Spectrum Map and its application in Resource Management in Cognitive Radio Networks

Saptarshi Debroy, Shameek Bhattacharjee and Mainak Chatterjee

Abstract—Measurements on radio spectrum usage have revealed an abundance of under-utilized bands of spectrum that belong to primary (licensed) networks. Prior knowledge about the occupancy of such bands and the expected achievable performance on those bands can help secondary (unlicensed) networks to devise effective strategies to improve utilization. Such prior spatio-temporal spectrum usage statistics can either be obtained from a database that is maintained by the primary networks or could be measured by customized sensors deployed by the secondary networks.

In this paper, we use Shepard's interpolation technique to estimate a spatial distribution of spectrum usage over a region of interest, which we call the *spectrum map*. The interpolation is achieved by intelligently fusing the data shared by the the secondary nodes considering their mutual distances and spatial orientation with each other. The obtained map is a two-dimensional interpolation function that is continuously differentiable and passes through all the spectrum usage values recorded at arbitrary locations; thus providing a reference for primary occupancy in that region. For determining the optimal locations for sensing primary activity, we use an iterative clustering technique that uses tree structured vector quantization. We use the spectrum map to estimate different radio and network performance metrics like channel capacity, network throughput, and spectral efficiency. As a comprehensive case study, we demonstrate how the spectrum map can be used for efficient resource allocation in TV white space. In particular, we consider an IEEE 802.22 based WRAN and show how the rendezvous probability can be improved for better radio resource allocation, thereby increasing the secondary spectrum utilization.

Index Terms—Cognitive Radio Network, Spectrum Map, Cooperative Spectrum Sensing, Resource Allocation, IEEE 802.22 WRAN

1 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 (i.e., licensed) nodes and opportunistically access the unused or under-utilized licensed bands of primaries [1]. However, the most important regulatory aspect of these networks is the fact 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. Due to the temporal and spatial fleetingness of primary spectrum occupancy, such reactive nature 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 the region of interest. In order to facilitate such preemptive primary usage knowledge acquisition, the recent ruling by FCC [2] necessitates the need for networks to create, manage, and refer to spectrum databases for efficient secondary access without providing any concrete framework. This has opened new discussions on the design, implementation techniques and capabilities of such spectrum databases.

Spectrum databases are usually manifested either through mandating primary transmitters to report their transmission activities to a central authority, or through building spectrum usage maps (e.g., TV whitespace database [3]). Spectrum usage maps give signal strength from primary transmitters on different channels in a particular geographical region. The stringent policy enforcement of reporting spectrum activities has had some roadblocks in terms of the underlying legal and policy issues, which are suspected to pose a barrier for wide-adoption of dynamic spectrum access technologies. Thus, primary usage prediction schemes and models have garnered much traction in recent times. These schemes mostly deal with either modeling spectrum utilization using statistical models [4] or using real-world measurement data to predict usage patterns [5]. Such mechanisms are mostly specific to primary network types (such as, TV, cellular etc.) and lack extensibility for generic primary networks. Thus, structuring design principles and implementation specifics for predicting spectrum usage at *any arbitrary location* for *any generic primary network* still remains a challenge. This work is motivated by the idea of radio frequency (RF) cartography, which suggests developing a *map* of the RF field in space by sampling this field at particular locations, which could be done by receivers or dedicated RF sensors. These samples are then interpolated or extrapolated to yield an estimate of the RF field at every point in the space. Although our work is independent of the underlying primary networks and secondary sensor devices, our solutions do consider their applicability to solve practical secondary networking problems.

In this paper, we create a *spectrum map* by defining a spatial distribution function for spectrum utilization. This map works as a reference database to predict the spectrum usage at any arbitrary location. We argue that such prediction can be achieved by fusing the information gathered by the various stationary secondary nodes operating at different locations. Therefore, we use a collection of such nodes at different locations that monitor the spectrum usage in a distributed manner and share their findings with others. By fusing the raw spectrum usage data from these nodes, we show how an interpolation function can be used to construct a continuously differentiable distribution function which governs the estimation of the spectrum utilization

Debroy is with the Department of Computer Science at the University of Missouri, Columbia, MO. Bhattacharjee is with the Department of Computer Science at the Missouri University of Science and Technology, Rolla, MO. Chatterjee is with the Department of Electrical Engineering and Computer Science at theUniversity of Central Florida, Orlando, FL. emails: debroysa@missouri.edu, bhattacharjeesh@mst.edu, mainak@eecs.ucf.edu.

at any location. The proposed scheme is independent of the outdoor fading and shadowing environment; only the sampling and reporting frequencies may vary depending on the environment. To find optimal locations for these monitoring nodes, we use an *iterative clustering* technique using tree structured vector quantization (TSVQ).

We evaluate our proposition by measuring the accuracy of estimation and also characterizing the nature of the spectrum map. We emulate an environment using the real-world spectrum data of RWTH Mobnets [6] and replicating their transmission pattern. We observe that the projected spectrum map mimics the real nature of spectral power density. We also demonstrate the effectiveness of iterative clustering in finding the sensing locations over random selection through simulation experiments. We use the prior knowledge of spectrum utilization to predict radio and network performance. We demonstrate how performance metrics like the channel capacity, spectrum efficiency, and system throughput, can be estimated for a secondary transmitter-receiver pair at any secondary network location. We also simulate the nature of these performance metrics.

We also apply our spectrum map for a secondary network use case: intra-cell channel allocation for IEEE 802.22 networks [7], [8] in TV white space. We show how using spectrum map, the rendezvous probability between secondary users increases in the absence of a common control cannel. Our spectrum map inspired channel allocation algorithm achieves close to perfect allocation and more than 50% utilization of free channels even in the worst case scenario with no intra-cell spatial reuse.

The rest of the paper is organized as follows. In Section 2, we discuss the notable prior work in this area. In Section 3, we show how the spectrum map is built with data obtained from arbitrary sensing locations. In Section 4, we use describe the iterative clustering technique for finding the ideal sensing locations . In Section 5, spectrum map-aided prediction of performances metrics are discussed. In Section 6, we present the IEEE 802.22 network channel allocation case study. Conclusions are drawn in the last section.

2 Related Work

Prior work in modeling the spectrum usage involves deriving the distribution of spectrum utilization in a primary network using both theoretical models and real-world data logs [4], [5], [6], [9], [10], [11], [12]. One of the earliest work in spectrum measurements was reported by NSF [9] where it was concluded that less than 1% of the spectrum opportunities, both in frequency and time, were utilized at the place of measurement. Authors in [5] measured the spectrum usage at different locations in Tokyo city and created 3-dimensional plots that showed the temporal distribution of frequency usage near the monitored points. In [4], the authors observed that the location distribution of primary TV transmitters in USA and Europe closely follow the Poisson model. A frequency distribution model to emulate the nature of noise from a primary transmitter on different channels was presented in [6]. The most notable study was carried out by Harrold et. al [10] where three years of continuous measurements for the city of Chicago were presented. Some notable work has also been done on geolocation databases. Authors in [13] highlight the benefits of geolocation database technology and Murty et. al in [14] went closest to implementing a database-driven secondary network for TV white space (TVWS).

Although most of the above real-world data based learning schemes have valid contribution in the area of primary usage estimation, as with most learning algorithms, there is a 'time to learn'. On a macro time scale where events occur on large time durations (e.g., limited spectrum utilization during weekends in a downtown area), such learning helps. However, research has shown that in order to capture the instantaneous spectrum occupancy coupled with the channel fading conditions and use that for scheduling at the MAC layer every 10 or 20 ms (4G networks), learning does not yield optimal results. Moreover, some of the primary traffic has been proven to obey Poisson distribution [15], [16]; making learning ineffective in such cases. Thus, decisions based on the ground truth obtained by the sensors are proving to be more effective.

Works based on such ground truth observation mostly propose interpolation or extrapolation functions that estimate new data points given a discrete set of observed data points. The most common in this category are linear, piecewise linear, polynomial, splines, multi-variant, Whittaker-Shannon, nearest neighbor, and Hermite interpolation based techniques. As far as constructing spectrum maps using interpolation are concerned, there have been a few notable efforts. In [17], a spline based method for field estimation was introduced that uses collaborative and adaptive sensing. In [18], an algorithm was developed for sensor placement on a map that is generated by Kriging and LiVE techniques. In [19], a linear unbiased predictor was developed to predict the value at any point on the map's surface. The reliability of the bounds of interpolation techniques that are inverse distance weighting based were analyzed in [20].

Although the above mentioned techniques allow prediction of spectrum usage at unknown locations, there is still little understanding on how to build a mathematical function that captures the spatial distribution for the spectrum utilization. Moreover, for these techniques, construction of a new data point from known ones is computationally intensive. Sometimes, convergence is slow for techniques that resort to numerical methods. For example, the Kriging method and its variants are among the most popular methods. However, the basic Kriging approaches suffer from poor scalability, requiring up to $O(n^3)$ operations for prediction at a single location, where n is the number of measurement points from which data is obtained [21]. Furthermore, the success of most of these proposed interpolation algorithms is restricted by weaker power levels as the efficiency of the interpolation scheme is a function of received signal strength. Thus, such schemes have limitations of their applicability for different types of primary networks. Therefore, there is a requirement for a generic spectrum usage estimation technique that is computationally lightweight and equally efficient for different types of primary networks (e.g., TVWS, cellular network, 4G etc.) with different transmission characteristics.

3 Building the Spectrum Map

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]. It can be noted that the estimation method is identical for both distributed and centralized sharing system.

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

3.1 Sharing raw spectrum data

The concept of sharing and fusing spectrum data among cognitive radio enabled devices is not new. The process of sharing channel usage information among different cognitive radio enabled devices to predict the existence of primary user in a channel is called cooperative sensing [22] or collaborative sensing [23], [24]. The nature of such cooperation can be *distributed* or *centralized* depending upon the location of the fusion center. Our proposed scheme borrows the essence of cooperative spectrum sensing by allowing the delegation secondary nodes to share their sensed spectrum data periodically with the base station or a fusion node. However, unlike conventional cooperative spectrum sensing where the devices share decision vectors (representing occupancy of channels), the sensing nodes share their raw spectrum data which are later fused to estimate spectrum usage at unknown points.

Although sharing such raw spectrum data requires a higher data rate than sharing binary vectors, the overhead and its impact to the secondary communication are negligible. Let us assume that there are N channels, then N bits are shared among the sensor nodes. Whereas, our proposed scheme requires $1byte/8\ bits$ to represent a power spectral density with 1 bit for the sign (positive/negative) and 7 bits for the value. Here we assume that $-128\ dBm/kHz$ is the minimum power spectral density required for most of the primary network operation where secondary access is permitted. Thus, in our scheme, the sharing overhead is $8 \times Nbits$. Now if we assume a particular primary network, such as TVWS with 50 channels, then the system requires 400 bits for raw spectrum data sharing. Even if we assume a faster than normal sensing frequency of 60 seconds, i.e., sensor nodes sense the spectrum every 60 seconds and share their raw spectrum data, the required data rate to support such mechanism is merely $400/60 \approx 7\ bps$. Such data rate can be supported by any secondary network with minimal overhead.

3.2 Interpolation function

An ideal spectrum map should be in 3-D as the devices accessing and using spectrum are not expected to be on the same height from the reference (sea) level. Though computing a 3-D map is theoretically possible using the extrapolation function, the number of computations will be one order of magnitude higher. For example, instead of computing 100×100 points, we will have to compute $100 \times 100 \times H$. Since this height H is expected to be very small compared to the distances in the planar 2-D region (i.e., the TV towers are separated by 5-10 KMs), we can project the 3-D map on to a 2-D map without any loss of generality. Thus, we consider a 2-D spectrum map.

Let us assume that we have $|\Delta|$ cognitive radio enabled secondary nodes monitoring the spectrum usage and let the co-ordinates of the ith secondary node δ_i be (x_i, y_i). Also, this node records some data value of z_i. The data value can correspond to one of the many radio parameters like SNR, duty cycle, or detected energy for a particular channel that the node is sensing. Now, given $|\Delta|$ such triplets (x_i, y_i, z_i), we seek to find a two dimensional interpolation function $f(x, y) = z$ that will be continuously 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 by 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 the 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 [25] method of interpolation for irregularly spaced data points in a two dimensional region.

3.3 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 $R^t = \{\delta_1, \delta_2, \cdots, \delta_n\}$ such that $0 \leq d_1^t \leq d_2^t \leq \cdots \leq d_n^t$ 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 their distance from the target. Such a weighing function dependent on the search radius is given by [25]:

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

The above function is defined to be continuously differentiable for 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 captive the effect of spatial orientation of the data points i.e., the relative angle they make with each other.

3.4 Direction of data points

First, let us demonstrate the effect of direction of data points with the help of an example. In Fig. 1, we see two different orientations of data points 1, 2 and 3 with respect to target (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). Thus, we consider *all* the possible set of angles that each data point makes with all other data points.

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

$$a_i^t = \frac{\sum_{\delta_j \in R^t} (p_j^t)[1 - cos\angle \delta_i\ t\ \delta_j]}{\sum_{\delta_j \in R^t} (p_j^t)} \forall j \neq i \tag{3}$$

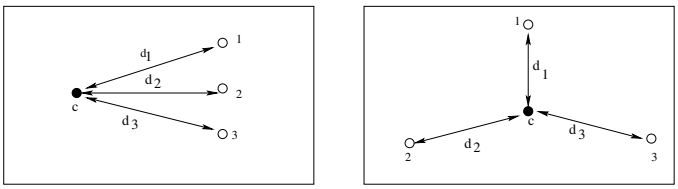


Fig. 1. Different orientations of data points influencing estimation at (x_t, y_t)

Equation (3) signifies when two sensors are on the same line, the algorithm only considers the contribution from the closer sensor. When the sensors are at right angles, the algorithm weights each of them equally, and for angles in between, the weighting follows a cosine.

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 [25]. Thus, the final interpolated received energy for channel ch_q considering the distance and direction factors is:

$$\phi_q^t = \frac{\sum_{\delta_i \in R^t} w_i^t e_q^i}{\sum_{\delta_i \in R^t} w_i^t} \tag{4}$$

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:

$$\Phi^t = [\phi_1^t, \phi_2^t, \cdots, \phi_N^t] \tag{5}$$

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 continuously differentiable) for computing the spectrum usage scenario at the target location (x_t, y_t). With the usage for the entire spectrum known, we can find the set of free channels at (x_t, y_t) for which we need to apply hypothesis analysis on every channel.

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 $O(N)$.

Our choice for using Shepard's method for constructing the spectrum map over other multi-dimensional interpolation functions are because of the following reasons:

1) Shepard's method is inherently suited for multi-dimensional *spatial* data.
2) It involves simple mathematical operations with low computational complexity.
3) Only a few data points are to be considered in R^2.
4) Time independence obviates effects of fading which can be countered with suitable sensing and collection frequency.

For building the spectrum map, we do not consider sensors reporting imprecise readings located behind obstacles. Such phenomenon can be corrected by using other known sensor localization techniques where the distance d to a sensor is corrected as d' [26]. Such discussions are beyond the scope of this paper.

Algorithm 1: Interpolation algorithm

Data: Set of sensors/data points $S^i = \{\delta_i\}$
Data: Target point (x_t, y_t)
Data: Sensor radius r
Data: Power spectral density e_q^i of data point i on channel q
Result: Estimated power spectral density in a region for each channel
for *all channels* q **do**
  for *all data points* i *in the set* S^i **do**
    if $0 \leq d_i^t \leq \frac{r}{3}$ **then**
      $disFact_i^t \Leftarrow \frac{1}{d_i^t}$
    else if $\frac{r}{3} < d_i^t \leq r$ **then**
      $disFact_i^t \Leftarrow \frac{27}{r}(\frac{d_i^t}{r} - 1)^2$
    for *all data points* $k \; \forall \; k \neq i$ **do**
      $angFact_i^t \Leftarrow angFact_i^t + disFact_i^t \times ((x_t - x_i)(x_t - x_k) + (y_t - y_i)(y_t - y_k))/disFact_i^t$
    end
    $fnlWght_i^t \Leftarrow (disFact_i^t)^2(1 + angFact_i^t)$
    $powSpecDnst_q^t \Leftarrow powSpecDnst_q^t + \frac{fnlWght_i^t \times e_q^i}{fnlWght_i^t}$
  end
end

3.5 Spectrum Map construction from real data

Now we demonstrate the spatial distribution of spectrum usage and measure the accuracy of estimation. We create a grid of 100x100 units and emulate primary behavior using real-world spectrum data archive of RWTH Mobnets used in [6]. The experiment performed considers 1000 channels of bandwidth 100 kHz each in the 2.4 GHz ISM band in Germany. The received power at any receiver from primary nodes was varied from -45 dBm to -80 dBm.

Figures 2 and 3, show the power spectral density of primary channel usage for one of the 1000 channels using 40 and 80 data points respectively. Locations of the data points used for interpolation are shown in darker shades. Expectedly, the surface plots become increasingly accurate as the number of data points increases. Similar results are observed for all the channels. Although it seems beneficial to use as many data points as possible to get more accurate spectrum map, there is a trade-off between the accuracy of estimation and complexity of such estimation. Such trade-off is dependent on other physical and environmental factors such as type of primary network (threshold signal strength to operate, probability of primary activity), accuracy of sensors involved in detecting signals, location of the sensors, physical environment in terms of terrain affecting fading and shadowing. Moreover, this trade-off is not only dependent on the number of sensors but also on the frequency of periodic sensing, making such trade-off modeling non-trivial and to our opinion, beyond the scope of the paper. However, we make an attempt to perform an experimental trade-off quantification in Section 4 for a specific approach on data point placement.

It is to be noted that spectrum usage reports from sensors located in heavily faded regions will not be very effective - even more so when there are few sensors. We argue that, there will be enough sensors to average out any bad readings. Moreover, the deployment of the sensors are not ad hoc. Rather, in our proposed method, their locations are determined considering topological effects and primary locations which will be discussed in Section 4.

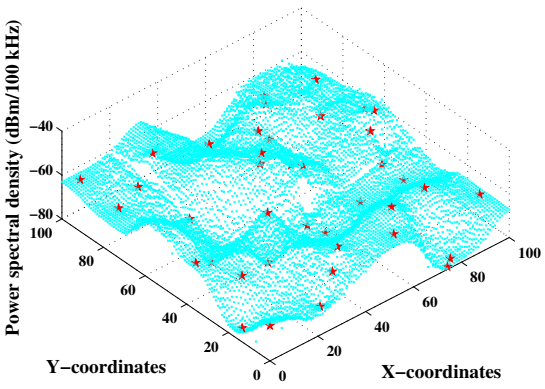


Fig. 2. Estimated power spectral density with 40 sensing locations

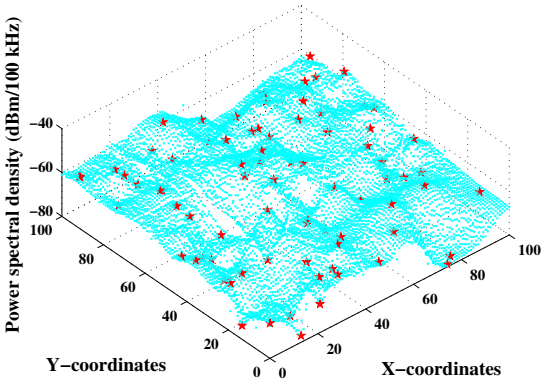


Fig. 3. Estimated power spectral density with 80 sensing locations

Once the estimates of channel usage for all the channels are known, we apply the energy detection hypothesis to decide the channel occupancy from the interpolated values (detected energy) and find the set of free channels at (x_t, y_t). Although the proposed interpolation scheme successfully estimates the received signal strength of a region of interest, there are certain assumptions and limitations in terms of the techniques' applicability. The are:

- The proposed model does not include channel variables for phenomena such as shadowing or fading to keep the algorithmic complexity low and lightweight to make it easily implementable on low-cost and power-constraint sensor devices.
- The proposed model works best in a slow fading and low shadowing environment. For fast fading and shadowing environment, the estimation accuracy decreases; however, the estimation is always conservative. Thus the estimation does not generate false negative in terms of channel occupancy.
- Since the transmitter and receiver heights are very small compared to the distances in the planar 2-dimensional region, we can project an actual 3-dimensional spectrum map on to a 2-dimensional map with minimal effect.

4 Selection of Ideal Sensing Locations

We argue that the accuracy of the Shepard's interpolation technique can be improved by selecting the data points at strategic locations. Be it centralized or distributed, either with a central repository or multiple fusion centers, strategically placing the sensing locations that depend on the primary locations have a profound effect on the construction of the Spectrum Map.

Thus, we propose an *iterative clustering* technique using tree structured vector quantization (TSVQ) to find these representation points, i.e., sensing locations. *Vector quantization* (VQ) is a powerful data compression technique where an ordered set of real numbers is quantized. The idea of such quantization is to find $|\Delta|$ representation points (distinct vectors) from a large set of vectors so that the average distortion is minimized. With iterative clustering, the size of the representation points grows from 1 to the desired value, $|\Delta|$. Given a set of primary transmitter locations, their centroid is the ideal representation point when $|\Delta| = 1$, as the sum of the Euclidean distances to all the primary transmitters is minimum at the centroid. We show an illustrated example in Fig. 4(a) where S_1 denotes the centroid of the primary transmitters represented as hollow dots creating a single cluster C_1.

4.1 Iterative clustering

With further iterations, S_1 is split into two points, S_1 and $S_1 + \epsilon$, where ϵ is a small Euclidean distance. Each of the primary locations is grouped on to the closer of the two representation points thereby creating two clusters with two representation points. The centroids of these two clusters are determined and the representation points, S_1 and $S_1 + \epsilon$, are updated with the centroids' position creating new representation points S_1 and S_2. Clustering is performed again on these two representation points till the desired size of $|\Delta|$ is achieved. Fig 4(b) shows the scenario with two such points corresponding to the two clusters C_1 and C_2. Further splitting into four and iteratively applying the clustering algorithm, we obtain the four centroids along with the four clusters as shown in Fig. 4(c). It is to be noted that such binary splitting yields 2^n clusters.

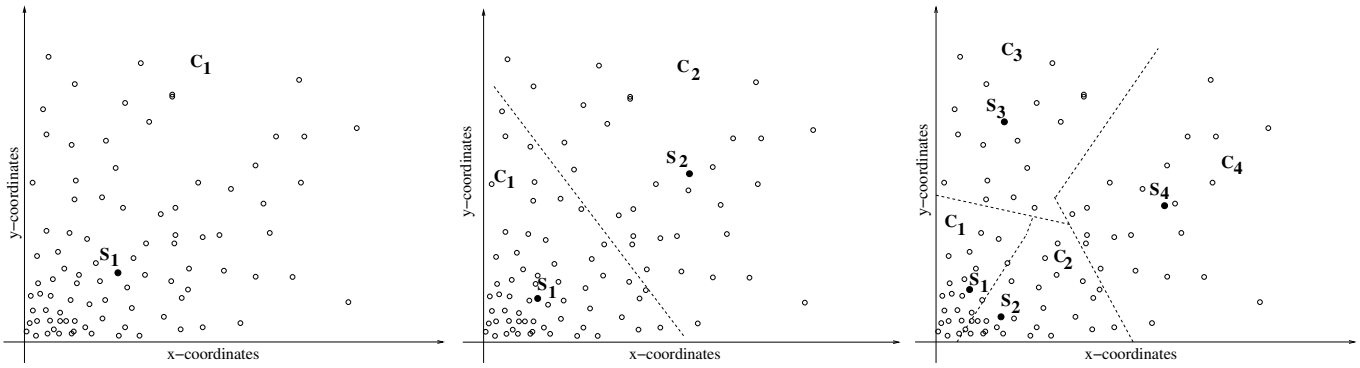
4.2 Choice of number of representation points ($|\Delta|$)

The requirement of $|\Delta|$ in the given scenario differs from the choice of $|\Delta|$ in conventional VQ. In conventional VQ, the main goal of designing a quantizer is to find the representation points and the cluster such that the average distortion is minimized for a fixed number of such points. However, in a cognitive radio network, the choice of $|\Delta|$ is governed by variables like the type of secondary network, number of primary transmitters and density of secondary nodes. More data points obviously result more accurate spectrum but with more sharing and complexity overhead. In next section, we shed more light on the accuracy of estimation with choice of $|\Delta|$.

4.3 Performance gains using Iterative Clustering

We conduct simulation experiments to demonstrate the benefits of using iterative clustering technique over deterministic and random selection of data points. The simulation is conducted using C and the figures are generated using MATLAB. A fixed number of primary users (70 in this case) are deployed randomly over a 100x100 area region. A certain number of primaries are chosen randomly to be active and assigned powers ranging from high to low based on various primary network's transmission patterns using RWTH Mobnets data [11].

In the first set of experiments, we compare the actual RSS values (calculated using a highly sophisticated and



(a) One cluster with a single representation point

(b) Two clusters with two points

(c) Four clusters with four points

Fig. 4. Iterative clustering and corresponding representation points

widely used path-loss model proposed in [27]) with the estimated values (calculated using our interpolation scheme). We choose the data points (representation points) according to iterative clustering and vary the number of data points from 20 to 60. We then omit 3 data points from the group of data points and reconstruct the map using our proposed interpolation scheme. The results of actual and estimated RSS at the missing data point locations are presented in Table 1. The same experiment is run 3 different times with different numbers of data points used each time to construct the map. We observe that in terms of dBm, the error is negligible and with more data points the accuracy increases.

In the next set of results, we use the number of mismatches as the metric to evaluate the benefits of using iterative clustering for data point location selection. A mismatch is defined for each channel where the the actual and the estimated binary vectors do not match. Fig. 5 depicts the effectiveness of iterative clustering in estimating the spectrum map for varied primary activity, i.e., the probability of primary being active on a particular channel. This probability is calculated by the number of channels where primaries are active over the total number of channels (100 in this case). First, we find the actual RSS in the secondary locations for all the channels and compare the values with a threshold to find actual binary decision vectors. Then, we use iterative clustering technique to find the suitable data point locations for Shepard's interpolation. From the estimated RSS, we find the estimated binary decision vectors using the same threshold. We observe that even with high primary activity, the iterative clustering performs fairly well in terms of mismatches when there are more than 16 data points selected for interpolation. For low primary activity, the average mismatch is less than 10% for any number of selected data points.

In Fig. 6, we compare the performance of the proposed iterative clustering technique with random and deterministic selection of data points. In deterministic selection, data points are chosen uniformly in a grid-like orientation to cover the entire region without considering the locations of the primaries. The performance comparison is made for low primary activity of value 0.3. From the figure, we observe that even for low primary activity, iterative clustering performs better than random and deterministic selections in terms of average mismatches for any number of selected data points for interpolation. We also observe that, especially for random selection the number of mismatches vary more than iterative and deterministic scenarios proving random selection being the most imprudent choice.

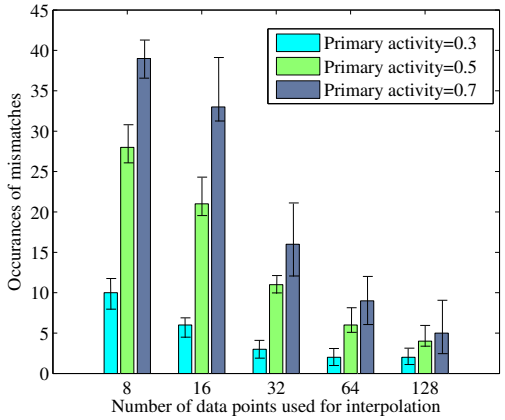


Fig. 5. Performance of iterative clustering for different primary activities

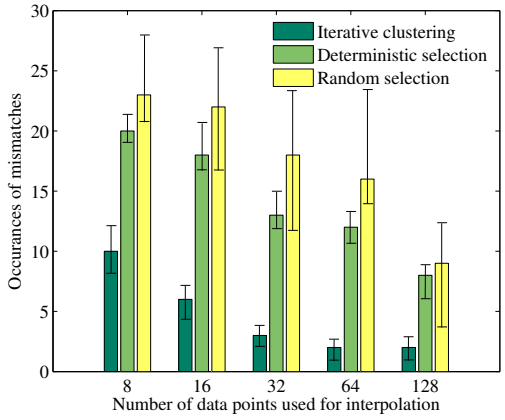


Fig. 6. Comparison of different data point selection strategies

We also observe that accuracy of estimation is not only a function of location of data points selected, but also on the number of data points and primacy activity. This makes the theoretical quantification of trade-offs between accuracy and complexity of estimation is non-trivial and in our opinion, an independent avenue of research altogether.

TABLE 1
Comparison between actual and estimated power spectral density (dBm/100 kHz) at missing data points for different number of data points

Missing data point locations	20 data points		40 data points		60 data points	
	Actual	Estimated	Actual	Estimated	Actual	Estimated
A	-42.3	-42.66	-42.3	-42.306	-42.3	-42.32
B	2.7	2.75	2.7	2.707	2.7	2.712
C	-28.6	-27.9	-28.6	-28.4	-28.6	-28.4

5 Predicting Channel Performance Metrics

We extend our spectrum usage prediction model to predict the performance of a secondary transmitter-receiver pair and also the secondary network as a whole. We assume a scenario where K secondary nodes are exposed to M primary users. A secondary receiver is interfered by potentially all transmitters on all possible channels.

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 receiver 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 another secondary communication in the cell using same channel ch_q. Then

$$I_q^t = \sum_{\forall j \in \kappa^q} P|h_q^j|^2 \tag{6}$$

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

With the above parameters defined, we show how the we can calculate critical channel performance metrics, such as, channel capacity, spectral efficiency, and secondary throughput. With each performance metric we provide simulation results of the corresponding metric using the same experimental environment described in Section 3.5.

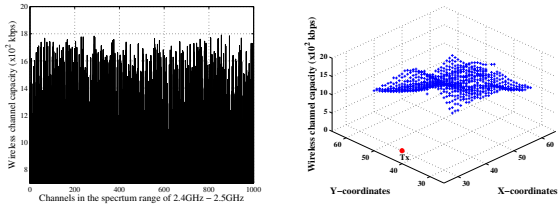
5.1 Channel Capacity

We use Shannon-Hartley's capacity model for a band-limited channel with additive white Gaussian noise (AWGN) [28]. Thus, we express the channel capacity C_q^t for channel ch_q as

$$C_q^t = B\ log_2(1 + \frac{P|h_q^t|^2}{\varphi_q^t + I_q^t}) \tag{7}$$

where B is the channel bandwidth. Note that the channel capacity is dependent on the channel used by the transmitter-receiver pair as the interference experienced from primary transmitters and other secondary communication are different on different channels.

In Fig. 7(a), we show the channel capacity characteristics for an arbitrary receiver location calculated using Equation (9). The figure shows different channel capacity values for 1000 channels of 100KHz each in the 2.4 GHz ISM band. Such representation of channel capacity can be exploited by a particular transmitter-receiver pair for selecting better channels in terms of achievable capacity when multiple such free channels are available for communication. Through another 3D representation shown in Fig. 7(b), we present the channel capacity values (calculated using Eqn. (7)) of a particular channel and for a particular transmitter and multiple potential receiver locations. Such representation can be further exploited for: a) optimal allocation of a particular channel to the best contending secondaries, and b) optimal location selection of a secondary node (in cases where secondary node installation is pre-planned) for statistically empty channels. Both approaches shown in figures 7(a) and 7(b) are highly effective for better secondary utilization. Later in Section 6.5 we will show how this expected channel capacity information can be utilized for efficient channel allocation in an IEEE 802.22 network.



(a) Channel capacity for different channels (b) Channel capacity of a region for same channel

Fig. 7. Channel capacity characteristics

5.2 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 receiver at (x_t, y_t), then the transmission rate considering all other secondary communication is given by

$$\pi_q^t = log(1 + \frac{P|h_q^t|^2}{I_q^t + \varphi_q^t + \sigma^2}) \tag{8}$$

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

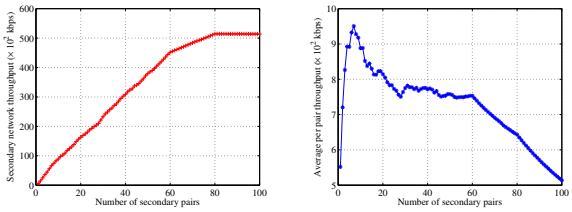
To obtain the network throughput for the channel ch_q, we sum the transmission rates of all the secondary pairs using ch_q as [29]:

$$\Pi_q = \sum_{\forall j \in \kappa^q} \pi_q^t = \sum_{\forall j \in \kappa^q} log(1 + \frac{P|h_q^t|^2}{I_q^t + \varphi_q^t + \sigma^2}) \tag{9}$$

In figures 8(a) and 8(b), we show the system throughput and per-pair throughput. The nature of system throughput is similar to a conventional wireless network hitting a plateau after a certain point. Interestingly, per-pair throughput characteristic shows a convex nature and there exists a number of secondary pairs that yields maximum per-pair throughput.

5.3 Spectral Efficiency

Spectral efficiency provides an indication of how efficiently a bandwidth-limited frequency spectrum is used. Spectral



(a) System throughput (b) Per-pair throughput

Fig. 8. Throughput estimation using spectrum map

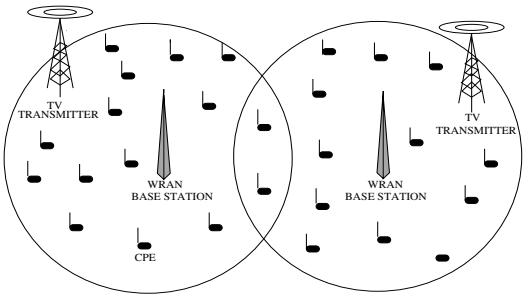


Fig. 9. Architecture of an IEEE 802.22 network

efficiency measured in bits/sec/Hz can be represented in two ways: link spectral efficiency and system spectral efficiency. The former is defined as the net bit-rate that can be achieved by a link per channel bandwidth (Hz). Similarly, system spectral efficiency is defined as the maximum throughput, summed over all nodes, divided by the channel bandwidth. It gives the number of secondary nodes that can be *simultaneously* supported by the available spectrum in a geographic area. Thus, link spectral efficiency for ch_q between the secondary transmitter and receiver is

$$\xi_q^t = \frac{1}{B} log(1 + \frac{P|h_q^t|^2}{I_q^t + \varphi_q^t} + \sigma^2) \quad (10)$$

Similarly, the system spectral efficiency is obtained as

$$\Xi_q = \sum_{\forall j \in \kappa^q} \xi_q^t = \frac{1}{B} \sum_{\forall j \in \kappa^q} log(1 + \frac{P|h_q^t|^2}{I_q^t + \varphi_q^t} + \sigma^2) \quad (11)$$

The nature of spectral efficiency is similar to that of channel capacity as the two parameters are proportional to each other.

6 Spectrum Map Application Case Study: Resource allocation in TVWS

We demonstrate the effectiveness of the spectrum map by using it for resource allocation in TV white space (TVWS). We choose TVWS as the primary network because the spectrum occupancy in the sub-900 MHz TV band is much investigated and IEEE has already proposed an initial draft standard (IEEE 802.22 WRAN) [7] to exploit these unused bands. The core components of an IEEE 802.22 WRAN are base stations (BS) and consumer premise equipments (CPE) as shown in Fig. 9. Secondary nodes (BSs and CPEs) opportunistically access unused or underutilized TV bands when not in use.

6.1 Network Model

6.1.1 Primary Network

The primary network consists of TV transmitters, TV receivers and wireless microphones with the latter taking only a small amount of band space. The TV transmitters are deployed depending on population density in a geographic region. An urban area has denser transmitters than rural regions. In Section 6.5.3, we try to emulate the location distribution of TV transmitters from existing literature surveys. We assume that the signal diffuses isotropically in the environment and is then received by x_i with intensity P multiplied by a loss factor of $l_{(x_i,x_j)}$ due to isotropic dispersion and absorption in the environment.

6.1.2 Secondary Network

IEEE 802.22 networks are centralized CRNs divided into cells, each having one BS. The BS communicates with the CPEs in its cell as well as with neighboring BSs. The BS is aware of the location of all the CPEs under it. We assume that there are *no pre-defined control channels* in the system, i.e., the only means of communication between the BS and the CPEs are the free channels that are currently not being used by the primary users. Each BS has two processing units responsible for two sets of functions. Periodic spectrum sensing, sharing the sensed spectrum data and creating the spectrum map is performed by one unit and the other is responsible for beacon broadcast to advertise free channels and QoS provisioned allocation of the free channels to the CPEs. The latter functions are performed after consulting the latest version of spectrum map. Details of the allocation process is discussed in Section 6.5.2.

CPEs are also equipped with sensing devices for finding unused channels; however, performing spectrum sensing is intelligent and selective depending on the state of the node discussed in Section 6.5.1. CPEs are allocated a pair of uplink and downlink channels which they are allowed to use till a primary arrives or unless they relinquish the channels themselves.

6.2 Resource Allocation Problem

The main challenge in allocating channels to the CPEs is the absence of pre-defined control channels between the BS and the CPEs. Upon detection of any transmission activity from a primary user on a channel, the secondary nodes are mandated to relinquish those channels [30]. The ability of the BS to assign the best possible channels to the CPEs depends on how well the BS is able to capture the spectrum availability at the various CPE locations. Since the BS cannot perform sensing at locations other than itself, it has to rely upon the sensed spectrum reports shared by the CPEs. If *all* the CPEs were to *continuously* share their spectrum usage reports, the BS would have the most accurate information. However, the communication overhead becomes a bottleneck as sharing of data has to be done on the same channels that the BS is supposed to allocate. Thus, channel assignment to CPEs becomes a challenge.

6.3 Our Solution using the Spectrum Map

We design an *on-demand* channel allocation scheme for IEEE 802.22 networks using our proposed spectrum map. We let a subset of CPEs to work as data points and feed their spectrum usage data to the BS for it to create the spectrum map. The spectrum map allows a two-fold solution for efficient resource allocation. First, the map is used for quicker communication with a candidate CPE and increases the probability of rendezvous between the CPE and the BS. Such

communication allows the BS to acquire the actual spectrum usage at the CPE and evaluate different performance metrics. Such proactive performance analysis not only identifies the best candidate channel for a CPE but also indicates the best possible CPE among candidate CPEs for a particular channel. Channel allocation scheme thus adopted increases the overall network throughput and maximizes spectrum usage.

6.4 Improving Rendezvous Probability using Spectrum Map

The means of initial communication between the BS and a CPE looking for channels is the beacons sent by the BS and subsequent handshaking process. The latest IEEE 802.22 based WRAN specifications [8] mandate the MAC layer to be able to adapt dynamically to changes in the environment by sensing the spectrum. Although the MAC layer is mandated to consist of specific data structures, details about the mechanism and involved channels (i.e., dedicated common control channel or dynamic channel rendezvous) for such rendezvous are not specified. However, in the absence of any control channel, this communication is probabilistic, i.e., the BS can send beacons with the specific data structures on the available channels and hope that the CPEs respond. Although viable, this traditional technique is ineffective and probability of rendezvous between the BS and a CPE is low. We propose an intelligent beacon broadcast scheme with the help of the spectrum map that can minimize the number of channels where beacons are sent and thus increasing the probability of dynamic rendezvous with the CPEs. This obviates the need for a any control channel between the BS and the CPEs making the secondary communication completely opportunistic. In the intelligent beacon broadcasting scheme, the beacons are sent only on *selected* channels depending on the requirement of idle CPEs. These selected channels are those that belong to the common set of the available channels at the BS and the set of channels which are estimated to be available for the idle CPEs (from spectrum map). The BS sends beacons in each of this common pool of channels and waits for channel allocation request from any node and after a stipulated time moves to the next channel. In case of successful reception of a channel allocation request, the BS logs the request before moving to next channel. At the end of the beacon cycle, the BS proceeds to allocate channels to the nodes whose channel allocation requests were successfully received during the beacon broadcast cycle. Such prediction of channel usage at CPE and sending beacons only on the potential free channels reduce interference to primaries.

Let us assume that there are K CPEs in a cell tuning periodically to N channels in order to receive a BS beacon. Therefore initially,

$$\text{Prob}\{\text{A CPE receiving a beacon}\} = \frac{1}{N}$$

Let us also assume that out of K CPEs, $|\Delta|$ have been already been delegated for sensing. From the remaining $K - |\Delta|$ CPEs, let the probability that a CPE is allocated at any instance be p_{alloc}. Therefore, if $|J|$ is the set of idle CPEs in the system at any time then,

$$|J| = K - (K - |\Delta|) \times p_{alloc} - |\Delta|$$

We also assume that γ^b_{free} and γ^i_{free} are the sets of free channels at the BS and at any CPE i as estimated by the BS using the spectrum map ($|\gamma^b_{free}|, |\gamma^i_{free}| \leq N$). Now from $\gamma^i_{free} \forall\ i$, the BS can estimate the entire set of free channels for the cell as $\gamma^{cell}_{free} = \bigcup_{i \in J} \gamma^i_{free}$ and use the channels belonging to the set $\gamma^b_{free} \bigcap \gamma^{cell}_{free}$ only; thus initiating the intelligent beacon broadcast scheme.

Now for a successful handshake between a BS and a CPE, ideally only a single CPE should reply to the beacon with a channel allocation request. More than one reply will cause interference resulting in unsuccessful reception. Thus, collision is only possible when more than one CPE listen to a beacon on the same channel at the same time. Therefore, the probability of a successful rendezvous in a beacon period, P_{ren} is expressed as:

$$\begin{aligned} P_{ren} &= \text{Prob}\{\text{Beacon is sent on a channel and only a} \\ &\quad \text{single CPE is tuned to the channel}\} \\ &= \text{Prob}\{\text{A single CPE tuned to a channel} \\ &\quad | \text{ beacon was sent on that channel}\} \end{aligned}$$

Since the tuning by a CPE and the beacon broadcast are mutually independent,

$$\begin{aligned} P_{ren} &= \text{Prob}\{\text{A single CPE is tuned to a channel}\} \\ &= |J| \times \frac{1}{N} \times (1 - \frac{1}{N})^{|J|} \end{aligned} \quad (12)$$

Therefore, the expected number of successful handshakes per beacon period is,

$$\text{E[No. of successful handshakes]} = |\gamma^b_{free} \bigcap \gamma^{cell}_{free}| \times P_{ren}$$

It is to be noted that the above analysis considers an aggressive transmission-reception strategy adopted by the CPEs, i.e., the CPEs do not wait before transmitting the channel allocation request. In the other alternative, i.e., the non-persistent strategy where CPEs wait for a time window before sending channel allocation request, the possibility of collision between CPEs decreases. Therefore, the above analysis considers the worst case performance in terms of probability of a successful rendezvous.

6.4.1 *Numerical Results*

To better elucidate the advantages of proposed intelligent beacon broadcast over conventional probabilistic technique, we perform numerical analysis in MATLAB capturing the spatial correlation among nodes, and interference among CPEs. The probability of a channel being free is kept at 0.3.

In Fig. 10, we compare the probabilities of a CPE receiving a beacon for conventional and proposed intelligent beacon broadcast schemes. We see that irrespective of the total number of channels in the spectrum, the probability is higher with the intelligent beacon broadcasting technique.

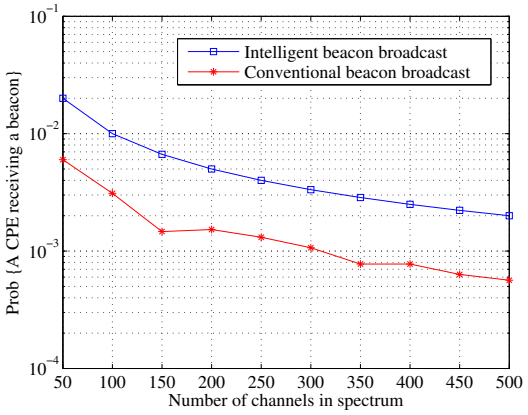


Fig. 10. Benefits of proposed beacon broadcast technique in terms of beacon reception

In Fig. 11, we observe the nature of P_{ren} from Eqn. (12). There exists a convexity where the probability of handshake increases with number of unallocated CPEs and gradually decreases afterwards. The gradual decrease is due to the limited channel availability for increasing unallocated CPEs. With more channels in the system, the saturation point is with higher value of unallocated CPEs. However, the maximum number of CPEs allowed in a cell is 255 [8] and the number of simultaneous unallocated CPEs will be even less.

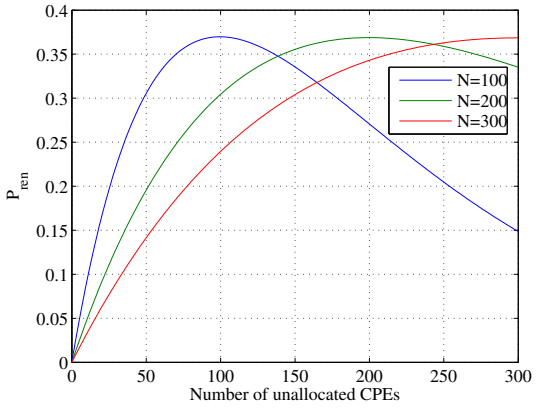


Fig. 11. Nature of probability of a successful handshake

Fig. 12 compares the performance of the proposed intelligent beacon broadcast with conventional probabilistic rendezvous technique in terms of expected number of successful handshakes in a beacon period. We observe that with any number of unallocated CPEs, the proposed technique performs better than conventional beacon broadcast.

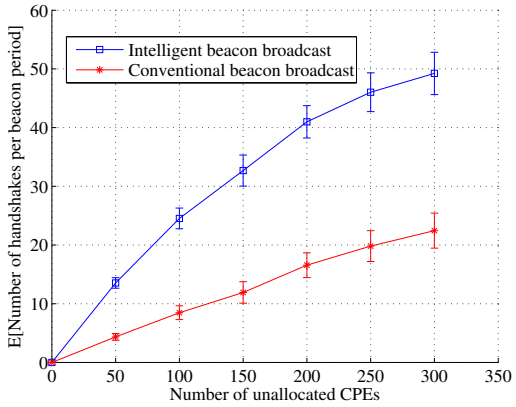


Fig. 12. Benefits of proposed beacon broadcast technique in terms of expected number of successful handshakes

6.5 Channel Allocation

In order to efficiently allocate channels to the CPEs, we use iterative clustering as was discussed in Section 4 to find delegation among the CPEs to work as data points.

6.5.1 Delegation CPEs with Iterative Clustering

We create zones are logical divisions of the coverage area of a BS. The zones are equivalent to the clusters discussed in Section 4. This division is proposed as a BS specific feature and therefore *does not* violate the specifications [30]. Each of such zones (clusters) has a cluster-head which is the designated data point for that zone. We call these data-points as *Delegation* CPEs. BS controls the selection of a particular CPE as a delegation member. The idea is to have almost equal representation from all zones and no zone is unrepresented as far as possible. The BS maintains three queues for each zone, namely: AQ (for allocated CPEs), DQ (for delegated CPEs) and HQ (for potentially delegate CPEs i.e., idle CPEs with channels). Selection of a *delegation* member takes place after termination of data transmission (either voluntarily or forcefully) with the channel used now serving as sharing medium. If not selected for delegation, a CPE can relinquish its channels or hold onto the channels in anticipation of channel requirement in the near future. Holding time depends on the demand for the particular channels by other CPEs. However, such holding onto channels can result in that CPE being selected as a delegation member (using the same channels for sharing) if instructed by the BS. A delegated CPE continues to serve until it is instructed otherwise by the BS. This obligates the CPE to use the allocated channels only as means to share its spectrum usage with BS even if the CPE wants to go to the transmission state. The decision of keeping such a CPE in sharing state solely depends on the importance of that CPE in building the spectrum profile. Soon after the BS finds a suitable alternative, it allows the CPE to switch to the transmission state and use the same channel for data communication.

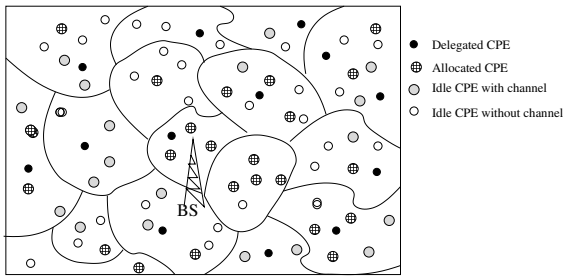


Fig. 13. IEEE 802.22 network with BS and different types of CPEs

Fig. 13 is an illustrative example that shows how the coverage area of the BS is divided into zones. The BS attempts to find a delegated CPE for every zone.

6.5.2 Allocation process

The allocation process by the BS is initiated by reception of a channel allocation reply from any CPE c on any of the $\nu_{est} \cap \gamma^b_{free} \cap \gamma^c_{free}$ channels. Such allocation reply is accompanied with raw spectrum usage at c, Υ^c=$\{\phi^c_j | \ \forall ch_j \in \hat{N}\}$. The BS creates the set of available channels ρ^c_{avl}=$\{j | \ \forall \phi^c_j < T\}$ allocable to c. Channels used by other allocated or delegated CPEs in the interference range of c can not be allocated to c for obvious reasons. In order to ensure this, the BS maintains the AQ, DQ and HQ of all the zones within the interference range of c. From the AQs, the BS creates the set $\widehat{AQ}^c$ with only those channels which are allocated to CPEs within the interference range of c. Similarly $\widehat{DQ}^c$ and $\widehat{HQ}^c$ are created. No channels in $\widehat{AQ}^c$ and $\widehat{DQ}^c$ can be allocated to avoid co-channel interference. BS updates the set of available channels for allocation ρ^c_{avl} excluding all such channels. For all the rest of the channels in ρ^c_{avl}, expected channel performance (can be any desirable parameter or a combination of some) is calculated and given a

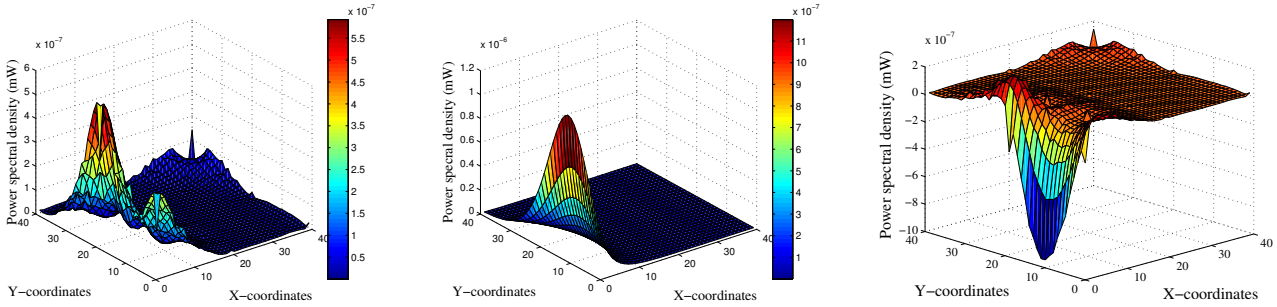


Fig. 14. (a) Estimated RSS (b) Actual RSS, and (c) Error in estimation of a channel in Region 1

score. ρ^c_{avl} is further updated with descending order of the channel scores. There is also the set $\widehat{HQ}^c$ whose channels are held by other idle CPEs in the interference range of c. The purpose for allowing the CPEs to hold onto channels is to expedite channel allocation by not making them go through the entire handshaking process. Therefore, BS always tries to allocate the best channel among $|\rho^c_{avl}|$ not belonging to $\widehat{HQ}^c$, if possible.

6.5.3 *Experimental Results*

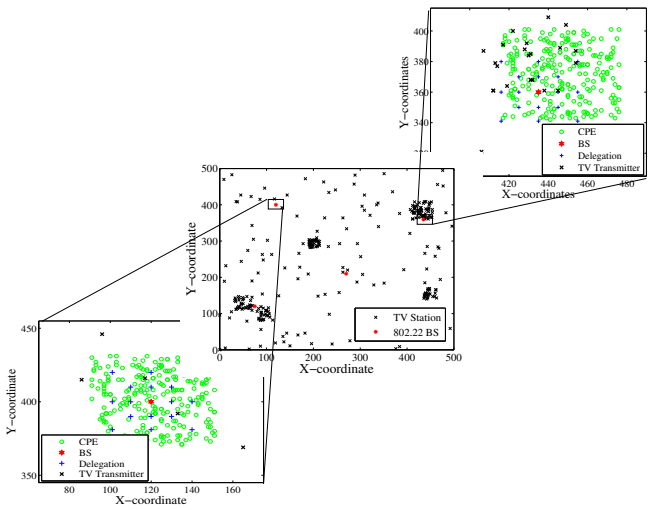


Fig. 15. Simulation topology with primary transmitters, secondary BSs and CPEs

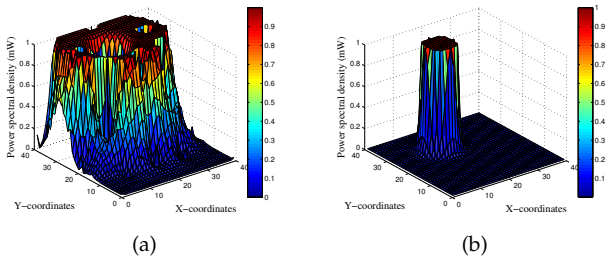


Fig. 16. (a) Estimated and (b) actual signal strengths of a channel in Region 2

In a completely new experimental set-up to the one discussed in Section 3.5, we simulate the locations of TV stations in a 500×500 sq. miles grid as shown in Fig 15 using TV station location distribution in [4]. The total number of channels is varied from 50-400 with corresponding bandwidth of 6 MHz-750 kHz. The power-profile of the TV stations ranges between 10kW-1MW. We identify two regions with dense and sparse primary densities (mimicking big cities and small towns) and deploy IEEE 802.22 networks (WRAN cells) shown in Fig. 15. Locations of different types of nodes and transmission patterns are different in the two scenarios while total number of CPEs, density, and the number of delegated CPEs are kept the same.

In figures 14(a), 14(b), and 16(a), 16(b), we show the estimated and actual signal strengths for the two regions. The figures reflect the success of the interpolation technique to capture the nature of the spectrum usage in those regions. By comparing the estimated and actual signal strength values for both region, we get the error in estimation (as shown in Fig. 14(c) for Region 1) of the order 10^{-6} mW which is too small to alter the occupancy decision vector.

Figures 17 and 18 show the channel occupancy vectors of two different channels at two CPEs in the same two regions. In both the cases the prediction of channel occupancy is highly accurate with only 6 and 3 false positives respectively. It is important to note that none of these variations are false negatives which is mandatory for secondary communication.

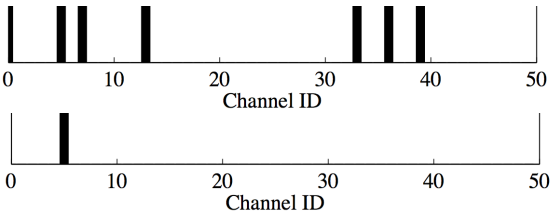


Fig. 17. Estimated and actual channel occupancy of CPE #38 in Region 1

Figures 19(a) and 19(b) collectively portray how the available channels in the spectrum are utilized by the proposed allocation technique. In Fig. 19(a), a line with 45° slope signifies 100% utilization which occurs when there are more channels available to allocate to the CPEs. However even with 50 channels with probability of 0.6, we see that 17 free channels are allocated resulting in 60% $(18/[50\text{x}0.6])$ utilization. We keep the number of idle CPEs at 50.

Nature of free channel utilization with varying number of available channels in the spectrum is better understood in Fig. 19(b) where we vary the total number of channels and the probability of getting white space; thereby indirectly

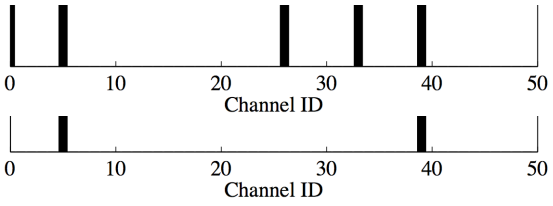
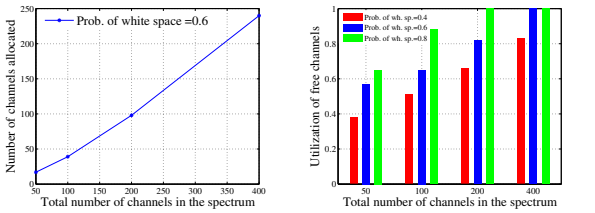


Fig. 18. Estimated and actual channel occupancy of CPE #106 in Region 2

varying number of free channels to utilize. We observe that even with a very low probability of white space 0.4 (with respect to TV spectrum), the proposed allocation scheme ensures more than 50% utilization of free spectrum in most occasions. The total number of allocable CPEs are kept at 255 for all the scenarios above which is the maximum number of CPEs that an IEEE 802.22 cell can have [30].



(a) Number of channel allocated from total available channels

(b) Utilization of available channels for different probabilities of white space

Fig. 19. Utilization of available channels in the spectrum

In Fig. 20, we show the efficiency of the proposed allocation technique by calculating the number of CPEs being allocated a pair of uplink and downlink channels. We observe that with 400 channels and a probability of getting white space of 0.8, we allocate 106 out of possible 160 CPEs.

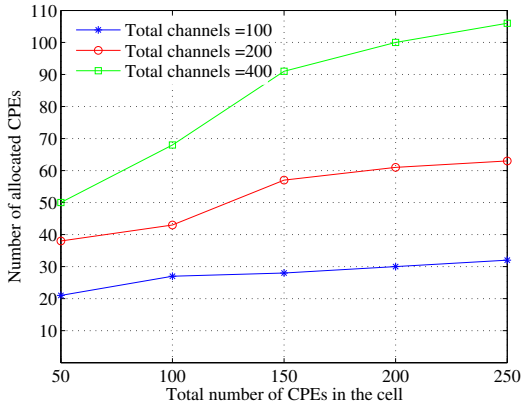


Fig. 20. Number of allocated CPEs for different number of channels

In Fig. 21, we compare the performances of channel allocation with and without using spectrum map for different number of free channels in terms of normalized supported data rate. We define normalized aggregate data rate as the aggregate bit rate supported by the allocated channels to the total achievable capacity of the free channels. We see that except the case with 50 free channels, the spectrum map

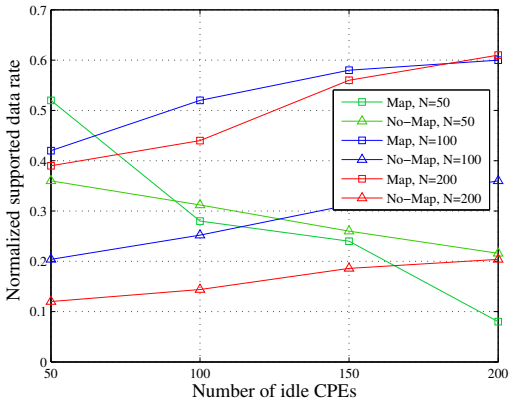


Fig. 21. Normalized supported data rate for different number of idle CPEs

initiated allocation performs better. For very low number of available channels, if there are more idle CPEs, it creates more collision during beacon broadcast phase resulting less utilization.

7 Conclusions

In this paper, we use Shepard's interpolation technique to create a spectrum map, i.e., an estimate of the spectrum usage at any arbitrary location in a secondary cognitive radio network. This is achieved by sharing and fusing raw spectrum data sensed at various strategic locations computed using iterative clustering. We demonstrate how the spectrum map can help in predicting channel performance metrics, such as, channel capacity, spectral efficiency, and secondary throughput. Through simulation experiments, we validate the correctness of the map and show how we can compute the distribution of these metrics at any arbitrary location, which can potentially be used by secondary networks for better channel selection to maximize spectrum usage. Finally, we demonstrate how the proposed map helps attain near perfect channel allocation in IEEE 802.22 networks and improve channel rendezvous probability.

References

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks Journal (ELSEVIER)*, vol. 50, pp. 2127–2159, 2006.

[2] "FCC adopts rule for unlicensed use of television white spaces." [Online]. Available: https://www.fcc.gov/document/tv-white-spaces-rule-changes/

[3] "Google spectrum database." [Online]. Available: https://www.google.com/get/spectrumdatabase/

[4] J. Riihijarvi and P. Mahonen, "Exploiting spatial statistics of primary and secondary users towards improved cognitive radio networks," in *Cognitive Radio Oriented Wireless Networks and Communications, CrownCom. 3rd International Conference on*, May 2008, pp. 1 –7.

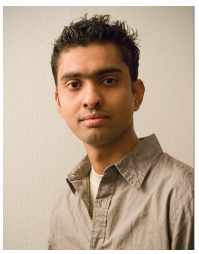
[5] Y. Li, T. T. Quang, Y. Kawahara, T. Asami, and M. Kusunoki, "Building a spectrum map for future cognitive radio technology," in *Proceedings of the 2009 ACM workshop on Cognitive radio networks*, ser. CoRoNet '09, 2009, pp. 1–6.

[6] J. Riihijarvi, P. Mahonen, M. Wellens, and M. Gordziel, "Characterization and modeling of spectrum for dynamic spectrum access with spatial statistics and random fields," in *Personal, Indoor and Mobile Radio Communications, PIMRC. IEEE 19th International Symposium on*, Sept. 2008, pp. 1 –6.

[7] "IEEE 802.22, working group on wireless regional area networks (WRAN)." [Online]. Available: http://grouper.ieee.org/ groups /802/22

[8] "IEEE draft standard for wireless regional area networks part 22: Cognitive wireless ran medium access control (MAC) and physical layer (PHY) specifications: Policies and procedures for operation in the TV bands - Amendment: Management and control plane interfaces and procedures and enhancement to the management information base (MIB)," *IEEE P802.22a/D2, October 2013*, pp. 1–551, 2013.
[9] M. A. McHenry, D. McCloskey, D. Roberson, and J. T. MacDonald, "Spectrum occupancy measurements - Chicago, Illinois," Shared Spectrum Company, Tech. Rep., 2005.
[10] T. Harrold, R. Cepeda, and M. Beach, "Long-term measurements of spectrum occupancy characteristics," in *New Frontiers in Dynamic Spectrum Access Networks, DySPAN, IEEE International Symposium on*, May 2011, pp. 83 –89.
[11] V. Atanasovski, J. van de Beek, A. Dejonghe, D. Denkovski, L. Gavrilovska, S. Grimoud, P. Mahonen, M. Pavloski, V. Rakovic, J. Riihijarvi, and B. Sayrac, "Constructing radio environment maps with heterogeneous spectrum sensors," in *New Frontiers in Dynamic Spectrum Access Networks, DySPAN. IEEE International Symposium on*, May 2011, pp. 660–661.
[12] B. Jayawickrama, E. Dutkiewicz, I. Oppermann, G. Fang, and J. Ding, "Improved performance of spectrum cartography based on compressive sensing in cognitive radio networks," in *Communications, ICC. IEEE International Conference on*, 2013, pp. 5657–5661.
[13] A. Stirling, "Exploiting hybrid models for spectrum access: Building on the capabilities of geolocation databases," in *New Frontiers in Dynamic Spectrum Access Networks, DySPAN. IEEE International Symposium on*, May 2011, pp. 47 –55.
[14] R. Murty, R. Chandra, T. Moscibroda, and P. Bahl, "Senseless: A database-driven white spaces network," *Mobile Computing, IEEE Transactions on*, vol. 11, no. 2, pp. 189 –203, Feb. 2012.
[15] M. Wellens, J. Riihijarvi, and P. Mahonen, "Modeling primary system activity in dynamic spectrum access networks by aggregated ON/OFF-processes," in *6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON) Workshops,*, June 2009, pp. 1–6.
[16] M. Wellens, J. RiihijäRvi, and P. MäHöNen, "Full length article: Empirical time and frequency domain models of spectrum use," *Phys. Commun.*, vol. 2, no. 1-2, pp. 10–32, Mar. 2009.
[17] J. Bazerque, G. Mateos, and G. Giannakis, "Group-lasso on splines for spectrum cartography," *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4648–4663, Oct 2011.
[18] H. Yilmaz, C.-B. Chae, and T. Tugcu, "Sensor placement algorithm for radio environment map construction in cognitive radio networks," in *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, April 2014, pp. 2096–2101.
[19] R. Mahapatra and E. Strinati, "Interference-aware dynamic spectrum access in cognitive radio network," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, Sept 2011, pp. 396–400.
[20] D. Denkovski, V. Atanasovski, L. Gavrilovska, J. Riihijarvi, and P. Mahonen, "Reliability of a radio environment map: Case of spatial interpolation techniques," in *Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), 2012 7th International ICST Conference on*, June 2012, pp. 248–253.
[21] J. Riihijarvi, J. Nasreddine, and P. Mahonen, "Demonstrating radio environment map construction from massive data sets," in *Dynamic Spectrum Access Networks (DYSPAN), 2012 IEEE International Symposium on*, Oct 2012, pp. 266–267.
[22] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio, part I: Two user networks," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 6, pp. 2204 –2213, June 2007.
[23] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *New Frontiers in Dynamic Spectrum Access Networks, DySPAN. IEEE International Symposium on*, 2005, pp. 131–136.
[24] E. Visotsky, S. Kuffner, and R. Peterson, "On collaborative detection of tv transmissions in support of dynamic spectrum sharing," in *New Frontiers in Dynamic Spectrum Access Networks, DySPAN. IEEE International Symposium on*, 2005, pp. 338–345.
[25] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM national conference*, ser. ACM '68, 1968, pp. 517–524.
[26] M. Sen, I. Banerjee, M. Chatterjee, and T. Samanta, "Sensor localization using received signal strength measurements for obstructed wireless sensor networks with noisy channels," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2015 IEEE*, March 2015, pp. 47–51.
[27] "Modelling and simulation of rayleigh fading, path loss, and shadowing fading for wireless mobile networks," *Simulation Modelling Practice and Theory*, vol. 19, no. 2, pp. 626 – 637, 2011.
[28] C. E. Shannon, "A mathematical theory of communication," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, Jan. 2001.
[29] S.-W. Jeon, N. Devroye, M. Vu, S.-Y. Chung, and V. Tarokh, "Cognitive networks achieve throughput scaling of a homogeneous network," *Information Theory, IEEE Transactions on*, vol. 57, no. 8, pp. 5103 –5115, Aug. 2011.
[30] "IEEE P802.22/d1.0 draft standard for wireless regional area networks part 22: Cognitive wireless RAN medium access control (MAC) and physical layer (PHY) specifications: Policies and procedures for operation in the TV bands, april 2008." April 2008.

Saptarshi Debroy received the BTech degree in information technology from West Bengal University of Technology, India, in 2006, the MTech degree in distributed and mobile computing from Jadavpur University, India, in 2008, and the PhD degree in computer engineering from the University of Central Florida, Orlando in 2014. He is a postdoctoral fellow at Virtualization, Multimedia and Networking (VIMAN) Lab in the Department of Computer Science, University of Missouri-Columbia. His research interests include cloud computing, big data networking, cognitive radio networks, and QoE provisioning for online multiplayer gaming. He received multiple awards including IEEE PIMRC Best Paper Award, multiple US National Science Foundation (NSF) and IEEE Comsoc Travel Grant Awards, and an academic year Gold medal at Jadavpur University. He has served as publicity co-chair for ACM MoViD 2013, TPC member of conferences such as IEEE GCNC, IEEE BigDataService, ICA3PP, IFIP/IEEE QCMan, IEEE ANTS, IEEE ICNC, and reviewer for several international conferences and journals.



Shameek Bhattacharjee received his Ph.D. and MS in Computer Engineering from University of Central Florida, Orlando in 2015 and 2011 respectively, and BTech degree in information technology from West Bengal University of Technology, India in 2009. He is currently a Post-Doctoral Research Fellow in CReWMAN Lab at the Missouri University of Science and Technology. His research interests include Dynamic Spectrum Access Networks, Trust and Reputation Scoring and Information Assurance in Networking Systems. He received IEEE PIMRC Best Paper Award. He also serves as TPC member and reviewer of several international conferences and peer-reviewed journals.

Mainak Chatterjee received the BSc degree in physics (Hons.) from the University of Calcutta, the ME degree in electrical communication engineering from the Indian Institute of Science, Bangalore, and the PhD degree from the Department of Computer Science and Engineering from the University of Texas at Arlington. He is an associate professor in the Department of Electrical Engineering and Computer Science, University of Central Florida, Orlando. His research interests include economic issues in wireless networks, applied game theory, cognitive radio networks, dynamic spectrum access, and mobile video delivery. He has published more than 150 conferences and journal papers. He got the Best Paper Awards in IEEE Globecom 2008 and IEEE PIMRC 2011. He received the AFOSR sponsored Young Investigator Program (YIP) Award. He co-founded the ACM Workshop on Mobile Video (MoVid). He serves on the editorial board of Elsevier Computer Communications and Pervasive and Mobile Computing Journals. He has served as the TPC co-chair of several conferences including ICDCN 2014, IEEE WoWMoM 2011, WONS 2010, IEEE MoVid 2009, Cognitive Radio Networks Track of IEEE Globecom 2009, and ICCCN 2008. He also serves on the executive and technical program committee of several international conferences.