Primary User Behavior in Cellular Networks and Implications for Dynamic Spectrum Access

Daniel Willkomm
Telecommunication Networks Group
Technische Universität Berlin,
Germany
willkomm@tkn.tu-berlin.de
Sridhar Machiraju
Sprint Applied Research
California, USA
machiraju@sprint.com
Jean Bolot
Sprint Applied Research
California, USA
bolot@sprint.com
Adam Wolisz
Technische Universität Berlin,
Germany
University of California, Berkeley, USA
wolisz@ieee.org

Abstract

*D*ynamic *S*pectrum *A*ccess (DSA) approaches, which propose to opportunistically use underutilized portions of licensed wireless spectrum such as cellular bands, are increasingly being seen as a way to alleviate spectrum scarcity. However, before DSA approaches can be enabled, it is important that we understand the dynamics of spectrum usage in licensed bands. Our focus in this paper is the cellular band. Using a unique dataset collected *inside* a cellular network operator, we analyze the usage in cellular bands and discuss the implications of our results on enabling DSA in these bands. 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 if and when spectrum is available, develop models of primary usage and understand the implications of these results on DSA techniques such as sensing.

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, however, well known that any static allocation leads unavoidably to underutilization – at least from time to time. Therefore, the option of reusing assigned spectrum when it is temporarily (and locally) available – frequently referred to as Dynamic Spectrum Access (DSA) – promises to increase the efficiency of spectrum usage. 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. Hence, understanding the way PUs use spectrum is very important to implementing DSA.

We present the results of a large-scale measurement-driven study of PUs (also, see [2]) in cellular bands and the implications of these results on enabling DSA in these bands. Our focus on cellular spectrum is important for several reasons. Apart from TV bands, cellular bands are viable band to implement DSA – both because they are widely used throughout the world and also because engineering devices and data applications for these bands are well understood. In fact, cellular femtocells, which have recently become popular, can be viewed as implementing a type of secondary usage that uses a naive mechanism, namely, reduced power, to avoid interference. We believe that future femtocells will likely incorporate more sophisticated mechanisms based on sensing that minimize such interference.

Our study is based on the analysis of a unique dataset consisting of call records collected *inside* a cellular network. Thus, we are able to provide insights on a call level that prior sensing-based studies [3], [4], [5] were unable to. Another advantage of our study is its scale – we are able to study 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 measure the "ground truth" as seen by the network, and, hence, are able to model call arrival processes as well as system capacity.

We provide insights into three different aspects relevant for enabling cellular DSA. First, we show that cellular DSA is viable and attractive, especially during nights and weekends. Hence, we recommend an emphasis on developing scenarios for secondary usage that operate during such non-peak hours. Second, we describe two models of primary usage. The first models the call arrival process but needs to account for the skewed distribution of call durations. The second model tracks the total number of calls and does not require any knowledge of call durations. However, it is less successful than the call-based model and is more applicable during peak hours when the number of calls is high. We also find that rare but significant spikes in usage exist and must be guarded against. Third, since the success of cognitive radios depends crucially on how readily spectrum bands can be sensed, we provide guidelines for sensing in cellular bands. This is much more challenging compared to sensing in TV bands, for example, because cellular voice usage exhibits frequent variations in time and space. Hence, SUs of cellular voice bands likely need to employ more agile DSA techniques than SUs of TV bands.

II. Methodology

The dataset we use in this paper was collected from hundreds of cell sectors[1] of a US CDMA-based cellular operator. The data captured 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 captured 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 was measured with a resolution of several milliseconds. The duration was 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 in prior work.

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 capture and what they do not.

The first limitation of our dataset is its lack of full information on mobility. We were able to record only the initial and final sector of each call. Thus, we are unable to account for spectrum usage in the other sectors that users may have visited during calls. 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 from which we collected our dataset – 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

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

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 only 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 analysis are likely to apply to other cellular voice networks.

Using either of the aforementioned approximations, we 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 beginning 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 begins and decreased by -1 when a call terminates.

III. Dynamics of Spectrum Availability

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.

The day/night dependence is exhibited system-wide as seen in Figure 2. Here, we ignore information about the individual cells to which calls are assigned and consider all calls as arriving to a single entity. For such a hypothetical system, we plot the normalized average call arrival rates during 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.

Figure 3, which plots average call durations as a function of time, 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. These are caused by a brief interruption in the data collection, which caused some short calls to not be recorded, thereby artificially inflating the mean duration of calls.

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 4. 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.

A. Implications

Knowing the spectrum occupancy of a PU, more precisely the dynamic change of the occupancy over time, is crucial to determining the degree to which secondary usage can be allowed, for example, as discussed in [6], [7]. 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. Knowing the future trends of occupancy further helps 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 4 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. Figure 4 highlights 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 more spectrum is being used.

IV. Modeling Primary Usage

Since secondary users opportunistically use spectrum not utilized by primary users, models of primary usage in individual cells play an important role in designing and deploying cellular DSA approaches. There are two simple models that fit the behavior of primary users well (see [2] for details). One such model is the *call-based model*. This model uses two random variables, T and D to describe the inter-arrival time between two calls and the duration of calls. An obvious and popular choice is to model call arrivals as a Poisson call process (independent and identically distributed exponential inter-arrival times) and call durations as being exponentially distributed.

It turns out that the distribution of call inter-arrival times is well described by an exponential distribution in more than 90% of the hours for most cells. We use the *Anderson-Darling* test with 95% confidence level as a goodness-of-fit test for exponential distribution. Note that, since we use a 95% confidence level for the Anderson-Darling test, we expect only 95% of our tests to succeed. The technical details of these tests can be found in [2].

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. Though Poisson processes have been used to model fixed telephone calls for a long time, our study is one of the first to show that this is largely true for *individual cells* in mobile systems.

A. Call Durations

Though the inter-arrival times of calls are well modeled as a Poisson process, the call-based model has a significant disadvantage, namely, the distribution of call durations. In Figure 5, we plot the empirically-observed histogram of 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 about 27 seconds. 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.

Further analysis shows that there are likely two different distributions of call durations – one during the day and the other during the night (11PM-5AM). Furthermore, the transition hours between day and night likely see a mixture of both these distributions. In Figure 6 **(Left)**, we compare the overall and nighttime distributions of call durations. Note that we use the log-log scale. We find that the nighttime 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 classic 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. In Figure 6 **(Left)**, 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 6 **(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 6 **(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 6 **(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.

B. Event-based Model

The skewed distribution of call durations is the primary disadvantage of the call-based model and can be eliminated by using an alternative *event-based* model. This model ignores details about individual calls and instead models only the load $X(\cdot)$, i.e., the total number of ongoing calls. 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:

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

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 otherwise. Since there is a $+1$ for every -1, p should be $\frac{1}{2}$. A Poisson process is the obvious choice to model the inter-event times.

It turns out that inter-event times are well-modeled as exponential distributions for only about 50% of the hours in most cells. As shown in Figure 7 **(Left)**, exponential modeling fails almost twice as often during the night (when the load is low) than during the day.

The skewed distribution of call durations is also responsible for the failures of the event-based model. This is because the $+1$ and -1 events correspond to the initiation and termination of a call and 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. Thus, 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.

A second component of the event-based 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 nighttime, 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 event-based 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 skewed distribution of call durations is the superior model.

C. Implications

Characterization and modeling of PU spectrum usage provides several insights that are crucial to enable secondary usage of spectrum. For example, the owners of spectrum need models of their PUs to determine how much secondary usage is feasible and how it can be priced. Models for call arrival and call duration are essential for optimal pricing strategies of auctioned spectrum. In [1], [8] 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.

V. Short-term Variability

One of the primary requirements of DSA-based approaches is that SUs should not affect PUs. Hence, it is critical that SUs in cellular networks frequently sense the spectrum and vacate it if new PUs are detected. Also, 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 the time T_s between consecutive sensing periods accordingly. Figure 8 **(Left)** 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 thereafter. Notice also that the variation at $T_\mathrm{s} = 30$ is often close to the variation for $T_\mathrm{s} = 5$ 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 unoccupied 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 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.

A. Implications on 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 must sensing 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 *C*umulative *D*istribution *F*unction (CDF) ($p_\mathrm{i} = P(X \leq T_\mathrm{s})$). Equation (2) shows the calculation of T_s assuming an exponential call arrival process.

$$p_\mathrm{i} = 1 - \mathrm{e}^{-\lambda T_\mathrm{s}} \Leftrightarrow T_\mathrm{s} = -\frac{\ln(1-p_\mathrm{i})}{\lambda} \tag{2}$$

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 $p_\mathrm{i} = 0.001$ this would result in a required inter-sensing time between $T_\mathrm{s} = 0.03\,\mathrm{s}$ and $T_\mathrm{s} = 18\,\mathrm{s}$. This huge gap clearly indicates the gains achievable by 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.

VI. Related Work

In recent years, many measurement studies have been carried out to show the underutilization of licensed spectrum. Some examples of wide-band measurement campaigns include the Chicago spectrum measurements [9], covering the spectrum range from $30\,\mathrm{MHz}$ to $3\,\mathrm{GHz}$ and $960\,\mathrm{MHz}$ to $2500\,\mathrm{MHz}$, respectively, and the New Zealand measurements [10] in the spectrum range from $806\,\mathrm{MHz}$ to $2750\,\mathrm{MHz}$. 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 [3], [4]. 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 [5] 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.

VII. Conclusions and Future Work

We presented a large scale characterization of primary users in the cellular spectrum and discussed the implications on enabling cellular DSA. We used a dataset that allowed us to compute the load of hundreds of base stations over three weeks. 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.

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)*, April 2007, pp. 535–542.

[2] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary Users in Cellular Networks: A Large-scale Measurement Study," in *Proceedings of the 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2008)*, October 2008.

[3] 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*, July 2007, pp. 1–5.

[4] 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.

[5] 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*, Nov. 2005, pp. 289–298.

[6] 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)*, April 2007, pp. 320–330.

[7] 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)*, November 2005, pp. 194–202.

[8] 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)*, April 2008, pp. 682–690.

[9] 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*, August 2006.

[10] 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)*, April 2007, pp. 3016–3020.

VIII. Biographies

DANIEL WILLKOMM (willkomm@tkn.tu-berlin.de) received his Diploma degree (with distinction) in electrical engineering from the Technische Universitt Berlin (TUB), Germany, in 2004. He is currently a Ph.D. student at the Telecommunication Networks Group (TKN), TUB, working in the area of cognitive radio networks. From 2005 to 2007 he received a scholarship from the DFG graduate program (Graduiertenkolleg) Stochastic Modeling and Quantitative Analysis of Complex Systems in Engineering, an interdisciplinary research group of three major universities in Berlin. Currently he is visiting Sprintlabs in California for a joint project between Sprint and TKN.

SRIDHAR MACHIRAJU (machiraju@sprint.com) is a research scientist in the Applied Research Group at the Sprint Advanced Technology Labs in Burlingame, California. He has been with Sprint since the summer of 2005. He received his M.S. and Ph.D. degrees from the University of California at Berkeley in 2003 and 2006 respectively. Prior to that, he obtained his Bachelor's degree from the Indian Institute of Technology, Madras (IIT-M) in Chennai, India. He is broadly interested in problems related to performance analysis and algorithm design in the area of networked computer systems especially mobile systems. His current research is focused on wireless resource allocation, wireless security and active measurements.

JEAN BOLOT (bolot@sprint.com) runs the research lab of Sprint located in the San Francisco Bay area. His research interests center around the measurement, analysis, and economics of the Internet, and in particular of the mobile Internet. Prior to joining Sprint, he was a founding team member of Ensim, a Silicon Valley company in the area of data center automation. Earlier, he did research at INRIA in France on Internet measurement and voice over the Internet. He received his MS and PhD in Computer Science from the University of Maryland at College Park in 1988 and 1991, respectively.

ADAM WOLISZ (wolisz@ieee.org) received his degrees (Diploma 1972, Ph.D. 1976, Habil. 1983) from Silesian Unversity of Technology, Gliwice, Poland. He joined TUB in 1993 where he is a chaired professor in telecommunication networks and executive director of the Institute for Telecommunication Systems. He is also an adjunct professor at the Department of Electrical Engineering and Computer Science, University of California, Berkeley. His research interests are in architectures and protocols of communication networks. Recently he has been focusing mainly on wireless/mobile networking and sensor networks. He was with the Polish Academy of Sciences until 1990 and GMD-Fokus, Berlin from 1990 to 1993.

FIGURES

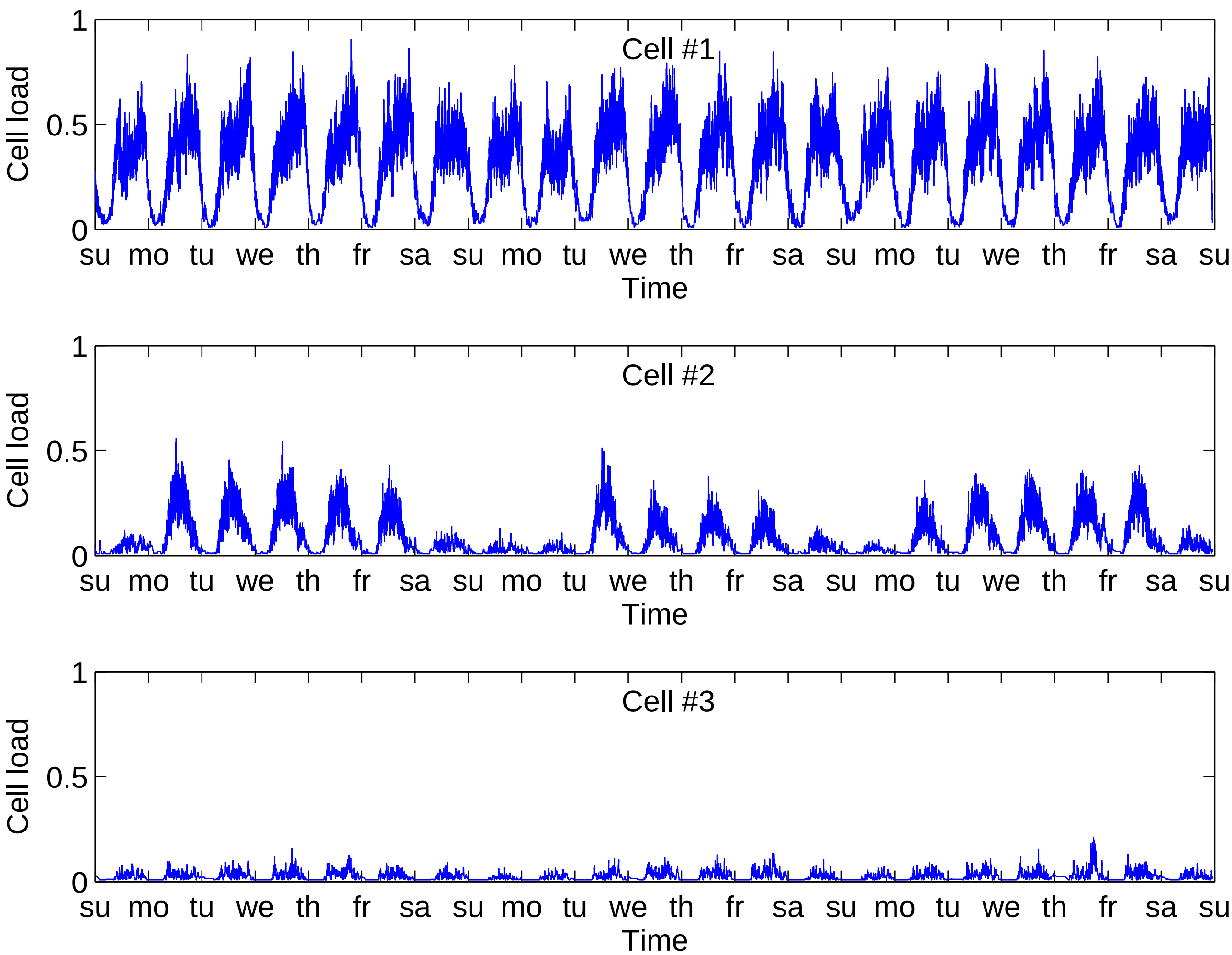

Fig. 1. Normalized load of three different cell sectors over 3 weeks. We plot the moving average of each cell over one second. The cells show high load **(Top)**, varying load **(Middle)**, and low load**(Bottom)**.

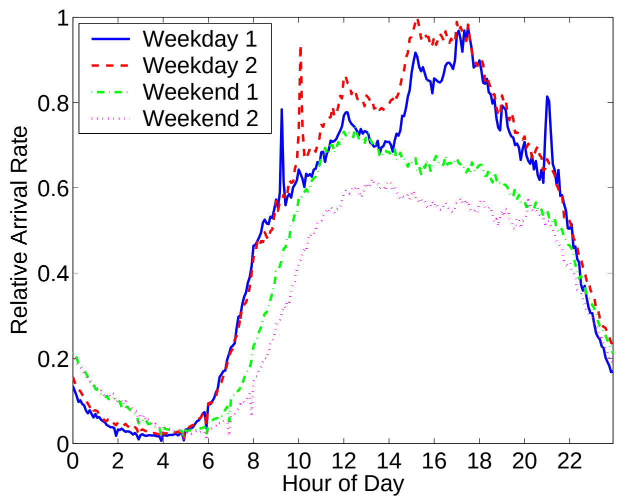

Fig. 2. Distribution of system-wide average call arrival rates during four different days. The arrival rates are averaged over 5-minute slots.

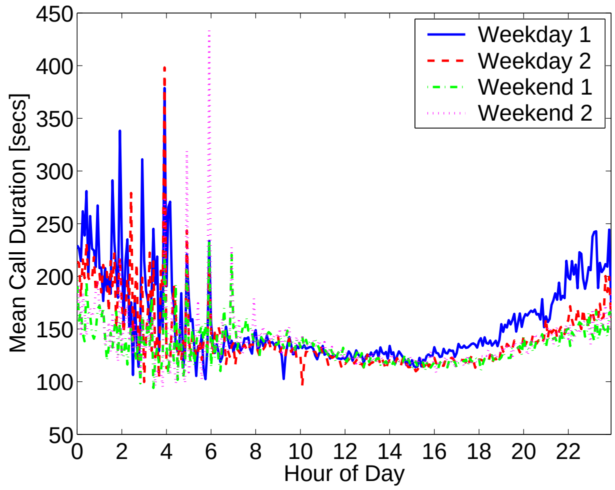

Fig. 3. Distribution of average call duration over 5-minute periods during four different days. The large spikes during the mornings are due to small gaps in collection.

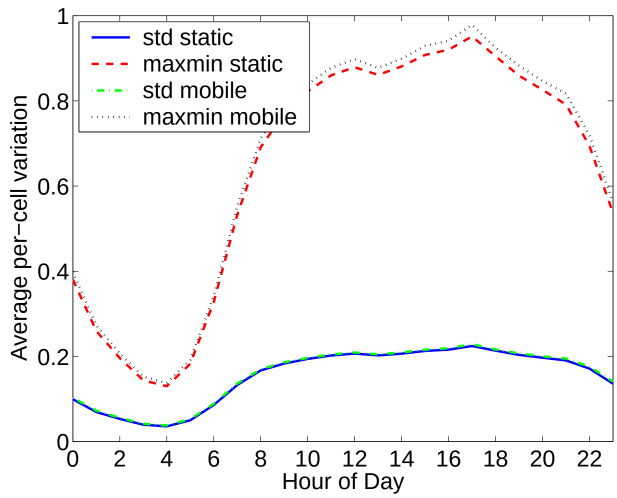

Fig. 4. Average per-cell variation of load on an hour-of-day basis. We calculate the variation using the standard deviation and the difference between maximum and minimum.

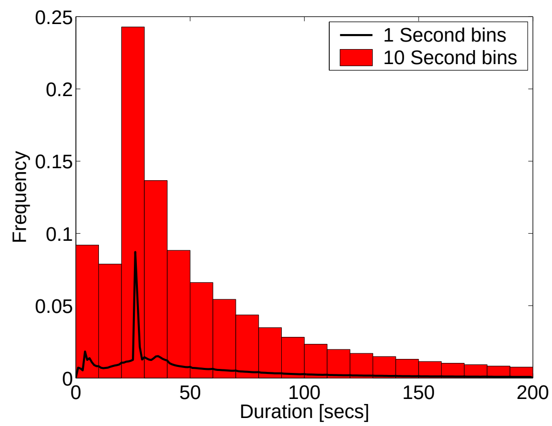

Fig. 5. Histogram of call durations. We plot the histogram using different bin sizes.

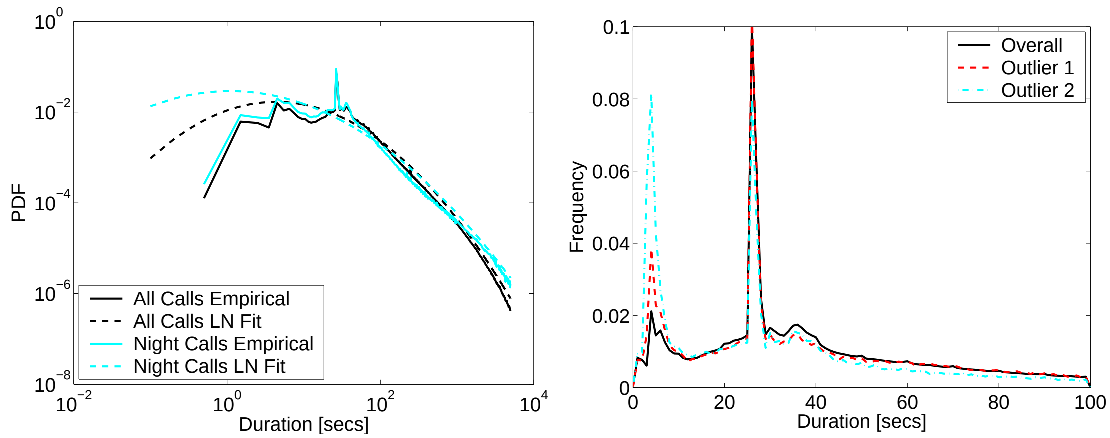

Fig. 6. **(Left)**: Duration distributions and lognormal fits. **(Right)**: Illustration of anomalous distributions during 2 hours.

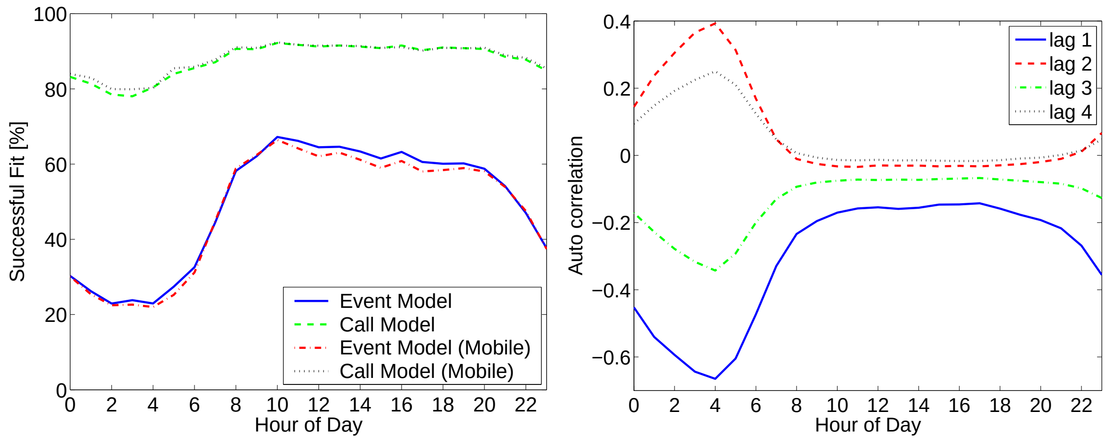

Fig. 7. **(Left)**: The percentage of successful fits (across all cells) averaged on a per-day basis. **(Right)**: The per-hour auto-correlation of the step sizes (Φ) in our event-based models averaged across all cells.

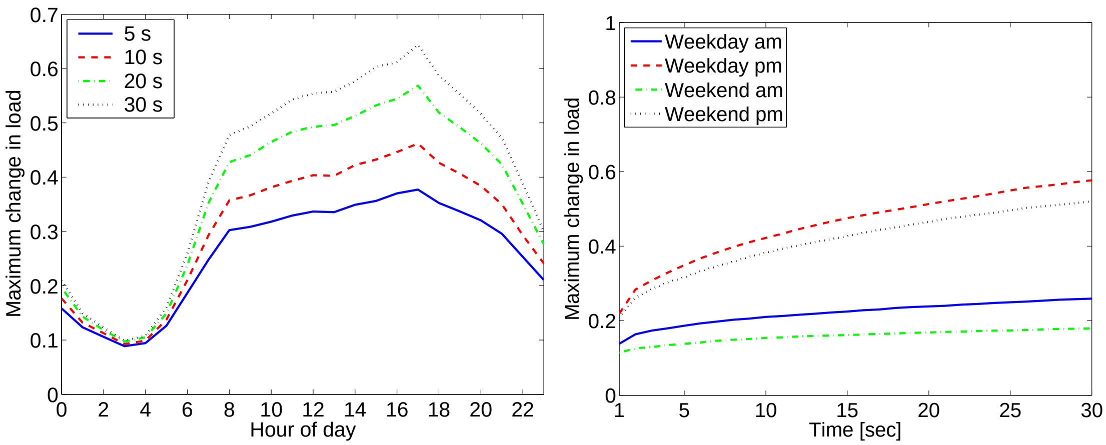

Fig. 8. **(Left)**: Maximum change in load, averaged across all cell sectors, plotted on an hourly basis. We use different time windows T_s over which the maximum change is calculated. **(Right)**: Maximum change in load, averaged across all sectors, plotted as a function of T_s for 4 different hours.