CogLEACH: A Spectrum Aware Clustering Protocol for Cognitive Radio Sensor Networks

Rashad M. Eletreby, Hany M. Elsayed and Mohamed M. Khairy
Department of Electronics and Electrical Communications Engineering,
Faculty of Engineering, Cairo University, Egypt
{reletreby,helsayed,mkhairy}@ieee.org

Abstract—The integration of cognitive radios and wireless sensor networks enables a new paradigm of communication in which, sensor nodes can avoid heavily crowded transmission bands by tuning their transmission parameters to less crowded bands thanks to the cognitive radio capabilities. In such setting, sensor nodes act as a secondary user, opportunistically accessing vacant channels within a band originally licensed to a primary user. In this paper, we discuss the problem of how to cluster cognitive radio sensor nodes in a dynamic frequency environment set by the primary users. We introduce Cognitive LEACH (CogLEACH), which is a spectrum-aware extension of the Low Energy Adaptive Clustering Hierarchy (LEACH) protocol. CogLEACH is a fast, decentralized, spectrum-aware, and energy efficient clustering protocol for cognitive radio sensor networks. CogLEACH uses the number of vacant channels as a weight in the probability of each node to become a cluster head. We show that CogLEACH improves the throughput and lifetime of the network compared to the regular LEACH protocol that is operating in the same settings.

Index Terms—clustering, cognitive radio, wireless sensor networks.

I. INTRODUCTION

Due to the scarcity of the available radio spectrum and the rapidly growing demand, cognitive radios emerged as a solution to increase radio spectrum utilization. Cognitive radios allow for the existence of a Secondary User (SU) system that is opportunistically accessing the vacant channels in a band originally licensed to a Primary User (PU) system [1].

The integration of cognitive radios with Wireless Sensor Networks (WSN) is introduced in [2]. In [2] authors identified that this integration could result in a new paradigm of sensor networks communication. In this paradigm, nodes can communicate (by tuning their radios to vacant channels) through a collision-free band instead of communicating through the heavily crowded unlicensed bands. This new communication paradigm opens the door to a new class of applications introduced by authors in [2], such as:

- Multimedia applications where high bandwidth is needed.
- Indoor sensing applications, where unlicensed bands are typically crowded.
- Multi-class heterogeneous sensing applications, where networks with different objectives coexist together.

However, this integration imposes a new protocol design constraint: spectrum-awareness along with the traditional Wireless Sensor Network (WSN) protocol design constraints (e.g. energy and computation constraints).

Due to the energy limitation of a sensor node along with its possible operation in harsh environments, sensors are typically deployed randomly with large population and form a network in an ad-hoc manner. Such a scenario requires an energy-efficient routing protocol that accounts for scalability. Clustering is shown to achieve such constraints and generally enlarges the network lifetime [3].

The resulting network structure is composed of a two level hierarchy: Cluster Head (CH) nodes and member nodes. Each member node is attached to one of the CH nodes according to application-specific objectives (e.g. minimum distance CH to reduce transmission power) resulting in a cluster-shaped network. Member nodes send their data to their corresponding CH nodes instead of sending the data directly to the Base Station (BS) (i.e. reduce communication energy). Subsequently, CH nodes send these packets to the BS either in single-hop [4] or multi-hop [5].

In this paper, we introduce CogLEACH, which is a probabilistic clustering algorithm that uses the number of vacant channels as a weight in the probability of each node to become a cluster head. CogLEACH performs clustering with minimum number of exchanged messages, supports different network and PU models, and accounts for network scalability. CogLEACH is a spectrum-aware extension of the legacy LEACH protocol that is introduced in [4]. We show that CogLEACH improves the throughput and lifetime of the network compared to the regular LEACH protocol that is operating in the same settings.

This paper is outlined as follows. In Section II, we survey the related research work. In Section III, we analytically introduce the CogLEACH Protocol. In Section IV, we describe our model parameters and assumptions, and evaluate the performance of CogLEACH through intensive Simulations. Finally, Section V concludes the paper.

II. BACKGROUND

Low Energy Adaptive Clustering Hierarchy (LEACH) [4] is considered the base line distributed clustering scheme for WSN. The operation of LEACH consists of cycles, where each cycle consists of $\frac{N}{k}$ rounds. In LEACH each node decides its state (CH or member node) autonomously through the
following probability \( P_i(t) \):

\[
P_i(t) = \begin{cases} 
  \frac{k}{N - k * (r \mod \frac{N}{r})} & : C_i(t) = 1 \\
  0 & : C_i(t) = 0
\end{cases}
\]  

(1)

Where \( N \) denotes the total number of nodes, \( k \) denotes the desired number of CHs (on average) per round, \( r \) denotes the current round number, and \( C_i(t) \) is an indicator function determining whether or not node \( i \) has been a CH for the most recent \( r \mod \left( \frac{N}{r} \right) \) rounds. This formula of \( P_i(t) \) maintains that each node is selected as a CH once per cycle of operation and thus balancing the load between nodes and extending time of the first node death. LEACH shows significant improvements to Minimum Transmission Energy (MTE) routing, but LEACH protocol is designed for a WSN operating on a fixed channel and thus not applicable for Cognitive Radio Sensor Network (CRSN).

In [5], CH election is based on nodes residual energy as a primary clustering parameter, and node proximity to its neighbors as a secondary parameter. Hybrid Energy Efficient Distributed (HEED) protocol shows significant improvements to LEACH in terms of lifetime and throughput, but it is designed for fixed channel settings and thus not applicable for CRSN.

In [6], a spectrum-aware clustering algorithm for a cognitive radio ad-hoc network is introduced. The solution is based on finding the optimal biclique graph at each node subject to set of constraints. Clusters are formed with the aim of maximizing the summation of the number of common idle channels with the number of nodes within the cluster, or maximizing the multiplication of the same two parameters, or maximizing the number of nodes within the cluster given a threshold on the number of common idle channels. This solution is deterministic and it provides a stable clustering scheme, but it is NP-Hard and violates WSN main design constraints: energy and computation.

In [7], A spectrum-aware clustering scheme for CRSN is proposed by introducing Group Wise Constraint (GWC) to the legacy K-means clustering algorithm. GWC states that nodes within each cluster must have common idle channels and be in physical proximity relative to each others. The protocol initially sets each node as a CH and then merges CHs within each iteration until the number of CHs reaches an optimal number that was obtained apriori by theoretical analysis. Information about physical proximity and common channels is sent from nodes to their corresponding CHs. Then, a CH node can determine which cluster should be merged with it. Such information requires intensive exchange of messages between nodes and CHs and thus not energy efficient.

### III. COGLEACH Protocol Structure

#### A. Primary User Activity Model

In this paper, we assume a Semi-Markov ON-OFF process, modeling the behavior of the PUs. In Semi-Markov ON-OFF process, the activity of a PU on a given channel alternates between ON state and OFF state, where the time duration of each state is an independent random variable. In [8] the authors verified the validity of modeling the time duration of ON and OFF intervals as two independent geometrically distributed random variables with parameters \( q_i \) and \( p_i \), respectively, as shown in Fig. 1. Authors of [8] showed that the stationary probability of idle channel, \( p_{f} \), for a channel \( i \) is calculated as follows:

\[
p_{f}^{i} = \frac{q_{i}}{p_{i} + q_{i}}
\]  

(2)

In [6], It is shown that the probability of having \( n \) idle channels out of a set consisting of \( m \) channels belongs to a binomial distribution:

\[
Pr[N = n] = \binom{m}{n} p_{f}^{n} (1 - p_{f})^{m-n}
\]  

(3)

assuming that all channels have the same \( p_{f} \).

#### B. Network Model

The network consists of \( N \) CRSN nodes that act as a secondary system existing with \( P \) PUs operating in the same band over a set \( C \) of orthogonal frequency channels. We define Interference Protection Range (IPR) for each Active PU as an area circulating each PU; no other radio inside the IPR of PU\( _i \) could be tuned to the channel used by PU\( _i \) to avoid interference. Thus any CRSN within the IPR of an active PU\( _i \) will sense the channel used by this PU as busy channel. It shall be noted that the terms interference protection range and coverage range are used interchangeably in this paper.

#### C. Spatial and Spectrum Similarity Models

We study different node deployment and operation scenarios in the context of spatial and spectrum similarity. By spatial similarity we mean that all nodes exist within active PU systems. By spectrum similarity we mean that each PU system is modeled after a Semi-Markov ON-OFF process with the same characteristics (i.e. same \( p_{f} \)) for all channels. This classification is important as it will enable us to identify the conditions under which a decentralization of our clustering algorithm will be possible.

1) Spatial and Spectrum Similarity Model: As shown in Fig. 2a, in this model the network is distributed over a large area such that:

- Spatial Similarity: All nodes exist within the coverage range of PU systems, and
Fig. 2: Circles represent WSN nodes, the transceiver represents WSN BS, the triangles represent PUs. a) areas encapsulated by solid lines represent each PU coverage range. b) the circles centered at each PU represent each PU coverage range.

- Spectrum Similarity: Each PU system has the same $p_f$ for each channel, thus the same ON-OFF characteristics for each channel (e.g. $PU_1$ uses the same $p_f$ for each channel), and
- All PU systems use the same $p_f$ (e.g. $p_f$ used by $PU_1$ is similar to that used by $PU_2$).

2) Spectrum Similarity Model: As shown in Fig. 2b, in this model:
- Spatial heterogeneity: some nodes exist within the coverage range of PU systems, and
- Spectrum Similarity: Each PU system has the same $p_f$ for each channel, thus the same ON-OFF characteristics for each channel (e.g. $PU_1$ uses the same $p_f$ for each channel), and
- PU systems use different $p_f$ (e.g. $p_f$ used by $PU_1$ is different to that used by $PU_2$).

This model, where all channels per each PU have the same traffic (ON-OFF) characteristics, is commonly used in literature to describe the behavior of a PU system in a cognitive radio environment to facilitate the analysis [6], [8], [9].

3) Arbitrary Model: In this model:
- Spatial heterogeneity: some nodes exist within the coverage range of PU systems, and
- Spectrum heterogeneity: Each PU system has different $p_f$ for each channel, thus different ON-OFF characteristics for each channel (e.g. $p_f$ used by $PU_1$ on channel 1 is different to that used by $PU_1$ on channel 2), and
- PU systems use different $p_f$ (e.g. $p_f$ used by $PU_1$ is different to that used by $PU_2$).

D. CogLEACH Protocol

As shown in Section II, in the regular LEACH [4], $P_i(t)$ is chosen such that $E\{\#CHs\} = k$.

Therefore:

$$P_i(t) = \begin{cases} \frac{k}{N - k \ast (r \mod \frac{N}{k})} & : C_i(t) = 1 \\ 0 & : C_i(t) = 0 \end{cases}$$ (4)

Now we want to employ the Number of sensed idle channels as a metric in choosing CHs. If node $i$ senses more idle channels than node $j$, node $i$ has better opportunity to find common channels with nearby nodes, which results in a better opportunity to setup a cluster with a common channel.

Our goal is to construct a new form of $P_i(t)$ such that:
- $E\{\#CHs\} = k$, and
- a node with higher number of idle channels is more likely to become a CH.

Let $P_i(t) = \min(k \alpha, 1)$ where $m$ denotes the total number of channels in the used band (i.e. $|C|$) and $c_i$ denotes number of idle channels available to node $i$.

We aim to make $E\{\#CHs\} = k$, but

$$E\{\#CHs\} = \sum_{i=1}^{N} P_i(t)$$ (5)

thus

$$k \alpha \frac{c_1 + c_2 + \ldots + c_N}{m} = k$$ (6)

that leads to:

$$\alpha = \frac{m}{c_1 + c_2 + \ldots + c_N}$$ (7)

thus

$$P_i(t) = \min(k \frac{c_i}{\sum_{j=1}^{N} c_j}, 1)$$ (8)

Now, our objective is to approximate the term $\sum_{j=1}^{N} c_j$ at each node, for each of the spatial and spectrum similarity models, so that each node can run the algorithm locally without the need of a central entity.

1) For Spatial and Spectrum Similarity Model: Let $N$ represents the total number of nodes. The number of channels that could be sensed idle by each node, belongs to a binomial distribution with mean $\mu = p_f \ast m$, thus each node senses $\mu$ idle channels on average.

Thus the summation term could be approximated to $N\mu$, which leads to:

$$P_i(t) = \min(k \frac{c_i}{N\mu}, 1).$$ (9)
Algorithm 1 CogLEACH Algorithm

Given: \( k \) clusters, \( N \) nodes and the set \( C = \{1, 2, ..., m\} \) of frequency channels.

**Phase 1: Sensing**
Each node \( (n_i) \) senses the available spectrum, determines the vacant channels (i.e. calculate the subset \( C_i \), which is a subset of \( C \), and represents the idle channels sensed by node \( n_i \)), and calculate \( c_i \), which denotes the number of elements in \( C_i \) (i.e. \(|C_i|\)).

**Phase 2: CH Decision**
Each node decides whether or not to become a CH according to \( P_i(t) \).

**Phase 3: Cluster Formation**
- Let \( n_i \) be one of the CHs of the current round and \( C_i \) be a subset of \( C \) and denotes the idle frequency channels sensed by node \( n_i \) in Phase 1. \( n_i \) broadcasts a \textit{CH_Tentative Announcement} message including Node ID and \( C_i \) over the Common Control Channel (CCC).
- Node \( n_j \) with a subset of frequency channels \( C_j \) within the transmission range of \( n_i \) and having one or more common frequency channels with \( n_i \) (i.e. \( C_i \cap C_j \neq \emptyset \)) listen to \textit{CH_Tentative Announcement} message over the CCC. Node \( n_j \) sends a \textit{Tentative Join Request} message including the ID of its sensed idle channels \( C_j \) to the CH with the lowest communication cost (based on the received signal strength) over the CCC.
- \( n_i \) (the CH) records all tentative join requests initiated from other non-CH nodes along with the IDs of frequency channels that are included in the message. Afterwards, CH node decides to use a channel \( C_{iq} \) out of its \( C_i \) subset where \( C_{iq} \) is the channel that the majority of requesting nodes has in common with \( n_i \). Subsequently, \( n_i \) broadcasts \textit{CH_Final Announcement} with the ID of the selected channel over the CCC.
- Other non-CH nodes (e.g. \( n_j \)) listen to \textit{CH_Final Announcement} and choose the CH with the lowest communication cost.
- The cluster head node sets up a TDMA schedule and transmits this schedule to the nodes in the cluster.

**Phase 4: Steady State Phase**
As introduced in LEACH [4], the steady-state operation consists of frames, in which nodes send their data to their CH within their allocated transmission slot. As many clusters may use the same channel for their intra-cluster communications, each cluster uses a unique Direct Sequence Spread Spectrum (DSSS) spreading code as introduced in [4].

*Reiterate* After round time, Clustering is triggered again and Phases 1-4 are re-initiated.

Then, each node can decide its role using the CogLEACH algorithm that is shown in algorithm 1, in a decentralized manner.

2) For Spectrum Similarity Model: Let \( N_1 \) represents a subset of nodes that are within the coverage range (that is of radius \( r \)) of any PU system with a predefined \( p_f \) that we call \( p_{f_1} \). (PU systems may have different \( p_f \). In this case we shall define \( N_1, N_2, ..., N_p \) each with its own \( p_f \), and \( N_2 \) represents the disjoint subset of nodes that aren’t within the coverage range of any PU system, thus \( p_{f_2} = 1 \).

Let the field be a square of side length equals to \( M \) meters, and area equals to \( M \times M \), and the number of PU systems be \( P \).

To approximate the summation term of the modified form of \( P_i(t) \) we shall define PU Effective Area as the area covered by the PU system (with no intersections)

\[
A_{eff} = P \times (\pi r^2)
\]

Thus any node can fall within the coverage range of any PU system with a probability \( p_{PU} = \frac{A_{eff}}{M^2} \). Thus the summation term can be approximated as:

\[
\sum_{j=1}^{N_c} c_j = N_1 \times p_{f_1} \times m + N_2 \times p_{f_2} \times m
\]

where \( N_1 = N \times p_{PU} \) and \( N_2 = N \times (1 - p_{PU}) \) which leads to:

\[
P_i(t) = \min(k \sum_{j=1}^{N_c} c_j \frac{c_j}{N_1 \times p_{f_1} \times m + N_2 \times p_{f_2} \times m} + 1)
\]

Then, each node can decide its role using the CogLEACH algorithm that is shown in algorithm 1, in a decentralized manner.

3) For the Arbitrary Model: In this model, nodes will not be able to estimate the summation term in a decentralized manner, and a centralized solution is necessary. In this model nodes obtain the summation term from the BS through two steps:

1) Only after Phase 1 of algorithm 1, each node sends the number of its sensed idle channels (i.e. \( c_i \)) to the BS.
2) The BS later broadcasts a public message containing the summation of all received \( c_i \)'s (i.e. \( \sum_{j=1}^{N_c} c_j \)).

Subsequently each node continues the execution of CogLEACH algorithm that is shown in algorithm 1.

It shall be noted that in CogLEACH, we assume the presence of a Common Control Channel (CCC) and network wide synchronization so that all nodes execute CogLEACH simultaneously. CCC is commonly used in literature to facilitate the exchange of control messages between nodes using a fixed channel (e.g. on other band) that can be accessed by all nodes [8], [10].

IV. PERFORMANCE EVALUATION

A. Simulation Parameters

We consider a network consisting of 100 CRSNs, 10 PUs, 5 channels, \( p_f \) of 0.3, square area of side length 100 meters, IPR radius of 20 meters, packet size of 4000 bits, BS is located in the middle of the area, and the initial energy of each node equals to 0.5 J. Due to the fact that the steady state operation of CogLEACH is similar to that of LEACH [4], the optimal number of CHs (\( k_{opt} \)) is set to 10 according to the derivation in [11], and the energy consumption model introduced in [4].
B. Simulation Results

We compare our results with the original LEACH [4] protocol operating in the same radio settings such that:

- LEACH nodes select any channel (out of the $m$ channels) to use for their transmissions.
- Existence of a nearby PU operating on the same channel leads to a packet drop, and thus reducing the throughput.

It shall be noted that CogLEACH_Model1 refers to the spectrum and spatial similarity model, CogLEACH_Model2 refers to the spectrum similarity model, and CogLEACH_Model3 refers to the arbitrary model.

1) Lifetime: Fig.3a shows the number of alive nodes per round for nodes running CogLEACH and LEACH protocol under different similarity models stated in Section III. Nodes running CogLEACH protocol lasts longer than those running LEACH protocol; this is because of the fact that if a node running CogLEACH protocol senses the whole spectrum as busy (i.e., blockage), it will not transmit any packet (to avoid collisions with PU systems) and thus it saves energy. Nodes running CogLEACH protocol on network model 1 suffer more blockage than those operating in model 2 and 3, as in network model 1 All nodes exist within PU system coverage range.

2) Throughput: Fig.3b shows the total throughput in bits obtained from different protocols and different similarity models. CogLEACH shows a significant improvement in the obtained throughput due to the Spectrum-Awareness property. Nodes running LEACH protocol suffer from packet drop due to colliding with PU systems operating on the same channel. CogLEACH nodes operating in network model 1 shows a remarkable enhancement in throughput relative to LEACH, because in network model 1 all LEACH nodes exist under PU IPR and thus the collision probability increases which results in a severe throughput degradation.

3) Average Number of Cluster Heads: Authors in [11] defined the stable period as the period until the first node dies, as afterwards the optimal election probability becomes invalid due to the change in the number of nodes. Fig. 4 shows the average number of CHs per round under different similarity models. All protocols have an average of 10 CHs during the stable period. The preceding results acknowledges that the new form of $P_i(t)$ produce the required $k_{opt}$ clusters on average.

C. Performance Gain with respect to $p_f$

Intensive simulations were conducted to measure the performance gain of CogLEACH versus that of LEACH [4] under different values of $p_f$ (the stationery probability of the channel being idle). As shown in Fig.5, maximum throughput gain is obtained when $0.3 \leq p_f \leq 0.4$ for the similarity models introduced in section III. It shall be noted that $p_f$ was set to 0.3 for the simulations introduced in Section IV.B.

D. Asymptotic Signalling Analysis

In the legacy LEACH protocol [4], each CH broadcasts an advertisement message to inform other nodes of its role as a CH. Afterwards, each node selects the CH that would require minimum transmission power (i.e. the closest CH). In this scheme, we have $O(k_{opt})$ CH announcements, followed by $O(N - k_{opt})$ join requests. Thus, the cluster formation phase requires $O(N)$ messages in total. Nonetheless, in CogLEACH protocol, that is shown in algorithm 1, we have $O(2k_{opt})$ CH announcements (i.e. tentative and final), followed by $O(2 * (N - k_{opt}))$ join requests. Thus, the cluster formation phase requires $O(2N)$ messages in total.
Fig. 4: Average Number of CHs per rounds Under Different Protocols and Similarity Models

Fig. 5: Performance Gain of CogLEACH versus that of LEACH Under Different Values of $p_f$, Different Protocols and Different Similarity Models

V. CONCLUSION

In this paper we introduce CogLEACH: a spectrum-aware and energy efficient clustering protocol for cognitive radio sensor networks. CogLEACH is based on the legacy LEACH protocol [4] and uses the number of vacant channels as a weight in the probability of each node to become a cluster head. The protocol is shown to satisfy wireless sensor networks limited computation and energy constraints by allowing each node to select its role without the need for intensive cooperation between nodes to setup the clusters. Semi-Markov ON-OFF process is used to model the PU activity, while several similarity models are introduced according to the level of spectrum and spatial similarity, also the conditions under which CogLEACH can operate in a decentralized manner are identified. It is shown that CogLEACH achieves higher throughput and life time compared to LEACH protocol coexisting with PU systems on the same band.

ACKNOWLEDGMENT

This publication was made possible by NPRP grant # [5-250-2-087] from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES


