Exploiting LTE White Space using Dynamic Spectrum Access Algorithms based on Survival Analysis

Timothy A. Hall, Anirudha Sahoo, Charles Hagwood, Sarah Streett
National Institute of Standards and Technology
Email: {tim.hall, anirudha.sahoo, charles.hagwood, sarah.streett}@nist.gov

***Abstract*—In this study, we design and implement two algorithms for dynamic spectrum access (DSA) that are based on survival analysis. They use a non-parametric estimate of the cumulative hazard function to predict the remaining idle time available for secondary transmission subject to the constraint of a preset probability of successful completion. To show that the algorithms are effective in real-world scenarios even at fine time scales, we evaluate their performance using data collected from an LTE band to model primary user activity. The algorithms are run in different configurations, i.e., they are trained and run on a few combinations of data sets. Our results show that as long as the cumulative hazard functions are fairly similar across datasets, the algorithms can be trained on one day's dataset and run on that of another day's without any significant degradation of performance. The algorithms achieve fairly high white space utilization and have a measured probability of interference which always stays below the preset threshold.**

I. Introduction

Dynamic spectrum access (DSA) seems poised to mitigate the problem of spectrum scarcity. In a typical DSA scenario, a primary user (PU) has priority access to a given band. A secondary user (SU) can transmit during unoccupied (idle) periods opportunistically but must vacate when the PU needs the band again. In order to make efficient use of the spectrum in a DSA environment, an accurate and useful model of spectrum occupancy is needed.

Spectrum occupancy refers to whether or not a particular channel or band is occupied. In this paper, we use the term channel to denote the smallest allocable range of frequencies within a particular communications technology, e.g., 180 kHz for LTE. A band is comprised of multiple channels and represents a single service, e.g., there are 50 channels in a 10 MHz LTE uplink band. We model the occupancy of a given channel as a two-state (binary) random process similar to that used by Spaulding and Hagn [1]:

$$X(t) = \begin{cases} 1 & \text{if } P_R(t) > P_{th} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $P_R(t)$ is the signal power observed at the receiver at time t and P_{th} is a threshold value. $X(t) = 1$ represents the occupied state and $X(t) = 0$ represents the unoccupied state.

A. Previous Work

Various models have been proposed in the literature for spectrum occupancy. A two state Discrete-Time Markov Chain (DTMC) has been used to model spectrum occupancy in [2]. However, stationary DTMC models have been found to be inadequate to represent idle and busy periods. Hence, authors in [2] have proposed a time-inhomogeneous DTMC model. Some authors have also used semi-Markov models for spectrum occupancy [3]. This study assumes a general distribution (rather than exponential) for the idle and busy periods of the spectrum. Further, since there are only two states (ON/OFF), the process is also analyzed as an Alternating Renewal Process [3], [4].

Continuous-Time Markov Chain (CTMC) based models have also been used to represent spectrum idle and busy periods. Since some measurement studies have shown that the ON and OFF periods of spectrum are not exponentially distributed, authors in [5], [6], [7] have used semi-Markov models for the purpose.

Model occupancy of adjacent channels has been modeled as a two-dimensional Markov chain by Gibson and Arnett in [8], [9].

Some studies have shown that busy and idle periods of spectrum exhibit negative correlation, i.e., the idle period following a long busy period is typically short and vice-versa [10]. In this study, the authors have proposed time-correlation models for periodic and non-periodic auto-correlation functions.

There have been few models proposed for predicting spectrum occupancy, which is critical to allocating spectrum to the secondary users. The Partially-Observable Markov Decision Process (POMDP) model has been proposed in [11]. The spectrum sharing scheme proposed in [12] is based on prediction of spectrum occupancy by the primary users in terms of the expected remaining OFF time. A two state semi-Markov model proposed in [3] is used to estimate the distribution parameters of ON/OFF periods. Some methodologies proposed in the literature indirectly predict spectrum occupancy by limiting the duration of transmission of the secondary user (SU) to some constraint. In [13], the transmission duration of an SU is based on the maximum bound on probability of interference to the primary user (PU). Residual idle time of an Alternating Renewal Process is used in [4] to indirectly predict reappearance of the PU. Some researchers have used a Restless Multiarm Bandit formulation for opportunistic channel access [14], [15]. Researchers have also looked at pattern mining of spectrum occupancy data to predict channel

Hagwood, Robert; Hall, Timothy; Sahoo, Anirudha; Streett, Sarah.
"Exploiting LTE White Space using Dynamic Spectrum Access Algorithms based on Survival Analysis."
Paper presented at 2017 IEEE International Conference on Communications (ICC), Paris, France. May 21, 2017 - May 25, 2017.

availability [16], [17].

B. Motivation for Present Work

The motivation behind the present work is threefold. We want to develop a prediction scheme that is robust, flexible and useful even for very fine time scales. We assume centrally coordinated scheduling for the SUs. The scheduler knows when the primary user is no longer active, and when an SU requests a transmission opportunity, the scheduler grants or denies the SU request. Our scheme is not limited to a centralized scheduling architecture, however. It can be used in a carrier sense multiple access (CSMA) system as well. In such a system, the SUs would sense the channel and use our algorithms to predict residual idle time before transmitting as a form of collision avoidance. Analysis and application of prediction schemes presented in this paper to a CSMA based system is beyond the scope of this study.

Most of the stochastic based schemes in the literature either assume a certain distribution (e.g., exponential) of spectrum occupancy data or require that a distribution be fitted to a set of observed data. This study does not have such a requirement. It uses a non-parametric estimate of the cumulative hazard function from historical data to grant dynamic access to the SUs. Hence, our scheme is much simpler to implement in practice.

Finally, most of the DSA schemes in the literature are run over simulated spectrum occupancy data. We ran our algorithms over real spectrum occupancy data to show that they are suitable for implementation on practical systems. We show the effectiveness of our DSA algorithms over LTE Band 17, which is centered at 709 MHz with a 10 MHz bandwidth in the uplink.

We also envision that our scheme (or some variation thereof) may be used in the Spectrum Sharing architecture proposed in the 3.5 GHz band [18]. In this architecture, there will be three tiers of users in the band. First tier users have the highest priority, but they use the band infrequently. The tier two users, called Priority Access Layer (PAL) users, will likely be LTE carriers and have medium priority. When tier 1 and tier 2 users are not present in the band, it can be used by tier 3 users called General Authorized Access (GAA) users. It is conceivable that a PAL user can *sell* its white spaces (idle times) to users who can make use of transmission opportunities of the order of hundreds of milliseconds as long as the interference to PAL users remains below an agreed threshold. These opportunistic users can implement our scheme to exploit PAL white spaces.

Let us now define the prediction problem upon which our DSA algorithms are based more precisely. We are concerned with how long the channel has been unoccupied by the PU and how much longer the channel will remain unoccupied. Specifically, given that the channel has been unoccupied by the PU for duration t and a request from an SU arrives to transmit for a duration τ, what is the probability that the SU will be able to complete the transmission before the PU appears on the channel? Figure 1 illustrates the relationship between the PU and SU.

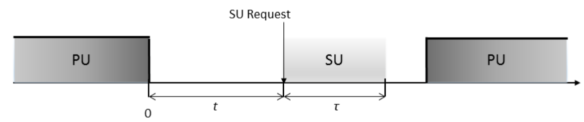

Fig. 1. SU request

The remainder of this paper is structured as follows. Section II formulates the prediction problem in terms of survival analysis, resulting in two algorithms for secondary channel requests. Section III describes the collected data, simulation environment and metrics we used to evaluate the algorithms. Section IV presents our results. Section V interprets the results and discusses future work.

II. Prediction Algorithms

A. Survival Analysis

Survival analysis has been used to analyze statistical properties of the duration of time until an event, such as failure in a mechanical system, occurs [19]. Our prediction problem can be solved by using survival analysis as presented below.

Let $T_1, S_1, T_2, S_2, \ldots$, represent the successive idle and busy periods of the spectrum. Thus T_i and S_i represent the i^{th} idle and busy periods respectively. The T_i's can be thought of as survival times. That is, an idle period survives only until the channel becomes busy again. Let random variable T represent an arbitrary survival time and $0 < p < 1$ an adjustable parameter. Assuming the T_i are independent and identically distributed as T, our prediction problem can be represented by the hypothesis testing problem given by

$$\begin{aligned} \mathcal{H}_0 :\ & P[T \geq t+\tau \mid T \geq t] > p \quad versus \\ \mathcal{H}_1 :\ & P[T \geq t+\tau \mid T \geq t] \leq p \end{aligned} \tag{2}$$

$\mathcal{H}_0$ holds if the idle period, having lasted t units of time, lasts τ more units of time with probability greater than p. Note that p represents the probability of successful transmission for duration τ, given that the channel has been idle for duration t.

The basic functions of survival analysis are the survival function and the hazard function. The survival function at time t is the probability of surviving at least t units of time and is given by

$$S(t) = P[T \geq t] = 1 - F(t) = \int_t^\infty f(s)ds \tag{3}$$

where $f(s)$ and $F(t)$ are the probability density function and cumulative distribution function of T, respectively. The hazard function is the probability of instantaneous failure at time t given survival up to time t and indicates the risk of failure at time t. The hazard function is given by

$$\begin{aligned} h(t) &= \lim_{\delta t \to 0} \frac{P[t \leq T < t+\delta t \mid T \geq t]}{\delta t} \\ &= \lim_{\delta t \to 0} \frac{P[t \leq T < t+\delta t]}{P[T \geq t] \cdot \delta t} \\ &= \frac{1}{P[T \geq t]} \cdot \lim_{\delta t \to 0} \frac{P[t \leq T < t+\delta t]}{\delta t} \\ &= \frac{f(t)}{S(t)} \end{aligned} \tag{4}$$

From (3), it is clear that the derivative of $S(t)$ is $-f(t)$. Hence, (4) can be rewritten as

$$h(t) = -\frac{d}{dt} log S(t) \tag{5}$$

Now integrating both sides of (5) from 0 to t, noting that $S(0) = 1$ and finally taking the exponential on both the sides, we have

$$S(t) = \exp\left(-\int_0^t h(s)ds\right) \tag{6}$$

The function important to us is the cumulative hazard function, defined by $H(t) = \int_0^t h(s)ds, t \geq 0$. Using (6) we have

$$\begin{aligned} P[T \geq t+\tau \mid T \geq t] &= \frac{P[T \geq t+\tau]}{P[T \geq t]} \\ &= \exp\left(-\int_0^{t+\tau} h(s) + \int_0^t h(s)ds\right) \\ &= \exp(-[H(t+\tau) - H(t)]). \end{aligned} \tag{7}$$

Thus, using (7) the hypotheses in (2) can be expressed as

$$\begin{aligned} &\mathcal{H}_0: \ \exp(-[H(t+\tau) - H(t)]) > p \quad versus \\ &\mathcal{H}_1: \ \exp(-[H(t+\tau) - H(t)]) \leq p \end{aligned} \tag{8}$$

Having observed a large sample $T_1, T_2, \ldots, T_n$ of n survival times, a non-parametric estimate of the survival function can be computed using the empirical distribution function, $F_n(t)$ of the data $T_i, i = 1, \ldots, n$, as shown below.

$$S_n(t) = 1 - F_n(t) = 1 - \frac{1}{n}\sum_{i=1}^{n} 1_{T_i < t} \tag{9}$$

where 1_A is the indicator function for event A.

Let $T_{(1)} \leq T_{(2)} \cdots \leq T_{(n)}$ be the ordered $T_i, i = 1, \ldots, n$. Then the survival function at any $T_{(i)}$ can be computed using (9) as follows.

$$\begin{aligned} S_n(T_{(i)}) &= 1 - \frac{1}{n}\sum_{j=1}^{n} 1_{T_j < T_{(i)}} \\ &= 1 - \frac{1}{n} \cdot (i-1) = \frac{n-i+1}{n} \end{aligned} \tag{10}$$

In the above derivation, we used the fact that exactly $(i-1)$ values of T_i are strictly less than $T_{(i)}$. Each $T_{(i)},\ 1 \leq i \leq n$, has an estimated probability of occurence of $\frac{1}{n}$. Hence,

$$f_n(T_{(i)}) = \frac{1}{n} \tag{11}$$

Using (10) and (11) in (4) we have

$$\begin{aligned} h_n(T_{(i)}) &= \tfrac{1}{n-i+1} \quad \text{for } i = 1, 2, \cdots, n \\ h_n(t) &= 0 \quad \text{for all other } t \end{aligned}$$

Using the definition of the cumulative hazard function, an estimate is given by

$$H_n(t) = \sum_{i:T_{(i)} \leq t} \frac{1}{n-i+1} \tag{12}$$

Our test statistic is based on the difference of the cumulative hazard function at two different times. An estimate for the difference of the cumulative hazard function at two different times is given by

$$H_n(t+\tau) - H_n(t) = \sum_{i:t \leq T_{(i)} \leq t+\tau} \frac{1}{n-i+1} \tag{13}$$

Note that this is a form of the well-known Nelson-Aalen estimator for the cumulative hazard function. We used a more general form of $H_n(t)$ to account for duplicate values of $T_{(i)}$, that is, multiple idle times of the same duration [20]. Therefore, after simple manipulation of $\mathcal{H}_0$ in (8), our prediction algorithms are formulated in terms of an approximate test statistic,

$$\text{Reject } \mathcal{H}_0 \text{ if}: \ H_n(t+\tau) - H_n(t) \geq (-\ln p). \tag{14}$$

B. Definition of Algorithms

Below are two formulations of the prediction algorithm. The first is a request to transmit on a channel for duration τ. If the channel is occupied at the time of request, the request is denied. If the channel is not occupied, then the algorithm grants the request if it determines that the probability of a successful transmission (i.e., the probability of completing the transmission without colliding with the PU) is above a given threshold.

Algorithm 1 Request channel for τ seconds

input:
τ - the transmit duration requested
parameters:
$H_n(t)$ - the estimated cumulative hazard function
t_0 - the time elapsed since end of last transmission
p - the probability of successful transmission
output: *Grant* or *Deny*

if *occupied* **then**
    return *Deny*
end if
$\theta := -\ln p$
$W_n := H_n(t_0+\tau) - H_n(t_0)$
if $W_n < \theta$ **then**
    return *Grant*
else
    return *Deny*
end if

The second algorithm returns the longest estimated duration available for transmission for a request made at a particular time. The time returned is the largest value for which the probability of successful transmission exceeds the given threshold.

Algorithm 2 Request maximum channel availability

parameters:
$H_n(t)$ - the estimated cumulative hazard function
$\{T_{(i)}\}$ - the n ranked idle times used to compute $H_n(t)$
t_0 - the time elapsed since end of last transmission
p - the probability of successful transmission
output: τ - the maximum transmit time available now

if *occupied* **then**
    return 0
end if
$\theta := -\ln p$
Find largest τ in $[0, T_{(n)}]$ such that $H_n(t_0+\tau) - H_n(t_0) < \theta$
return τ

III. Evaluation

To evaluate the algorithms, we used real LTE uplink spectrum occupancy data to represent our primary user occupancy. Secondary user requests for spectrum were simulated using a Poisson arrival model, i.e., the SU request inter-arrival times were exponentially distributed.

A. Data Collection

Data was collected in Band 17, a 10 MHz uplink LTE band centered at 709 MHz. A small 10.78 cm rubber duck antenna was connected to an Ettus Universal Sofware Radio Peripheral (USRP) [1] running USRP hardware driver (UHD) version 003.009.001 and GNU Radio version 3.7.9rc1. The output is a 56 point power spectrum computed every 100 ms. Each power spectrum coefficient is an 8 bit signed integer representing a decibel (dB) value rounded to the nearest integer. Each coefficient corresponds to peak power in dB over a 180 kHz range. The middle 50 coefficients correspond to the 50 LTE channels. We applied a noise threshold power value to produce a binary occupancy sequence for each of the 50 channels. We looked at all the different power values collected and picked the $75th$ percentile value as the noise threshold. For the data collected, the threshold turned out to be -67 dB. The idle time distributions for day1 and day2 are presented in Table I.

Data was collected for two continuous 48 hour periods. The first 48 hours ran from 2:00 PM UTC (9:00 AM local time) Monday, February 1, 2016 to 2:00 PM UTC on Wednesday, February 3, 2016. The second dataset covers weekend hours, 2:00 PM UTC Saturday, February 27, 2016 to 2:00 PM UTC Monday, February 29, 2016. Each 48 hour data set is split into two parts, each containing 24 hours of data. Thus, the data captured from Monday 9 AM to Tuesday 9 AM is designated as $day1$ data, and the data captured from Tuesday 9 AM to Wednesday 9 AM is designated as $day2$. Similarly, the weekend data is termed as $wknd1$ and $wknd2$.

B. Simulation

As stated above, an idle period of the spectrum occupancy is a set of one or more consecutive zeros. Each zero represents an idle period with a duration of one sampling interval (100 ms for our experiments). Thus, the T_i values (in terms of sampling interval) are represented as the number of consecutive zeros. Similarly, a busy period is a set of one or more consecutive ones. In our experiments, we have used the occupancy of LTE uplink channel number 5 as our PU traffic. After building the idle and busy periods, we then compute the cumulative hazard function as per (13). We have set the probability of successful transmission (p) to 0.9. Thus, the interference threshold, which is equal to $(1-p)$, is set to 0.1. Note that the PU expects its measured probability of interference (PoI) to be less than this preset interference threshold.

[1]The identification of any commercial product or trade name does not imply endorsement or recommendation by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

We have evaluated the performance of Algorithm 1 and Algorithm 2 in different configurations as described below. The configurations are denoted as *train_run*, where *train* is the data used for training the algorithm (i.e., the cumulative hazard function is built using this data) and *run* represents the data which is used to run the algorithm. We have four data sets, each of 24 hours duration.

As an example, in configuration $day1\_day1$, the algorithms are trained using $day1$ data, i.e., the cumulative hazard function is built using $day1$ data and then the algorithm is also run on $day1$ data. Results from this configuration validate the effectiveness of survival analysis for opportunistic spectrum access.

When using configuration *day1_day2* the algorithms are trained using day1 data but run on day2 data. This configuration helps us understand how the algorithms perform when the training and running data are from different week days. Note that in practice, the *day1_day1* configuration does not correspond to a realistic scenario, since the training has to happen on some historical data and then the algorithm would run on different data. Hence, this configuration is useful in practice.

Yet another example is configuration *wknd1_day1*. This configuration helps us determine if it is feasible to train the algorithm on a weekend data set and run it on a week day data set.

C. Metrics

We used the following metrics to measure performance of the two algorithms. The first two metrics are common to both the algorithms whereas the remaining four are defined for Algorithm 1 only.

- White Space Utilization (WSU): Given the spectrum occupancy of a channel, White Space Utilization (WSU) of the channel by a secondary user is defined as the fraction of total idle time used by the secondary user for its own transmission. In another words, it is the ratio of total duration of idle time used by the secondary user for its own transmission to the total idle time duration in the spectrum occupancy of the channel.
- PoI: For a given channel, the PoI of the secondary user is defined as the probability that a transmission of the SU collides with that of the PU. Thus, it is the ratio of the number of times an SU transmission collides (or runs into a busy period) with a PU transmission to the total number of SU transmissions over a statistically long observation period.
- Desirable Accept Ratio (DAR): This is defined as the fraction of requests that were accepted and the corresponding transmissions were successful. In these cases the algorithm correctly predicted the remaining idle time.
- Undesirable Accept Ratio (UAR): This is defined as the fraction of requests that were accepted and the corresponding transmissions were not successful, i.e., these transmissions resulted in collision with PU transmission. In these cases the algorithm incorrectly predicted the remaining idle time.

Length (sample interval)	Number of occurrences day1 (%)	Number of occurrences day2 (%)	Length (sample interval)	Number of occurrences day1 (%)	Number of occurrences day2 (%)
1	10 331 (39.75 %)	9866 (35.98 %)	(10, 20]	760 (2.92 %)	1093 (3.99 %)
2	8523 (32.78 %)	8784 (32.03 %)	(20, 30]	281 (1.08 %)	447 (1.63 %)
3	2174 (8.36 %)	2475 (9.03 %)	(30, 40]	165 (0.63 %)	258 (0.94 %)
4	957 (3.68 %)	969 (3.53 %)	(40, 50]	124 (0.48 %)	149 (0.54 %)
5	664 (2.55 %)	796 (2.9 %)	(50, 500]	546 (2.1 %)	783 (2.86 %)
6	431 (1.66 %)	491 (1.79 %)	(500, 5000]	64 (0.25 %)	133 (0.48 %)
7	308 (1.18 %)	361 (1.32 %)	(5000, 10 000]	5 (0.02 %)	18 (0.07 %)
8	264 (1.02 %)	302 (1.1 %)	(10 000, 17 871]	3 (0.01 %)	10 (0.04 %)
9	240 (0.92 %)	295 (1.08 %)	(17 871, 132 171]	4 (0.01 %)	0 (0 %)
10	152 (0.58 %)	193 (0.7 %)			

TABLE I
IDLE TIME DISTRIBUTION OF DAY1 AND DAY2 DATA

- Desirable Reject Ratio (DRR): This is defined as the fraction of requests that were rejected and would have resulted in collision with the PU if they were accepted. So, in these cases the algorithm correctly predicted the remaining idle time and rejected the requests.
- Undesirable Reject Ratio (URR): This is defined as the fraction of requests that were rejected and would have resulted in successful transmission if they were accepted. In these cases the algorithm incorrectly predicted the remaining idle time and rejected the requests. This metric represents lost opportunities for the SU.

IV. RESULTS

Figure 2 shows the performance of Algorithm 1 in terms of WSU as average request inter-arrival time varies. As request inter-arrival time increases, WSU decreases since the offered load from the SU decreases. It is interesting to note that the performance of all the configurations in terms of WSU are almost the same. Since the cumulative hazard functions of the different days have almost the same slope for most of the values between $n = 0$ to $n = 100$, the decision to accept or reject a request is almost the same regardless of which day's data is used for training. This leads to nearly the same WSU for different configurations. Thus, Algorithm 1 can be trained using data from any of the days without significantly affecting the WSU of the system.

From Figure 3, we observe that the PoI is always less than the set threshold of 0.1. When we compare the PoI of Algorithm 2 (see Figure 5) we notice that the PoI is an order of magnitude less than that of Algorithm 2. Algorithm 1 only transmits for a fixed duration when the request is granted. In our experiment the fixed duration is 200 ms, which is a relatively short duration. In other words, when the requested duration is short, Algorithm 1 is less aggressive than Algorithm 2. Hence, the probability of an SU transmission colliding with the PU is very low.

Figure 4 shows the performance of Algorithm 2 in terms of WSU. As the average request inter-arrival time increases, the SU exploits less white space for transmission. Hence, the WSU decreases. We also notice that WSUs for the set of configurations which are run on day1 (e.g., day1_day1, day2_day1, wknd1_day1) are higher than those run on day2. This can be explained by studying the cumulative hazard functions of the two days. The cumulative hazard functions of the data set are shown in Figure 6 and Figure 7. Since the range of idle period is very large, we show the $H(\cdot)$ function up to 100 sampling intervals in Figure 6, whereas Figure 7 shows the entire range of idle period. The cumulative hazard functions of day1 and day2 have almost the same slope for idle periods less than 100 sampling intervals. Hence, for a given time of arrival of an SU request, the maximum duration granted would be almost the same for the two days. So, the idle time distribution of the two days has more influence than the $H(\cdot)$ function on the WSU for the two days for grants less than 100 sampling intervals. From Table I, we observe that day1 has some very large idle times. For example, day1 has four idle times in the range $(17\,871, 132\,171]$, which are very large idle times, whereas the maximum idle time of day2 was $17\,871$. This leads to more transmission opportunities for the SU when running over day1 data and gives rise to higher WSU for day1 than for day2. Now for idle periods greater than 100 sampling periods (refer to Figure 7), for day2, the slope of $H(\cdot)$ is very steep, whereas for day1 the slope flattens due to the extremely long run lengths of idle periods. Thus, when the time of arrival of a request falls into an idle period that has lasted longer than 100 sampling intervals, day1 grants longer transmission opportunities. Furthermore, when the elapsed idle time is more than the longest idle period in the training data, the transmission opportunity granted is set to the longest idle period of training data. These two factors also contribute to higher WSU for day1 than day2.

Since the cumulative hazard functions of the four days have almost equal slope for most of the n values less than 100 sampling intervals, training the algorithm on data from any day produces nearly the same WSU for that given day. Thus, the curves for day1_day1, day2_day1 and wknd1_day1 are close to each other.

When we compare the WSU performance of Algorithm 1 with Algorithm 2 for a given request inter-arrival time, we notice that the WSU of Algorithm 1 is much lower than that of Algorithm 2. The fundamental design of the two algorithms gives rise to this behavior. For a given request, Algorithm 2 maximizes the SU transmission duration, whereas Algorithm 1 only checks to see if it can grant a request for a constant transmission duration (200 ms in our experiment). Thus, Algorithm 2 is able to achieve higher WSU.

Hagwood, Robert; Hall, Timothy; Sahoo, Anirudha; Streett, Sarah.
"Exploiting LTE White Space using Dynamic Spectrum Access Algorithms based on Survival Analysis."
Paper presented at 2017 IEEE International Conference on Communications (ICC), Paris, France. May 21, 2017 - May 25, 2017.

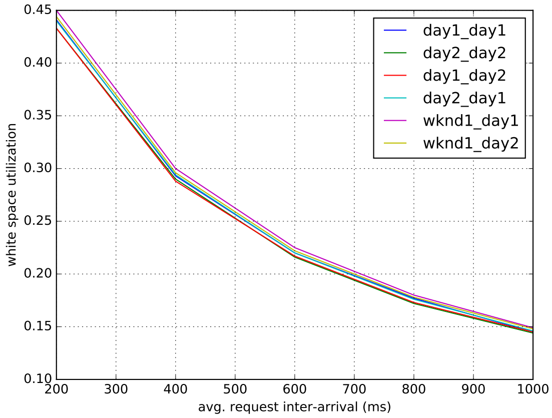

Fig. 2. WSU vs inter-arrival time for Algorithm 1

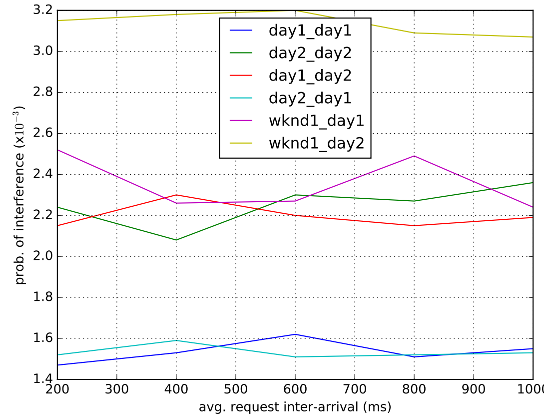

Fig. 3. PoI vs inter-arrival time for Algorithm 1

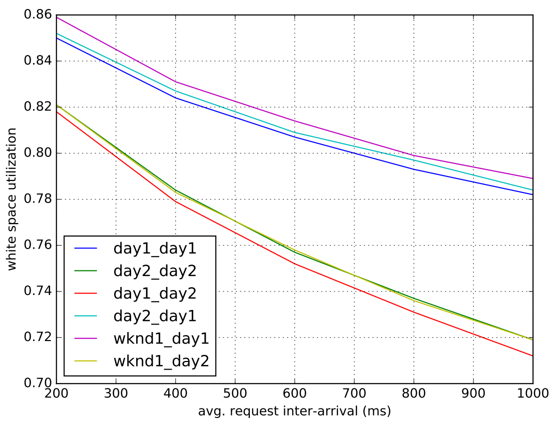

Fig. 4. WSU vs inter-arrival time for Algorithm 2

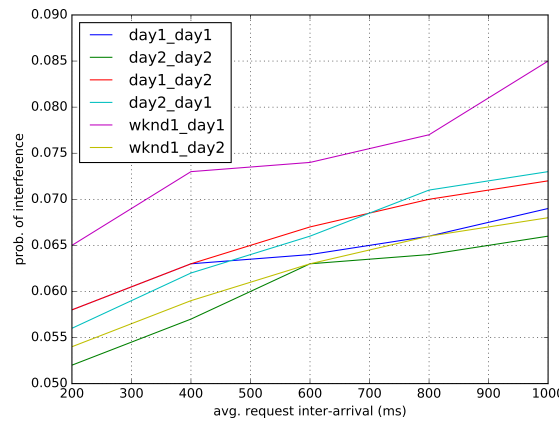

Fig. 5. PoI vs inter-arrival time for Algorithm 2

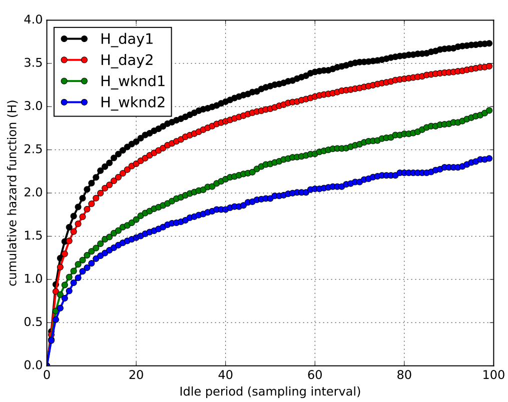

Fig. 6. Cumulative Hazard Function (up to 100 sampling interval)

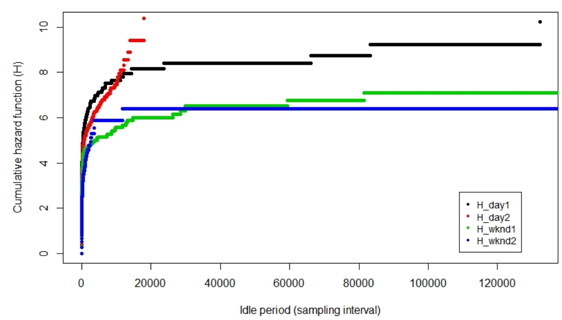

Fig. 7. Cumulative Hazard Function (for the entire range of idle period)

Figure 5 shows the variation of PoI as the inter-arrival time of the requests increases. For all configurations, the PoI is below the set threshold (0.1), thus satisfying the interference constraint of the PU.

We also ran our two algorithms on the weekend data (wknd1 and wknd2 data) sets. We are not able to present the results here due to space limitation, however, the results look very similar to the week day results presented in this paper.

Our results indicate that the algorithms can be trained using any data set and run on another data set as long as the cumulative hazard functions are similar (in terms of slope).

We show the desirable and undesirable accept and reject ratios of Algorithm 1 when the SU request inter-arrival time is 200 ms. The URR is zero for all configurations. Thus

Configuration	DAR (%)	UAR (%)	DRR (%)	URR (%)
day1_day1	78.7	0.2	21.1	0
day2_day1	78.6	0.2	21.1	0
wknd1_day1	81.4	0.4	18.2	0
day2_day2	76.2	0.3	23.5	0
day1_day2	76.2	0.3	23.5	0
wknd1_day2	79.6	0.5	19.9	0

TABLE II
VARIOUS ACCEPT AND REJECT RATIOS FOR REQUEST INTER-ARRIVAL TIME 200 MS FOR ALGORITHM 1

Algorithm 1 has no lost opportunities in all the configurations. The algorithm also has very low UAR, which is good, since this metric shows how well the algorithm avoids making bad decisions in accepting a request. Although we have results for other request inter-arrival times, we are not able to present them due to space limitations. However, those results are equally good.

All our experiment runs were for a very long duration (approximately twenty-four hours). Hence, the number of SU requests were very large. So, the computed performance metrics of the two algorithms (e.g., WSU and PoI) had very little variation across different runs.

V. CONCLUSION AND FUTURE WORK

We introduced DSA algorithms based on survival analysis that make efficient use of white space in an LTE band, even at very fine time scales. They are stochastic but non-parametric and therefore do not require the assumption of a particular distribution. This makes the implementation simple. The tuning parameter for the algorithm is the probability of successfully completing a transmission or, equivalently, the PoI. Thus, it is easy to interpret and directly reflect desired system performance metrics. We used real LTE band occupancy data for the PU activity in our simulations. Our results show that if the cumulative hazard functions are fairly similar (in terms of slope) across different datasets, the algorithms can be trained on one day's dataset and run on another day's dataset without significant degradation of performance. This is a very important property of the algorithms, since in practice, the algorithms will be trained on historical data and then run in real-time. We expect that in actual spectrum sharing systems the PUs will be wary of sharing their spectrum with SUs for fear of too much interference. This is addressed in our algorithms by showing that the PoI is always below the preset threshold in all configurations.

This paper provides an initial performance analysis of the algorithms in an LTE band. Evaluation using datasets collected in different bands at varying locations with other traffic characteristics needs to be done. Other time scales need to be investigated to show the range over which the algorithms are effective. A theoretical performance analysis and comparison with other prediction schemes are needed as well.

Depending on the SU application, alternative forms of the algorithms presented in this paper can easily be developed using the same fundamental approach. One can imagine a form of spectrum requests that includes a maximum or desired transmit time and a minimum acceptable time. The algorithm would then either deny the request or return a grant duration in the requested range. Another form could have the user requesting a minimum initial grant and then the scheduler can add additional follow-on transmission time, if available, once the initial request has elapsed. An adaptive version of the algorithm may be more attractive for implementation on practical systems. It would update the estimated cumulative hazard function as new idle periods appear in the spectrum.

REFERENCES

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

[2] M. Lopez-Bentez and F. Casadevall, "Discrete-time spectrum occupancy model based on markov chain and duty cycle models," in *2011 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, May 2011, pp. 90–99.

[3] H. Kim, and K. Shin, "Efficient Discovery of Spectrum Opportunities with MAC-layer Sensing in Cognitive Radio Networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 533–545, May 2008.

[4] M. Sharma and A. Sahoo, "Stochastic Model Based Opportunistic Channel Access in Dynamic Spectrum Access networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1625–1639, July 2014.

[5] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in WLAN channels: Empirical model and its stochastic analysis," in *TAPAS '06 Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum*, August 2006.

[6] ——, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 66–72, May 2007.

[7] L. Stabellini, "Quantifying and modeling spectrum opportunities in a real wireless environment," in *2010 IEEE Wireless Communication and Networking Conference*, April 2010, pp. 1–6.

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

[9] ——, "Measurements and statistical modelling of spectrum occupancy," in *HF Radio Systems and Techniques, 1994., Sixth International Conference on*, July 1994, pp. 150–154.

[10] M. L. Lopez-Benitez and F. Casadevall, "Modeling and simulation of time-correlation properties of spectrum use in cognitive radio," in *International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCom)*, June 2011, pp. 1–5.

[11] Q. Zhao, L. Tong, A. Swami and Y. Chen, "Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 2, pp. 589–600, April 2007.

[12] K. W. Sung, S. Kim and J. Zander, "Temporal Spectrum Sharing Based on Primary User Activity Prediction," *IEEE Transactions on Wireless Communications*, vol. 9, no. 12, pp. 3848–3855, December 2010.

[13] A. Plummer, M. Taghizadeh and S. Biswas, "Measurement based bandwidth scavenging in wireless networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 1, pp. 19–32, January 2012.

[14] C. Tekin and M. Liu, "Online learning of rested and restless bandits," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5588–5611, August 2012.

[15] Y. Gai and B. Krishnamachari, "Decentralized online learning algorithms for opportunistic spectrum access," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, December 2011, pp. 1–6.

[16] S. Yin, D. Chen, Q.Zhang, M. Liu and S. Li, "Mining Spectrum Usage Data: A Large-Scale Spectrum Measurement Study," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 1033–1046, June 2012.

[17] P. Huang, C-J. Liu, X. Yang, L. Xiao and J. Chen, "Wireless Spectrum Occupancy Prediction Based on Partial Periodic Pattern Matching," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 7, pp. 1925–1934, July 2014.

[18] "Further Notice of Proposed Rule Making," https://apps.fcc.gov/edocs_public/attachmatch/FCC-13-100A1.pdf, April 2014, [Online; accessed 25th March 2015].

[19] R. G. Miller Jr, *Survival analysis*. John Wiley & Sons, 2011, vol. 66.

[20] O. Aalen, O. Borgan, and H. Gjessing, *Survival and event history analysis: a process point of view*. Springer Science & Business Media, 2008.