Primary Users in Cellular Networks: A Large-scale Measurement Study

Daniel Willkomm
Telecommunication Networks Group
Technische Universität Berlin,
Germany
willkomm@tkn.tu-berlin.de

Sridhar Machiraju
Sprint
California, USA
machiraju@sprint.com

Jean Bolot
Sprint
California, USA
bolot@sprint.com

Adam Wolisz
Technische Universität Berlin,
Germany
University of California, Berkeley, USA
wolisz@ieee.org

Abstract—Most existing studies of spectrum usage have been performed by actively sensing the energy levels in specific RF bands including cellular bands. In this paper, we provide a unique, complementary analysis of cellular primary usage by analyzing a dataset collected *inside* a cellular network operator. One of the key aspects of our dataset is its scale – it consists of data collected over three weeks at hundreds of base stations. We dissect this data along different dimensions to characterize and model primary usage as well as understand its temporal and spatial variations. Our analysis reveals several results that are relevant if *D*ynamic *S*pectrum *A*ccess (DSA) approaches are to be deployed for cellular frequency bands. For instance, we find that call durations show significant deviations from the often-used exponential distribution, which makes call-based modeling more complicated. We also show that a random walk process, which does not use call durations, can often be used for modeling the aggregate cell capacity. Furthermore, we highlight some applications of our results to improve secondary usage of licensed spectrum.

I. INTRODUCTION

The prevailing approach to wireless spectrum allocation is based on statically allocating long-term licenses on portions of the spectrum to providers and their users. It is widely believed that approaches based on DSA can lead to more efficient use of wireless spectrum than static approaches. A multitude of DSA-based approaches have been proposed for secondary spectrum usage in which *S*econdary *U*sers (SUs) use parts of the spectrum that are not being used by the licensed *P*rimary *U*sers (PUs). PUs can enable such secondary usage, for instance, by using short-term auctions of underutilized spectrum [1]. Alternatively, SUs can sense and autonomously use parts of the spectrum that are currently not being used by (licensed) PUs. A key technical component of such approaches are *C*ognitive *R*adios (CRs), which enable spectrum sensing. Apart from detecting idle spectrum, the sensing done by CRs is also needed by SUs to vacate the spectrum again when PUs resume their usage.

The feasibility and success of secondary spectrum usage depends crucially on if, when, and how idle spectrum becomes available, as well as how easily it can be sensed. For instance, SUs operating in the spectrum licensed to PUs with predictable and well-defined usage, e.g., TV broadcasting, can easily identify spectrum that will be idle over long periods of time. In contrast, cellular voice spectrum usage tends to exhibit much more variations both in time *and* space. Hence, SUs of cellular voice spectrum likely need to employ more agile DSA techniques than SUs of TV broadcasting spectrum. Characterization of primary usage in specific spectrum ranges is therefore crucial to the deployment of DSA in those spectrum ranges.

In this paper, we present a large-scale measurement-driven characterization of primary usage in cellular networks. Our focus on cellular spectrum is important for several reasons. Apart from TV bands, cellular frequencies represent one of the most viable ways of achieving DSA – both because they are widely used throughout the world and also because engineering devices and data applications for these frequencies are well understood. On the other hand, compared to TV bands, cellular spectrum usage is expected to be much more dynamic. Thus, data-driven studies are necessary to design DSA systems that optimize such usage. Additionally, looking from a completely different perspective, the low bit rates involved in wireless voice transmission make it an attractive application for secondary usage. Hence, understanding the nature of this application can also help drive secondary usage markets.

A few prior studies [2], [3], [4], [5] have characterized primary usage in cellular (and non-cellular) bands. All of these studies – but [5] – are based on active sensing of energy. In contrast, we provide a unique analysis of cellular primary usage that is based on call records collected *inside* a cellular network. Thus, we are able to provide insights on a call level that other studies are unable to. Another specific aspect is the scale of our study – we are able to study the usage at hundreds of base stations simultaneously. In contrast, sensing-based studies are usually based on only a few spectrum analyzers and, typically have limited spatial resolution. Moreover, we are able to study the entire spectrum band used by a cellular operator. Sensing-based studies take time to "sweep" such a band and, hence, have to tradeoff the sampling frequency with the width of a band. The temporal diversity of our data is also large - we use measurements of tens of millions of calls over a period of three weeks. Finally, by looking at call records, we are able to measure the "ground truth" as seen by the network, and, hence, are able to model call arrival processes as well as system capacity.

The aforementioned advantages of our analysis make the results useful in developing specific guidelines for DSA-based approaches. For example, we are able to (in)validate assumptions that spectrum pricing studies have used in the past. We are also able to capture average-case and worst-case behavior thereby providing insights into when secondary users should sense and capture/vacate spectrum.

This paper is organized as follows. In Section II, we introduce the goals of our work, describe our measurement methodology, the limitations of our data and the inaccuracies that these can cause. In Section III, we analyze our dataset to gain insights into what models may apply. In this section, we ignore information on cells and consider all calls as arriving to a single entity representing the combined coverage area of all of our base stations. In Section IV, we perform per-cell analyses of primary usage and investigate the temporal variations of such usage. In Section V, we extend our analysis to gain insights into the spatial dependence of spectrum usage. In Section VI, we discuss the implications of our results on specific DSA design decisions such as sensing. We survey prior work related to spectrum sensing and primary usage characterization in Section VII. In Section VIII, we present our conclusions and discuss future directions.

II. METHODOLOGY

In this section, we present our measurement methodology. We start by describing our dataset and discuss its pros and cons. Then, we discuss how we analyzed this dataset and the kinds of questions we tackle in the rest of this paper.

A. Dataset

The dataset we use in this paper was collected from hundreds of cell sectors[1] of a US CDMA-based cellular operator. The data captures voice call information at those sectors, which were all located in densely-populated urban areas of Northern California, over a period of three weeks. In particular, our dataset captures the start time, the duration, the initial and final sector of each call. Note that the call duration reflects the RF emission time of the data transmission for the call, i.e., the duration of time for which a data channel was assigned. This is precisely what is relevant for DSA questions. The start time of the call is measured with a resolution of several milliseconds. The duration is measured with a resolution of 1 millisecond. Overall, our data consists of tens of millions of calls and billions of minutes of talk time. To our knowledge, such a large-scale *network* viewpoint of spectrum usage has not been analyzed before in prior work.

B. Limitations

As with any measurement-based study, our dataset has certain limitations. We state these up-front since it is important to understand what our results state and what they do not.

The first limitation of our dataset is its lack of full information on mobility. We are able to record only the initial and final sector of each call. Thus, we are unable to account for
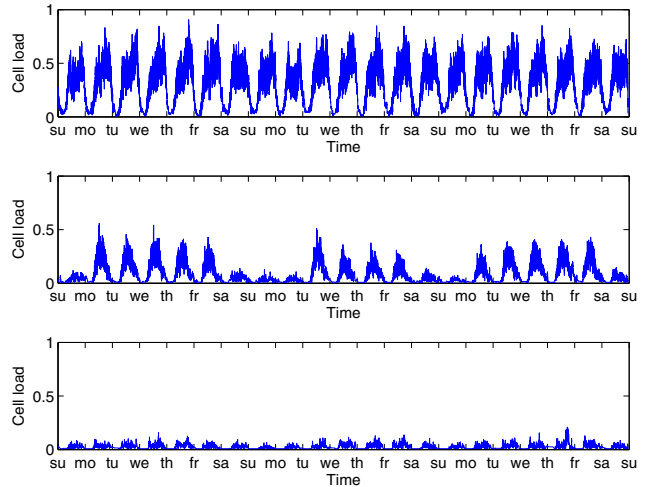
---

[1]We do not give the specific number for proprietary reasons.

Fig. 1. Normalized load of three cell sectors over 3 weeks. We plot the moving average over one second. The cells show high load **(Top)**, varying load **(Middle)**, and low load**(Bottom)**.

the spectrum usage in the other sectors that the user may have visited during the call. To address the resulting incompleteness of information, we use two types of approximations. In the first approximation, we assign the entire call as having taken place in the initial sector. We use this approximation by default. In the second approximation, we assign the first (last) half of the call to the initial (final) sector. We refer to this as the *mobile* approximation. Throughout the paper, we provide results using both approximations and find that our conclusions do not change. These results indicate that the results are not sensitive to our approximations and would likely not change with full mobility information.

The second limitation relates to the cellular system we collected our dataset from – a CDMA-based network. Without additional knowledge from the base stations, the precise CDMA system capacity cannot be easily calculated. Hence we implicitly assume that each voice call uses the same portion of a cell capacity. This assumption, which is correct for TDMA based systems like GSM, is obviously not precise for CDMA. Due to the critically important power control loop, individual CDMA calls may require different portions of the cell capacity, which cannot be easily expressed just in the number of calls. Nevertheless, since user calling behavior is unlikely to depend on the underlying technology except under rare overload conditions, many aspects of our study are likely to apply to other cellular voice networks.

C. Preliminary Observations

Using either of the aforementioned approximations, we can compute the total number of ongoing calls in each cell sector during the entire time period of our study. To do so, we split the call records based on the sector. We create two records for each call – corresponding to the begin and end of each call. Then, we sort these records in order of their time. We maintain a running count that is increased by +1 when a call
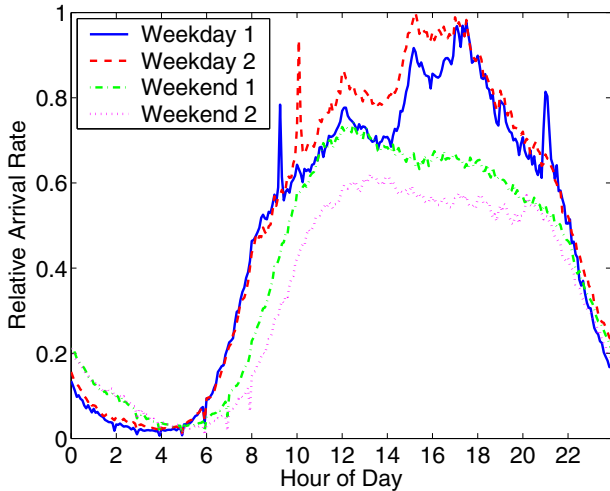
Fig. 2. Distribution of system-wide average call arrival rates during four different days. The arrival rates are averaged over 5-minute slots.
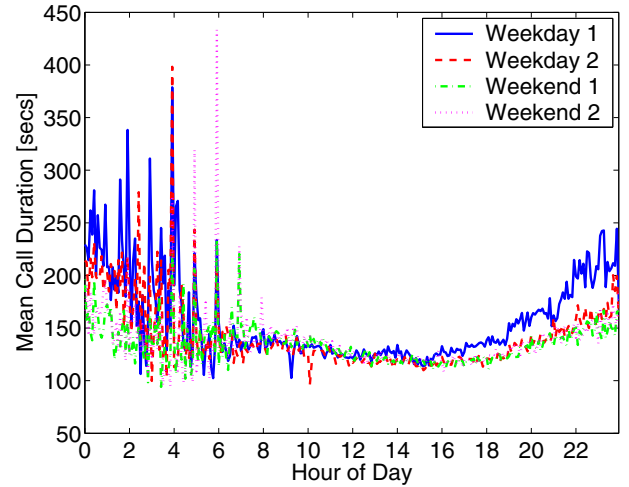

Fig. 3. Distribution of average call duration over 5-minute periods during four different days. The large spikes during the mornings are due to small gaps in collection.

begins and decreased by -1 when a call terminates. Our data consists of less than 0.1% of the calls whose initial or final cell sectors were not captured. We ignore all such calls since they represent a relatively insignificant fraction of the calls.

Our primary goal is to use the computed information on load and call characteristics to characterize spectrum usage. Broadly speaking, we seek to answer questions that are relevant for one of three DSA-related categories:

- **Primary User Behavior:** How can primary user behavior be modeled? How does it vary within and across cell sectors[2]? How does it vary across space and time?
- **Secondary Usage and Pricing:** When and how much secondary usage is feasible? How conservative should such usage be so that primary users are not affected?
- **Cognitive Radio Sensing:** How often should sensing be done? What can be assumed? How does it depend on space and time?

We plot the obtained "load" of three representative cells in Figure 1. For proprietary reasons, we normalize the values of load by a constant value such that only the relative change is seen. The top cell has low load only at night whereas the middle cell has low load during the weekends too (note that the second Monday in the observation period was a public holiday). The bottom cell always has low load, i.e., during both day and night. Our plots in Figure 1 show that spectrum usage varies widely over time *and* space – an illustration of the challenges that are likely to be faced with cellular DSA.

III. SYSTEM-WIDE MODELS

In this section, we ignore the information related to the individual cells and consider all calls as arriving to a single entity consisting of all the monitored base stations. Under this system-wide formulation, the capacity refers to the space-aggregated capacity of all cells.

[2]We use cells and cell sectors interchangeably.

A. Call-based Model

Often, in prior work [1], [6], primary usage has been modeled as a Poisson process (exponential inter-arrival times) and exponential call durations. This motivates us to explore a *call-based model*, using two random variables, T and D, to understand the dynamics of spectrum usage. T is the random variable corresponding to the inter-arrival time between two calls and D corresponds to the duration of calls. Under this model, the system can be described using only T and D.

We now explore how well the above model describes the spectrum usage of the entire system of cells. A key question we are faced with is the timescales over which the call arrival process and the call duration distribution can be viewed to be *stationary*, i.e., the time-span during which the parameters of the model remain constant. As with any modeling of empirical data, we want to aggregate data into time periods smaller than timescales of stationarity. At the same time, we want to choose timescales that are as large as possible so that we have minimal estimation variance. To help us choose the aggregation timescales for our analysis, we plot the normalized average call arrival rates and average call durations calculated over 5-minute slots in Figure 2 and Figure 3. We plot these for four different days.

Figure 2 illustrates three key effects regarding the dynamics observed in the system. First, there are two distinct periods, which roughly correspond to day and night and have high and low arrival rates respectively. Moreover, the steepest change in arrival rates occurs in the morning and late in the evening, which correspond to the transition between the day and night periods. Second, the system characteristics are unlikely to remain stationary at timescales beyond an hour. Except for the transition hours, the mean arrival rates do not vary significantly during an hour. Third, weekdays and weekends appear to show distinct trends. This is not wholly unexpected since many cellphone pricing plans provide unlimited calling in the weekend.
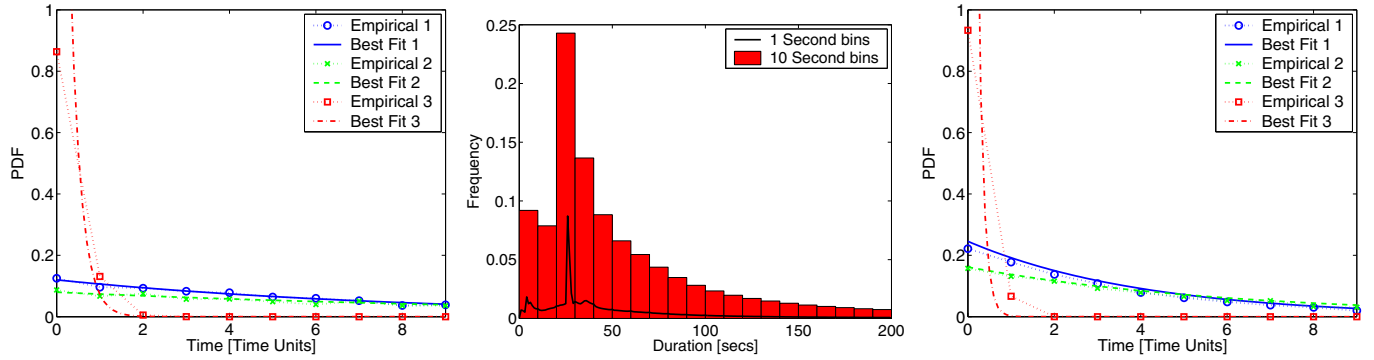
Fig. 4. **(Left)**: Distribution and exponential fits to system-wide call interarrival times during three representative hours. **(Middle)**: Histogram of call durations. We plot the histogram using different bin sizes. **(Right)**: Distribution and exponential fits to system-wide inter-event times during three representative hours.

Figure 3 illustrates similar trends as Figure 2. However, we find that the range of variability in mean call duration is much smaller than that of arrival rates. Note that there are a few large spikes in Figure 3. During these 5-minute periods, there were dips in arrival rates, too as can be seen in Figure 2. We believe that these are caused by a brief interruption in the data collection at certain sectors. This caused some short calls to not be recorded thereby artificially inflating the mean duration of calls. Figure 2 and 3 confirm that an hour is a good timescale of stationarity (except during the day-night transition periods). Hence, by default, we choose one hour as our timescale for aggregation since it is neither too large nor too small to yield little data for individual cells.

In Figure 4 **(Left)**, we plot the actual distribution (*P*robability *D*ensity *F*unction (PDF)) of inter-call arrival times, for three representative hours. We use the standard approximation, which is true for all $b > 0$:

$$
\begin{aligned}
P(D = d) &\approx \frac{1}{b} \int_{d-\frac{b}{2}}^{d+\frac{b}{2}} P(D = x)dx \\
&= \frac{1}{b} \left[CDF(d + \frac{b}{2}) - CDF(d - \frac{b}{2}) \right] \quad , (1)
\end{aligned}
$$

where CDF is the *C*umulative *D*istribution *F*unction. Using the above approximation, we estimate the PDF of D using the empirically observed histogram:

$$
\hat{P}(D = d) = \frac{\#(D \le d + 0.5) - \#(D \le d - 0.5)}{\#(D)} \quad (2)
$$

Note that the right-hand side calculates the fraction of inter-call arrival times that were centered around d. In Figure 4 **(Left)**, we plot the inter-arrival times of calls as a function of the minimum clock resolution of our measurement infrastructure. This was the order of several milliseconds. Here and in the rest of the paper, for proprietary reasons, we do not provide the exact values of inter-arrival times and arrival rates. We also plot the *M*aximum *L*ikelihood *E*stimate (MLE) exponential fits for these curves in Figure 4 **(Left)**. The best exponential fit is obtained easily as the distribution with parameter $\hat{\mu}^{-1}$ where $\hat{\mu}$ is the sample mean of the empirical data [7].

For two of the three empirical distributions in Figure 4 **(Left)**, the inter-arrival times of calls are visually well modeled as exponential distributions. The curve labeled "Empirical 3" does not appear to be well-modeled by the exponential best-fit. We believe that this is only an artifact of our measurement setup. In particular, our measurement resolution is so large that most of the distribution is centered around the 0 and 1 values. This causes the approximation in Equation (1) to be violated and leads to the seeming discrepancy between the empirical and best-fit curve. We also find the auto-correlation at lag 1 of the sequence of inter-arrival times to be significant. For example, the sequences during the three hours of Figure 4 **(Left)** had values of 0.32, 0.29 and 0.14 respectively. This would imply that the sequence of inter-arrival times is dependent. We believe that this is not the case and that the high auto-correlation is also due to the discretization caused by coarse-grained time resolution.

But for such measurement resolution, we claim that the sequence of system-wide call inter-arrival times are well modeled by an i.i.d. exponential distribution and, hence, form a Poisson process. Our claim is further supported in Section IV when we verify the Poisson nature of call arrivals of individual cell sectors. Since the combination of two Poisson processes of intensity λ_1 and λ_2 is a Poisson process of intensity $\lambda_1 + \lambda_2$, Poisson arrival processes for individual cells imply a system-wide Poisson arrival process.

In Figure 4 **(Middle)**, we plot the histogram of all call durations. The histogram is quite unlike that of an exponential distribution. In fact, the histogram is not even monotonic. We see about 10% of calls having a duration of around 27 seconds. We verified that these correspond to calls during which the called mobile users did not answer and the calls were redirected to voice-mail. However, RF voice channels were allocated during these calls. This illustrates that call durations can be significantly skewed towards smaller durations due to non-technical failures, e.g., failure to answer. Also, note that the variance of the call durations is more than three times the mean, which is significantly higher than that of exponential distributions.
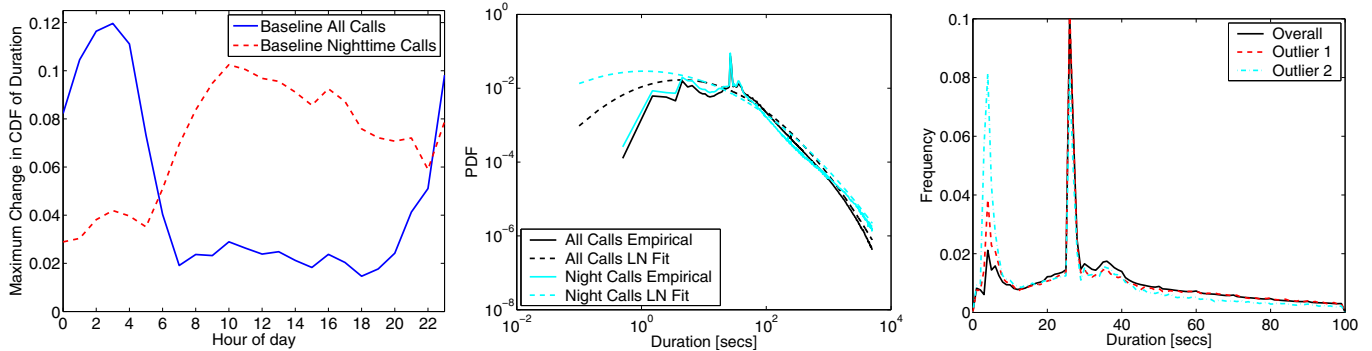
Fig. 5. **(Left)**: Variation of the hourly call duration CDF compared to the overall CDF and the CDF of only night-time calls. **(Middle)**: Duration distributions and lognormal fits. **(Right)**: Illustration of anomalous distributions during 2 hours.

B. Event-based Random Walk Model

The call-based model is complicated by the non-exponential nature of the call durations. We now explore a *random walk* model that ignores details about individual calls and instead models only the load $X(\cdot)$, i.e., total number of ongoing calls. We refer to this model as the *event based* model. Under this model, the load is considered to be a one-dimensional, continuous-time random walk where steps are either $+1$ or -1 corresponding to the initiation and termination events of a call:

$$X(t + E) = X(t) + (-1)^{\Phi}. \qquad (3)$$

Here E is a random variable representing the time between consecutive steps/events and Φ is a Bernoulli random variable, which takes the value $+1$ with probability p and 0 with probability $1 - p = q$. Since there is a $+1$ for every -1, p should be $\frac{1}{2}$. If valid, such a model has advantages over the call-based model because it abstracts out the non-exponential distribution of call durations. We plot the distribution of inter-event times in Figure 4 **(Right)**. At the system level, we find that these are visually well modeled by exponential distributions though the limited resolution continues to cause problems with PDF estimation and leads to significant auto-correlation at non-zero lags.

To check for consistency with the random walk model, we also need Φ to be a Bernoulli random variable. A necessary (but not sufficient) condition for this to be true is that the auto-correlation of the sequence of step sizes at non-zero lags be close to zero. We do find that the absolute value of auto-correlation is always smaller than 0.041 for the first lag and smaller than 0.043 for all lags.

C. Variations

We now investigate the distribution of call durations in more detail. We start by investigating how the *Cumulative Distribution Function* (CDF) of durations, calculated hour-wise (across all days) varies as compared to the overall distribution. Specifically, we compute the *Kolmogorov-Smirnov statistic* [8], which is the maximum difference between the hourly CDF and overall CDF. We plot this statistic as a function of the hour of day in Figure 5 **(Left)**. We find

that there is little variation during the day. However, there is significant variation during the night-time especially in the hours from 11PM to 5AM. Hence, we generate the CDF of durations during these hours and plot the variation around this CDF. The resulting plot appears to be roughly complementary to the variation around the overall CDF. We conclude that there are likely two different distributions of call durations – one during the day and the other during the night. Furthermore, the transition hours between day and night likely see a mixture of both these distributions.

In Figure 5 **(Middle)**, we compare the overall and night-time distributions of call durations. Note that we use the log-log scale. We find that the night-time distribution has more short calls as well as a heavier tail compared to the overall distribution. Both distributions have a "semi-heavy" tail and are not well-modeled by classical short-tailed distributions such as Erlang (results not shown). However, the shape of the above distributions is reminiscent of the lognormal distribution, which is parabolic in log-log scale. Recall that D is lognormally distributed with parameters μ and σ^2 if $log(D)$ is normally distributed with the same parameters. The best lognormal fit can be obtained using MLE, in which case the parameters are simply estimated as the sample mean ($\hat{\mu}$) and variance ($\hat{\sigma}^2$) [9]. In Figure 5 **(Middle)**, we also plot the best lognormal fits for the distributions of call durations. The head of the empirical distribution shows significant deviation from the best lognormal fit. Although the tails of the empirical and best fit agree better, they too diverge at large values.

Not only is the distribution of call durations hard to model, there can also be significant deviations during certain hours. We plot two such "outlier hours" in Figure 5 **(Right)**. The two outlier plots correspond to the weekday hours plotted in Figure 2 – the spikes in the arrival rate correspond to the spikes of Figure 5 **(Right)**. Both hours see a sudden spurt in short calls. We verified that at least one of these is caused by a large number of calls to a popular television show, whose telephone lines are often busy. Figure 5 **(Right)** thus demonstrates that social behavior and external events, which may not be easily predicted, can and do have significant short-term impact on spectrum usage.
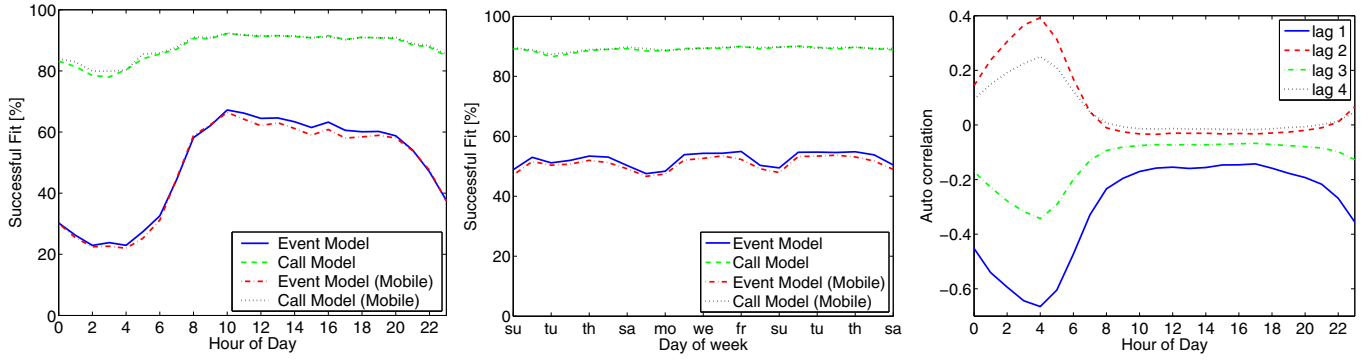
Fig. 7. **(Left)**: The percentage of successful fits (across all cells) averaged on a per-day basis. **(Middle)**: The percentage of successful fits (across all cells) averaged on a per-hour basis. **(Right)**: The per-hour auto-correlation of the step sizes (Φ) in our event-based models averaged across all cells.
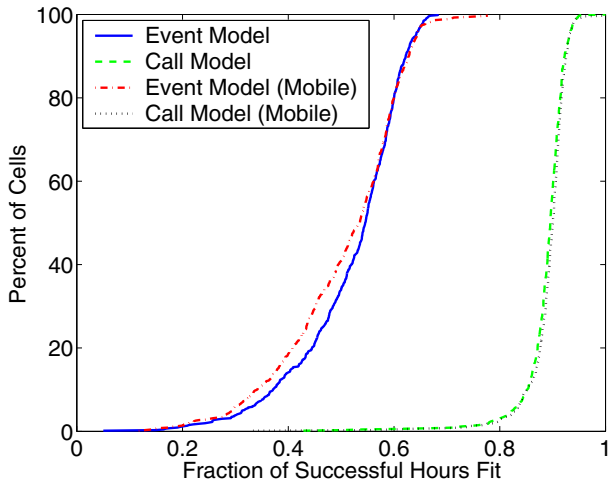

Fig. 6. The CDF of the per-cell fraction of hours that were successfully described using our models. The call model is mostly applicable whereas the random walk model is successful only in half the hours for all cells on average.

IV. TEMPORAL CHARACTERISTICS

In the previous section, we developed two models of primary usage – the call-based model and the event-based (random walk) model. We explored the efficacy of these models in describing the primary usage when all calls are considered to arrive at a single entity. In this section, we analyze individual cells, to understand if the same models can be used and how these results vary with time.

Note that the per-cell call durations do not vary significantly from the overall call durations. We, thus, focus on presenting per cell results for the inter-arrival times and the cell load in this section.

A. Modeling

We start by trying to fit our models to the primary usage in each cell sector on an hourly basis. Note that we assign a call to the hour in which it started. Since durations rarely span an hour, this has little impact on our results. Since there are hundreds of sectors and hundreds of hours, we resort to an automatic goodness-of-fit test to determine if our

models apply. Specifically, for the call-based model, we use the *Anderson-Darling* test [10] with 95% confidence level on the empirical series of call inter-arrival times. Given a sample set $\{x_i\}$ of size n sorted in ascending order, this test rejects the hypothesis of an exponential distribution if a test statistic exceeds 1.32 (for 95% confidence level). The test statistic, A^2, is computed as:

$$A^2 = -n - \sum_{k=1}^{n} \frac{2k-1}{n} \left[log(F(x_k)) + log(1 - F(x_{n+1-k})) \right].$$

Here, $F(\cdot)$ is the CDF of the exponential distribution (with mean equal to the sample mean). To accommodate cases where n is small, we use the standard adjustment for A^2 by multiplying it with $1 + \frac{0.6}{n}$. We also use the Anderson-Darling test to determine whether the random walk model of exponential inter-event times is valid or not.

For each cell sector, we obtain the fraction of hours during which inter-arrival times of calls and events pass the Anderson-Darling test. In Figure 6, we show how this fraction is distributed across all cells for both of our models. Recall from Section II that we have two ways of approximating user mobility. We use both approximations. Figure 6 shows that the call arrival process is well described by an exponential process in more than 90% of the hours for most cells. Note that, since we use 95% confidence level for the Anderson-Darling test, we expect only 95% of our tests to succeed. Thus, the call-based model is almost always applicable. In Figure 6, we also verify that our results are not sensitive to our mobile approximation.

We also calculate the auto-correlation coefficient for each per-cell per-hour sequence of call inter-arrival times. We find that only 20% of these sequences have auto-correlation coefficients (at non-zero lags) higher than 0.16. Though not conclusive, such low auto-correlation is consistent with independence. Hence, we believe that call inter-arrivals are well-modeled as an exponentially distributed i.i.d. sequence. In other words, call arrivals can be viewed as Poisson processes.

Recall that the above results also support our claim of *system wide* Poisson call arrivals since the addition of two Poisson processes of intensity λ_1 and λ_2 is a Poisson process of intensity $\lambda_1 + \lambda_2$.
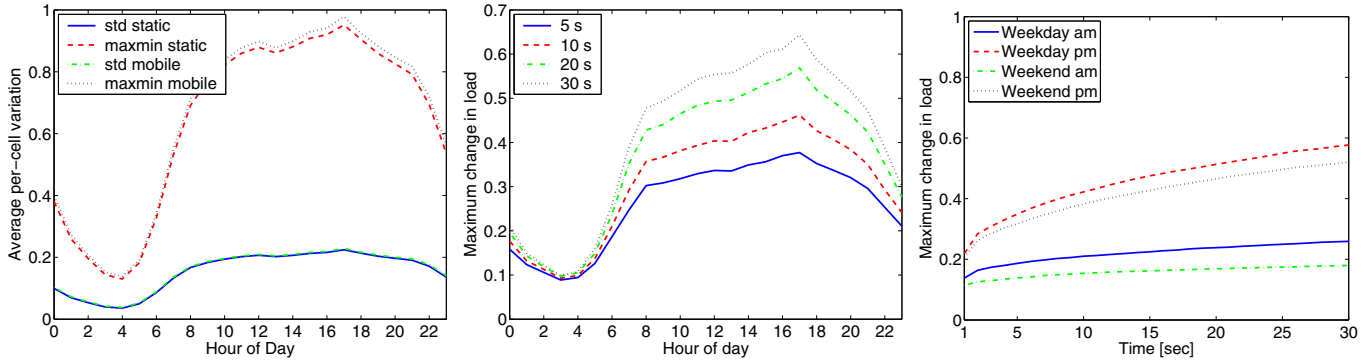
Fig. 8. **(Left)**: Average per-cell variation of load on an hour-of-day basis. We calculate the variation using the standard deviation and the difference between maximum and minimum. **(Middle)**: Maximum change in load, averaged across all cell sectors, plotted on an hourly basis. We use different time windows T_s over which the maximum change is calculated. **(Right)**: Maximum change in load, averaged across all sectors, plotted as a function of T_s for 4 different hours.

However, Figure 6 shows that the event-based random walk model is inferior and succeeds only in about $50 - 60\%$ of the hours for most cells. This can be explained thus. In the random walk model, the $+1$ and -1 events corresponding to the initiation and termination of a call are separated by the duration of that call, which is not exponentially distributed. If there are no additional events during the duration of that call, the duration itself will be an inter-event time. In general, call durations or portions thereof will be part of the inter-event times. We see the impact of this effect in Figure 7 **(Left)** where we plot the success fraction across all cells versus the hour-of-day. During the hours of the night, when the system load is low, the non-exponential distribution of call durations has a significant impact on the distribution of inter-event times. During the day, this impact is reduced.

In Figure 7 **(Middle)**, we plot the success fraction aggregated on a per-day basis. We find no significant variation across days. This is consistent with our intuition that low load is the main reason for the failures. Note that, at low load, the failure rate of the call-based model also increases slightly.

A second component of the random walk model is that the ± 1 events form a Bernoulli process. A necessary (but not sufficient) condition for this to be true is that the sequence of ± 1s should have close to zero auto-correlation at non-zero lags. To understand if this is true, we plot the mean autocorrelation at non-zero lag values on an hour-of-day basis in Figure 7 **(Right)**. We see a similar effect as above. During the night time, when the load is lower, the $+1$ of a call is more likely to be followed by the -1 of that call. This causes negative correlation at odd lags. Accordingly, we can also see the positive correlation at even lags. During the day this effect is reduced.

The above discussion shows that the random walk model is more applicable when the load is high though the Bernoulli assumption is not strictly valid. However, when the load is low, the call-based model with a non-exponential distribution of call durations appears to be the superior model.

B. Variability in Usage

Secondary usage requires the availability of free spectrum. Assuming secondary users are immobile, the best scenario is one in which free spectrum is available for as long as possible in any given cell. In other words, variability in per-cell spectrum availability is not desirable. We quantify this variability by computing the variation in load of each cell during each hour. We calculate the "average-case" variation using the standard deviation and the "worst-case" variation as the difference between maximum and minimum 1-minute load in a cell during each hour. We average these over all cells and plot them on an hour-of-day basis in Figure 8 **(Left)**. As before, we normalize the metrics by a constant factor for proprietary reasons. Notice that both metrics show the same trends. Not surprisingly, the variation is larger during the day, when the load is higher.

The above discussion illustrates that significant intra-cell variations can exist within an hour. Therefore, SUs need to frequently sense the spectrum they are using so as to not impact PUs. We use T_s to denote the time between two consecutive sensing periods. Since the available spectrum could change between two consecutive sensing periods, SUs must be aware of the extent of such short-term variations, and choose T_s accordingly. Figure 8 **(Middle)** provides insights into this by plotting the maximum increase in load averaged over all cells and plotted for different values of T_s. We plot the variation during a representative day of our dataset. The low variations at night are seen again. We see the peak variations late afternoon and a steep reduction in variation thereafter. Notice also that, the variation at $T_s = 30$ is often close to the variation for $T_s = 5$ and never more than twice. This indicates that $20 - 30$ seconds may provide a better tradeoff between sensing overhead and the spectrum that SUs need to leave unutilized for a sudden arrival of PUs.

We take a detailed look at the variation with T_s for four representative hours in Figure 8 **(Right)**. We see less variation during the weekend, possibly due to the reduced average load. We also see that during the AM hours, a small T_s $(1 - 2$ seconds) does not pay off, since the maximum change in load
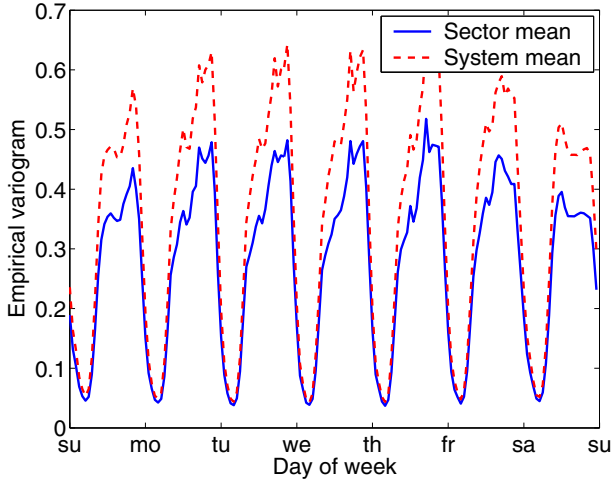
Fig. 9. Empirical variograms illustrating the spatial variation of mean per-cell load over the hours of a week. The sector variograms plot the variation between different sectors of the same cell. The system variograms plot the variation between any two sectors in our dataset.

only increases slightly. We found this to be true for all morning hours (before 10 AM). In the afternoon hours, however, there might be benefits to using a small T_s.

The above results also highlight a significant challenge for cellular DSA - when there is less spectrum available, the availability is more variable, too. Hence, more spectrum should be left unused when such spectrum is most necessary.

V. SPATIAL VARIATIONS

Our analysis so far has been restricted to that of individual cells over time. However, one of the unique advantages of our dataset is that it simultaneously captures primary usage at many base stations. We now report a few results on the spatial variations of primary usage.

It turns out that *variograms* [11] are a well-understood way of encapsulating spatial variations. Consider a random variable A that is defined over space. Then, the variogram is defined as a function of the spatial separation $(\Delta x, \Delta y)$ between two points:

$$\gamma(\Delta x, \Delta y) = \frac{1}{2} E\left[(A(x+\Delta x, y+\Delta y) - A(x,y))^2\right] \quad (4)$$

For random variables that are isotropic, i.e., do not change based on direction, the variogram can be simply defined as a function of the distance h between two points:

$$\gamma(h) = \frac{1}{2} E\left[(A(x',y') - A(x,y))^2\right] \quad (5)$$

where (x', y') and (x, y) are separated by a distance h.

Since our data essentially samples the usage at the locations of base stations, which are not uniformly distributed, we often have only 1 pair of locations separated by any given h. Hence, we use (two) artificial definitions of distance to calculate our variograms. The first artificial distance we use defines two sectors as being separated by distance 1 if they belong to the same cell. We call this the *sector variogram*. The second

distance we use defines any two distinct sectors, regardless of the cell they belong to, as being separated by distance 1. This is our *system variogram*. Note that, the former captures the average variation in usage between sectors of the same cell whereas the latter captures the average variation across the entire system. For our empirical variograms, we ignore the constant factor and report the square root of $\gamma(\cdot)$.

Figure 9 shows the empirical sector and system variograms for the hours of one week. We plot empirical variograms corresponding to per-cell random variables A representing the mean load in each cell sector. As always, we normalize using a constant to show only the relative variations. Surprisingly, we find that the sector variograms are similar to the system variograms. This implies that there is significant local variation in usage. In fact, the local variation is close to the variation we observe across the entire system. We also find that the variograms during the weekend are similar to that seen during the weekday. We also generate variograms for the maximum load in each cell sector. The results are similar and not shown here.

The above results show that, during any hour, there is significant local variability. We now explore the correlation in usage between pairs of cell sectors. To do so, we take the sequence of 1-minute maximum load in each sector and compute the cross-correlation coefficient of pairs of these sequences. About 20% of the pairs had coefficients greater than 0.2 and 20% had coefficients less than −0.2. We cluster the cell sectors into 10 groups so that the cell sectors in each group are more correlated with each other. To obtain such a clustering, we use the k-means [12] algorithm with the pairwise cross-correlation coefficients as the features of each sector. We show pairs of sectors that have significant positive and negative correlation in Figure 10. The thick boxes of black dots along the diagonal in the **(Left)** plot show that intra-cluster sectors have much more positive correlation of their load values among themselves - an indication that our clustering is meaningful. The **(Right)** plot shows that there can be significant negative correlation between sectors of different clusters.

We are currently in the process of using the actual locations of the base stations to understand why certain sectors are clustered together. One preliminary result, for example, is that the most correlated cluster consists of cell sectors at locations that often encounter traffic jams during peak hours. These insights would be quite valuable in determining if and where certain types of secondary usage, e.g., short distance multi-hop wireless communications, are viable.

VI. IMPLICATIONS FOR SECONDARY USAGE OF LICENSED SPECTRUM

Characterization and modeling of PU spectrum usage provides several important insights that are crucial to enable secondary usage of spectrum: spectrum auctions, i.e., coordinated secondary access and spectrum sensing, i.e., opportunistic spectrum access. The owners of spectrum need models of their PUs to determine how much secondary usage is feasible
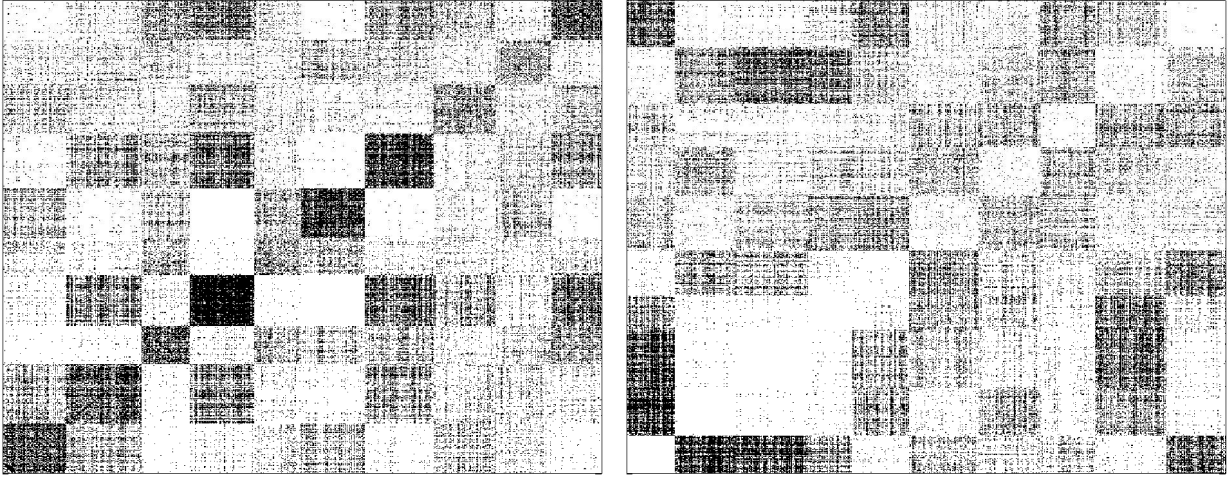
Fig. 10. **(Left)**: Plot showing pairs of cell sectors that have significant positive cross-correlation (> 0.2) of their per-minute load during a representative hour. Each row and column represents one cell sector. The intersection of a row and column has a black dot if the load of those two cell sectors are well correlated during the hour. The cell sectors are ordered for good visual effect. **(Right)**: Plot showing pairs of cell sectors that have significant negative cross-correlation (< -0.2) of their per-minute load during a representative hour.

and how it can be priced. Models of primary usage also help SUs optimize the sensing process and adapt it to the specific characteristics of the given PU. We now discuss insights that our work provides in the specific case of cellular spectrum.

A. Spectrum Auctions

Knowing the spectrum occupancy of a PU, more precisely the dynamic change of the occupancy over time, is crucial to determine the degree to which secondary usage can be allowed, for example, as discussed in [13], [14]. First of all, the instantaneous occupancy sets an upper limit on the resources available for SUs. Thus, our results in Figure 2 indicate that significant secondary usage is possible during the night until almost 7AM regardless of the location. Additionally, in some locations, spectrum can become available during the weekends and weekdays, too. Knowing the future trends of occupancy further help spectrum owners optimize their auction process without impairing PUs. For instance, if the primary spectrum occupancy tends to vary significantly (as can be observed in Figure 8 for the afternoon hours) secondary usage has to be allowed more conservatively, such that enough resources are available for new PUs. On the other hand, if the PU occupancy tends to decrease, spectrum can be rented more aggressively.

Models for call arrival and call duration are essential for optimal pricing strategies of auctioned spectrum. In [1], [6] the authors develop optimal pricing strategies for secondary usage of cellular CDMA networks. The strategies only depend on the call arrival and call duration distributions, which are both assumed to be exponential. Our results show significant deviations of call durations from exponential distributions. Hence, these strategies may have to be revised. The precise implications are subject to further studies.

B. Spectrum sensing

From a CR perspective, there are two fundamental questions to be answered for the development of sensing techniques for SUs: (1) How often does sensing have to be performed? (2) What is the required observation time of a single channel to reliably detect potential PUs? Answers to these questions determine how much time and resources are needed for detecting PUs.

The first question is usually answered by the PU, which specifies the so called *maximum interference time*, i.e., the maximum time a SU is allowed to interfere with a PU communication. Clearly, the maximum interference time sets an upper limit on the periodic time interval after which a channel used by a SU network has to be sensed (T_s). Knowing the probability distribution of the arrival process of the primary communication (in our study the call arrivals), and given a target probability p_i that the SU interferes with the PU, T_s can be simply calculated using the CDF ($p_i = P(X \leq T_s)$). Equation (6) shows the calculation of T_s assuming an exponential call arrival process.

$$p_i = 1 - e^{-\lambda T_s} \Leftrightarrow T_s = -\frac{\ln(1 - p_i)}{\lambda} \qquad (6)$$

The knowledge of the arrival process, thus, enables us to adjust the time (T_s) after which a channel needs to be sensed. For our investigation, the mean call inter-arrival time (over one hour) per cell varies from the sub-second range to tens of minutes. Assuming a maximum of 30 calls per cell and a probability of interference of $p_i = 0.001$ this would result in a required sensing time between $T_s = 0.03$ s and $T_s = 18$ s. This huge gap clearly indicates the achievable gains for adaptively choosing T_s based on the call inter-arrival time, which can itself be gleaned by sensing. Results such as those in Figure 8 also provide insights into good tradeoffs for sensing strategies.

An answer to the second question, i.e., determining the time needed for sensing a *single* channel, is much more complex and depends on various factors such as the sensitivity requirements of the PU, the specific sensing technique used, distributed/cooperative sensing aspects, etc. However, regardless of the time the sensing process takes for a specific system, it is desirable not to waste this time for sensing an occupied channel. Here a model of the duration of a PU communication can help to determine the time after which a channel sensed to be occupied by a PU should be sensed again. In particular, our analysis of the call durations shows that there are many short calls and the remaining are spread over a "semi-heavy" tail. Hence, a conditional sensing process is well-motivated: the SU initially uses a rapid sensing frequency for the case that a new call is short. After a few tens of seconds, rapid sensing is likely to yield little benefit. Hence, slower sensing is justified.

VII. RELATED WORK

In [15], general guidelines are given on how to perform spectrum measurements. The authors performed a measurement campaign in the US from 1992 [16] where they actually found the usage of the *I*ndustrial *S*cientific and *M*edical (ISM) band at $2.4\,$GHz very low. Fleurke et al. [17] differentiate between the actual occupancy and the mean occupancy. They describe different sampling methods (random and systematic sampling). In [18] a detailed modeling proposal for *W*ireless *L*ocal *A*rea *N*etwork (WLAN) and microwave ovens is given.

Measuring and modeling the spectrum occupancy in the *H*igh *F*requency (HF) range $(3 - 30\,$MHz) had drawn a lot of research attention since the late 1970's. Various models of the boolean variable that represents whether or not a channel is used have been proposed. Most of these are tailored to the specific characteristics and usage patterns in the HF spectrum. The Laycock-Gott logit model [19], [20], [21], [22] requires 25 parameters to model the measured data. Measurements are taken during the solstices, where atmospheric and galactic noise is low. Goultelard et al. [23] have developed a model for estimating the time variations of the so called congestion (occupancy) using channel availability statistics. In [24], the authors develop guidelines on sensing times and estimate confidence limits for these times. They also propose using first-order Markov chains to model channel availability. This work has been extended in [25], [26] to two-dimensional first-order Markov chains and in [27] to cyclostationary two-dimensional first-order Markov chains.

In recent years, a lot of measurement studies have been carried out to show the under-utilization of the licensed spectrum. Some examples of wide-band measurement campaigns include the Chicago spectrum measurements [28], [29] covering the spectrum range from $30\,$MHz to $3\,$GHz and $960\,$MHz to $2500\,$MHz, respectively, and the New Zealand measurements [30] in the spectrum range from $806\,$MHz to $2750\,$MHz. Further studies can be found on the Shared Spectrum's website [31]. Though these studies show the abundance of temporally unused spectrum, they give little insight into the dynamic behavior of the licensed users legally operating in those bands. A measurement campaign focusing on the cellular voice bands was carried out during the soccer world-cup 2006 in Germany [2], [3]. The authors show the differences in spectrum occupancy in the GSM and UMTS bands before, during, and after a match. However, similar to the wide-band measurements mentioned above, little insight into call dynamics such as call arrivals or call durations is gained. The authors of [4] analyze the spectrum utilization in the New York cellular bands (CDMA as well as GSM). The CDMA signals are demodulated to determine the number of active Walsh codes (i.e., the number of ongoing calls). To determine the number of calls in the GSM bands, image processing of the spectrogram snapshots is used. Although this analysis provides more detailed results for the utilization of the cellular bands, call arrivals and durations are also not examined.

In [5], similar to our study, the authors analyze the call logs of a switch of a cellular GSM provider. The study was conducted in Qingdao, China. In contrast to our study, they analyze and model only the call durations. Additionally, the amount of data used for the study is much more limited, namely two selected hours of successive days. The authors match the call durations to a log-normal distribution. We also find that lognormal has some advantages but it still does not model our durations fully.

VIII. CONCLUSIONS AND FUTURE WORK

We presented a large scale characterization of primary users in the cellular spectrum. We used a dataset that allowed us to compute the load of hundreds of base stations over three weeks. Using this dataset, we investigated a call arrival model and a random walk model that directly models the aggregate load. We derived several results, some of which are summarized below:

- Often, the duration of wireless calls (and the time for which voice channels are allocated) are assumed to be exponentially distributed. We find that the durations are not exponential in nature and possess significant deviations that make them hard to model.
- An exponential call arrival model (coupled with a non-exponential distribution of call durations) is often adequate to model the primary usage process.
- A simpler random walk can be used to describe primary usage under high load conditions.
- Spectrum usage can exhibit significant variability. We found that the load of individual sectors varies significantly even within a few seconds in the worst case. We also find high variability even across sectors of the same cell.
- Our spatial analysis revealed that there are clusters of sectors whose intra-cluster usage patterns are more correlated.

We believe that our work provides a first-step proof-point to guide both policy and technical developments related to DSA. In this paper, we made no use of sensing data and relied wholly on network data. In future work, we intend to

perform simultaneous sensing and in-network data collection. This would allow us to investigate how accurate a sensing-based approach is and also validate the results in this paper.

REFERENCES

[1] A. Al Daoud, M. Alanyali, and D. Starobinski, "Secondary pricing of spectrum in cellular CDMA networks," in *Proceedings of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2007)*, 2007, pp. 535–542.

[2] T. Renk, C. Kloeck, F. K. Jondral, P. Cordier, O. Holland, and F. Negredo, "Spectrum measurements supporting reconfiguration in heterogeneous networks," in *Proceedings of the 16th IST Mobile and Wireless Communications Summit*, 1–5 July 2007, pp. 1–5.

[3] O. Holland, P. Cordier, M. Muck, L. Mazet, C. Klock, and T. Renk, "Spectrum power measurements in 2G and 3G cellular phone bands during the 2006 football world cup in germany," in *Proceedings of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2007 (DySPAN 2007)*, Apr. 2007, pp. 575–578.

[4] T. Kamakaris, M. M. Buddhikot, and R. Iyer, "A case for coordinated dynamic spectrum access in cellular networks," in *Proceedings of the 1st IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN) 2005*, Baltimore (MD), Nov. 2005, pp. 289–298.

[5] J. Guo, F. Liu, and Z. Zhu, "Estimate the call duration distribution parameters in GSM system based on k-l divergence method," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCom 2007)*, 2007, pp. 2988–2991.

[6] H. Mutlu, M. Alanyali, and D. Starobinski, "Spot pricing of secondary spectrum usage in wireless cellular networks," in *Proceedings of the 27th IEEE Conference on Computer Communications 2008 (INFOCOM 2008)*, 2008, pp. 682–690.

[7] NIST/SEMATECH, "e-handbook of statistical methods," Online, Aug. 2008. [Online]. Available: http://www.itl.nist.gov/div898/handbook/

[8] F. J. Massey, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[9] E. Limpert, W. A. Stahel, and M. Abbt, "Log-normal distributions across the sciences: Keys and clues," *BioScience*, vol. 51, no. 5, pp. 341–352, May 2001.

[10] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes," *Annals of Mathematical Statistics*, vol. 23, pp. 193–212, 1952.

[11] G. Bohling, "Introduction to geostatisticsa and variogram analysis," Online, 2005. [Online]. Available: http://people.ku.edu/~gbohling/cpe940/Variograms.pdf

[12] H. Steinhaus, "Sur la division des corp materiels en parties," *Bulletin L'Acadmie Polonaise des Science C1*, vol. IV, no. C1. III, pp. 801–804, 1956.

[13] A. P. Subramanian, H. Gupta, S. R. Das, and M. M. Buddhikot, "Fast spectrum allocation in coordinated dynamic spectrum access based cellular networks," in *Proceedings of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2007 (DySPAN 2007)*, 17-20 April 2007, pp. 320–330.

[14] O. Ileri, D. Samardzija, and N. Mandayam, "Demand responsive pricing and competitive spectrum allocation via a spectrum server," in *Proceedings of the 1st IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN) 2005*, 2005, pp. 194–202.

[15] R. Matheson, "Strategies for spectrum usage measurements," in *Proceedings of the IEEE International Symposium on Record Electromagnetic Compatibility, 1988*, 1988, pp. 235–241.

[16] J. Wepman, R. Matheson, K. Allen, and E. Pol, "Spectrum usage measurements in potential pcs frequency bands," in *Proceedings of the 1st International Conference on Universal Personal Communications, 1992 (ICUPC '92)*, 1992, pp. 01.03/1–01.03/6.

[17] S. R. Fleurke, H. G. Dehling, H. K. Leonhard, A. D. Brinkerink, and R. den Besten, "Measurement and statistical analysis of spectrum occupancy," *European Transactions on Telecommunications*, vol. 15, no. 5, pp. 429–436, 2004.

[18] D. A. Roberson, C. S. Hood, J. L. LoCicero, and J. T. MacDonald, "Spectral occupancy and interference studies in support of cognitive radio technology deployment," in *Proceedings of the 1st IEEE Workshop on Networking Technologies for Software Defined Radio Networks (SDR '06)*, 25–25 Sept. 2006, pp. 26–35.

[19] P. Laycock, M. Morrell, G. Gott, and A. Ray, "A model for HF spectral occupancy," in *Proc. Fourth International Conference on HF Radio Systems and Techniques*, 11–14 Apr 1988, pp. 165–171.

[20] C. Pantjiaros, P. Laycock, G. Gott, and S. Chan, "Development of the laycock-gott occupancy model," *IEE Proceedings-Communications*, vol. 144, no. 1, pp. 33–39, Feb. 1997.

[21] G. Gott, S. Chan, C. Pantjiaros, and P. Laycock, "High frequency spectral occupancy at the solstices," *IEE Proceedings-Communications*, vol. 144, no. 1, pp. 24–32, Feb. 1997.

[22] L. Economou, H. Haralambous, C. Pantjairos, P. Green, G. Gott, P. Laycock, M. Broms, and S. Boberg, "Models of HF spectral occupancy over a sunspot cycle," *IEE Proceedings-Communications*, vol. 152, no. 6, pp. 980–988, 9 Dec. 2005.

[23] C. and Goutelard and J. Caratori, "Time modelling of H.F. interferences," in *Proceedings of the Fifth International Conference on HF Radio Systems and Techniques*, J. Caratori, Ed., 1991, pp. 343–348.

[24] A. Spaulding and G. Hagn, "On the definition and estimation of spectrum occupancy," *IEEE Transactions on Electromagnetic Compatibility*, vol. 19, no. 3, pp. 269–280, 1977.

[25] A. J. Gibson and L. Arnett, "Statistical modelling of spectrum occupancy," *Electronics Letters*, vol. 29, no. 25, pp. 2175–2176, 1993.

[26] A. J. Gibson and L. Arnett, "Measurements and statistical modelling of spectrum occupancy," in *Proceedings of the 6th International Conference on HF Radio Systems and Techniques*, 1994, pp. 150–154.

[27] D. Percival, M. Kraetzl, and M. Britton, "A markov model for HF spectral occupancy in central australia," in *Proceedings of the 7th International Conference on HF Radio Systems and Techniques*, 7–10 July 1997, pp. 14–18.

[28] 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 *TAPAS '06: Proceedings of the first international workshop on Technology and policy for accessing spectrum*. New York, NY, USA: ACM Press, 2006, p. 1.

[29] J. T. MacDonald, "A survey of spectrum occupancy in chicago," Illinois Institute of Technology, Tech. Rep., 2007.

[30] R. Chiang, G. Rowe, and K. Sowerby, "A quantitative analysis of spectral occupancy measurements for cognitive radio," in *Proceedings of the 65th IEEE Vehicular Technology Conference 2007, (VTC Spring 2007)*, 2007, pp. 3016–3020.

[31] Shared Spectrum Company, "Recent spectrum measurements." [Online]. Available: http://www.sharedspectrum.com/measurements/recent.html