Spectrum Map: Toward Predicting the Spatial Distribution of Spectrum Usage in CRNs

Saptarshi Debroy Shameek Bhattacharjee and Mainak Chatterjee
Department of Electrical Engineering & Computer Science
University of Central Florida
Orlando, FL 32816
Email: {saptarsh, shameek, mainak}@eecs.ucf.edu

***Abstract*—Recent measurements on radio spectrum usage have revealed the abundance of under-utilized bands of spectrum that belong to licensed users. Prior knowledge about occupancy of such bands and corresponding achievable performance metrics can potentially help secondary networks to devise effective strategies to improve utilization. In this paper, we use Shepard's method of interpolation to create a *spectrum map* that provides a spatial distribution of spectrum usage over a region of interest. It is achieved by intelligently fusing the spectrum usage reports shared among the secondary nodes at various locations. We further use this spectrum usage distribution to estimate different radio and network performance metrics like channel capacity, network throughput, spectral efficiency and bit error rate. Through simulation experiments we show the correctness of the prediction model and how it can be used by secondary networks for strategic positioning of secondary pairs and selecting candidate channels.**

I. Introduction

In a cognitive radio network (CRN), secondary nodes (i.e., unlicensed users) equipped with cognitive radio enabled devices continuously monitor the presence of primary nodes (i.e., licensed users) and opportunistically access the unused or under-utilized licensed bands of primary nodes [1]. However, the most important regulatory aspect of these networks is that the secondary nodes must not interfere with primary transmissions. Thus, when secondary nodes detect transmissions from primaries, they are mandated to relinquish those interfering channels immediately and switch to other non-interfering channels. Such reactive nature of working principle of secondary networks is insufficient for desired utilization of under-used licensed spectrum. However, a prior knowledge of the possible transmission activities of the primaries can allow the secondary nodes to effectively access the available channels and predict the expected radio and network performances for quality of service (QoS) provisioning. Thus, there is a need to proactively estimate the spectrum usage at any *arbitrary* location and predict the nature of spectrum utilization in a region of interest. This can be achieved by fusing the information gathered by the various stationary secondary nodes operating at different locations. Such fused information in turn helps in strategic positioning of secondary devices for best utilization of under-utilized radio resource.

Prior work in this domain involves modeling the distribution of spectrum utilization in a primary network using both theoretical models and real-world data logs [4], [5], [7], [8]. One of the earliest work in spectrum measurements was reported by NSF [5]. Authors in [4] measure the spectrum usage at different locations in Tokyo city and creating 3-dimensional plots that show the temporal distribution of frequency usage near the monitored points. In [8], authors observed that the location distribution of primary TV transmitters in USA and Europe closely follow Poisson model. A frequency distribution model to emulate the real nature of noise from primary transmitter on different channels was presented in [7]. Though the above mentioned techniques allow us to predict the spectrum usage at known locations, there is still little understanding on how to build a mathematical function to capture the spatial distribution for the spectrum utilization at any arbitrary location based on the observations at known locations.

In this paper, we create a *Spectrum Map* by defining a spatial distribution function for spectrum utilization. This map works as a reference to predict the spectrum usage at any arbitrary location. We use a collection of networked cognitive nodes at different locations that monitor the spectrum usage in a distributed manner and share their findings with others. Through cooperatively fusing the raw usage data, we show how an interpolation function can be used to construct a continuous and differentiable distribution function which governs the estimation of the spectrum utilization at any location. We use this prior knowledge of spectrum utilization to predict radio and network performance. We also demonstrate how performance metrics like the channel capacity, system throughput, spectrum efficiency, and bit error rates, can be estimated for a secondary transmitter-receiver pair at any potential secondary network location. We evaluate our model through a simulation environment emulating the real-world spectrum data of RWTH Mobnets [7] and duplicating their transmission pattern. We graphically represent the predicted spatial nature of spectrum usage and compare the accuracy of spectrum usage estimation by using the results with actual received signal strength values. We also simulate the nature of key performance metrics.

The rest of the paper is organized as follows. Section II discusses how the secondary nodes share their raw spectrum data. Section III describes the proposed technique to characterize a spatial distribution function. In Section IV, we show how the spectrum distribution function can be used to calculate the performance metrics. The simulation results are presented in Section V. Conclusions are drawn in the last section.

This research was sponsored by National Science Foundation, under award

CROWNCOM 2011, June 01-03, Osaka, Japan

DOI 10.4108/icst.crowncom.2011.245923

II. Cooperative Sharing of Raw Spectrum Data

We consider a collection of secondary cognitive nodes that are randomly deployed at different locations in a region of interest. These nodes continuously sense the transmissions by primary nodes and record the noise from primary transmitters for every channel for the entire spectrum band under consideration. We do not focus on any particular sensing technique since there are several techniques that could possibly be used [3], [12], [14]. The simplest form of these detectors can be thought of being energy detectors [14] which have proved to be simple, efficient, and cost-effective. Although any standalone sensing technique can be used to detect the presence of primary nodes on a channel, we are more interested in fusing the power spectral density at any location.

Cooperative sharing of sensed data: Cooperative sharing of sensed spectrum data between secondary transmitters employed in our model is functionally different from conventional Cooperative Spectrum Sensing [2], [13]. Cooperative spectrum sensing is a technique where multiple secondary nodes share their findings (binary decision vectors indicating occupancy of primary nodes on each channel) in a cooperative manner to better detect the presence of primary transmitters and reduce false positives. However, our approach using collaborative sharing is different– we allow the sensing nodes to share raw spectrum data (i.e., power spectral density) instead of just the decision (binary) vectors. Sharing of such raw spectrum data although increases the broadcasting payload, the eventual fusion of all such data minimizes individual detection error and increases interpolation efficiency.

III. Characterizing Spectrum Usage Distribution

The basis for modeling a spatial distribution of spectrum usage is to estimate the activity on every channel. Such estimation of spectrum usage at any arbitrary point from a given set of points is non-trivial. In essence, we seek to define a continuously differentiable two-dimensional interpolation function which passes through all the given irregularly-spaced data points[1].

Let us consider that we have $|\Delta|$ cognitive radio enabled secondary nodes monitoring the spectrum usage and let the co-ordinates of the ith secondary node be δ_i be (x_i, y_i). Also, this node records some data value of z_i (in our case detected energy). Now, given δ such triplets (x_i, y_i, z_i), we seek to find a two dimensional interpolation function $f(x, y) = z$ that will be continuous and differentiable, passing through all the data points i.e., $f(x_i, y_i) = z_i$, and should conform to real life values. Such an interpolation function will allow us to evaluate the spectrum usage (i.e., the data value) at any arbitrary target location say (x_t, y_t).

We start with a basic approach to interpolate values using weighed averages. Let e_q^i be the value of the detected energy at δ_i for channel ch_q . If d_i^t is the Euclidean distance between δ_i and (x_t, y_t), then the estimated received energy in channel ch_q can be interpolated as:

$$\phi_q^t = \frac{\sum_{i=1}^{|\Delta|} (d_i^t)^{-k} e_q^i}{\sum_{i=1}^{|\Delta|} (d_i^t)^{-k}} \tag{1}$$

Here k is the power of the distance weighing factor.

Although this technique of finding expected received energy at an arbitrary point is easy to compute, it overlooks some key aspects: the distance between the data points and the secondary receiver, and the relative positions of the known data points with respect to that receiver. In this regard, we make use of the Shepard's [11] method of interpolation for irregularly spaced data points in a two dimensional region.

A. Distance of data points

Let r be the radius of circle drawn centering (x_t, y_t) and the furthest of the data points being at the edge of the circle. The value of r depends upon choice of (x_t, y_t) and the number of monitoring secondary nodes. Let us define the set $R^t = \{\delta_1, \delta_2, ..., \delta_n\}$ such that $0 \leq d_1^t \leq d_2^t \leq .... \leq d_n^t$ which gives the data points in an ascending order of their distances from (x_t, y_t). As the data points have varying distances from (x_t, y_t), they ought to have a weighing function that reflects the effect of distance of a data point. Such a weighing function dependent on the search radius is given by [11]:

$$p_i^t = \begin{cases} \frac{1}{d_i^t} & if \ 0 < d_i^t \leq \frac{r}{3} \\ \frac{27}{4r}(\frac{d_i^t}{r} - 1)^2 & if \ \frac{r}{3} < d_i^t \leq r \end{cases}$$

The above function is defined to be continuously differentiable over all $d_i^t > 0$. It can easily be argued that more data points will yield a better estimation; however, they will also increase the computational complexity.

Considering the effect of distance of the data points, the estimated received energy value can be modified as:

$$\phi_q^t = \frac{\sum_{\delta_i \in R^t} (p_i^t)^2 e_q^i}{\sum_{\delta_i \in R^t} (p_i^t)^2} \tag{2}$$

However, this interpolation function does not reflect the effect of direction of those data points i.e., the relative angle they make with each other.

B. Direction of data points

In Fig. 1, we see two different orientations of data points 1, 2 and 3 with respect to point (x_t, y_t). In both cases, the distances of the points from (x_t, y_t) are d_1, d_2 and d_3 respectively. In the first orientation, all the points are on the same side of (x_t, y_t) whereas in the second orientation they are on different directions with respect to (x_t, y_t). The disparate spatial orientations in these two cases yield different effects on (x_t, y_t).

[1]We use the monitoring secondary nodes as the data gathering points, hence we also refer to them as 'data points'.

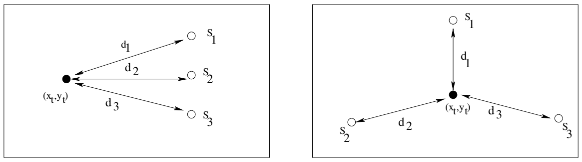


Fig. 1. Different orientations of data points influencing estimation at (x_t, y_t)

The directional weighting term for each selected data point δ_i near (x_t, y_t) is given as

$$a_i^t = \frac{\sum_{\delta_j \in R^t} (p_j^t)[1 - cos\angle\delta_i \ t \ \delta_j]}{\sum_{\delta_j \in R^t} (p_j^t)} \forall j \neq i$$

Now, considering the effect of number, distances, and directions of data points on (x_t, y_t), we define the weighing factor as $w_i^t = (p_i^t)^2(1 + a_i^t)$. It is to be noted that in the directional weighing term a_i^t, the distance weighting factor p_j^t is included in the numerator and the denominator because points near (x_t, y_t) should be more important in shadowing than distant points [11]. Thus, the final interpolated received energy for channel ch_q considering the distance and direction factors is:

$$\phi_q^t = \frac{\sum_{\delta_i \in R^t} w_i^t e_q^i}{\sum_{\delta_i \in R^t} w_i^t} \tag{3}$$

Note, this is the estimated channel usage of a *particular* channel ch_q. To get the values for the entire spectrum range, we simply repeat the computations for all the channels (let there be N channels in the spectrum) concerned. Therefore the estimated spectrum usage at (x_t, y_t) is given as:

$$\Phi^t = \sum_{q=1}^{N} \frac{\sum_{\delta_i \in R^t} w_i^t e_q^i}{\sum_{\delta_i \in R^t} w_i^t} \tag{4}$$

The process of computing Φ^t is shown in Algorithm 1. The interpolation technique meets all the requirements (i.e., defined at every point and is continuous, differentiable) for computing the spectrum usage scenario at the target location (x_t, y_t). With the usage for the entire spectrum known, we apply the energy detection hypothesis principle to find the set of free channels at (x_t, y_t).

The complexity of Algorithm 1 is $O(m^2N)$ where N is the number of channels and m is the number of data points (monitoring secondary nodes) involved in the interpolation. Since the number of monitoring secondary nodes is a constant, the average case complexity is $O(N)$.

IV. Estimating performance metrics

With the spectrum usage distribution known, we demonstrate how this distribution function can be used to predict the performance of a secondary transmitter-receiver pair and also the secondary network as a whole. We assume a network with K secondary nodes exposed to M primary users.

Algorithm 1 Interpolation Algorithm

Find radius r
Find the set of data points $R^t = \{\delta_i | 0 \leq d_1^t \leq ... \leq d_n^t\}$
for all channels q in the spectrum **do**
  for all data points i in the set R^t **do**
    if $0 \leq d_i^t \leq \frac{r}{3}$ **then**
      $p_i^t \Leftarrow \frac{1}{d_i^t}$
    else $\{\frac{r}{3} \leq d_i^t \leq r\}$
      $p_i^t \Leftarrow \frac{27}{r}(\frac{d_i^t}{r} - 1)^2$
    end if
    for all data points k where $i \neq k$ **do**
      Find p_k^t
      $angle_{ik}^t \Leftarrow (x - x_i)(x - x_k) + (y - y_i)(y - y_k)$
      $s_{ik}^t \Leftarrow num_{ik}^t + p_k(1 - angle_{ik}^t)$
      $t_{ik}^t \Leftarrow den_{ik}^t + p_k^t$
    end for
    $a_i^t \Leftarrow \frac{s_{ik}^t}{t_{ik}^t}$
    $w_i^t \Leftarrow (p_i^t)^2(1 + a_i^t)$
    $num_i^t \Leftarrow w_i^t e_q^i + num_i^t$
    $den_i^t \Leftarrow w_i^t + den_i^t$
  end for
  $\phi_q^t \Leftarrow \frac{num_i^t}{den_i^t}$
end for

The interference experienced by a secondary receiver at (x_t, y_t) is due to the primary transmitters as well as other receivers using the same channel. Let us suppose that the interference experienced at (x_t, y_t) from all primary transmitters is ϕ_q^t.

The received signal power at (x_t, y_t) is given by $P|h_q^t|^2$, where P is the transmit power of the corresponding secondary transmitter and h_q^t is the channel gain between the secondary transmitter-receiver pair. The channel gain between the two separated by a distance D_t is given by $h_q^t = \frac{A}{D_t^{\alpha/2}}$; where A is a frequency dependent constant and α is the path loss exponent.

Let I_q^t is the interference the receiver at (x_t, y_t) experiences from other secondary communications in the cell using same channel ch_q. Then,

$$I_q^t = \sum_{\forall j \in \kappa^q} P|h_q^j|^2 \tag{5}$$

where κ^q is the set of all other secondary pairs using channel ch_q.

With the above parameters defined, we show how various performance metrics can be estimated. We discuss four such metrics e.g., channel capacity, spectral efficiency, network throughput, and bit error rate.

Channel Capacity : The channel capacity C_q^t for channel ch_q is calculated using Shannon-Hartley theorem [10].

$$C_q^t = B \ log_2(1 + \frac{P|h_q^t|^2}{\phi_q^t + I_q^t}) \tag{6}$$

where B is the channel bandwidth.

Secondary Network Throughput : Secondary network throughput depends on the number of secondary pairs in a network using the same channel. If the secondary transmitter transmits with power P to a receiver at (x_t, y_t), then the transmission rate considering all other secondary communication is given by

$$\pi_q^t = log(1 + \frac{P|h_q^t|^2}{I_q^t + \phi_q^t + \sigma^2}) \tag{7}$$

where the received signals are corrupted by zero-mean additive white Gaussian noise of power σ^2.

To obtain the network throughput for the channel ch_q, we sum the transmission rates of all the secondary pairs using ch_q as [9]:

$$\Pi_q = \sum_{\forall j \in \kappa^q} \pi_q^t = \sum_{\forall j \in \kappa^q} log(1 + \frac{P|h_q^t|^2}{I_q^t + \phi_q^t + \sigma^2}) \tag{8}$$

Spectral Efficiency : Spectral efficiency provides an indication of how efficiently a bandwidth-limited frequency spectrum is used. Link spectral efficiency in bits/sec/Hz is defined as the net bit-rate that can be achieved by a link per channel bandwidth (Hz). Thus, link spectral efficiency for ch_q between the secondary transmitter and receiver is

$$\xi_q^t = \frac{1}{B} log(1 + \frac{P|h_q^t|^2}{I_q^t + \phi_q^t} + \sigma^2) \tag{9}$$

The system spectral efficiency also in bits/sec/Hz is defined as the maximum throughput, summed over all nodes, divided by the channel bandwidth. Therefore, system spectral efficiency is

$$\Xi_q = \sum_{\forall j \in \kappa^q} \xi_q^c = \frac{1}{B} \sum_{\forall j \in \kappa^q} log(1 + \frac{P|h_q^c|^2}{I_q^c + \phi_q^c} + \sigma^2) \tag{10}$$

Bit error rate : The bit error rate (BER) for secondary communication is dependent on the modulation and coding schemes for communication. For DPSK, BER is given by [6]

$$BER_q^{DPSK} = \frac{1}{2} e^{-\frac{E_b}{\eta}} \tag{11}$$

where E_b is the average energy per bit transmitted and $\eta/2$ is the noise power spectral density. The former is expressed as $E_b = P/\pi_q^t$. Noise power spectral density is given by $\eta = \phi_q^t/B$. Substituting the values in equation (11), we get,

$$BER_q^{DPSK} = \frac{1}{2} e^{-\frac{PB}{\pi_q^t \phi_q^t}} \tag{12}$$

If there are Q bits in a transmitted packet then the packet error rate is given by $PER = 1-(1-BER)^Q$. BER and PER can be calculated for any modulation scheme.

V. Simulation Model and Results

We conducted extensive simulation experiments to model the spatial distribution of channel usage and verify its applicability to predict important performances metrics. We use a grid size of 100x100 square units and to emulate primary behavior, we use the model and the range of values from the real-world spectrum data archive of RWTH Mobnets [7]. We consider 50-1000 channels of bandwidth 2MHz-100 kHz each in the 2.4 GHz ISM band. The received power at any receiver from primary nodes was varied from -45 dBm to -80 dBm. We seek to evaluate system performance in three fronts: i) we focus on the channel usage distribution, ii) we measure the correctness of the predicted channel occupancy vectors, and iii) we predict the nature of some key performance metrics.

Figures 2(a) and 2(b), show the power spectral density of channel usage for one of the 1000 channels. Locations of the data gathering points used for interpolation are shown in darker shades. Expectedly, the surface plots become increasingly accurate as the number of data points increases; however there is a trade-off between the accuracy of estimation and complexity of calculation.

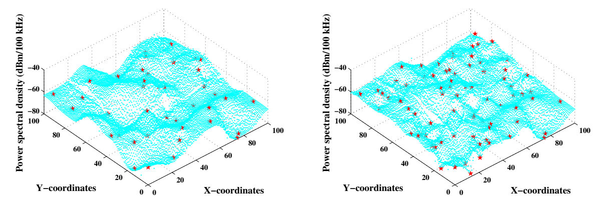
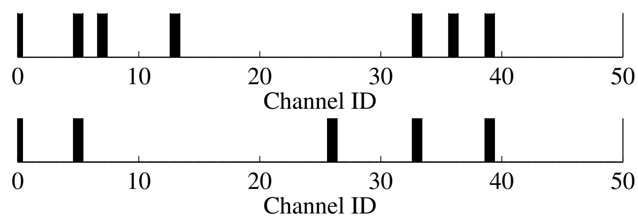
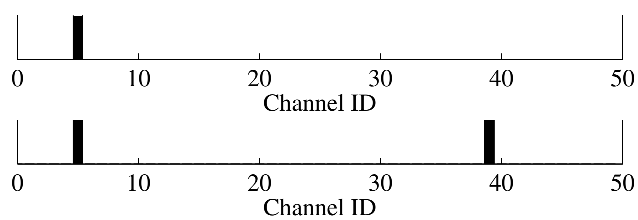


Fig. 2. Power spectral density of a channel with 40 and 80 sensing locations



(a) Estimated spectrum occupancy vectors

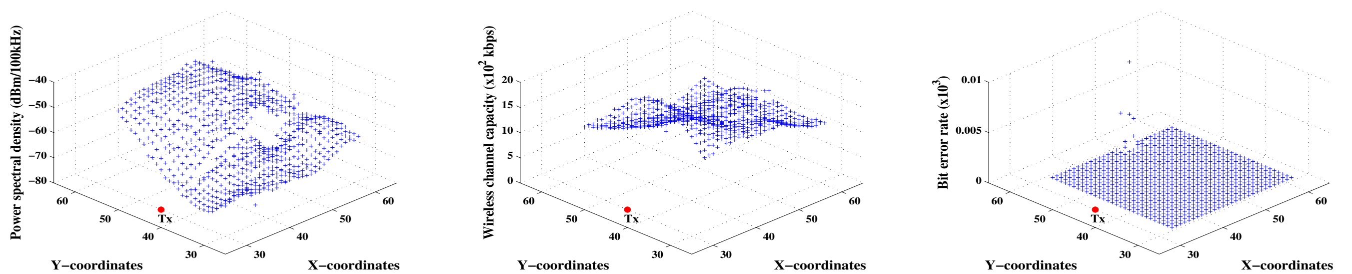


(b) Actual spectrum occupancy vectors

Fig. 3. Spectrum occupancy vectors for two different scenarios

Fig. 3 shows the estimated and actual channel occupancy vectors of two secondary receivers in two different network scenarios for 50 channels. The two scenarios are created with different density and orientation of primary and secondary nodes but with similar transmission characteristics. The actual spectrum occupancy vectors are computed by received signal strength calculation at the secondary receiver for all interfering primary transmissions. It can be seen that in both the cases the prediction of channel occupancy is highly accurate with 6 and 3 false predictions respectively. It is important to note that none of these variations are false negatives which is essential for secondary communication. Our results show that on an average the probability of false positive is as low as 0.06.

Figure 5 depicts the channel capacity for all 1000 channels at an arbitrary receiver location. This result can be exploited by a transmitter-receiver pair in selecting the better channels for communication when multiple options are available.



(a) (b) (c)

Fig. 4. Spatial distribution of estimated noise, channel capacity, and bit error rate

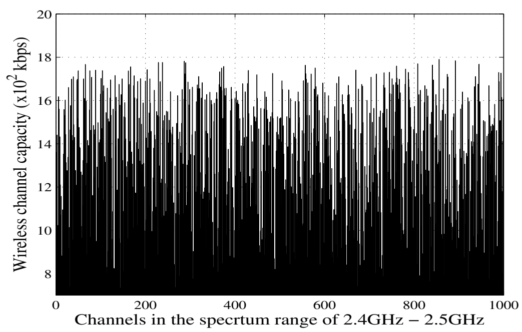
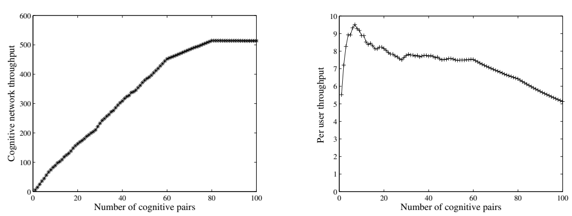


Fig. 5. Channel capacity for different channels in the spectrum band



(a) (b)

Fig. 6. (a) System throughput (b) Per-user throughput

Figures 4(a), 4(b), and 4(c) show the received power from primaries, channel capacity, and bit error rate on the same channel, for the same geographic region. The transmitter located at (30,45) is assumed to transmit at 1mW. The region shown in the plots are the potential locations for the receivers. Figure 4(b) shows the spatial distribution of channel capacity calculated using equation (6). As expected, the channel capacity gradually decreases when the receiver moves away from the transmitter. Figure 4(c) shows the spatial characteristic of estimated bit error rate from equation (11) in the same region. Most of the region showing zero bit error rate, actually have very small non-zero value. It is to be noted that channel capacity is minimum and bit error rate is maximum in regions where noise from primary is higher.

In figures 6(a) and 6(b), system throughput and per-user throughput are shown. The nature of system throughput is similar to a conventional wireless network hitting a plateau after a certain point. Interestingly the per-user throughput has a peak value around 10 secondary pairs and then decreases with more secondary pairs.

VI. Conclusions

In this paper, we use Shepard's interpolation technique to estimate the spectrum usage distribution function over space based on the distributed sensing and sharing by secondary nodes in a dynamic spectrum access network. We demonstrate how such estimation can help in predicting channel performance metrics like channel capacity, spectral efficiency, network throughput, and bit error rate. Through simulation experiments, we validate the correctness of our prediction and show how to compute the distribution of these metrics at any arbitrary location which can potentially be used for network installation and candidate channel selection.

References

[1] Akyildiz, I. F.; Lee, W. Y.; Vuran, M.C.; Mohanty, S.; , "NeXt Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey," Elseview Journal of Computer Networks, Vol. 50, pp. 2127-2159, September 2006

[2] Ganesan, G.; Li Ye; , "Cooperative Spectrum Sensing in Cognitive Radio, Part II: Multiuser Networks," Wireless Communications, IEEE Transactions on , vol.6, no.6, pp.2214-2222, June 2007

[3] Gardner, W.A.; , "Signal interception: a unifying theoretical framework for feature detection," Communications, IEEE Transactions on , vol.36, no.8, pp.897-906, Aug 1988

[4] Li, Yun; Quang, T.T.; Kawahara, Yoshihiro; Asami, Tohru; Kusunoki, Masanori; , "Building a spectrum map for future cognitive radio technology," Cognitive radio networks, 2009. CoRoNet 2009. ACM workshop on , pp.1-6, 2009

[5] McHenry, M. A.; , "NSF Spectrum Occupancy Measurements Project Summary," 2005

[6] Proakis, J. G.; , *Digital Communications*, 4th ed. McGraw-Hill, 2001

[7] Riihijarvi, J.; Mahonen, P.; Wellens, M.; Gordziel, M.; , "Characterization and modelling of spectrum for dynamic spectrum access with spatial statistics and random fields," Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on , vol., no., pp.1-6, 15-18 Sept. 2008

[8] Riihijarvi, J.; Mahonen, P.; , "Exploiting Spatial Statistics of Primary and Secondary Users towards Improved Cognitive Radio Networks," Cognitive Radio Oriented Wireless Networks and Communications, 2008. CrownCom 2008. 3rd International Conference on , vol., no., pp.1-7, 15-17 May 2008

[9] Sang-Woon Jeon; Devroye, N.; Mai Vu; Sae-Young Chung; Tarokh, V.; , "Cognitive networks achieve throughput scaling of a homogeneous network," Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2009. WiOPT 2009. 7th International Symposium on , vol., no., pp.1-5, 23-27 June 2009

[10] Shannon, C. E.; , "A mathematical theory of communication," Bell Syst. Tech. J., vol. 27, pp. 379-423, 623-656, 1948

[11] Shepard, D.; , "A Two-Dimensional Interpolation Function for Irregularly-Spaced Data," *ACM National Conference*, 1968

[12] Tian, Zhi; Giannakis, Georgios B.; , "A Wavelet Approach to Wideband Spectrum Sensing for Cognitive Radios," Cognitive Radio Oriented Wireless Networks and Communications, 2006. 1st International Conference on , vol., no., pp.1-5, 8-10 June 2006

[13] Unnikrishnan, J.; Veeravalli, V.V.; , "Cooperative Sensing for Primary Detection in Cognitive Radio," Selected Topics in Signal Processing, IEEE Journal of , vol.2, no.1, pp.18-27, Feb. 2008

[14] Urkowitz, H.; , "Energy detection of unknown deterministic signals," Proceedings of the IEEE , vol.55, no.4, pp. 523- 531, April 1967