Clustering-based spectrum sensing In cognitive radio wireless sensor networks

Vybhav N [#1], Uma S [*2]

[#]*M. Tech Student, Dept. of ECE, SJB Institute of Technology, Bangalore, India*
[1]*jannat.vybhav@gmail.com*

**Assistant Professor, Dept. of ECE, SJB Institute of Technology, Bangalore, India*
[2]*uma.shivanna@gmail.com*

Abstract—**A cognitive radio wireless sensor network (CR -WSN) where each sensor node is equipped with cognitive radio. A typical concern in CR-WSN is energy consumption due to resource-constrained nature of sensor nodes. Moreover, additional energy is consumed in a CR-WSN to support CR-exclusive functionality such as spectrum sensing and switching, which could shorten sensor node lifetime. However, some sensor nodes could receive similar signal due to similar channel condition such that they probably have same spectrum sensing results. A clustering based scheme for spectrum sensing in CR-WSN, which reduces energy consumption by involving less nodes in spectrum sensing. With improved clustering algorithm, sensor nodes are grouped into different sets based on their similarity in sensing result. In order to identify the optimal cluster number, a new objective function, based on new intra-cluster and inter-cluster proximity measures is proposed which can effectively reduce the energy consumption of sensor node and improve global detection probability. Cluster analysis or Clustering is said to be a collection of objects. It is used in various application in the real world. Such as data/text mining, voice mining, image processing, web mining and so on. It is important in real world in certain fields.**

Keywords: **Cognitive radio, Clustering analysis, Spectrum sensing, Wireless sensor networks.**

I. INTRODUCTION

Wireless sensor network (WSN) is one of the recent rising technologies in wireless communication with a wide range of applications, such as environment surveillance, health care, intelligent buildings and battle field control, etc. A key feature for current WSN solutions is operating in unlicensed frequency bands. Cognitive radio has been considered as a promising way to alleviate spectrum scarcity. In CR, unlicensed users or secondary users (SUs) can access the allocated band (licensed or Primary Users (PUs)) opportunistically without interfering with the PUs.

Cognitive radio system is a technique which overcomes that spectrum underutilization. Cognitive radio is a technique where secondary user looks for a free band to use when primary user is not in use of its licensed band. A function of cognitive radio is called Spectrum sensing which enables to search for the free bands and it helps to detect the spectrum hole (frequency band which is free enough to be used) which can be utilized by secondary user with high spectral resolution capability. In CR-based WSNs, when a SU starts transmission, it must sense potentially vacant bands to seek for available channels. If the SU detects PU appearance on its operation channel, it must move to another one as soon as possible.

Clustering is a common technique for statistical data analysis, used in a variety of fields, including machine learning, pattern recognition, image analysis, information retrieval, etc.,. With clustering techniques, data objects are grouped into clusters based on their similarity. The greater the similarity within a group and the greater difference between groups, the better or more distinct the clustering. There are different types of clustering analysis, such as partitioning methods and hierarchical methods. Hierarchical clustering techniques are one of the most important categories of clustering methods. By applying hierarchical clustering analysis in CR-WSN, we can group sensor nodes into different clusters based on similarity between them. In each cluster, we can select one sensor node for spectrum sensing instead of all since their similarity are very high, and then share its spectrum sensing information with others. Thus, the energy consumption can be reduced greatly because each SU will not consume additional energy to detect PU's presence every time.

II. OBJECTIVE

The main objective is to assess the performance of the energy detection method for spectrum sensing. The values with same distance will be placed under a specific cluster in their respective cluster range. Finally a hierarchic structure is formed which shows the actual output. Usually the distance between two clusters A and B is one of the following:

- The maximum distance between elements of each cluster (also called complete-linkage clustering):
  $$\max \{d(x,y) : x \in A, y \in B\}.$$
- The minimum distance between elements of each cluster (also called single-linkage clustering):
  $$\min \{d(x,y) : x \in A, y \in B\}.$$
- The mean distance between elements of each cluster (also called average linkage clustering
  $$\frac{1}{|A| \cdot |B|} \sum_{x \in A} \sum_{y \in B} d(x,y)$$
- The sum of all intra-cluster variance.
- The increase in variance for the cluster being merged (Ward's method).
- The probability that candidate clusters spawn from the same distribution function (V-linkage).

III. PROBLEM FORMULATION

A. New Evaluation Criteria

In most methods of hierarchical clustering, measure of similarity between sets of observations achieved by use of an appropriate metric and a linkage criterion which specifies the similarity of sets as a function of the pairwise distances of observations in the sets. In most hierarchical clustering, Euclidean distance, Squared Euclidean distance and Manhattan distance are the most common similarity measure. However, these measures may not appropriate in CR-WSN because of the similarity between sensor nodes focus on spectrum sensing results. Consequently, new similarity measures which are more suitable for CR-WSN sensor nodes are as follows

- SIM_{ij} = The similarity between ith and jth sensor node, which represents average similarity rate under c channels. It is defined as the probability that two sensor nodes have the same spectrum sensing results

$$SIM_{ij} = \frac{1}{c \cdot times} \sum_{p=1}^{c} count\ (d_i = d_j)$$

where c is the number of channels, times is the number of spectrum sensing using energy detection, and count(d==0) is the number of same spectrum sensing results of the ith and jth SU under kth channel using energy detection.

- WCS_k = Intra-Cluster similarity of kth cluster, which is defined as the average SIM_{ij} between each pair of sensor nodes within kth cluster. when only one sample in a cluster, its WCSk will be zero.

$$WCS_k = \begin{cases} 0 & n_k = 1 \\ \frac{1}{n_k(n_k-1)/2} \sum_{i=1}^{n_k} \sum_{j=i+1}^{n_k} SIM_{ij} & n_k > 1 \end{cases}$$

where n_k is the number of sensor nodes in kth cluster

- BCS_k = Inter-Cluster similarity of kth cluster. This criterion determines the similarity between clusters.

$$BCS_k = \begin{cases} 0 & m = 1 \\ \sum_{l=1, l \neq k}^{m} \frac{1}{n_k(n_k-1)/2} \sum_{i=1}^{n_k} \sum_{j=i+1}^{n_k} SIM_{ij} & m > 1 \end{cases}$$

where m is the number of clusters, n_k and n, is the number of sensor nodes in k_{th} and l_{th} cluster, respectively.

- Objective Function: In general, the results of hierarchical clustering are usually presented in a dendrogram, but we desire to determine a reasonable number of clusters to return from any hierarchical clustering algorithm, so we propose the new objective function in this paper. The new objective function is given below

$$OC(m) = \frac{1}{m} \sum_{k=1}^{m} WCS_k + \frac{1}{m(m-1)} \sum_{k=1}^{m} BCS_k$$

$$m_{opt} = \max(OC(m))$$

where m is the number of clusters, n_k is the number of sensor nodes in kth cluster. A is the threshold of intra-cluster similarity, which promises WCS_k each satisfy the minimum similarity after clustering. In the above equation, m_{opt} is the optimal number of clusters, which maximizes the OC(m) objective function.

Cognitive Radio

Cognitive radio is a form of wireless communication where a transceiver can intelligently detect the channels for communication which are in use and which are not in use, and move into unused channels while avoiding occupied ones. A spectrum hole (Fig 1) is generally a concept of spectrum as non-interfering, considered as multidimensional areas within frequency, time, and space. For secondary radio systems, the main challenge is to be able to sensing spectrum hole when they are within such frequency bands.

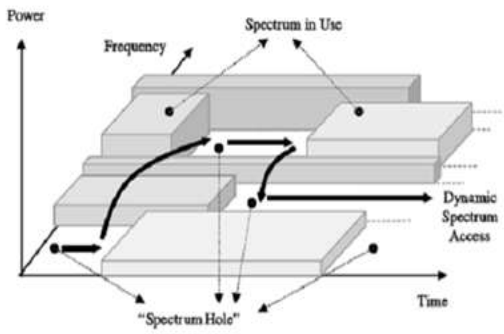

Fig. 1 Spectrum holes concept

A. Types of CR

There are two types of Cognitive Radios:

- Full Cognitive Radio: Full Cognitive Radio (CR) considers all parameters. A wireless node or network can be conscious of every possible parameter observable.

- Spectrum Sensing Cognitive Radio: Detects channels in the radio frequency spectrum. Fundamental requirement in cognitive radio network is spectrum sensing. To enhance the detection probability many signal detection techniques are used in spectrum sensing. Spectrum sensing determines the presence of primary user on a band. The cognitive radio is able to share the result of its detection with other cognitive radios after sensing the spectrum. The goal of spectrum sensing is to find out the spectrum status and activity by periodically sensing the target frequency band.

B. Characteristics of CR

There are two main characteristics of the cognitive radio and can be defined

- Cognitive capability: Cognitive Capability defines the ability to capture or sense the information from its radio environment of the radio technology.
- Re-configurability: refers to radio capability to change the functions, enables the cognitive radio to be programmed dynamically in accordance with radio environment (frequency, transmission power, modulation scheme, communication protocol).

C. Functions of CR

There are four major functions of CR Fig 2 shows the basic cognitive cycle in which spectrum sensing is discussed above

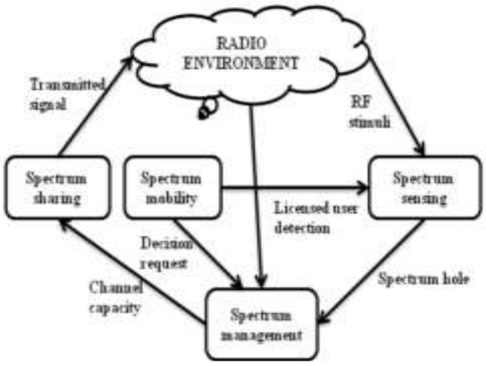

Fig. 2 Basic cognitive cycle

1) *Spectrum Management:* Spectrum Management provides the fair spectrum scheduling method among coexisting users. Spectrum sensing, spectrum analysis, and spectrum decision fall in spectrum Management. Spectrum Analysis makes possible the characterization of different spectrum bands, which is exploited to get the spectrum band appropriate requirements of the user. Spectrum decision refers to a cognitive radio decides the data rate, determines the transmission mode, and the transmission bandwidth.

2) *Spectrum Sharing:* Cognitive Radio assigns the unused spectrum (spectrum hole) to the secondary user (SU) as long as primary user (PU) does not use it. This property of cognitive radio is described as spectrum sharing.

- Underlay Spectrum Sharing: Underlay spectrum sharing is the availability of the radio spectrum access with minimal transmission power that the interference temperature above its pre-designed thresholds wouldn't be raised.
- Overlay Spectrum sharing: Unlicensed users can utilize a spectrum band for the fraction of time where this band is under-utilized by the licensed users in Overlay Spectrum sharing technique.

3) *Spectrum Mobility:* When a licensed (Primary) user is detected the Cognitive Radio (CR) vacates the channel. This is the process that allows the Cognitive Radio user to change its operating frequency. Cognitive Radio networks try to use the spectrum dynamically to operate in the best available frequency band and maintain the transparent communication.

D. CRS Applications

More flexible and efficient use of spectrum in the future open up exciting opportunities for cognitive radio to enable and support a variety of emerging applications, ranging from smart grid, public safety and broadband cellular, to medical applications. This section presents a brief view on how cognitive radio would support such applications, the benefits that cognitive radio would bring, and also some challenges that are yet to be resolved. Some of applications of CR are as follows

- Smart grid networks.
- Public safety networks.
- Cellular networks.
- Wireless medical networks.

Clustering

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

A. Types of Clustering

Cluster is said to be "Collection of data objects". There are two types of similarities of clusterings are:

• Intra-class similarity - Objects are similar to objects in same cluster.

• Inter-class dissimilarity - Objects are dissimilar to objects in other clusters.

B. Methods of Clustering

• Partitioning methods

• Hierarchical methods

• Density-based methods

1) Partitioning Methods: The partitioning methods generally result in a set of M clusters, each object belonging to one cluster. Each cluster may be represented by a centroid or a cluster representative. The precise form of this description will depend on the type of the object which is being clustered. If the number of the clusters is large, the centroids can be further clustered to produces hierarchy within a dataset.

Single Pass: A very simple partition method, the single pass method creates a partitioned dataset as follows:

- Make the first object the centroid for the first cluster.
- For the next object, calculate the similarity, S, with each existing cluster centroid, using some similarity coefficient.
- If the highest calculated S is greater than some specified threshold value, add the object to the corresponding cluster and re determine the centroid; otherwise, use the object to initiate a new cluster. If any objects remain to be clustered, return to step 2.

Advantages

- Relatively scalable and simple.
- Suitable for datasets with compact spherical clusters that are well-separated.

Disadvantages

- Severe effectiveness degradation in high dimensional spaces as almost all pairs of points are about as far away as average the concept of distance between points in high dimensional spaces is ill-defined.
- Poor cluster descriptors.
- Reliance on the user to specify the number of clusters in advance.
- High sensitivity to initialization phase, noise and outliers.
- Frequent entrapments into local optima.
- Inability to deal with non-convex clusters of varying size and density.

2) Hierarchial Methods: Connectivity based clustering, also known as hierarchical clustering, is based on the core idea of objects being more related to nearby objects than to objects farther away. As such, these algorithms connect "objects" to form "clusters" based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster. At different distances, different clusters will form, which can be represented using a dendrogram. Hierarchical techniques produce a nested sequence of partitions, with a single, all inclusive cluster at the top and singleton clusters of individual points at the bottom. Each intermediate level can be viewed as combining two clusters from the next lower level (or splitting a cluster from the next higher level). The result of a hierarchical clustering algorithm can be graphically displayed as tree, called a dendogram. This tree graphically displays the merging process and the intermediate clusters.

There are two basic approaches to generating a hierarchical clustering:

• Agglomerative approach

• Divisive approach

a) **Agglomerative**: Start with the points as individual clusters and, at each step, merge the most similar or closest pair of clusters. This requires a definition of cluster similarity or distance.

b) **Divisive**: Start with one, all-inclusive cluster and, at each step, split a cluster until only singleton clusters of individual points remain.

Advantages

- Embedded flexibility regarding the level of granularity.
- Well suited for problems involving point linkages.

Disadvantages

- Inability to make corrections once the splitting/merging decision is made.
- Lack of interpretability regarding the cluster descriptors.
- Vagueness of termination criterion.
- Prohibitively expensive for high dimensional and massive datasets.
- Severe effectiveness degradation in high dimensional spaces due to the curse of dimensionality phenomenon.

3) Density Based Clustering: Density-based clustering methods group neighbouring objects into clusters based on local density conditions rather than proximity between objects. These methods regard clusters as dense regions being separated by low density noisy regions. Density-based methods have noise tolerance, and can discover non-convex clusters. Similar to hierarchical and partitioning methods, density-based techniques encounter difficulties in high dimensional spaces because of the inherent scarcity of the feature space, which in turn, reduces any clustering tendency. Objects in these sparse areas - that are required to separate clusters - are usually considered to be noise and border points.

Density Reachability - A point "p" is said to be density reachable from a point "q" if point "p" is within ε distance from point "q" and "q" has sufficient number of points in its neighbors which are within distance ε.

Density Connectivity - A point "p" and "q" are said to be density connected if there exist a point "r" which has sufficient number of points in its neighbors and both the points "p" and "q" are within the ε distance. This is chaining process.

Advantages

- Discovery of arbitrary-shaped clusters with varying size.
- Resistance to noise and outliers

Disadvantages

- High sensitivity to the setting of input parameters.
- Poor cluster descriptors.
- Unsuitable for high-dimensional datasets because of the curse of dimensionality phenomenon.

C. Issues in Clustering technique

We have basically three issues in WSN clustering.

- Distance: As distance between the nodes increases the number of nodes in a cluster decreases and it may lead to higher consumption of energy.

- Energy: The energy consumption within a cluster can be reduced by decreasing the number of transmitting messages. Lesser the energy consumption leads to the longer lifetime of network.

- Density: The increase in sensors density may overload the network. Such overload might cause latency in communication and inadequate tracking of events.

D. Clustering Parameters

In WSNs clustering algorithms, it is worth reporting on some important parameters with regard to the whole clustering procedure in WSN.

- Number of clusters (cluster count): In most recent probabilistic and randomized clustering algorithms the CH election and formation process lead naturally to variable number of clusters. The number of clusters is usually a critical parameter with regard to the efficiency of the total routing protocol.

- Intra-cluster communication: In some initial clustering approaches the communication between a sensor and its designated CH is assumed to be direct (one-hop communication).However, multi-hop intra-cluster communication is often (nowadays) required, i.e., when the communication range of the sensor nodes is limited or the number of sensor nodes is very large and the number of CHs is bounded.

- Cluster-head selection: The leader nodes of the clusters (CHs) in some proposed algorithms (mainly for heterogeneous environments) can be pre-assigned.

- Algorithm complexity: In most recent algorithms the fast termination of the executed protocol is one of the primary design goals. Thus, the time complexity or convergence rate of most cluster formation procedures proposed nowadays is constant (or just dependent on the number of CHs or the number of hops).

IV. SYSTEM MODEL

Consider a wireless sensor network composed of n randomly distributed nodes and m channels; also each sensor node is equipped with cognitive radio. Sensor nodes are assumed to be static or move infrequently. Due to low-power and resource constraints of sensor nodes, such a CR-WSN has limited the lifetime. Moreover, additional energy could be consumed to support CR technology. Therefore, we propose a clustering analysis based energy efficient model m CR-WSN, as shown in Fig 3.

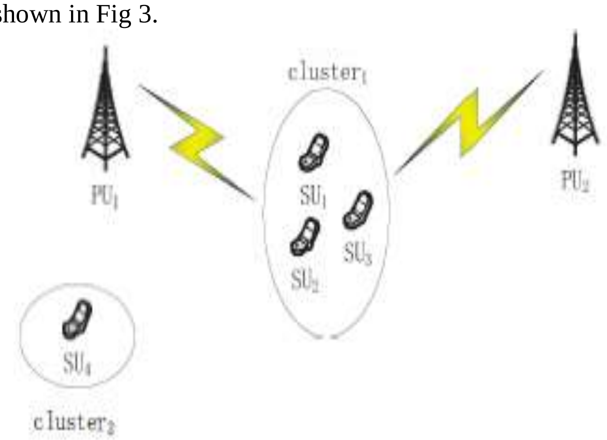

Fig. 3 System model with two licensed channels and four sensor nodes

Suppose there are four SUs (i.e. four sensor nodes) sense licensed channels state before transmitting data and PUs (i.e. licensed channels) which are subject to Rayleigh fading effects in Fig 3. In this network, every each SU receives signal from primary user with different received SNR. We assume that the SUI, SU2 and SU3 has highly similar spectrum sensing results for the two licensed channels, so we can apply hierarchical clustering methods to group them into cluster based on high similarity between them, and SU4 form another cluster2 alone which has no similar results. After clustering, in cluster1, we can choose one SU which has high probability of detection instead of all to sense the spectrum and decide PU's absence; finally, share its spectrum sensing results to the rest of SUs. Thus, SUs which belongs to cluster does not need to sense licensed channel to make decision before transmitting data. Consequently, the overall spectrum sensing energy in the CR-WSN can be saved.

The basic idea is collecting the dataset from the user and input those datasets to the hierarchic algorithm and process it to produce the output

- Start by assigning each item to a cluster. Let the distances (similarities) between the clusters the same as the distances (similarities) between the items they contain.
- Find the closest (most similar) pair of clusters and merge them into a single cluster, so that now you have one cluster less.
- Compute distances (similarities) between the new cluster and each of the old clusters.
- Repeat steps 2 and 3 until all items are clustered into a single cluster of size N.
- Step 3 can be done in different ways, which is what distinguishes single-linkage from complete-linkage and average-linkage clustering.

In single-linkage clustering we consider the distance between one cluster and another cluster to be equal to the shortest distance from any member of one cluster to any member of the other cluster. If the data consist of similarities, we consider the similarity between one cluster and another cluster to be equal to the greatest similarity from any member of one cluster to any member of the other cluster. In complete-linkage clustering we consider the distance between one cluster and another cluster to be equal to the greatest distance from any member of one cluster to any member of the other cluster.

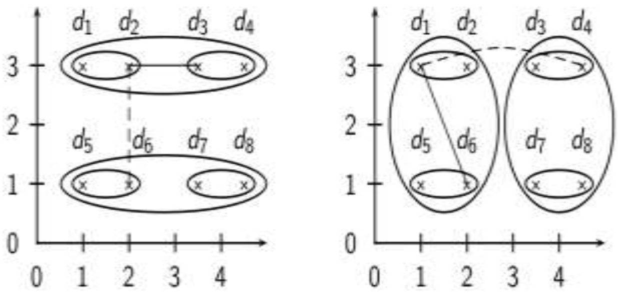

Fig. 4 Single-link vs. Complete link clustering

V. SIMULATION RESULT

We evaluate the performance of our proposed clustering based energy efficient scheme using energy saving percent as the criterion. In addition, we also investigate the impact of the detection probability P_d and the false alarm probability P_f

Consider 50 sensor nodes and 3 PUs randomly distributed in a square field with a length of 100m as shown below .We model the wireless channel between the cognitive sensor and the PU using a free-space path loss model. In order to measure the energy consumption for spectrum sensing, we used the same energy model. It is observed from Fig 5 that the optimal number of clusters is 5, which maximums the objective function. Fig 6 shows the sensor nodes distributed in 100 x 100m after using clustering, different color shows different cluster and the 3 red squares represent PUs. Intra-cluster similarity of each cluster is presented in Fig 7 where we can see each WCSk is greater than 90%. The energy saving percent is plotted in Fig 8 which shows that the total spectrum sensing energy can be reduced more than 70%. Each sensor node may not sense the licensed channel state every time since after clustering, only the sensor node which has highest P_d to share its sensing information to others in same set is chosen for spectrum sensing. However, the similarity between the chose node and others cannot attain 100%, such that each node may spend a little energy on spectrum sensing when collision happens. Spectrum sensing is to ensure a high detection probability. Therefore, the detection probability and false alarm probability are investigated with our proposed scheme. Fig 9 and Fig 10 show the P_d and P_f comparison between before and after clustering for various detection thresholds.

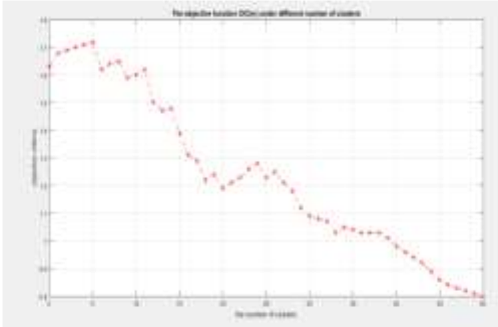

Fig. 5 Objective function OC(m)

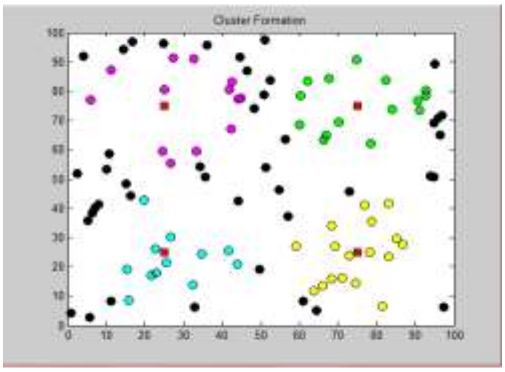

Fig. 6 Optimal Clustering

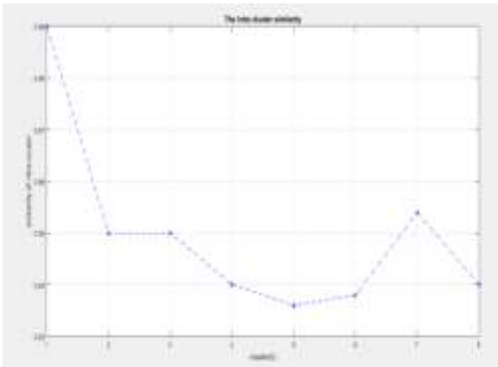

Fig. 7 Intra-Cluster similarity

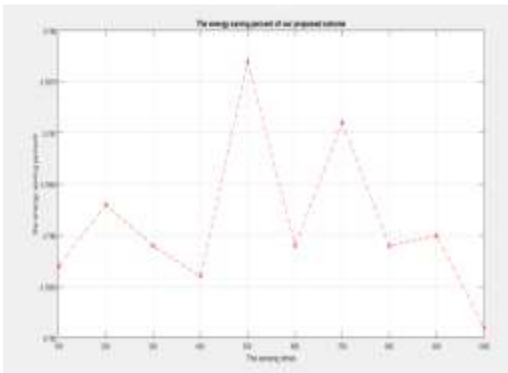

Fig. 8 Energy saving percent

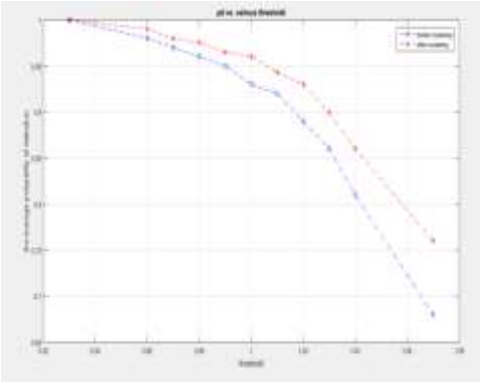

Fig. 9 P_d vs thresholds

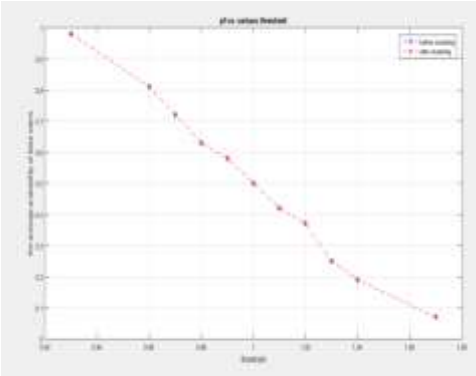

Fig. 10 P_f vs thresholds

VI. CONCLUSION

We propose a clustering based spectrum sensing scheme in CR-WSN, which can significantly reduce energy consumption by involving only one sensor node in a cluster for spectrum sensing instead of all, and an improved hierarchical clustering algorithm, as well as new similarity metric has also been presented. Moreover, we have introduced new objective function, which has been shown to work reasonably well in determining the number of clusters for our hierarchical clustering algorithm. On the basis of power spectral density of the channel which can be used cognitively to search the available spectral gaps those can be used to new incoming users (SU) thus improving the overall channel's throughput. In this work the energy detection spectrum sensing using FFT within the specified frequency band is performed.

REFERENCES

[1] F. Akyildiz, W. Su, Y. Sankarasubramanian, and E. Cayirci, "A survey on sensor work," IEEE Communications Magazine, vol. 40, pp. 102-114, Aug. 2002.

[2] Cavalcanti, D.; Das, S.; Jianfeng Wang; Challapali, K.; , "Cognitive Radio Based Wireless Sensor Networks," Computer Communications and Networks, 2008. ICCCN '08. Proceedings of 17th International Conference on , vol., no., pp.I-6, 3 -7 Aug. 2008.

[3] S. Maleki, A. Pandharipande, and G. Leus, "Energy efficient distributed spectrum sensing for cognitive sensor networks,"IEEE Sensors J., vol. II, no. 3, pp. 565-573, Mar. 2011.

[4] S. Gao, L. Qian, and D. R. Vaman, "Distributed energy efficient spectrum access in cognitive radio wireless adhoc networks,"IEEE Trans. Wireless Commun., vol. 8, no. 10, pp. 5205-5213, Oct. 2009.

[5] S. Izumi, K. Tsuruda, T. Takeuchi, H. Lee, H. Kawaguchi, and M. Yoshimoto, "A low-power multi resolution spectrum sensing architecture for a wireless sensor network with cognitive radio," in Proc. SENSORCOMM, Athens/Glyfada, Greece, Jun. 2009.

[6] Xu Rui, Donald Wunsch II, Survey of Clustering Algorithms, IEEE Transactions on Neural Networks, 2005, 16(3): 645-678. [II] Guha, S., R. Rastogi & K. Shim. CURE: An Efficient Clustering Algorithm for Large Databases. In Proc. Of ACM SIGMOD Intl. Conf. on Management of Data,pp. 73 -82, 1998.