https://doi.org/10.31449/inf.v50i2.9819 Informatica **50** (2026) 247–260 **247**

K-means Clustering for Cluster-Based Cooperative Spectrum Sensing under Rayleigh Fading in Low SNR Conditions

Aakanksha Sharma, Shweta Pandit, Rajiv Kumar
Department of Electronics and Communication Engineering, Jaypee University of Information Technology, Solan, India
E-mail: sharmaak16@gmail.com, shweta.pandit@juit.ac.in, rajiv.kumar@juit.ac.in

Keywords: Cluster-based collaborative spectrum sensing, K-means clustering, rayleigh fading, inter-fusion, intra-fusion, overhead, latency, ROC curve

Received: June 21, 2025

The paper discloses the analysis of K-means clustering based cooperative spectrum sensing using energy-based detector for Rayleigh fading channel under low SNR (0 dB to -25 dB) using ROC curves with different rules for fusion "OR-OR, OR-AND, AND-OR and AND-AND fusion rules". As the current communication systems such as IEEE 802.22 WRAN need to operate with better performance parameters in low SNR conditions, therefore our proposed system considers the similar practical environment. In the proposed system, K-means clustering is used to cluster the secondary users with secondary user closest to centroid selected as cluster head. The clustering technique has provided 70-85% of data overhead reduction in comparison to cooperative or collaborative spectrum sensing. The probability-of-detection doubles in comparison to non-cooperative spectrum sensing at 0.1 probability-of-false-alarm. Further, the comparative analysis is done with other clustering techniques (DBSCAN and LEACH) in terms of data overhead, latency and probability-of-detection. The proposed system has provided an advantage of reduced data overhead, deterministic cluster selection and better performance in terms of detection with OR-AND rule-based fusion in low SNR conditions in comparison to other clustering techniques. Further, the performance variation is also analysed with respect to number of clusters and it has been observed that the detection efficacy of OR-OR and AND-AND rule-based fusion remains unaffected by the variation in number of clusters. However, the detection capability of OR-AND/AND-OR rule-based fusion decreases/increases with the increase in clusters. Further, the OR-AND rule-based fusion performed better than AND-OR rule-based fusion till a certain number of clusters in the network and if clusters increase further, AND-OR rule provided superior performance compared to OR-AND rule. All the simulations were conducted in MATLAB 2024b using TCP-IPv6 model assumptions.

Povzetek: Članek predstavlja metodo sodelovalnega zaznavanja spektra s K-means združevanjem, ki pri nizkem razmerju signal–šum zmanjša podatkovno obremenitev in izboljša zaznavanje signalov.

1 Introduction

The next generation communication networks pose a high burden on the available spectrum. The current spectrum distribution is not optimized as the primary users (PUs) that own the license for spectrum do not always use it. Such under usage of licensed spectrum is fully benefited using radio network having cognitive capabilities, where secondary or unlicensed users (SUs) sense the spectrum for inactive PUs and obtain opportunistic access to the ideal spectrum without obstructing PU communication. The spectrum detection or sensing is therefore a decisive step for cognitive radio network (CRN) as the communication efficiency depends on the accuracy of spectrum detection. There are various methods available for spectrum detection such as matched filter, energy-based detection, eigen-value detection, cyclo-stationary feature detection, wavelet detection and entropy-based detection [1-4]. However, energy-based detection is commonly applied method for spectrum detection because of its straightforward implementation and less cost. In energy-based detector, the energy of signal received at a node is quantified over a defined bandwidth and time.

Further, the spectrum sensing with single SU is not that reliable because of issues like shadowing and/or concealed terminal and thus cooperative spectrum sensing (CSS) is introduced to address the above-mentioned limitations in single SU based sensing [2-7]. In CSS, many SUs participates and collaborate with each other to perform spectrum sensing [8]. The SUs reports the sensing result to fusion center (FC) using either of these combining methods:

(i) Data combining: Each SU forwards amplified received signal from PU to the FC for final decision [3-7].

(ii) Decision combining: Every SU decides whether or not PU is present, and the individual findings are communicated to the FC for final conclusion on the existence of PU [2,9].

Considering decision combining, which we have considered in our work, the decision of each SU on PU presence is communicated to FC and the FC gives the final verdict on spectrum availability. The decision combining at FC can be done by either of the following fusion rules [10,11]:

(i) OR fusion rule: Under this rule, FC indicates "status active" of the PU when at the minimum one SU notices the PU signal.
(ii) AND fusion rule: In the AND rule, the FC declares "status active" of the PU signal only if each SU independently detects the PU's presence.

The CSS with high user density faces data overhead issue. This data overhead can be reduced by grouping the SUs into clusters as only cluster-head (CH) communicates with the FC instead of all the collaborating SUs. This scenario is named as cluster-based cooperative/collaborative spectrum sensing (CB-CSS) and various cluster formation techniques have been discussed in literature. Clustering approach is more energy efficient and minimize the resource need by reducing the SUs reporting to the FC [12-15]. Further, there are several machine learning techniques that are used for more efficient spectrum sensing. Most of the literature uses machine learning techniques for making the final decision on PU existence [16-22]. The ideal number of clusters to obtain maximum throughput for CB-CSS over faded channels has been determined by the authors in [20]. Moreover, the selection of optimal amount of SUs from the entire network and using affinity-propagation to cluster the selected SUs is mentioned in [21]. Further, the performance evaluation of proposed algorithm in [21] based on energy consumption, energy efficiency and throughput are done. In addition, in [22], a distributed CSS using clustering approach over several fading channels is proposed. Authors in [23] have investigated CB- distributed CSS over Nakagami channel with four fusion rules (AND-AND, OR-OR, OR-AND and AND-OR) and compared its performance to centralised CSS. The CH selection is done using low-energy adaptive clustering hierarchy (LEACH) protocol. Further, authors have provided performance improvement by employing diversity combining schemes. In addition, authors in [23] have evaluated the performance of centralised CSS and distributed CSS over Nakagami channel where the reporting channel noise is considered. A CB-CSS based on the Bayesian learning to reduce rate loss and cooperative overhead while simultaneously enhancing spectrum sensing performance is proposed in [24]. Authors in [24] have done classification of the sensing nodes and CH selection using K-means learning based clustering algorithm. By reducing the overall Bayesian cost, authors in [24] have obtained the optimal sensing cut-off or threshold for the intra-CSS. Further, authors in [25] uses centralised LEACH protocol for cluster formation and proposed a multi-hop protocol that reduces power consumption, increase network lifetime and improve sensing performance. However, authors in [26] used K-means clustering to form clusters in the rings that are made across the radius of the network. The performance evaluation is done with respect to optimal number of rings and not in reference to number of clusters in [26]. Further, the memorial K-means clustering based CSS analysis is done under low SNR regime in [27]. Authors in [27] have proposed Memorial K-means Clustering using Pearson-Correlation-Coefficient to acquire feature values for Rayleigh faded channel. Moreover, authors in [28] used K-mean clustering for cluster formation and proposed an algorithm with sensor exclusion (temporary and permanent) and dynamic selection of CH that reduces energy consumption. However, in [28], the performance analysis is done in terms of number of sensing cycles versus probability-of-detection and number of live SUs. The mentioned state of art is also summarized below in Table 1. The K-means learning based algorithm classify or group sensing nodes, form clusters and select CHs, however performance analysis based on receiver operating characteristics (ROC) curve for different fusion rules under low SNR is not mentioned in the literature.

Table 1: Summary of the key prior works

Ref. No.	Technique used	Channel Model	Performance metric	Key findings
[20]	Not mentioned	Nakagami, Rician, Rayleigh and Weibull fading channels	Optimum number of clusters, Throughput	The ideal amount of clusters to obtain maximum throughput depends on the fading environment and fusion rules.
[21]	Affinity-propagation (AP)	Rayleigh fading	Energy consumption, Energy efficiency, Throughput	AP based clustering provides better performance as compared to usual clustering methods (fuzzy c-means clustering and LEACH) in terms of energy efficiency.
[22]	Not mentioned	Nakagami, Rayleigh and Rician fading channels	Detection probability, ROC curve, optimum threshold	The distributed OR-OR fusion rule outperforms conventional fusion rules; the likelihood of detection rises as the number of clusters grows; optimum detection threshold for minimum error probability is determined.
[23]	LEACH	Nakagami fading	Detection probability	The OR-OR fusion rule has superior performance than other fusion rules and centralized CSS for distributed CSS; the detection ability of distributed CSS can be enhanced by utilizing

				diversity, and the detection likelihood rises with increase in CR users.
[24]	Bayesian learning, K-means learning	Rayleigh fading	Rate loss, cooperative overhead, optimum threshold	In addition to reducing rate loss and cooperative overhead, Bayesian learning-based intelligent clustering CSS enhances sensing performance for both imperfect and perfect sensing reports.
[25]	LEACH-C	Rayleigh fading	Power consumption, network lifetime, sensing performance	Multi-hop CSS provides reduced transmission energy consumption. As the number of hops is increased, sensing accuracy is improved but power consumption is increased. Thus, trade-off between accuracy, delay and energy consumption is required.
[26]	K-means clustering, In-network data aggregation	Not mentioned	Optimal number of rings, Energy consumed, Network lifespan, Collision rate, Latency	Network lifetime is extended by selecting optimal number of clusters and applying in-network data aggregation.
[27]	Memorial k-means clustering	Rayleigh fading	Detection probability	Proposed memorial K-means clustering provides 5% to 12% increase in probability of detection in low SNR regime in comparison to other methods.
[28]	K-means clustering	Rayleigh fading	Energy consumption	Energy consumption is reduced without compromising detection probability by proposed algorithm, thus increasing the network's lifespan. The proposed algorithm combines sensor exclusion (temporary and permanent) along with the dynamic CH selection.
Proposed	K-means clustering	Rayleigh fading	ROC curve, Data overhead, Latency	K-means clustering based CSS provides reduced data overhead, better detection probability in low SNR regime. The detection performance of OR-OR and AND-AND fusion rule remains unaffected by the variation in amount of clusters and detection ability of OR-AND/AND-OR decreases/increases with the increase in amount of clusters. Further, OR-AND rule performs better than AND-OR fusion rule till a certain value of number of clusters. This value of number of clusters may vary with the change in network configuration even with same number of SUs.

The novelty of our article lies in the ROC curve-based performance analysis of K-means clustering technique used for grouping the SUs in CRN over Rayleigh fading channel in low SNR regime using different rules for fusion (AND-AND, OR-OR, AND-OR and OR-AND). The ROC curve variation in relation to number of clusters was studied with an objective to understand the influence of varying number of clusters on detection probability using specific fusion rules in such low SNR scenarios.
Therefore, proposed study presents the use of K-means clustering in formation of clusters for CB-CSS and analyse the network's detection performance for Rayleigh fading channel under low SNR regime using ROC curve. The comparison of proposed K-means clustering to other clustering algorithms (LEACH and Density-Based Spatial Clustering of Applications with Noise (DBSCAN)) is also provided.

The remaining paper is categorized as follows. Section 2 briefly discusses system model for energy detection-based spectrum sensing with CB-CSS. Section 3 gives the performance metrics and their mathematical expressions. Section 4 is dedicated to the analysis of CB-CSS for various fusion rules. Section 5 talks through the simulation results and at last Section 6 concludes the sudy with future scope.

2 System model

The system comprises a radio network including cognitive capabilities with M number of SUs and N_c clusters. M SUs are located randomly over a geographical area. These M SUs are grouped into various clusters using un-supervised machine learning technique (K-means clustering) and the distribution of SUs in each cluster depends upon the clustering algorithm such that first cluster has P_1 number of SUs, second cluster has P_2 and so on. In each of the cluster, the SU closest to centroid is selected as CH. Figure 1 shows the system model with 14 SUs (M=14) grouped into three clusters.

Cluster 1, 2, and 3 have P_1=2, P_2=4 and P_3=8 SUs respectively including the CHs.

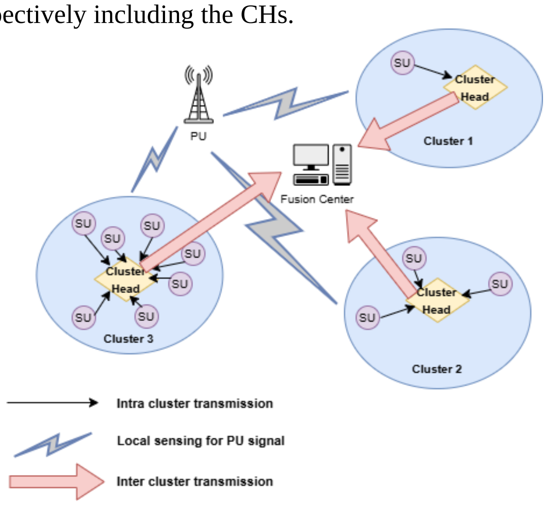


Figure 1: System model for CB-CSS

The SUs in each cluster senses the PU activity and shares the decision with CH which is further communicated to the FC. The communication in such networks mostly uses TCP-IP communication protocol and thus we have considered TCP-IPv6 communication protocol for communication between all nodes [30]. The spectrum detection of the PU is done using energy-based detection, where the incoming signal's energy is calculated and compared to a preset detection threshold. If the calculated energy is below the detection threshold, then the PU signal absence is detected and vice versa. This binary hypothesis for spectrum detection or sensing can be given as: hypothesis H_0 (PU unavailable) and hypothesis H_1 (PU available). The received signal $\mathrm{y}(t)$can be given as [2]

$$\mathrm{y}(t) = \begin{cases} n(t) & : H_0 \\ h * s(t) + n(t) & : H_1 \end{cases} \quad (1)$$

where $s(t)$is the signal transmitted by the PU, h is the sensing channel gain and $n(t)$ is the noise signal (AWGN).

There will be decision combining requirement at two levels. First the decision combining will be at CH, where decision from all the SUs present in that cluster is combined and is known as intra-cluster fusion (here we have used OR fusion rule or AND fusion rule). The second decision combining will be at FC, where decision from each CH is combined using the same or other fusion rule and is known as inter-cluster fusion [20].

3 Performance metrics

In this study, the performance analysis is done based on following performance metrics.

Data overhead: The data overhead depends upon the size of sensing report, control and/or acknowledgement messages in the communication protocol. In this paper, we have considered TCP-IPv6 communication protocol (RFC 8200, [30]). Therefore, the data overhead can be given as

$$\text{Total data overhead} = \text{Application data size} + \text{TCP header size} + \text{IP header size} \quad (2)$$

Latency: The latency in making final decision on spectrum sensing consists of (i) intra-cluster reporting time, (ii) CH processing/aggregation time, and (iii) CH→FC backhaul time. Further, setting up clusters will also add to the latency. Therefore, total latency can be modeled as:

$$T_{total} \approx T_{set-up} + T_{processing} + T_{intra} + T_{backhaul} \quad (3)$$

where, T_{set-up} is the time required for setting up the clusters. $T_{processing}$ is the processing time and is considered constant, i.e, 1 ms. T_{intra} is the reporting time from SUs in the cluster to their CH and is given as

$$T_{intra} = \max_i T_{report,i} \quad (4)$$

where

$$T_{report,i} = T_{propagation,i} + T_{transmission,i} \quad (5)$$

and further can be given as

$$T_{report,i} = \frac{d_i}{c} + \frac{L}{B\log_2\left(1+\frac{P_t {d_i}^{-\alpha}}{N_o}\right)} \quad (6)$$

where $d_i$is the geographical spacing from i[th] SU to CH, c is speed of light, L is the sensing report size in bits, B is the channel-bandwidth, P_t is the transmit power of SU, N_o is the noise power and α is the path loss exponent.

Further, $T_{backhaul}$ is the reporting delay from CH to FC and is given as

$$T_{backhaul} = \max_j T_{report,j} \quad (7)$$

and further can be given as

$$T_{report,j} = \frac{D_j}{c} + \frac{L}{B\log_2\left(1+\frac{P_t {D_j}^{-\alpha}}{N_o}\right)} \quad (8)$$

where D_j is the geographical spacing from j[th] CH to FC.

Further, for ROC curve, following performance metrices are given [2-5].

Probability-of-false-alarm ($\mathrm{P_{fa}}$): The likelihood of a SU detecting that the PU is available when it is unavailable or when the spectrum is actually idle and is given as

$$\mathrm{P_{fa}} = \mathrm{P}(\mathrm{H_1}|\mathrm{H_0}) = \mathrm{P}(\mathrm{y} > \lambda|\mathrm{H_0}) \quad (9)$$

where λ is the cut-off or threshold for detection.

Probability-of-detection ($\mathrm{P_d}$): The likelihood of SU detecting that the PU is available when it is indeed present or when the spectrum is occupied and is given as

$$\mathrm{P_d} = \mathrm{P}(\mathrm{H_1}|\mathrm{H_1}) = \mathrm{P}(\mathrm{y} > \lambda|\mathrm{H_1}) \quad (10)$$

Probability-of-missed-detection ($\mathrm{P_{md}}$): The likelihood of SU detecting that the PU is unavailable when it is actually available and is given as

$$\mathrm{P_{md}} = \mathrm{P}(\mathrm{H_0}|\mathrm{H_1}) = \mathrm{P}(\mathrm{y} \le \lambda|\mathrm{H_1}) \quad (11)$$

and can also be given as

$$\mathrm{P_{md}} = 1 - \mathrm{P_d} \quad (12)$$

Probability-of-error (P_e): The probability of number of errors in detecting the PU signal. It is the sum of $\mathbf{P_{fa}}$ and $\mathbf{P_{md}}$ [23] and is given as

$$P_e = P(H_1)\,P_{md} + P(H_0)\,P_{fa} \quad (13)$$

where $P(H_1)$ is prior-probability that the spectrum is occupied and $P(H_0)$ is prior-probability that the spectrum is idle. It is assumed that there is 50% chance that spectrum is idle and thus, $P(H_0) = P(H_1) = 0.5$.

The received signal from PU follows an iid random distribution with zero mean and variance ${\sigma_s}^2$. The received SNR at the detector for the given channel gain h and with noise variance ${\sigma_w}^2$ is given as [2]

$$\gamma = \frac{|h|^2 {\sigma_s}^2}{{\sigma_w}^2} \quad (14)$$

Further, for large samples value (N), and using central limit theorem, the probability-distribution function of received signal energy follows normal distribution with $N{\sigma_w}^2$ mean and $N{\sigma_w}^4$ variance, in case PU is absent. In case PU is present, and the signal is modulated using phase-shift-keying (PSK) modulation and is having complex values, then the probability distribution function of received signal energy follows normal distribution with $N{\sigma_w}^2(1+\gamma)$ mean and $N{\sigma_w}^4(1+2\gamma)$ variance, where γ is the SNR. For given h, the P_{fa} and P_d is given as [2]

$$P_{fa} = Q\left(\left(\frac{\lambda}{N{\sigma_w}^2} - 1\right)\sqrt{N}\right) \quad (15)$$

$$P_d(\gamma) = Q\left(\left(\frac{\lambda}{N{\sigma_w}^2} - 1 - \gamma\right)\sqrt{\frac{N}{(1+2\gamma)}}\right) \quad (16)$$

where $Q(.)$ is the standard Q-function.

This study focuses on very low SNR regime (0 to -25 dB) and under low SNR assumption ${\sigma_s}^2 \approx {\sigma_w}^2$, that is, the variance of the decision-statistic under hypothesis H_1 is not affected by the signal. The probability distribution function of received signal energy follows Gaussian distribution with $N{\sigma_w}^2(1+\gamma)$ mean and $N{\sigma_w}^4$ variance [2]. For such low SNR, the average P_d for Rayleigh fading channel is given along with its derivation in [29] and the final expression is given below.

$$\overline{P_d^{Rayleigh}} = 1 - \frac{1}{2}\left[\mathrm{Erfc}\left(\frac{N\sigma^2-\lambda}{\sqrt{2N}\sigma^2}\right) - e^{\frac{\frac{1}{\bar{\gamma}^2}+\frac{4}{\bar{\gamma}}\left(\frac{N\sigma^2-\lambda}{\sqrt{2N}\sigma^2}\right)\sqrt{\frac{N}{2}}}{2N}}\mathrm{Erfc}\left(\frac{N\sigma^2-\lambda}{\sqrt{2N}\sigma^2} + \frac{1}{\bar{\gamma}\sqrt{2N}}\right)\right] \quad (17)$$

4 Cluster-based cooperative spectrum sensing

As discussed earlier in section 1 and 2, CB-CSS is introduced to reduce data overhead in CSS by grouping the SUs into clusters and only CH communicates the sensing decision to the FC. The likelihood of PU activity detection for different fusion rules can be given as follows [20, 24].

(i) OR Fusion Rule: For M collaborating SUs, the probability-of-detection with OR rule-based fusion is given as:

$$Q_d^{OR} = 1 - (1 - P_d)^M \quad (18)$$

where, P_d is the probability-of-detection of each SUs reporting to FC and considered equal for all collaborating SUs.

(ii) AND Fusion Rule: For M number of collaborating SUs, the probability-of-detection with AND rule-based fusion is given as

$$Q_d^{AND} = {P_d}^M \quad (19)$$

As the number of SUs collaborating for decision on PU presence increases, there is large amount of data collection and processing required at the FC. In order to mitigate this aforementioned issue, the SUs are grouped using K-means clustering and within each cluster, the SU closest to centroid is selected as CH. Further, the selected CH from each cluster then forwards the sensing decision to the FC instead of all the SUs sending individual decision to FC. The proposed clustering algorithm used in this manuscript based on K-means clustering is given as

Algorithm: K-means clustering for efficient cluster formation and CH selection within the considered CB-CSS framework

1. Input parameters: number (M) and location of SUs, number of clusters (N_c)
2. Random N_c number of centroids will be selected.
3. Distance between each SU and centroids is calculated.
4. Assignment of the SU to the cluster whose centroid is closest is performed.
5. Once all SUs are assigned to clusters, centroids are recalculated.
6. Steps 3 to 5 are repeated until centroids no longer change or convergence is achieved.
7. The algorithm outputs the final cluster and assignment of SUs to a cluster.
8. Calculate the distance of each SU in a cluster to respective centroid and select closest SU as CH.

The distance mentioned in above algorithm is Euclidean distance and algorithm completes clustering process when the centroid converges or is fixed for repeated iteration. Here, we have done the analysis of CB-CSS over Rayleigh fading channel under very low SNR regime with following four fusion rules [20,23]:

(i) OR-OR Fusion rule

In this rule, OR rule-based fusion is considered within the clusters (intra-fusion) and also between FC and CHs (inter-fusion). All the collaborating SUs are independent. The detection likelihood for this rule is given as

$$Q_d^{OR-OR} = 1 - \prod_{j=1}^{N_c}\left(1 - Q_d^{OR(j)}\right) \quad (20)$$

where N_c is the amount of clusters, j is the amount of SUs in a cluster and $Q_d^{OR(j)}$ is the probability-of-detection for OR rule-based fusion for j collaborating SUs and is assumed to be equal for all collaborating SUs.

(ii) OR-AND Fusion rule

In this rule, OR rule-based fusion is used for intra-fusion and AND rule-based fusion is used for inter-fusion. The likelihood of detection for OR-AND rule-base fusion is given as

$$Q_d^{OR-AND} = \prod_{j=1}^{N_c} Q_d^{OR(j)} \quad (21)$$

(iii) AND-OR Fusion rule

In this rule, AND rule-based fusion is used for intra-fusion and OR rule-based fusion is used for inter-fusion. The detection likelihood for AND-OR rule-based fusion is given as

$$Q_d^{AND-OR} = 1 - \prod_{j=1}^{N_c}\left(1 - Q_d^{AND(j)}\right) \quad (22)$$

where $Q_d^{AND(j)}$ is the probability-of-detection for AND fusion rule for j number of collaborating SUs and is assumed equal for all collaborating SUs.

(iv) AND-AND Fusion rule

In this rule, AND rule is used for both intra-fusion and inter-fusion combining. The likelihood of detection for OR-AND rule-based fusion is given as:

$$Q_d^{AND-AND} = \prod_{j=1}^{N_c} Q_d^{AND(j)} \quad (23)$$

The value of Q_d^{OR} and Q_d^{AND} for j^{th} SU can be taken from equation (18) and equation (19), respectively.

The P_d for Rayleigh fading under very low SNR regime and in CB-CSS using above mentioned four fusion rules (AND-AND, OR-OR, OR-AND and AND-OR) can be computed by placing the values of P_d for Rayleigh fading channel as obtained from equation (17) in equation (18) and (19) to get the Q_d for OR and AND fusion rules, respectively. Afterwards, the obtained values for Q_d for OR and AND fusion rules are used to compute the Q_d for OR-OR, OR-AND, AND-OR, AND-AND fusion rules as per the expressions given in equations (20)-(23).

5 Simulation results and discussion

We have performed the simulation on MATLABR2024b. Figure 2 shows data overhead with M=20 SUs for CSS and CB-CSS (with 3, 4, 5 and 6 number of clusters) over TCP-IPv6 communication protocol. The SUs communicates in the form of IP packets over the network. The total data overhead includes the payload size and the header size. The TCP header include 160 bits and IPv6 header include 320 bits and for application data we have considered only one bit as only decision bit (0 or 1) is transmitted from the SUs or CHs to the FC. The data overhead is calculated using equation 2. It is quite evident from figure 2 that data overhead at FC is significantly reduced (70 to 85%) for CB-CSS in comparison to CSS as only CHs communicate with the FC rather than all the collaborating SUs. Further, the data overhead increases with increase in clusters as shown in figure 2. The clusters to be formed can be decided on the basis of data handling capacity of FC.

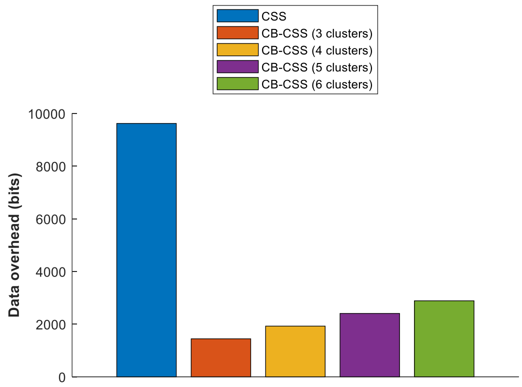
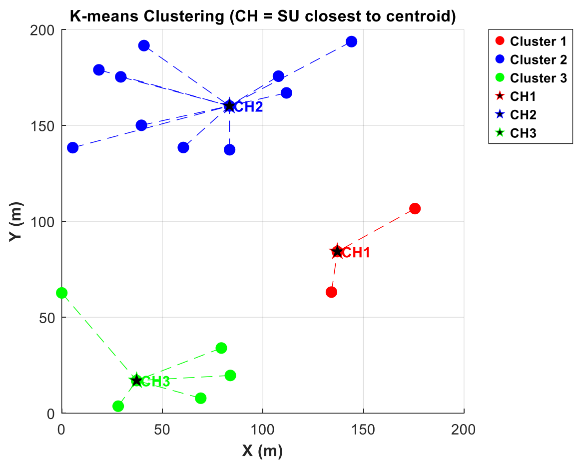
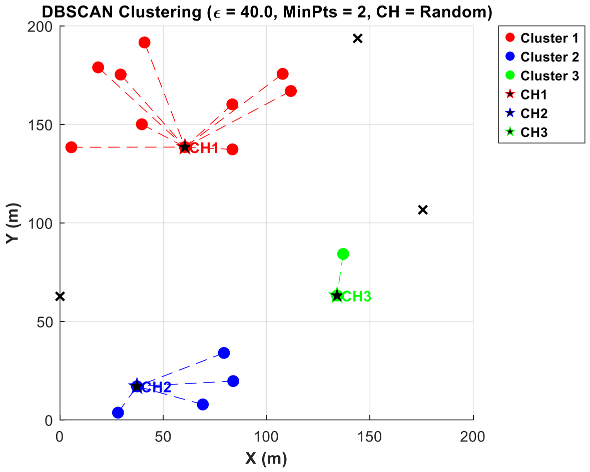


Figure 2: Data overhead comparison for CSS and CB-CSS (with various number of clusters)

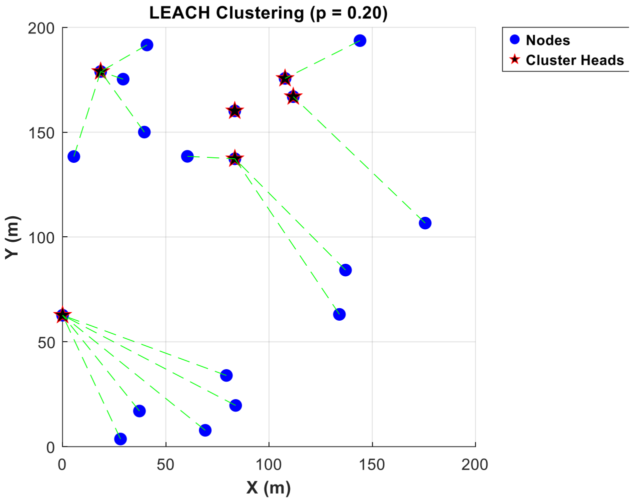
Figure 3 (a) to 3 (c) shows clustering of M=20 SUs with same spatial configuration using K-means, DBSCAN and LEACH protocol, respectively. Figure 3 (a) shows clustering of M=20 SUs using K-means clustering into three clusters. Figure 3 (b) shows clustering of M=20 SUs using DBSCAN clustering having neighborhood radius ε = 40 m and Minpts (minimum amount of SUs necessary to form a cluster) as 2. The neighborhood radius ε is the maximum distance within which SUs are considered neighbors and are grouped into a specific cluster. Further, figure 3 (c) shows clustering of M=20 SUs using LEACH protocol having probability of CH selection, p = 0.2. Furthermore, figure 3 (d) shows data overhead at FC comparison for K-means clustering, DBSCAN (for both cases where outlier SUs are dropped out from CSS and where each outlier SU is considered as CHs) and LEACH respectively. It is clear from these figures that K-means clustering is the better option for clustering as it has more balanced clusters, less data overhead at FC and is deterministic in nature as it provides option to choose the number of clusters.



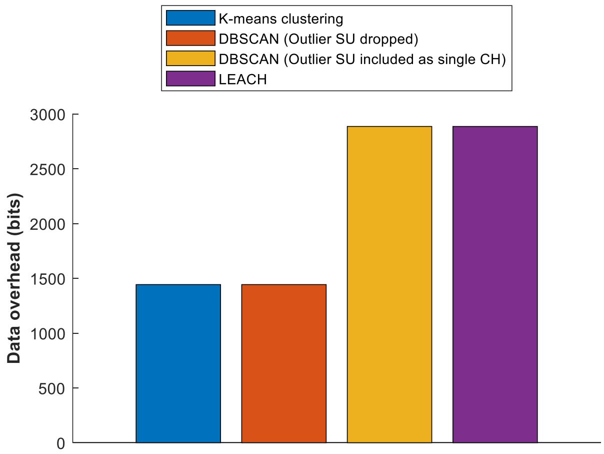
(a) Clustering of M=20 SU using K-means clustering



(b) Clustering of M=20 SU using DBSCAN



(c) Clustering of M=20 SU using LEACH protocol



(d) Data overhead comparison for different clustering techniques

Figure 3: Comparison of various clustering techniques

Figure 4 shows ROC curve for K-means clustering, DBSCAN (when outlier SUs are dropped out of CSS and when outlier SUs are considered as individual CHs) and LEACH for network configuration as given in figures 3 (a), 3 (b) and 3 (c) and SNR of -20 dB. It is validated from figure that K-means clustering technique provides best detection performance with OR-AND fusion rule in comparison to DBSCAN and LEACH in such low SNR.

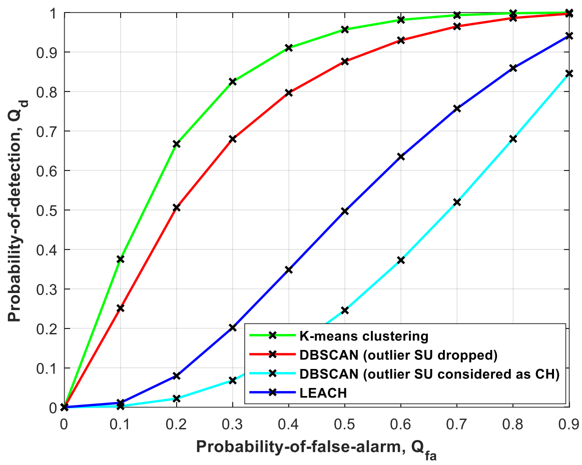


Figure 4: ROC curve-based comparison of different clustering techniques

Table 3 shows the latency and cluster set-up time for K-means clustering, DBSCAN (when outlier SUs are dropped out of CSS and when outlier SUs are considered as individual CHs) and LEACH for network configuration given in figures 3 (a), 3 (b) and 3(c) respectively. The processing time is considered 1ms and set up time is calculated from elapsed time from MATLAB. The latency is calculated using equation (3). The value of other parameters used in the calculation of latency are given in table 2.

Table 2: Simulation parameters for latency calculation

Parameter	Value
speed of light, c	$3*10^8$ m/s
sensing report size, L	161 bits
Channel bandwidth, B	1 MHz
Noise power, N_o	-100 dBm
Transmit power, P_t	0 dBm
Path loss component	3.5

Table 3: Latency values for K-means clustering, DBSCAN and LEACH

Clustering Technique	Cluster Set-up time (seconds)	Latency (seconds)
K-means clustering	0.197118	0.2422
DBSCAN (Outlier is dropped)	0.167157	0.2093
DBSCAN (Outlier considered as CH)	0.167157	0.2095
LEACH	0.140014	0.18946

The results shows that K-means clustering has higher latency as compared to other clustering techniques. Thus, a tradeoff is there between the detection performance and latency.

Further, we investigated performance of K-means clustering based CSS with different fusion rules. The variables used for simulation outcomes are given in Table 4. The wireless channels between SUs, PU and FC are modeled using Rayleigh distribution. The detection threshold (λ) is computed from the values of P_{fa} using equation (15) which is further used to find the P_d / Q_d in all the results presented further in this section.

Table 4: Simulation parameters

Parameter	Value
Total geographical area, A	40,000 m^2
Total number of SUs, M	20
Number of clusters, N_c	3
Number of samples, N	2000
Sensing SNR, γ	0 to -25 dB
Probability-of-false-alarm, P_{fa} / Q_{fa}	0 to 1
Detection threshold	2057.3 to 1942.7 approx.

Further, figure 5 shows clustering of 20 SUs into three clusters using K-means clustering technique and where SU closest to centroid in each cluster is selected as CH. The SUs are randomly located within an area of 40,000 m^2.

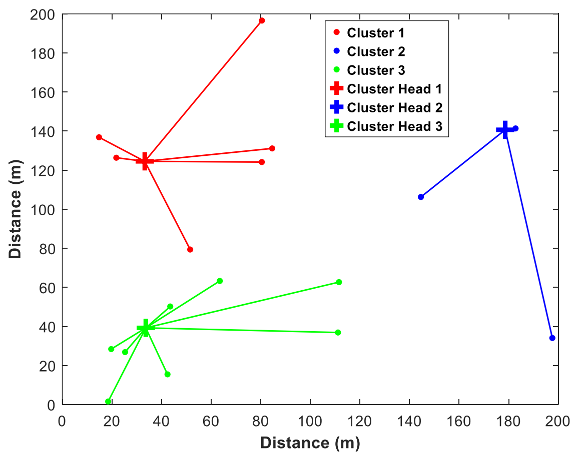


Figure 5: Clustering of M=20 SUs into three clusters using K-means clustering

figure 6 compares the detection performance for non-CSS, CSS (OR and AND rule-based fusion) with 20 SUs and CB-CSS (OR-OR, OR-AND, AND-OR and AND-AND rule-based fusion) with 20 SUs grouped into 3 number of clusters as indicated in figure 5. In figure 6, the probability-of-detection for non-CSS is designated as P_d and as Q_d for other two cases. As is illustrated in Figure 6, the cooperative OR rule with 20 SUs and cluster based OR-OR fusion rule with 20 SUs and 3 clusters are performing exactly same and has outperformed other sensing techniques because even if a single SU out of all the collaborating SUs detects the PU activity, the FC gives final decision of "PU present" leading to higher P_d / Q_d. Further as shown in figure 6, OR-AND fusion rule in CB-CSS has better performance than non-CSS. However, the non-CSS performs better than AND-OR fusion rule with an exception at high values of Q_{fa} ($Q_{fa} \geq 0.87$). The CSS with AND rule-based fusion and CB-CSS with AND-AND rule-based fusion have worst detection performance as FC gives final decision of PU presence only if all SUs collaborating for detection finds the PU activity.

The CB-CSS has advantage of significant reduction in data overhead at FC as shown in figure 2.

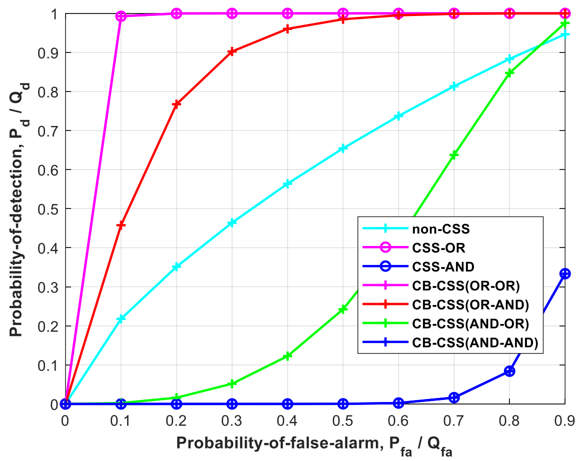


Figure 6: Comparison of ROC curves for non-CSS, CSS and CB-CSS

Further, figure 7 shows the variation of P_e with detection threshold for non-CSS, CSS with OR and AND rule-baed fusion and CB-CSS with AND-AND, OR-OR, OR-AND and AND-OR rule-based fusion for three clusters and spatial configuration as shown in figure 5. The simulation outcomes shows that CSS with OR fusion rule and CB-CSS with OR-OR fusion rule has least P_e and CSS with AND fusion rule and CB-CSS with AND-AND fusion rule has maximum P_e. Thus, the optimum value of threshold for least P_e for a fusion rule can be determined.

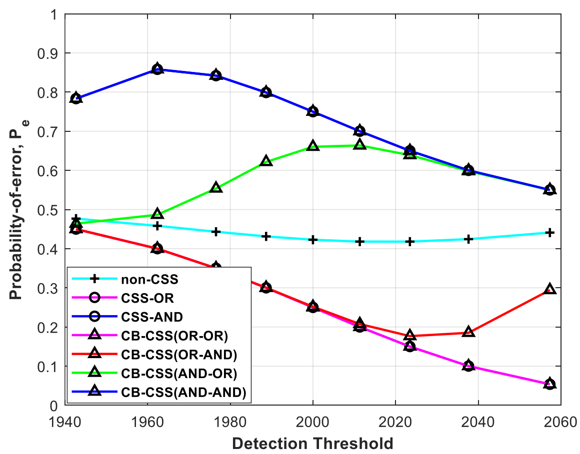


Figure 7: Variation of P_e with threshold for non-CSS, CSS and CB-CSS

Figure 8 shows variation of P_e with detection threshold for different number of clusters in CB-CSS and various fusion rules. It is clear from the figure that P_e is not affected by number of clusters for OR-OR and AND-AND fusion rule. Further, P_e increases/decreases with increase in clusters for OR-AND/AND-OR rule-based fusion, respectively. Moreover, P_e for OR-OR and OR-AND fusion rule first decreases and then increases with increase in threshold. Also, P_e first increases and then decreases with increase in threshold.

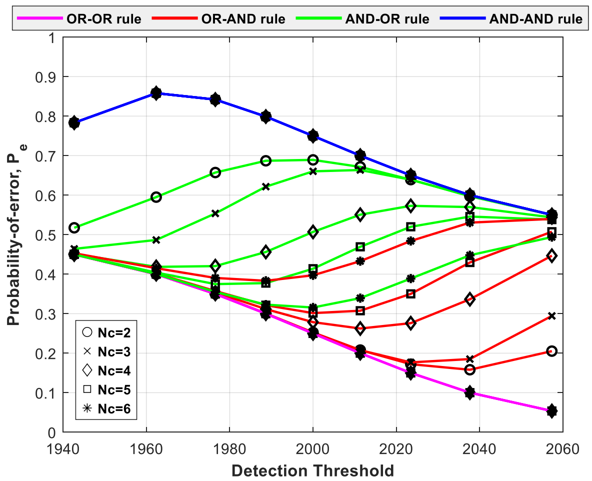
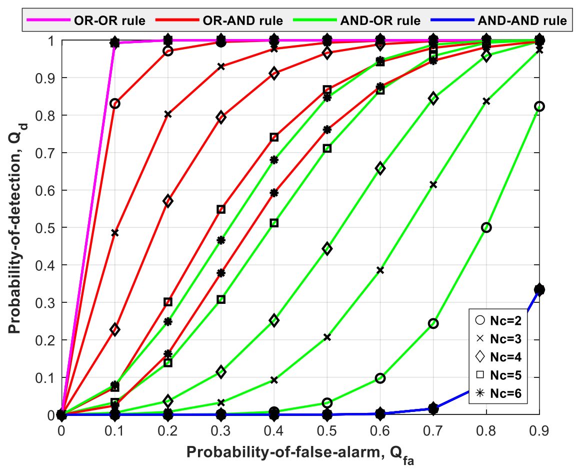


Figure 8: Variation of P_e with threshold and number of clusters for CB-CSS with different fusion rules

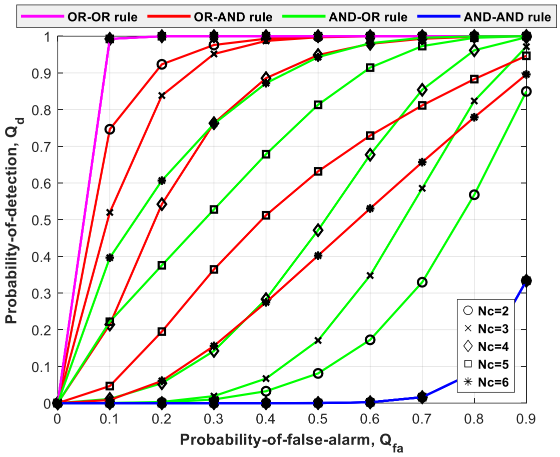
Further, we wanted to look into the effect of number of clusters on detection performance with different fusion rules (AND-AND, OR-OR, OR-AND and AND-OR fusion rule). The detection performance for two different spatial configurations of 20 SUs is shown in figures 9(a) and (b). The spatial configurations are given in Appendix section (figure 11 and 12). The parameters that were used in simulation are same as shown in Table 4. From figures it is quite evident that the detection performance for OR-OR and AND-AND rule-base fusion remain unaffected by the change in number of clusters. The OR-OR rule-base fusion has superior performance in terms of detection than other fusion rules and the AND-AND rule-based fusion show the worst detection performance. Further, the P_d for OR-AND/AND-OR rule-based fusion decreases/increases with the increase in number of clusters, respectively. This is explicable with the previous studies [29] which shows that in OR combining, the detection capability improves with the rise in number of SUs and in AND rule, detection performance decreases with the rise in amount of SUs.

Moreover, as the clusters increases for fixed number of SUs, the CH count also increases and the amount of SUs in each cluster decreases. Since, in OR-AND/AND-OR rule-based fusion, the intra-cluster combining is done with OR/AND rule and inter-cluster combining uses AND/OR rule, therefore, with increase in clusters, the amount of SUs in each clusters decreases and detection performance also decreases/increases as it uses OR/AND rule-based fusion, respectively. Further, with increase in clusters, the number of CHs collaborating for spectrum sensing increases and detection performance decreases/increases as it uses AND/OR rule-based fusion, respectively. Thus, in conclusion the detection performance for OR-AND/AND-OR rule-based fusion decreases/increases with the rise in clusters, respectively.

Further, it is observed that the detection performance of OR-AND rule-based fusion is better than AND-OR rule-based fusion till a certain value of clusters. This can be explained as both the fusion (OR-AND and AND-OR rule-based fusion) shows opposite behaviour to each other with increase in clusters, therefore there will be a threshold after which AND-OR fusion rule will have superior detection performance than OR-AND fusion rule. However, this threshold is not same even if the number of SUs (in our case 20 SUs) are same because of different spatial configurations that vary the amount of SUs in clusters. This is validated from figures 9(a) and (b). For first spatial configuration (as shown in figure 11 of Appendix section), the OR-AND rule shows enhanced detection capability in comparison to AND-OR rule till cluster count reaches five and for six or more clusters, the performance behaviour is reciprocated as shown in figure 9 (a). However, for second spatial configuration (as shown in figure 12 of Appendix section) with same number of SUs (M=20) grouped into same number of clusters (two, three, four, five and six number of clusters, respectively), the cluster count at which AND-OR rule-based fusion surpasses the performance of OR-AND rule-based fusion is different than that seen in figure 9 (a). The OR-AND rule shows improved performance in detection as compared to AND-OR rule till number of clusters reaches four and for five or more clusters, AND-OR rule has better detection ability as shown in figure 9 (b). Therefore, we can say that optimum amount of cluster for a given fusion rule (OR-AND, AND-OR) in order to obtain better detection performance, vary on spatial configuration and is not fixed for a given amount of SUs. Thus, it can be concluded that both the factors, i.e., cluster count and the count of SUs in each cluster affect the detection capability of OR-AND and AND-OR fusion rules.



(a) ROC curve for first spatial configuration



(b) ROC curve for second spatial configuration

Figure 9: ROC curve for CB-CSS for different number of clusters with various fusion rules.

The above simulations were carried out with an assumption that all SUs have same sensing SNR thus having same detection probability. Further, we studied the effect of varying sensing SNR at different SUs. Figure 10 shows ROC curve for four cases: (i) all SUs have same SNR values (i.e. -20 dB), (ii) SUs have varying SNR values of -8 dB, -10dB, -15 dB, -17 dB and -20 dB (i.e. SNR ⩾-20 dB), (iii) SUs have varying SNR values of -20 dB, -21 dB, -23 dB, -24 dB and -25 dB (i.e. SNR ⩽-20 dB) and (iv) SUs have varying SNR values of 0 dB, -8 dB, -20 dB, -23 dB and -25 dB (i.e. SNR ⩾-20 dB and ⩽-20 dB).It is evident from the figure that performance of SNR ⩾ -20 dB scenario is best as compared to other cases followed by the SUs having SNR ⩾ and <-20 dB because detection performance is enhanced with increase in SNR. Further, it is evident that if SUs has varying SNR which is less than -20 dB, then detection performance of SUs with varying SNR will be low in comparison to SUs having same SNR (-20 dB) because detection performance degrades with decrease in SNR. This variation of detection probability with SNR was validated in our previous paper as well [29].

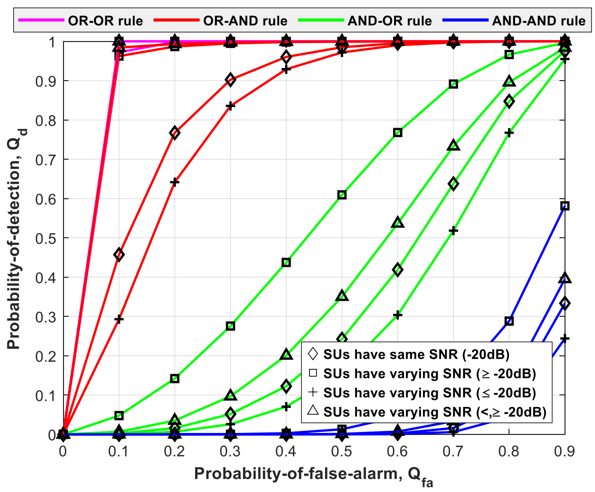


Figure 10: ROC curve for SUs having same SNR values and varying SNR values.

6 Conclusion

In this manuscript we have employed the unsupervised machine learning technique (K-means clustering) for clustering of SUs in the network. The CB-CSS has provided significant data overhead reduction (70 to 85%) at FC in comparison to CSS.

The comparison of K-means clustering with other techniques (DBSCAN and LEACH) is done in terms of data overhead, latency and detection probability. It is concluded that the suggested method is optimal option for low SNR conditions as K-means clustering provides better detection probability as compared to DBSCAN and LEACH in low SNR conditions using OR-AND fusion rule and reduced data overhead at FC. However, latency is large in K-means clustering as compared to DBSCAN and LEACH. Thus, a trade-off is required between detection performance and latency. Further, K-means clustering provides structured grouping and is deterministic, i.e., in K-means clustering, we can choose the number of clusters based on SNR conditions, data overhead handling capacity of FC and SU distribution in the network.

Further, this study examines the detection or sensing performance of CB-CSS for various decision fusion rules for Rayleigh fading channel over low SNR. The OR-OR rule-based fusion has outperformed other fusion rule, however, AND-AND rule-based fusion has shown the worst performance. In addition, OR-AND rule-based fusion outperforms AND-OR rule-based fusion till a certain value of clusters. This value of clusters may vary with the shift in network configuration as the number or distribution of SUs in each cluster may vary. Since, the current communication system is dynamic in nature and therefore it is not necessary to have same amount of SUs in each cluster. The amount of SUs in each cluster may vary with the change in spatial configuration. Thus, we can select the type of fusion rule to be employed depending upon the spatial configuration as mapped to the simulation results.

In conclusion, based on our study, it can be decided that whether to use AND-OR combining or OR-AND combining for a specific cluster count and with a specified number or distribution of SUs in each cluster. Also, number of clusters to be formed can be decided on the basis of data handling and processing capability of FC.

The P_e based analysis in comparison to threshold is done to evaluate the optimum threshold for least P_e. Further, it indicates that P_e increases/decreases with increase in clusters for OR-AND/AND-OR rule-based fusion, respectively and P_e for OR-OR and AND-AND rule-based fusion is not affected by number of clusters.

Further, influence of varying sensing SNR values at SUs in comparison to same SNR at each SU assumption is studied and it is concluded that detection performance increases in case SNR values are greater than same SNR assumption and vice versa.

This study provides the detection performance analysis for K-means clustering based grouping of SUs in CB-CSS. The fuzzy-based adaptive clustering can be seen as a potential future extension as fuzzy-logic–based adaptive clustering can dynamically tune clustering parameters (e.g., number of clusters) in response to fluctuating SNR conditions or SU mobility. Further, computational complexity of clustering algorithm was not addressed in the current study and can be taken into account in future studies.

References

[1] Nepal, N., Shakya, S., & Koirala, N. (2014). Energy detection-based techniques for Spectrum sensing in Cognitive Radio over different fading Channels. *Cyber Journals: Multidisciplinary Journals in Science and Technology. Journal of Selected Areas in Telecommunications (JSAT), 4* (2), 15-22.

[2] Atapattu, S., Tellambura, C., & Jiang, H. (2011). Spectrum Sensing via Energy Detector in Low SNR. *IEEE International Conference on Communications (ICC), Kyoto, Japan*, 1-5. doi: 10.1109/icc.2011.5963316

[3] Atapattu, S., Tellambura, C., & Jiang, H. (2011). Energy Detection Based Cooperative Spectrum Sensing in Cognitive Radio Networks. *IEEE Trans. Wireless Commun., 10* (4), 1232-1241. doi: 10.1109/TWC.2011.012411.100611

[4] Muzaffar, M. U., & Sharqi, R. (2024). A review of spectrum sensing in modern cognitive radio networks. *Telecommun Syst, 85*, 347–363. https://doi.org/10.1007/s11235-023-01079-1

[5] Ganesan, G., & Li, Y. (2007). Cooperative spectrum sensing in cognitive radio,part I: two user networks. *IEEE Trans. Wireless Commun., 6* (6), 2204–2213. doi: 10.1109/TWC.2007.05775

[6] Fan, R., & Jiang, H. (2010). Optimal multi-channel cooperative sensing in cognitive radio networks. *IEEE Trans. Wireless Commun., 9* (3), 1128–1138. doi: 10.1109/TWC.2010.03.090467

[7] Zhang, W., Mallik, R., & Letaief, K. (2009). Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks. *IEEE Trans. Wireless Commun., 8* (12), 5761–5766. doi: 10.1109/TWC.2009.12.081710

[8] Cabric, D., Mishra, S. M., & Brodersen, R. W. (2004). Implementation issues in spectrum sensing for cognitive radios. *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 1*, 772-776. doi: 10.1109/ACSSC.2004.1399240

[9] Letaief, K. B., & Zhang, W. (2009). Cooperative Communications for Cognitive Radio Networks. *Proceedings of the IEEE. 97*, (5), 878–893. doi: 10.1109/JPROC.2009.2015716

[10] Dikmese, S., Lamichhane, K., & Renfors, M. (2021). Novel filter bank-based cooperative spectrum sensing under practical challenges for beyond 5G cognitive radios.*J Wireless Com Network, 30*, https://doi.org/10.1186/s13638-020-01889-w

[11] Dharmapuri, C. M., & Reddy, B. V. R. (2023). Performance analysis of different combining rules for cooperative spectrum sensing in cognitive radio communication. *10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India*,1061-1064.

[12] Joshi, G.P., & Kim, S. W. (2016). A survey on node clustering in cognitive radio wireless sensor networks, *Sensors, 16* (9), 1465, https://doi.org/10.3390/s16091465

[13] Nardis, L. D., Domenicali, D. & Benedetto, M. G. (2009). Clustered hybrid energy-aware cooperative spectrum sensing (CHESS). *4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM'09), Hanover, Germany*, 1–6. doi: 10.1109/CROWNCOM.2009.5189147

[14] Hussain, S., & Fernando, X. (2012). Approach for cluster-based spectrum sensing over band-limited reporting channels, *IET Commun., 6* (11), 1466–1474. https://doi.org/10.1049/iet-com.2010.0510

[15] Guo, C., Peng, T., Xu, S., Wang, H., & Wang, W. (2009). Cooperative spectrum sensing with cluster-based architecture in cognitive radio networks. *IEEE 69th Vehicular Technology Conference (VTC Spring 2009), Barcelona, Spain*, 1–5. doi: 10.1109/VETECS.2009.5073471

[16] Thilina, K. M., Choi, K. W., Saquib, N. & Hossain, E. (2013). Machine Learning Techniques for Cooperative Spectrum Sensing in Cognitive Radio Networks. *IEEE Journal on selected areas in communications, 31*(11), 2209-2221. doi: 10.1109/JSAC.2013.131120

[17] Khamayseh, S., & Halawani, A. (2020). Cooperative Spectrum Sensing in Cognitive Radio Networks: A Survey on Machine Learning-based Methods. *Journal of Telecommunications and Information Technology, 81* (3), 36-46. https://doi.org/10.26636/jtit.2020.137219

[18] Janu, D., Singh, K., & Kumar, S. (2022). Machine learning for cooperative spectrum sensing and sharing: A survey. *Transactions on Emerging Telecommunications Technologies, 33* (1). https://doi.org/10.1002/ett.4352

[19] Kumar, V., Kandpal, D. C., Jain, M., & Debnath, S. (2016). K -mean Clustering based Cooperative Spectrum Sensing in Generalized κ - μ Fading Channels. *22nd National Conference on Communications (NCC), Guwahati, India*, 1-5. doi: 10.1109/NCC.2016.7561130

[20] Sharma, Y., Sharma, R., & Sharma, K. K. (2023). Enhancing Throughput Over Varied Fading Channels for Cluster-Based Cooperative Spectrum Sensing in Cognitive Wireless Networks.*Int. J. Wireless Inf. Networks, 30*, 227–240. https://doi.org/10.1007/s10776-023-00601-1

[21] Bhatti, D. S., Ahmed, S., Chan, A. S., & Saleem, K. (2020). Clustering formation in cognitive radio networks using machine learning. *AEU - International Journal of Electronics and*

Communications, *114*, https://doi.org/10.1016/j.aeue.2019.152994

[22] Sharma, G., & Sharma, R. (2017). Performance evaluation of distributed CSS with clustering of secondary users over fading channels. *International Journal of Electronics Letters*, *6* (3), 288-301. https://doi.org/10.1080/21681724.2017.1357762

[23] Sharma, G. & Sharma, R. (2019). Cluster-based distributed cooperative spectrum sensing over Nakagami fading using diversity reception. *IET Networks*, *8* (3), 211-217. https://doi.org/10.1049/iet-net.2018.5002

[24] Liu, X., Zhang, X., Ding, H., & Peng, B. (2019). Intelligent clustering cooperative spectrum sensing based on Bayesian learning for cognitive radio network, *Ad Hoc Networks*, *94*. https://doi.org/10.1016/j.adhoc.2019.101968.

[25] Kozal, A., Merabti, M., Bouhafs, F. (2014). Multi-hop Cluster Based Cooperative Spectrum Sensing Approach for Cognitive Radio Networks. *International Journal of Scientific & Engineering Research*, 5 (5), 760-769.

[26] Mortada, M. R., Nasser, A., Mansour, A., Yao K. C. (2021). In-Network Data Aggregation for Ad Hoc Clustered Cognitive Radio Wireless Sensor Network. *Sensors (Basel)*, *21*(20), 6741. doi: 10.3390/s21206741.

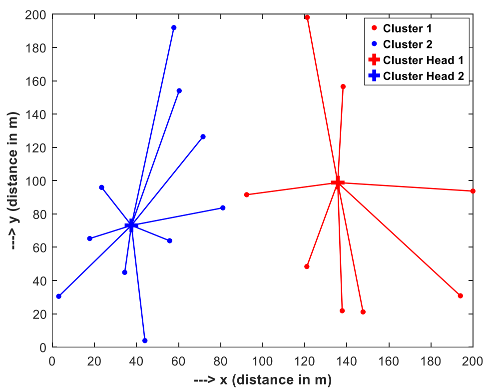
[27] Tao, B., Wu, J., Dou, X., Wang, J., Xu, Y. (2025). Memorial K-means clustering for cooperative spectrum sensing in cognitive wireless sensor networks at low SNR regimes. *Sensor Review*, *45*(3), 443–452, doi: https://doi.org/10.1108/SR-10-2024-0883

[28] Tomás, A., Brito, J.(2025). Algorithm for Sensor Exclusion and Dynamic Cluster Head Selection in Cognitive Radio Networks. *The Twentieth International Conference on Digital Telecommunications*, Nice, France, 5-10.

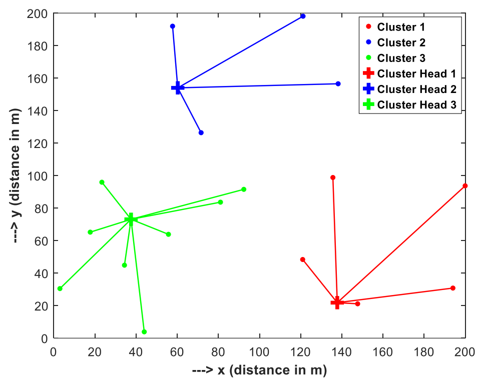
[29] Sharma, A., Pandit, S., Kumar, R. (2024). Cooperative Spectrum Sensing Using Energy-Based Detection for Low SNR Regime over Rayleigh Fading Channel. *2024 International Conference on Integrated Circuits, Communication, and Computing Systems (ICIC3S), Una, India*, 1-6. doi: 10.1109/ICIC3S61846.2024.10602980

[30] Deering, S. E., & Hinden, B. (2017). Internet Protocol, Version 6 (IPv6) Specification, RFC 8200, *Internet Engineering Task Force (IETF)*, https://datatracker.ietf.org/doc/html/rfc8200

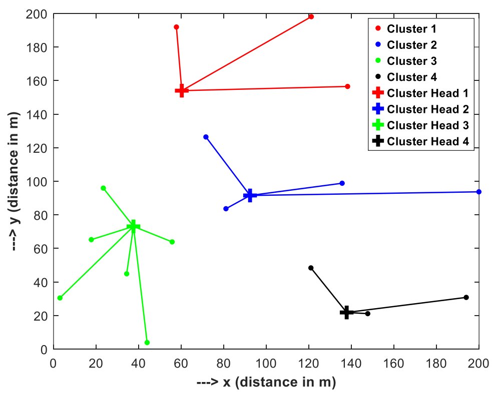
Appendix



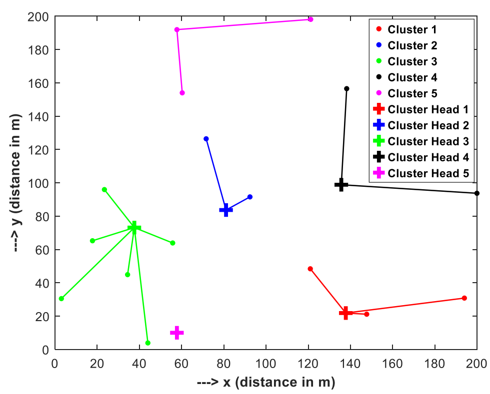
(a)



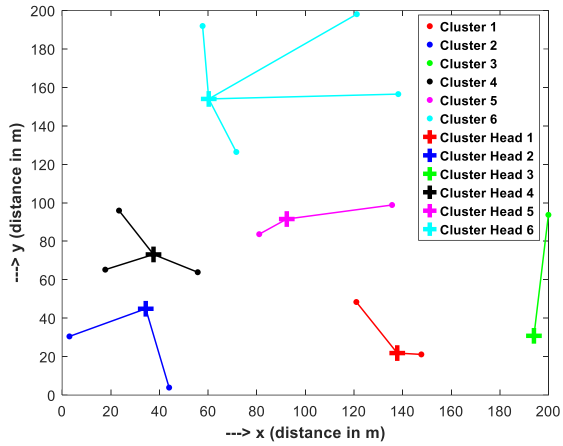
(b)



(c)

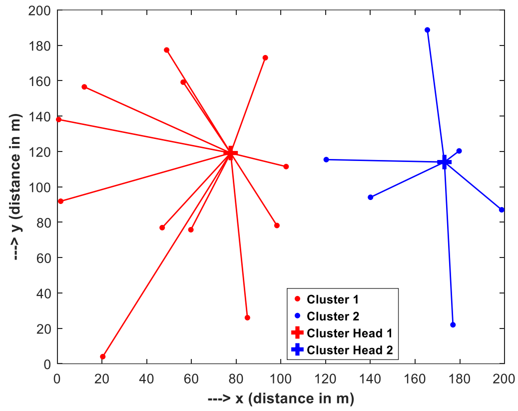


(d)

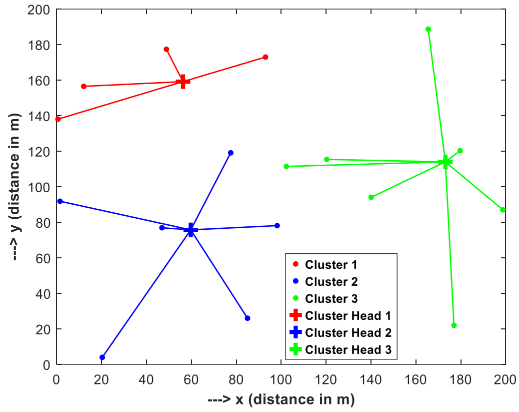


(e)

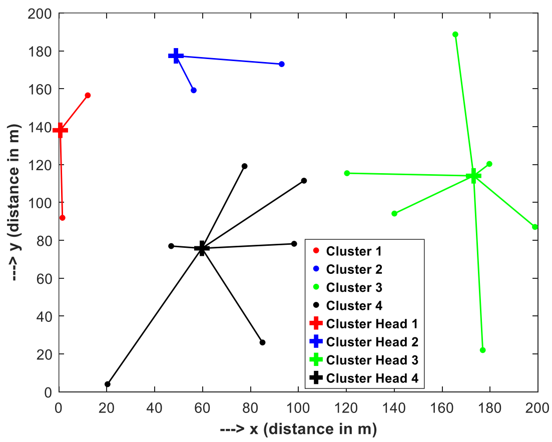
Figure 11: First spatial configuration: Cluster formation of 20 SUs using K-means clustering technique into a) N_c =2, b) N_c =3, c) N_c =4, d) N_c =5 and e) N_c =6 clusters.



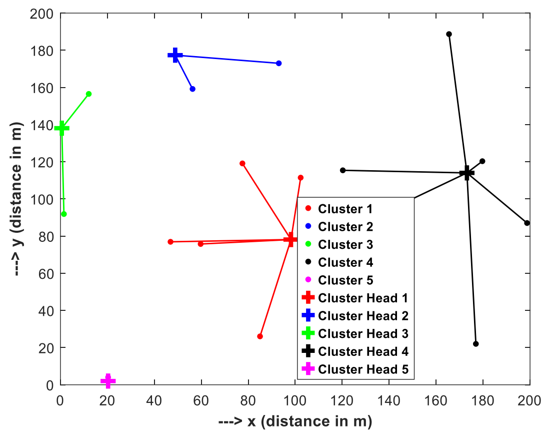
(a)



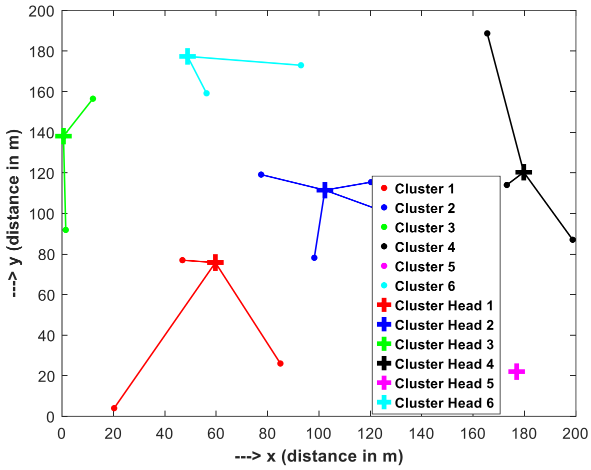
(b)



(c)



(d)



(e)

Figure 12: Second Spatial configuration: Cluster formation of 20 SUs using K-means clustering technique into a) N_c =2, b) N_c =3, c) N_c =4, d) N_c =5 and e) N_c =6 clusters.