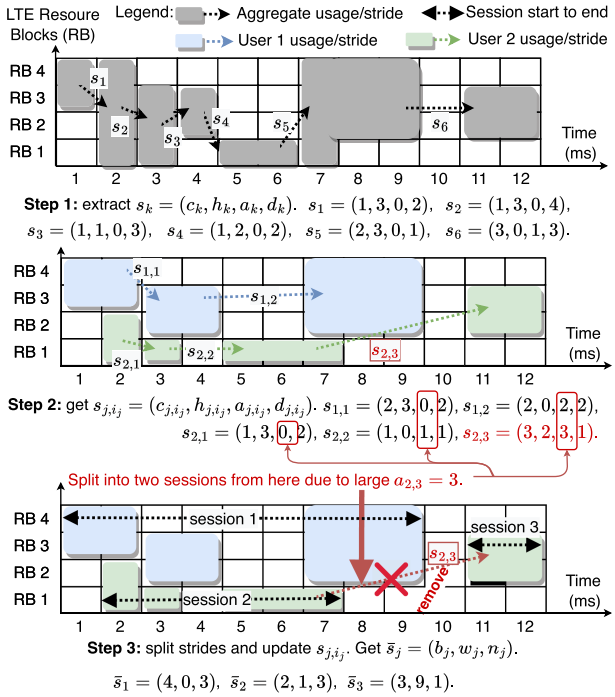


Fig. 7. An example of the three-step data processing.

are updated at long intervals ranging from tens of seconds to hours, the spectrum usage with the same RNTI may span over multiple user traffic sessions. For example, user traffic sessions, such as HTTP sessions, typically time out under 600 seconds [34]. Thus, we need to determine the end of spectrum mobility sessions to split a long sequence of spectrum usage with the same RNTI into multiple sessions so that the spectrum mobility model better reflects the spectrum usage due to user data transmissions. To separate long spectrum strides with the same RNTI into shorter sessions, we first analyze the individual inter-access times without session boundaries, denoted as A^*. Then, we deem that samples larger than the 90th percentile of A^* are formed by the spectrum accesses on the ends of two adjacent spectrum mobility sessions with the same RNTI. From our statistical analysis, we find $\mathbb{P}(A^* \leq 116) = 0.9$, and we thus split the spectrum usage with the same RNTI at the accesses after which the immediate next one is 116 ms or longer away.

Now that the two challenges in post-processing have been addressed, we introduce the three steps in our top-down approach to process the more than 37.8 million aggregate spectrum mobility strides $s_k, k \in \{1, \ldots, K\}, K \geq 37.8 \times 10^6$. As illustrated in Fig. 7, the three steps to extract the desirable features for *FLuMe* are as follows. First, we extract the aggregate spectrum mobility strides, that is, the overall usage of all LTE RBs. Then we locate the center frequency channels and extract the aggregate camping times, hopping distances, inter-access times, and the tenancy width, as shown in gray blocks in step 1in Fig. 7 without differentiating individual users, such as User 1 and User 2in the example. Second, we extract the individual-access variables s_{j,i_j} without cutting the sessions with the same RNTIs. The individual-access samples are processed before session variables because we cut sessions with the same RNTI between tenancy with long inter-access times as previously explained in the practical challenges. In the example in Fig. 7, the five inter-access times in the two sessions are 0,0,1,2, and 3. In the last step, we split the spectrum accesses of the same RNTI based on the inter-access time samples, a_{j,i_j}. For instance, if we choose the 80th percentile of inter-access time samples as the cutting length, then the user 2 session in step 2 is separated into two sessions at the point in step 3 shown in the figure, because $\min_a(\mathbb{P}(A^* \leq a) \geq 0.8) = 3$. As a result, three sessions are identified in the example in Fig. 7.

For the spectrum mobility of individuals, we obtain over one million samples of session variables $\bar{s}_j, j \in \{1, \ldots, J\}, J \geq 10^6\}$ in which there are nearly 200 million samples of individual-access variables $s_{j,i_j}, i_j \in \{1, \ldots, I_j\}, \sum_j I_j \geq 200 \times 10^6$.[1] The size of the data is 8 GB when zipped. Given the data size and the fact that they are collected in carefully chosen hours with diverse traffic levels, our data set exhibits substantial statistical significance even when compared with large scale measurements, such as the 2GB of 14900 × 8058 readings in [35]. Next, we present data analytics for the *FLuMe* variables.

IV. Observations With *FLuMe* Model

In this section, we analyze the data traces of the spectrum mobility and investigate the statistical characteristics of the *FLuMe* variables. We evaluate the measurement data analytics using four types of functions: exponential, power, quadratic, and power over logarithm.

A. Preliminaries

Our objective is to discover new observations of dynamics in spectrum usage, with a focus on new features of *FLuMe* spectrum mobility. To achieve this, we are considering the following issues.

Traffic loads: We analyze the distributions of *FLuMe* variables based on different levels of spectrum usage. The usage level, denoted as U, is defined as the ratio of occupied RBs to the total number of RBs during a camping time in one stride or session. The distributions are studied under three usage levels: $U \in [0, 0.3]$ for low traffic, $U \in (0.3, 0.7]$ for medium traffic, and $U \in (0.7, 1]$ for high traffic.

Fitting functions. The goal of the data-driven modeling approach is to identify fitting functions that can be used for proactive system design and performance evaluation, both with and without runtime measurements. To achieve this, we consider six functions and select the best three for summarizing our observations, that is,

$$f(x) = \lambda_1 e^{\lambda_2 x}, \tag{1}$$

$$f(x) = \lambda_1 e^{\lambda_2 x} + \lambda_3 e^{\lambda_4 x}, \tag{2}$$

$$f(x) = \lambda_1 x^{\lambda_2}, \tag{3}$$

$$f(x) = \lambda_1 x^{\lambda_2} + \lambda_3, \tag{4}$$

$$f(x) = \lambda_1 x^2 + \lambda_2 x + \lambda_3, \tag{5}$$

$$f(x) = \lambda_1 x^{\lambda_2} / \log(x + \lambda_3), \tag{6}$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are constant parameters estimated by Maximum Likelihood Estimation (MLE). These rudimentary functions have few fitting parameters and have been proved to capture the network traffic models characteristics [36]. As we will show later, these functions accommodate well to the spectrum mobility variables with limited fitting constants, rendering the FLuMe framework practical and easily applicable.

Quality of statistical models. We adopt the Akaike Information Criterion (AIC) for evaluating fitting performance, considering both the number of parameters in the function and the goodness of fit. The AIC measures the relative quality of statistical models, with smaller AIC values being preferred as they indicate models that capture the distributions effectively with fewer parameters. Since absolute AIC values are influenced by sample size and are primarily used for comparison, we use the more intuitive and meaningful AIC weights for performance comparisons [37]. The AIC weight for the v-th fitting function is calculated as $\omega_v = \frac{e^{-\Delta_v/2}}{\sum_{v'=1}^{V} e^{-\Delta_{v'}/2}}$, where V is the total number of candidates and $\Delta_v = AIC_v - AIC_{min}$ is the difference from the smallest AIC value. AIC weights range from 0 to 1, with the best fitting function among the V candidates having an AIC weight closest to one. The AIC value is

$$AIC = 2r - 2\log(L), \tag{7}$$

where L is the optimized scalar value of log-likelihood objective function, and r is the number of parameters that are estimated for the model. The likelihood function of Q observations $x_1, \ldots, x_Q$ of a random variable X with the parameter θ in its probability mass function $f(x|\theta)$ is

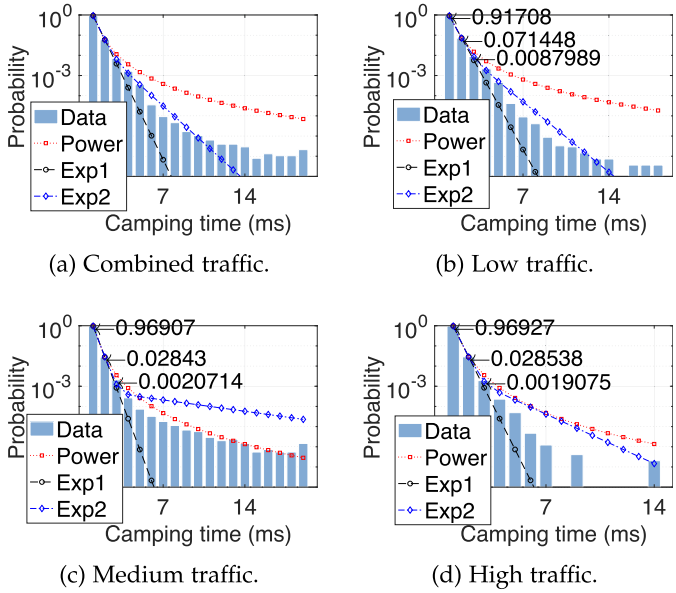
$$L(\theta) = \prod_{q=1}^{Q} f(x_q|\theta). \tag{8}$$

The proposed *FLuMe* framework offers both *individual* and *aggregate* spectrum mobility. The former captures the spectrum usage of each individual device, and more importantly, it reveals a *local* view of spectrum utilization in non-centralized systems. On the other hand, the latter describes the total spectrum usage, and tells a *global* view of spectrum usage, which is best suited for centralized resource management and scheduling in cellular systems, and is consistent with existing works on spectrum database and spectrum activity maps [23], [38].

B. Observations of Aggregate Spectrum Tenancy

Here we present our findings on camping time, inter-access time, hopping distance, and the tenancy width of aggregate spectrum mobility.

1) Aggregate Camping Time: We find that the three best fitting functions for the distributions of aggregate camping time are power law, exponential (Exp1) and two-term exponential (Exp2), which are studied under low, medium, and high traffic loads and shown in Fig. 8. The y-axis is the probability in log scale. Specifically, Fig. 8(a) plots the distributions of all camping times (ms) without conditioning on spectrum usage, and the probability of camping for one millisecond takes the major share close to one. The two-term exponential function



(a) Combined traffic. (b) Low traffic. (c) Medium traffic. (d) High traffic.

Fig. 8. Distributions of the *aggregate* camping time: low, medium, and high traffic loads.

achieves the best fitting performance with the AIC weight ω of 1 in most traffic conditions. We observe that the two-term exponential function captures both the initial components when the probabilities decrease sharply and the later components when the declining trend dies down. For example, the exponential decay factor is $\lambda_2 = -3.543$ for the fast decrease in the beginning, and then turns to $\lambda_4 = -0.1845$ in the two-term exponential function for the medium traffic case for $\mathbb{P}(C = x|0.3 < U < 0.7)$.

By taking a close look at the camping time distributions, regardless of traffic loads, they become more weighted towards the first few components, i.e., short camping times of a several milliseconds. This differs from previous studies using two-state Markov chain [39], where the camping time increases with the traffic load. We believe this is due to the frequency granularity, because the finer-grained spectrum usage may yield shorter camping time on each RB. Multiple RBs are used to meet the traffic demands, instead of occupying the same RB for a longer time. That also explains that in case of high traffic, slicing spectrum resources into narrow frequency channels is flexible in providing higher data rate on-demand, which is consistent with the intuition of spectrum division and multiplexing. This result also suggests that spectrum sensing may be more often for cognizant users under high traffic, in that LTE RBs are scheduled more often at larger quantities such that the center frequency changes more often than that under lower traffic.

2) Aggregate Inter-Access Time: We plot the distribution of inter-access time with both x and y axes in the log scale, and based on the AIC weights, the best regression functions are the power function, the power function with a constant term, and the power over logarithm function, which are marked by 'Power1', 'Power2', and 'POL', respectively in Fig. 9. Recall the inter-access time is the time interval between using spectrum bands, which conceptually, should be heavily dependent on the traffic loads; nonetheless, we do not observe a dramatic

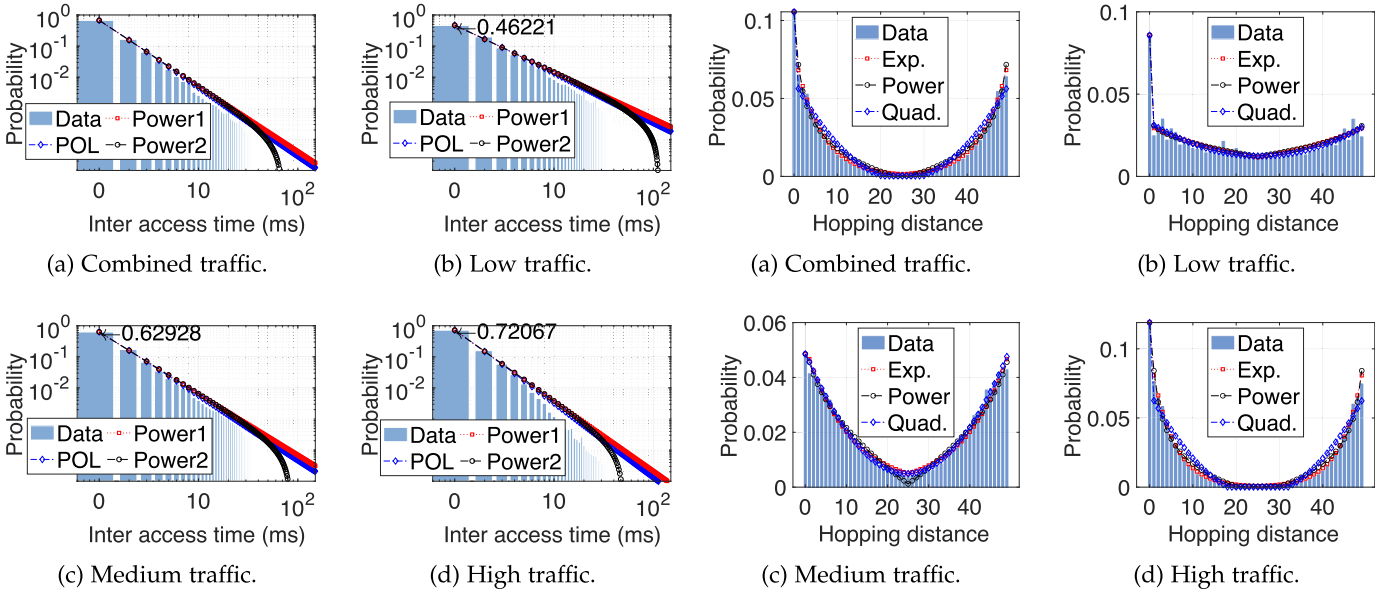


Fig. 9. Distributions of *aggregate* inter-access time.

Fig. 10. Distributions of *aggregate* hopping distance.

difference with the trends of the negative exponents growing smaller to steepen the decrease of possibilities as the inter-access times increase. The major difference is the probability of the most likely component, i.e., $\mathbb{P}(a = 0)$, which is 0.4622, 0.6293, 0.7207 in the three different traffic conditions, respectively. The more wireless traffic there exists, the inter-access time is more likely to be small, which is in line with the intuition that idle time is shorter during high traffic. The importance of measurement resolution can be appreciated from the fact that most inter-access times are under 10 ms, which cannot be observed in coarse resolutions. Therefore, both camping time and inter-access time distributions reveal high probability of small values under 10 ms, advocating that the dynamics of spectrum mobility favor spectrum measurements in fine granularity.

3) Aggregate Hopping Distance: We observe the distribution of aggregate hopping distance and discover that the probability of hopping distance zero is a singular point that stands out from the rest of the probabilities in Fig. 10. Moreover, the distributions of hopping distances other than zero are symmetric with respect to the hopping distance of $k/2 = 25$. Based on these observations, we fit the functions to the hopping distance distributions when the distance is in the range of [1, 25], and then flip the curve with respect to the line of $x = 25$.

The three functions that best fit the distributions of aggregate hopping distances are the exponential function, the power law function with a constant term, and the quadratic function. The fitting performance of the three functions varies in different spectrum usage conditions. When the traffic level is low, the power law function with a constant achieves the best fitting result; the quadratic function best fits the aggregate hopping distances when the traffic load is medium; the distribution is best fitted by the exponential function in the high traffic case.

When the traffic load is low, the spectrum mobility tends to be idle, i.e., staying on the current center frequency. If a frequency hop happens in this case, its next center frequency channel is similar to an RB chosen uniformly random due to the following two reasons. First, the number of occupied RBs in a time slot during low traffic is small, so the available choice of the center frequency is close to all the RBs. Second, the noise-to-signal ratio in the 50 RBs are similar with one another when examined at long time scales [40]. Since there is no constraint on the choice of frequency channel in low traffic, the hop distance tends to be uniformly chosen from all RBs. As the traffic load increases, the hopping distance starts to favor nearby spectrum bands. When the traffic load is high, the probabilities of further hops to bands around 25 are close to 0. This is because the majority of the spectrum bands are occupied each time slot under high traffic, so the hopping destinations of the center frequency are more limited to the middle of the system bandwidth after taking the average of the index of occupied RBs, allowing only short hopping distances. The hopping distributions provide valuable guidance for spectrum tenancy predictions based on spectrum mobility. For example, the large hops around 25 are most likely to happen during low traffic, while hops from 20 to 30 have close to zero possibility to appear.

4) Aggregate Tenancy Width: The distribution of the aggregate spectrum tenancy width is depicted in Fig. 11. The three functions that best fit the measurement data are the power law with a constant, the two-term exponential, and the quadratic function. The tenancy width exhibits a strong inclination to fully utilize all 50 RBs and the remaining widths are distributed evenly among the unused RBs. In terms of energy consumption, having all RBs in a time slot occupied with data is more efficient, as the same energy is used as long as one RB carries data. Thus, operators have an incentive to maximize the number of time slots with all RBs occupied to save energy.

Packing LTE traffic into all RBs is also beneficial for the coexistence of various wireless networks as it leaves more blank time slots to avoid collisions. The main distinction in the width distribution among the three traffic levels is that the

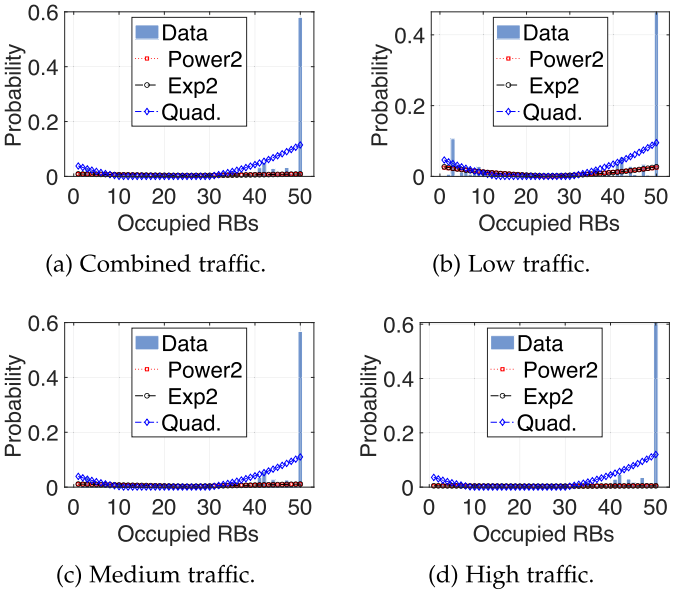


Fig. 11. Distributions of *aggregate* tenancy width.

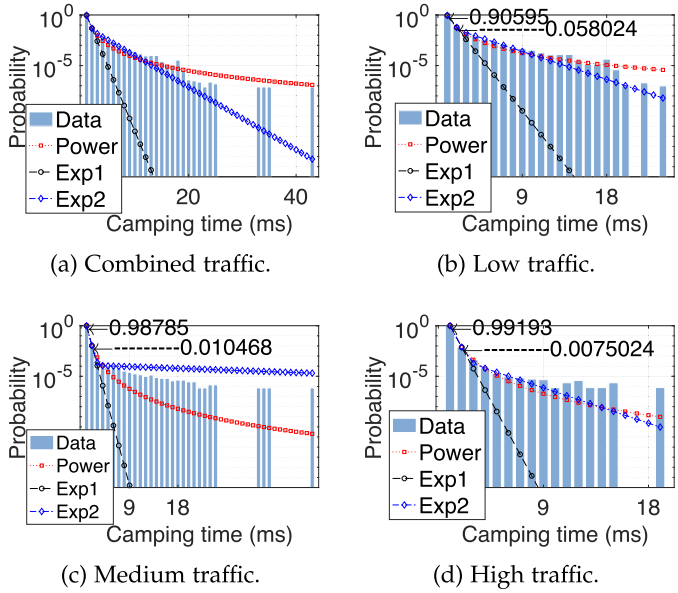


Fig. 12. Distributions of *individual* camping time.

probability of using all 50 RBs increases with traffic, while the even distribution of the remaining probabilities among the other widths remains constant.

C. Observations of Individual Spectrum Mobility

Different from the variables defined for the aggregate spectrum mobility, the individual spectrum mobility is extracted by tracing the spectrum mobility sessions of single mobile users. In this subsection, we explain the observations of camping time, inter-access time, and hopping distance from the perspective of individual users.

1) Individual Camping Time: We observe a similar pattern in the distribution of *individual* camping times as seen in the *aggregate* model. Specifically, when plotted with a logarithmic vertical axis, the best three fitting functions are the power function, exponential (Exp1), and two-term exponential (Exp2), as illustrated in Fig. 12. Under medium traffic conditions, the camping time is best fitted by the power function, whereas, at all other traffic levels, the two-term exponential function provides the best fit, with $\omega = 1$. It is noteworthy that the *maximum* individual camping time is approximately twice the length of that in *aggregate* camping time. Moreover, the distribution of differential camping times is more sensitive to traffic levels than its *aggregate* counterpart, as indicated by the more diverse values of the exponents λ_2 and d. This is because *aggregate* camping time captures the activities of all users, making it more stable than those of individuals. Under high traffic load, the spectrum resources assigned to individual users are very likely, with a probability exceeding 99%, to change every millisecond.

Remark 1: High traffic loads may yield more volatile spectrum assignments for individual devices. Therefore, the design of spectrum sensing and measurement sampling need to be highly adaptive for prediction of spectrum tenancy and achieve high spectrum utilization. When runtime spectrum usage cannot be provided, a fine-grained spectrum mobility model will be very helpful.

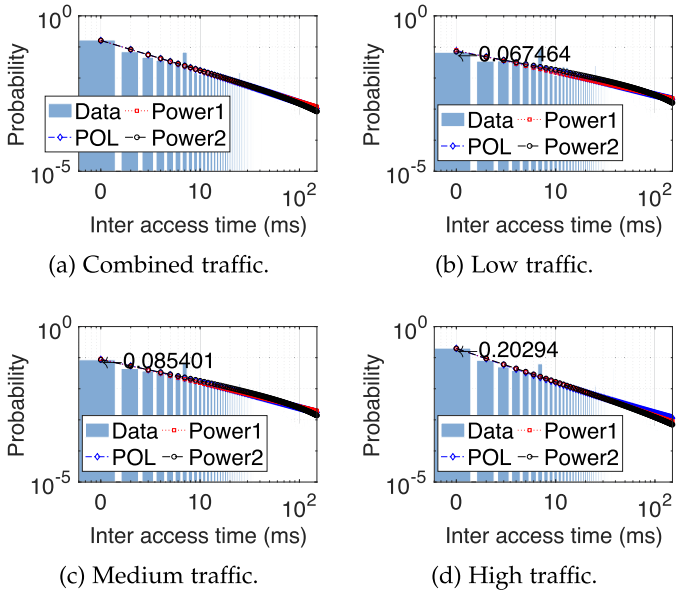


Fig. 13. Distributions of *individual* inter-access time under different traffic levels.

2) Individual Inter-Access Time: We find that the inter-access times bear long tails of small probabilities and use log scales on both axes to plot their distributions that are best fitted by the power (Power1), the power with a constant term (Power2), and the power over logarithm (POL) in Fig. 13 under different traffic loads. In particular, by comparing with the inter-access distribution of *aggregate* spectrum mobility, the *individual* inter-access time is best fitted with the power over logarithm function (POL) other than power with a constant term (Power2). This is because the individual inter-access times have larger exponents around -0.7 to -0.6 compared with those around -2.2 to -1.5

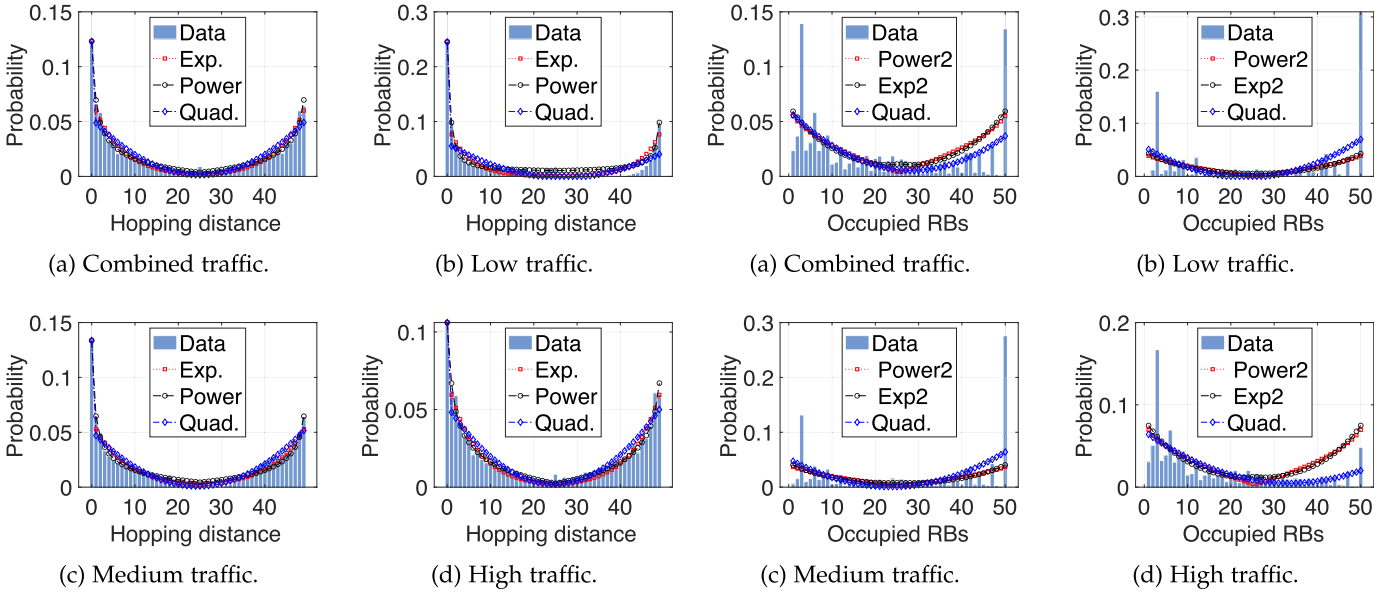


Fig. 14. Distributions of *individual* hopping distance under different traffic levels.

Fig. 15. Distributions of *individual* tenancy width.

in the *aggregate* case, showing a heavier tail than the power law. In other words, the *aggregate* inter-access time is less variable in comparison with *individual* devices, which captures a better long-term spectrum usage in contrast to a relatively *short-term*, local view of spectrum usage observed by *individual* devices.

Furthermore, it is noticed that the probability of zero inter-access time is much lower in the differential case than that in the *aggregate* model. For example, the probability of *aggregate* inter-access time being 0 is $\mathbb{P}(a = 0|U > 0.7) = 0.7207$ in heavy traffic, but the *individual* counterpart is $\mathbb{P}(A = 0|U > 0.7) = 0.209$. This is because the *aggregate* inter-access times are similar to taking the minimum of the inter-access times of all users.

Remark 2: On one hand, neither of the inter-access time distribution functions nor the probabilities of the major components in *aggregate* and *individual* spectrum share similar properties. On the other hand, camping time of *individual* devices presents a similar one- or two-term exponential distribution with the *aggregate* or global spectrum usage.

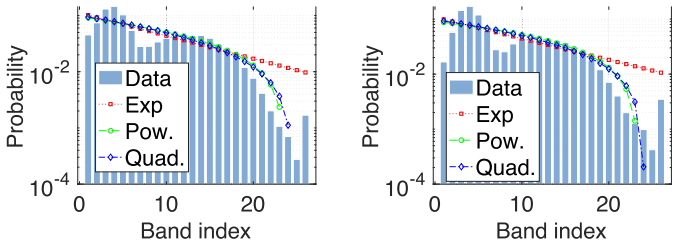
3) Individual Hopping Distance: We observe that the distributions of the individual hopping distances are best captured by the exponential, the power law, and the quadratic functions. The distributions and the fitted curves are illustrated in Fig. 14.

User perspective hopping distances in the low traffic can be best fitted by the power function, and the quadratic function achieves the best fitting in both medium and high traffic cases. This fitting result is different from that of the aggregate hopping distances, because individual hopping distances in the high traffic case demonstrates a more smooth decrease from the two sides to the center. This is due to the fact that individual users take up fewer amount of spectrum bands than the entire system, so there are more choices for the hopping destinations than only the few RBs near the center frequency of the system. Similar to the aggregate hopping distance distributions, hopping distances in the user perspective also have an outlier point at zero, and the rest of the probabilities are symmetric around the center. Thus, the next spectrum access is most likely to be centered on the same frequency channel, and this is true for both aggregate and individual spectrum mobility. Different from the aggregate spectrum hops, the probabilities of remaining in the previous RB $\mathbb{P}(H^u = 0)$ decreases as the traffic volume grows, because the spectrum accesses of more mobile users need to be accommodated and the average hopping distances are increased as a compromise. The different changes of hopping distance distributions caused by traffic intensities in aggregate and individual mobility can only be observed with spectrum measurement tagged by user IDs.

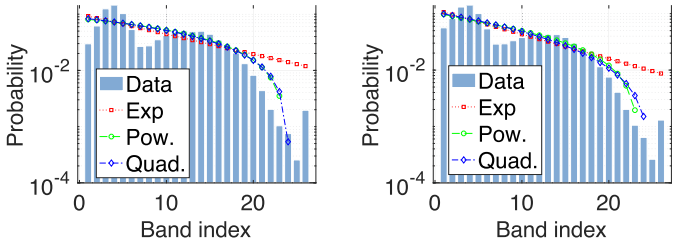
4) Individual Tenancy Width: The tenancy width for individual spectrum mobility is plotted in Fig. 15. Compared with the aggregate tenancy width, the individual tenancy width shows two different features. One is that the tenancy width distribution of individual users is symmetric with respect to the middle RB and grows larger towards the ends. The second feature of the distribution is that the probability of being assigned all the 50 RBs decreases as the traffic grows, which is contrary to the trend for aggregate spectrum mobility. This is intuitive because each user is assigned less spectrum resources when the cell traffic volume grows and the base station fairly allocates the limited spectrum to handle increased requests.

Next, we discuss the observations derived for *session* variables. Our aim is to determine whether the *initial* spectrum bands behave as random selections, as commonly assumed, or if they are influenced by inherent factors. This analysis leverages our high granularity measurements, which include associated user IDs.

5) Starting Channel Index: We index the total $m = 50$ channels, i.e., LTE RBs from 1 to 50 for which the probability

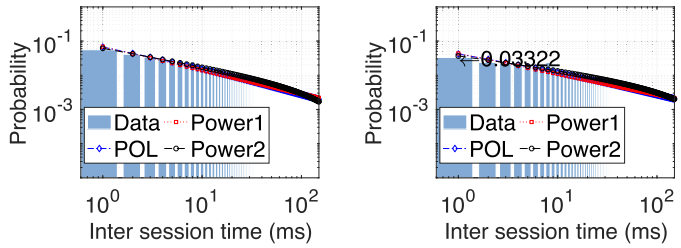


(a) Combined traffic. (b) Low traffic.

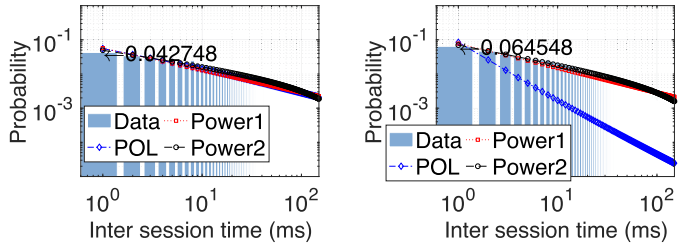


(c) Medium traffic. (d) High traffic.

Fig. 16. Distributions of starting channel index under different traffic levels.



(a) Combined traffic. (b) Low traffic.



(c) Medium traffic. (d) High traffic.

Fig. 17. Distributions of inter-session time under different traffic levels.

mass function of the frequency index where spectrum mobility sessions start is illustrated in Fig. 16, for various traffic levels. Among the three fitting functions, the power function with a constant term (Power2) achieves the best AIC weights, and the resulting distributions under different traffic levels are very similar, meaning that sessions of user spectrum mobility start on certain bands with similar probabilities regardless of the traffic conditions. Thus, the starting channel index in *individual* spectrum mobility is the only parameter that is independent of the wireless traffic conditions in *FLuMe*. Another observation is that the spectrum mobility sessions tend to start on channels in the lower frequency bands, and no sessions start on bands with index over 26, which is no arbitrarily random selection.

Remark 3: One reason for this design choice could be to ensure the successful establishment of initial communications in a session. Transmissions in lower frequencies experience less path loss, which is particularly advantageous for mobile devices with limited power. However, our discovery of the uneven distribution of starting channel indices contradicts the assumptions made in most channel access designs for multi-channel scenarios, where frequency channels are typically selected uniformly at random [41].

6) Inter-Session Time Distribution: The distributions of inter-session times and the fitting curves are illustrated in Fig. 17. Among the six functions, the three best fitting functions are the power function, the power function with a constant term, and the power over logarithm function. Since the inter-session times have long tails with small probabilities, both the horizontal and the vertical axes in the figure are shown in the log scale. Most of the inter-session times are under 100 ms in the three different traffic conditions under which the trend of the distribution functions are very similar. According to the AIC weights, the power function with a constant term best fits the inter-session time distributions.

Compared with the inter-access time distributions in aggregate and individual spectrum mobility, inter-session times not only are best fitted by a different distribution, they have a much smaller major component of $\mathbb{P}(W = 1)$. This probability is $\mathbb{P}(W = 1|U \leq 0.3) = 0.0332, \mathbb{P}(W = 1|0.3 < U \leq 0.7) = 0.0427, \mathbb{P}(W = 1|U \geq 0.7) = 0.0645$ for the low, medium, and high traffic conditions, respectively. Though the probability of inter-session times also decreases exponentially as the time length grows, the exponent is significantly smaller than that of the inter-access times. For example, the exponents of the best fitting functions of the distributions for inter-access time in aggregate and individual cases under high traffic are -2.268 and -0.7049, respectively, while the exponent in the function for the inter-session time distribution under the same traffic condition is -0.5353. Since a spectrum access session is composed of multiple spectrum accesses and much longer than a single spectrum access, the inter-session time is also much less likely to be as short as one.

Remark 4: Similar to the inter-access times that are curtailed during heavy traffic, inter-session times are more likely to be shorter during heavy traffic since spectrum access sessions arrive at higher rates. Though the traffic arrivals in important routers in data networks have been shown to follow power law distributions [42], inter-arrivals of spectrum access sessions are the first reported to follow power law shape distributions in this paper to the best of our knowledge.

7) Counting Spectrum Access: The spectrum access count specifies the number of discontinuous accesses, i.e., the number of individual spectrum mobility strides plus one during a session. Three functions, the power function, the power function with a constant, and the power over logarithm function, best fit the spectrum access counts in different traffic conditions. The fitting results and the empirical distributions of the data are presented in Fig. 18. Similar to the distribution of inter-session times, the

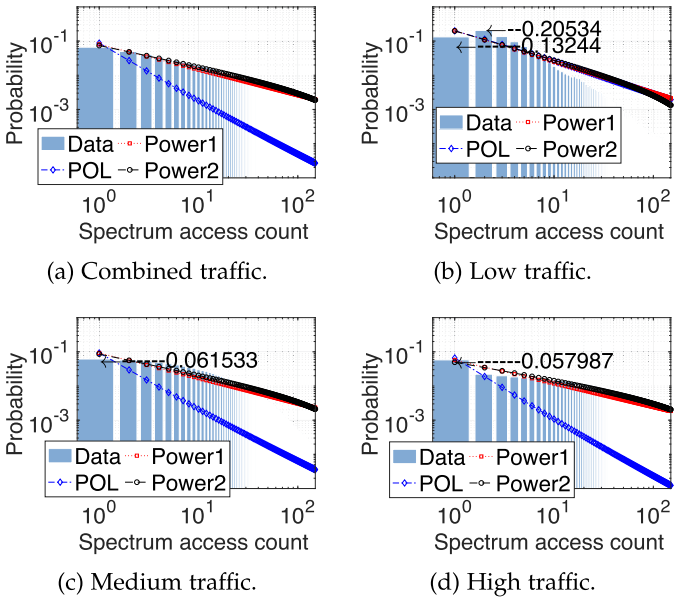


Fig. 18. Distributions of access count under different traffic.

distribution of spectrum access count also features a long tail with small probabilities, so both the vertical and the horizontal axes are plotted in the log scale. In low traffic, the spectrum access count is best fitted by power over logarithm function with the AIC weight of one, and the power function with a constant term best fits the distribution in other cases.

Since the major component in the heavy tail distributions takes up the largest probability, we examine the odds of $\mathbb{P}(N = 1)$. This probability is $\mathbb{P}(N = 1|U \leq 0.3) = 0.1324, \mathbb{P}(N = 1|0.3 < U \leq 0.7) = 0.0615, \mathbb{P}(N = 1|U \geq 0.7) = 0.0580$ for the low, medium, and high traffic conditions, respectively. Thus, the probability of having short mobility sessions decreases as the traffic volume increases. This is also corroborated by the increasing exponents in the fitting functions as the traffic intensity grows. Based on the distributions of the spectrum access counts, we observe that mobile users have more spectrum accesses during a session under heavy traffic.

Remark 5: According to the distribution studies on the number of spectrum accesses in a session and the inter-access times, a typical spectrum access session is composed of tens of accesses separated by a few ms, resulting in spectrum access sessions lasting from tens to a few hundred ms. Compared with wireless traffic sessions defined as the continuation of user activities which usually persists for hundreds to thousands of seconds, such as web-browsing or video streaming [43], the spectrum access sessions are much shorter and require measurement with higher resolutions.

V. Synthetic Tenancy and Validations

Besides the new observations enabled by the analysis of *FLuMe* variables in the high resolution spectrum tenancy, other key applications of the *FLuMe* framework is to synthesize spectrum usage data and predict spectrum tenancy.

A. *Synthesize Spectrum Usage With* FLuMe

Based on the versatile and light-weight *FLuMe* model, both individual and aggregate spectrum usage can be generated. In this subsection, we elaborate two ways to generate aggregate spectrum tenancy of an entire LTE cell, and one way to generate individual spectrum tenancy.

The steps to generate the aggregate spectrum usage of an LTE cell with *FLuMe* is as follows. First, choose the wireless traffic level that consumes the synthetic spectrum usage, from low, medium, or high, as defined by the utilization U. Then, select the index for the first center RB according to the distribution of B, the starting channel index. The aggregate camping time can be generated by the distributions in Fig. 8, and the tenancy width is governed by the distributions in Fig. 11. With the given tenancy time, width, and center RB, we can accurately specify the occupied spectrum resources in the time-frequency domain. The next tenancy can be found by generating an instance of the aggregate hopping distance and the inter-access time, H and A, and its size is again decided by the camping time and tenancy width in the same fashion as we do for the first spectrum usage. More four tuples of aggregate spectrum usage, (c_k, h_k, a_k, d_k), can be synthesized until the desired amount.

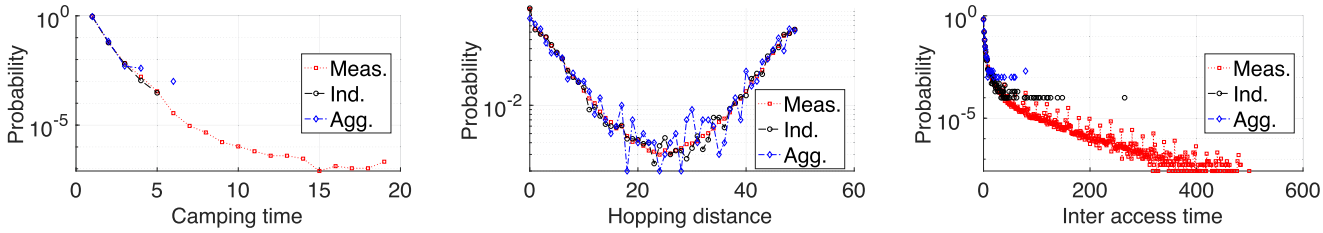
Besides generating the cell wide spectrum usage with aggregate *FLuMe* variables, the other way to synthesize aggregate tenancy is utilizing the individual tenancy and session variables. First, generate instances of the session variables (B, W, N), according to the distributions in Section IV-C. Within each session, we then generate individual spectrum usage strides for the number of times specified by the access count N, by selecting values of the four tuples for individual spectrum usage according to the distributions for random variables (C, H, A, D). After all the individual spectrum accesses for a session have been synthesized, the start time and the center RB position for the next spectrum usage session can be again decided by the session variables W and B.

In addition to the above two ways to generate aggregate spectrum usage, the *FLuMe* variables for individual spectrum mobility are capable of generating the spectrum tenancy of individual wireless devices. Similar to synthesizing aggregate spectrum mobility, we first select the center RB index based on the distribution of B. Then, the size of the tenancy is determined by the camping time C and the tenancy width D according to their distributions. The next individual tenancy starts after an instance of the inter-access time A on the center RB that is at a hopping distance H.

B. *Validate Distributions of* FLuMe *Variables*

To validate that the proposed model is able to synthesize authentic spectrum usage, we compare the distributions of three *FLuMe* variables in both measurement and synthetic data in this subsection.

We select the spectrum tenancy during high wireless traffic ($U \geq 0.7$). Then, we generate two data sets of aggregate spectrum tenancy in high traffic in the two ways explained in the previous subsection, using aggregate *FLuMe* variables and the individual variables with sessions, respectively. We analyze



(a) Camping time. (b) Hopping distance. (c) inter-access time.

Fig. 19. Model validations and unification.

the distributions of the aggregate camping time C, hopping distance H, and inter-access time A, in all the three data sets, the measurement data and the two synthetic data sets. The distributions for the three variables are illustrated in Fig. 19. The probability mass functions of the aggregate variables for the measurement data are shown as red squares, and the two synthetic datasets are depicted as green circles and blue diamonds. We observe that the distributions of camping time, hopping distance, and inter-access time in all three datasets are very similar. A two-sample Kolmogorov-Smirnov test confirms that the measurement data and synthetic datasets have the same distribution with a significance level of 0.05.

This result confirms that our spectrum mobility model captures real-world spectrum mobility from both aggregate and individual perspectives. Since the aggregate and the individual perspectives of the *FLuMe* model cross validate each other, and the number of parameters in the models are quite limited compared to the size of the dataset, the risk of over fitting *FLuMe* to measurement data is minimal [44].

To sum up, the *FLuMe* model is able to synthesize spectrum tenancy with the same variable distributions as the high resolution measurement. The model is also versatile to be adapted to model the spectrum usage data presented in other studies, such as those in [12], [18], [19].

C. Spectrum Prediction

After validating the *FLuMe* framework, we showcase its application to spectrum prediction, which plays a crucial role in facilitating the simultaneous usage of multiple wireless networks on overlapping frequency channels [45]. Although there are many spectrum prediction algorithms, some of which are model-free, the impact of spectrum tenancy models can still be seen in all prediction algorithms [46]. This is because some algorithms rely on the distributions of on/off intervals, and most predictions require historical tenancy as input data [47].

Since the on/off model is a classical spectrum model and the busy and the idle interval patterns have been adopted in spectrum predictions, we compare the prediction capabilities of our proposed spectrum mobility with the on/off model. To achieve fair comparisons, the prediction algorithm is chosen as the Multi-Layer Perceptron (MLP) [48], and the prediction inputs are the on/off period lengths and the *FLuMe* aggregate variables. As history usage has been adopted as the prediction inputs for MLP based predictions, we compare the prediction

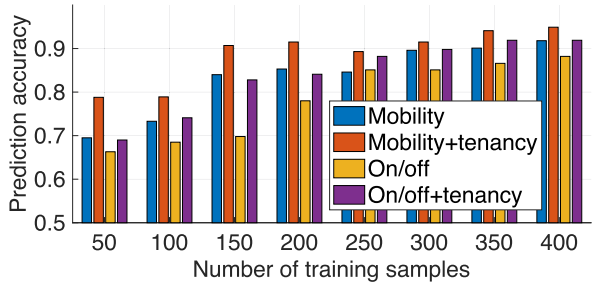


Fig. 20. Prediction accuracy comparisons.

accuracy in four cases with different prediction inputs, the aggregate parameters, the aggregate parameters plus history tenancy, on/off parameters, and on/off together with history tenancy.

For each case, we train the MLP network using different amounts of samples ranging from 50 to 400, and the performance is tested on 100 new data samples never used in the training stage. We adopt the prediction accuracy as the performance metric, and it is defined as the number of correctly predicted spectrum tenancy over the total amount of spectrum slices. For spectrum mobility based predictions, we supply the MLP with the camping times, the hopping distances, and the inter-access times in two previous strides defined in the aggregate variables of *FLuMe*. When the on/off time lengths or the history tenancy are employed as prediction features, we also consider two on/off cycles or history tenancy. All four cases with different prediction inputs are validated in ten randomly chosen spectrum tenancy segments, and the average prediction accuracy versus training sample sizes are plotted in Fig. 20.

The following findings can be observed from Fig. 20. First, MLP based spectrum tenancy prediction can be trained with a small amount of data, since the accuracy stabilizes after training sample sizes reach 400. If the prediction inputs are composed of both spectrum mobility parameters and the tenancy, the prediction accuracy achieves 90% with even smaller sizes of training sample such as 150 and 200. Thus, it is important to capture spectrum tenancy in LTE scheduling intervals of 1 ms, since measurement in coarse time resolutions in the seconds or hundreds of ms may miss the prediction inputs lasting only around 200 ms. Compared with on/off parameters, spectrum mobility parameters serve as more capable prediction features, because the prediction accuracy based on inputs of spectrum mobility is higher than that based on on/off time periods. This

is true when spectrum mobility and on/off parameters are used alone or together with history tenancy.

VI. Related Work

The spectrum tenancy model and its impacts on the spectrum prediction and the design of spectrum access schemes have been extensively investigated. However, these existing studies focus on the overall spectrum occupancy in a wide range of bands, lacking the characterization of the spectrum usage of individual mobile devices. In comparison, the proposed *FLuMe* framework provides a comprehensive modeling of spectrum usage from both the aggregate and the individual perspectives.

An ARMA model is proposed for spectrum consumption based on extensive measurement campaigns [45]. Although the model characterizes changes in energy levels in a wide spectrum band, it fails to capture the occupancy of individual users. A study in [49] models the spectrum occupancy along the time, frequency, and space domains, but does not incorporate the spectrum tenancy of single users or the aggregate usage of all users in a cell. The study in [18] observes both deterministic and stochastic patterns in the spectrum mobility of WiFi and LTE bands, however, it only focuses on the spectrum usage of all users in a single cell and not that of individual users.

In addition, our models are based on raw spectrum occupancy measurement data collected with the same time and frequency granularity as the spectrum resource scheduling of underlying mobile networks. However, other measurement-based models have a much coarser granularity that cannot reflect the actual spectrum usage. A study in [50] focused on a narrow bandwidth of 300 kHz, but had a time resolution of 5 to 30 seconds, much larger than the 1 ms scheduling interval in LTE systems. A cloud-based measurement system is proposed in [51], but it also has a time granularity of around 10 seconds. The survey [24] provides a summary of previous studies on spectrum usage, including measurement granularities. According to the table, time granularity is typically around tens of seconds, with a frequency resolution of around 200 kHz, larger than the scheduling unit of LTE spectrum resource allocation."

VII. Conclusion

Studies on spectrum usage are pivotal in wireless network design; however, many existing models lack the ability to accurately characterize dynamic changes in spectrum usage at fine time-frequency granularity for both aggregate and individual spectrum tenancy. This paper introduces the novel *FLuMe* framework based on measurement data that includes information such as user IDs, initial frequency channel, camping time, and inter-access time for both single accesses and multiple accesses in sessions. Our findings reveal a range of interesting observations, including the distributions of spectrum usage under different traffic loads and the similarities between aggregate and individual spectrum usage. We also discuss potential applications of FLuMe, including the generation of synthetic datasets for use in spectrum prediction, highlighting the significance and versatility of our data-driven modeling approach.

References

[1] S. Aerts et al., "In-situ measurement methodology for the assessment of 5G NR massive MIMO base station exposure at sub-6 ghz frequencies," *IEEE Access*, vol. 7, pp. 184 658–184 667, 2019.

[2] R. Zou and W. Wang, "Downlink decoding based accurate measurement of LTE spectrum tenancy," *IEEE Trans. Mobile Comput.*, vol. 22, no. 5, pp. 2613–2627, May 2023.

[3] 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 *Proc. 1st Int. Workshop Technol. Policy Accessing Spectr.*, 2006, Art. no. 1.

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

[5] 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 Trans. Mobile Comput.*, vol. 17, no. 9, pp. 2141–2154, Sep. 2018.

[6] W. Zhang, C.-X. Wang, X. Ge, and Y. Chen, "Enhanced 5G cognitive radio networks based on spectrum sharing and spectrum aggregation," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6304–6316, Dec. 2018.

[7] P. Gkotsiopoulos, D. Zorbas, and C. Douligeris, "Performance determinants in loRa networks: A literature review," *IEEE Commun. Surv. Tut.*, vol. 23, no. 3, pp. 1721–1758, Third Quarter 2021.

[8] L. Wan, Z. Guo, and X. Chen, "Enabling efficient 5G NR and 4G LTE coexistence," *IEEE Wirel. Commun.*, vol. 26, no. 1, pp. 6–8, Feb. 2019.

[9] L. C. Alexandre, A. L. De Souza Filho, and A. C. Sodré, "Indoor coexistence analysis among 5G new radio, LTE-A and NB-IoT in the 700 MHz band," *IEEE Access*, vol. 8, pp. 135 000–135 010, 2020.

[10] R. Zou and W. Wang, "Effi-Ace: Efficient and accurate prediction for High-Resolution spectrum tenancy," in *Proc. IEEE Conf. Comput. Commun.*, Vancouver, Canada, 2024, pp. 2199–2208.

[11] C.-H. Liu, W. Gabran, and D. Cabric, "Prediction of exponentially distributed primary user traffic for dynamic spectrum access," in *Proc. IEEE Glob. Commun. Conf.*, 2012, pp. 1441–1446.

[12] T. A. Hall, A. Sahoo, C. Hagwood, and S. Streett, "Exploiting lte white space using dynamic spectrum access algorithms based on survival analysis," in *Proc. IEEE Int. Conf. Commun.*, 2017, pp. 1–7.

[13] Y. Wu, F. Hu, Y. Zhu, and S. Kumar, "Optimal spectrum handoff control for CRN based on hybrid priority queuing and multi-teacher apprentice learning," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2630–2642, Mar. 2017.

[14] A. Koushik, F. Hu, and S. Kumar, "Intelligent spectrum management based on transfer actor-critic learning for rateless transmissions in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 5, pp. 1204–1215, May 2018.

[15] V. Raj, I. Dias, T. Tholeti, and S. Kalyani, "Spectrum access in cognitive radio using a two-stage reinforcement learning approach," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 20–34, Feb. 2018.

[16] Y. Wu et al., "A learning-based QoE-driven spectrum handoff scheme for multimedia transmissions over cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 2134–2148, Nov. 2014.

[17] K. Preetha and S. Kalaivani, "Analysis of spectrum handoff schemes for cognitive radio networks considering secondary user mobility," *Int. J. Grid Utility Comput.*, vol. 11, no. 4, pp. 443–456, 2020.

[18] K. Umebayashi, M. Kobayashi, and M. López-Benítez, "Efficient time domain deterministic-stochastic model of spectrum usage," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1518–1527, Mar. 2018.

[19] S. Geirhofer, L. Tong, and B. M. Sadler, "A measurement-based model for dynamic spectrum access in wlan channels," in *Proc. IEEE Mil. Commun. Conf.*, 2006, pp. 1–7.

[20] K. V. Vishwanath and A. Vahdat, "Realistic and responsive network traffic generation," in *Proc. Conf. Appl. Technol. Archit. Protoc. Comput. Commun.*, 2006, pp. 111–122.

[21] K. V. Vishwanath and A. Vahdat, "Swing: Realistic and responsive network traffic generation," *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 712–725, Jun. 2009.

[22] H. Mosavat-Jahromi, Y. Li, L. Cai, and J. Pan, "Prediction and modeling of spectrum occupancy for dynamic spectrum access systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 3, pp. 715–728, Sep. 2021.

[23] Y. Chen and H.-S. Oh, "A survey of measurement-based spectrum occupancy modeling for cognitive radios," *IEEE Commun. Surv. Tut.*, vol. 18, no. 1, pp. 848–859, First Quarter 2016.

[24] M. Höyhtyä et al., "Spectrum occupancy measurements: A survey and use of interference maps," *IEEE Commun. Surv. Tut.*, vol. 18, no. 4, pp. 2386–2414, Fourth Quarter 2016.

[25] E. Research, "x310-kit," Feb. 2023. [Online]. Available: https://www.ettus.com/all-products/x310-kit/
[26] E. Research, "Sbx120," Feb. 2023. [Online]. Available: https://www.ettus.com/all-products/SBX120/
[27] E. Research, "UHD," Feb. 2023. [Online]. Available: https://kb.ettus.com/UHD
[28] S. Kumar, E. Hamed, D. Katabi, and L. Erran Li, "LTE radio analytics made easy and accessible," in *Proc. Conf. ACM SIGCOMM Comput. Commun. Rev.*, 2014, pp. 211–222.
[29] N. Bui and J. Widmer, "OWL: A reliable online watcher for LTE control channel measurements," in *Proc. 5th ACM All Things Cellular MobiCom Workshop*, 2016, pp. 25–30.
[30] 3rd Generation Partnership Project (3GPP), "Evolved universal terrestrial radio access (E-UTRA); Medium access control (MAC) protocol specification," 3GPP, Technical Specification (TS) 36.321, Apr. 2017.
[31] S. X. Zhou, "Investigation of lte privacy attacks by exploiting the paging mechanism," Master's thesis, NTNU, 2018.
[32] J. Zhang et al., "Power consumption analysis of video streaming in 4G LTE networks," *Wirel. Netw.*, vol. 24, no. 8, pp. 3083–3098, 2018.
[33] R. P. Jover, "LTE security, protocol exploits and location tracking experimentation with low-cost software radio," 2016, *arXiv:1607.05171*.
[34] X. Lin, L. Quan, and H. Wu, "An automatic scheme to categorize user sessions in modern HTTP traffic," in *Proc. IEEE Glob. Telecommun. Conf.*, 2008, pp. 1–6.
[35] S. Yin, D. Chen, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: A large-scale spectrum measurement study," *IEEE Trans. Mobile Comput.*, vol. 11, no. 6, pp. 1033–1046, Jun. 2012.
[36] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the levy-walk nature of human mobility," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 630–643, Jun. 2011.
[37] D. J. Morin et al., "Is your ad hoc model selection strategy affecting your multimodel inference?," *Ecosphere*, vol. 11, no. 1, 2020, Art. no. e02997.
[38] Y. Xiao et al., "Decentralized spectrum access system: Vision, challenges, and a blockchain solution," *IEEE Wirel. Commun.*, vol. 29, no. 1, pp. 220–228, Feb. 2022.
[39] M. Lopez-Benitez and F. Casadevall, "Empirical time-dimension model of spectrum use based on a discrete-time Markov chain with deterministic and stochastic duty cycle models," *IEEE Trans. Veh. Technol*, vol. 60, no. 6, pp. 2519–2533, Jul. 2011.
[40] B. J. Cavalcanti, G. A. Cavalcante, L. M. D. Mendonça, G. M. Cantanhede, M. M. de Oliveira, and A. G. D'Assunção, "A hybrid path loss prediction model based on artificial neural networks using empirical models for LTE and LTE-A at 800 MHz and 2600 MHz," *J. Microw., Optoelectron. Electromagn. Appl.*, vol. 16, pp. 708–722, 2017.
[41] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for distributed dynamic spectrum access," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 310–323, Jan. 2019.
[42] C. Liu, Z. Cao, G. Xiong, G. Gou, S.-M. Yiu, and L. He, "MaMPF: Encrypted traffic classification based on multi-attribute markov probability fingerprints," in *Proc. IEEE/ACM 26th Int. Symp. Qual. Service*, 2018, pp. 1–10.
[43] I. Tsompanidis, A. H. Zahran, and C. J. Sreenan, "Mobile network traffic: A user behaviour model," in *Proc. 7th IFIP Wirel. Mobile Netw. Conf.*, 2014, pp. 1–8.
[44] X. Ying, "An overview of overfitting and its solutions," *J. Phys., Conf. Ser.*, vol. 1168, 2019, Art. no. 022022.
[45] H. Mosavat-Jahromi, Y. Li, L. Cai, and J. Pan, "Prediction and modeling of spectrum occupancy for dynamic spectrum access systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 3, pp. 715–728, Sep. 2021.
[46] G. Ding et al., "Spectrum inference in cognitive radio networks: Algorithms and applications," *IEEE Commun. Surv. Tut.*, vol. 20, no. 1, pp. 150–182, First Quarter 2018.
[47] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 2, pp. 257–265, Jun. 2018.
[48] V. K. Tumuluru, P. Wang, and D. Niyato, "A neural network based spectrum prediction scheme for cognitive radio," in *Proc. IEEE Int. Conf. Commun.*, 2010, pp. 1–5.
[49] M. Lopez-Benitez and F. Casadevall, "Spectrum usage in cognitive radio networks: From field measurements to empirical models," *IEICE Trans. Commun.*, vol. 97, no. 2, pp. 242–250, 2014.
[50] S. Iliya, E. Goodyer, J. Gow, and M. A. Gongora, "Spectrum occupancy survey in Leicester, UK, for cognitive radio application," *Int. J. Sci. Eng. Res.*, 2015.
[51] L. Li et al., "A cloud-based spectrum environment awareness system," in *Proc. IEEE 28th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun.*, 2017, pp. 1–6.

Rui Zou (Member, IEEE) received the BS degree in electrical engineering from the joint program provided by the Beijing University of Posts and Telecommunications (BUPT) and Queen Mary University of London in 2012 , the MS degree in electrical engineering from BUPT in 2015, and the PhD degree in computer engineering from NC State in 2023. He is a professor of Computer Information System in Central Arizona College. His current research interests range from measurement, analysis and modeling of spectrum tenancy to resource scheduling and security issues in dynamic spectrum access systems.

Wenye Wang (Fellow, IEEE) received the MSEE and PhD degrees in computer engineering from the Georgia Institute of Technology, Atlanta, GA, 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. She 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.