A Routing Protocol for Cognitive Radio Ad hoc Network

I. INTRODUCTION

Due to the expansion of wireless communications, it is almost impossible to cope with allocation of spectrum as it is limited and finite resource. Therefore, it is vital to ensure the best usage of this limited resource to meet the recent demand of radio frequency. The traditional spectrum allocation methods that are authorized by different regulators are not intelligently handled. It is observed that the licensed spectrum remains unoccupied most of the time [1] where the regulatory body is unable to allocate the frequency for upcoming new wireless devices [2]. This temporarily unused portions in the licensed spectrum are called spectrum holes [3]. Spectrum hole is define as the frequency band that has been allocated to a licensed user but it is unutilized at a particular time or location. To utilize these unoccupied band, resolve the spectrum scarcity for new wireless application and make spectrum allocation more dynamic, cognitive radio (CR) technology is introduced [4]. This intelligent technology has introduced opportunistic access of unlicensed user, known as secondary user (SU), to use licensed band without interfering the licensed user, known as primary user (PU) [5].

Cognitive radio networks (CRNs) can be classified as infrastructure-based CRN and cognitive radio ad hoc networks (CRAHNs) [5]. The infrastructure-based CRN has a central network entity to manage the network such as base-station (BS) in cellular networks. Instead, CRAHN does not have any infrastructure backbone and each CR user can communicate with other CR users through ad hoc connection [6]. Most of the work in the field of CRAHN has focused on channel scarcity problem at the lower layer (PHY, MAC), while routing in CRAHN is largely unexplored. Routing in CRAHN is very important task that have a great effect on the overall performance of the network. Routing in CRAHN differs with routing in traditional ad hoc network as it has to adopt to dynamic changes of spectrum due to stochastic behavior of PU and SU. Moreover, routing protocol in CRAHN must deal with heterogeneity of resources (available channel and available energy).

In this paper, we designed an efficient clustering mechanism and on top of that we proposed a robust routing algorithm to ensure the certain level of QoS in the network. A novel spectrum-aware clustering algorithm is introduced that jointly considers the available spectrum and the power level of the nodes to form the clusters. The proposed clustering algorithm enables the network to be more robust to PUs’ activities as well as node mobility. The proposed routing protocol is a hybrid protocol that uses a proactive method for intra-cluster routing and a reactive method for inter-cluster routing. The path selection in proposed protocol is defined as a multi-objective optimization problem, where proposed approach select the path with the low density. The idea is that packet will face long delay due to collision is highly dense area.

II. SYSTEM MODEL

Fig. 1 shows considered system architecture, in which we assume that there are $N_{SU}$ cognitive radios and $N_{PU}$ primary users deployed in CRSN. A given number of non-overlapping orthogonal channels $\{Ch|Ch, i=1, 2, \ldots n\}$ are available, and each channel has a unique channel ID. Each node is aware of its location and each node has a single half-duplex cognitive radio transceiver, which is capable of detecting and utilizing spectrum holes in a distributed and efficient way. Cognitive radios or SUs coexist with PUs and opportunistically and conditionally access the channels.

A discrete-time Markov chain is employed to PUs’ channel-usage patterns [7], which means that the PU’s may change their state (i.e., channel usage) after each process or step. Channel availability for each node is related to the node’s physical location. Similar to IEEE 802.22 standard, SUs use an available channel only when it is not occupied by PUs. By detecting PU’s presence, SU vacates the channel. The overlay spectrum sharing model [8], is used in our network as simplified interference avoidance model.

A dedicated control channel has been considered in the network. Nodes forming the cluster become cluster heads (CHs), which are responsible for inter-cluster communication as well as intra-cluster channel access control. Inter-cluster communication is relayed by gateway nodes (GW). GWs are the nodes which are in the border of two neighboring clusters and can hear both cluster beacons as shown in Fig. 1. GW inform CH about its status through control channel.

III. CLUSTER FORMATION

The proposed clustering mechanism divides the network into clusters based on three values: spectrum availability, node power level and node current speed. In the proposed clustering scheme, clusters are formed with the neighboring nodes in an ad hoc topology.

Nodes in the proposed architecture exchange their available channels list (ACL$_i$) based on the spectrum sensing information. Each node generates its own neighbor list $N_i$, where $i=1, 2, 3, \ldots, n$ using neighbor discovery mechanism. Afterwards, the cluster formation phase starts. Cluster formation is defined
is in the available channels of CR

creates an undirected bipartite graph $G$ between vertices $x \in (V, E)$ is called bipartite if vertices set $V$ can be split into two disjoint sets $A$ and $B$ where $A \cup B = V$, such that all edges in $E$ connect vertices from $A$ to $B$. Here, $A = V \setminus N$, and $B = C$, where $C$ is the ACL of CR. An edge $(x, y)$ exists between vertices $x \in A$, and $y \in B$, if $y \in C$, i.e., channel $y$ is in the available channels of CR.

To choose an optimal CHs among all nodes, we define a parameter named a cluster head election value (CHEV). In this paper we formulate the choosing of a CH as a maximization problem, which can be defined as follow:

$$i_j^* = \max_{i_j} (CHEV_{i_j}), \quad 1 \leq i_j \leq CM_j, \quad (1)$$

$$CHEV_{i_j} \propto N_{i_j}^{Ch_{i_j}}, \quad (2)$$

where $i_j$ indicates the node $i$ in cluster $j$, $N_{i_j}$ is the total number of neighboring nodes to node $i$ in cluster $j$, and $Ch_{i_j}$ is the total number of common channels for node $i$ in cluster $j$.

The idea behind choosing the CHEV value as it is given in (1) and (2) is to choose the node with the highest number of common channels and the highest number of neighbors to be the cluster head. This gets the cluster more flexible to the PU appearance and spectrum mobility as well as avoiding having a high number of clusters in the network.

Since CH is responsible about cluster stability, it should be the most powerful among the cluster nodes. Also, for producing a high mobility aware MAC protocol by avoiding frequent reclustering due to CH movement, the CH should be the lowest speed node among the cluster nodes in the network. By combining these two important features of CHs, we define the constant of the relationship given in (2) as follows:

$$CHEV_{i_j} = W_{i_j} \times N_{i_j}^{Ch_{i_j}}, \quad (3)$$

$$W_{i_j} = \frac{\gamma_{i_j}}{\sum_{i_j} \gamma_{i_j}} \quad where \quad \gamma_{i_j} = \frac{E_{ij}^\alpha}{V_{ij}^\beta}, \quad \alpha, \beta \in R^+, \quad (4)$$

where $W_{i_j}$ is a normalization factor that indicates how well node $i_j$ is powerful and static in relation to the other nodes in the cluster. $\gamma_{i_j}$, which is always a positive value, is the proposed parameter to indicate the relationship between the node energy and speed. Meanwhile, $\alpha$ and $\beta$ are design parameters for prioritising the speed and energy based on the application requirement.

To avoid having a large CHEV value, we take the log of CHEV as a final selection metric of CHs. Thus, the maximization problem in (1) can be written as follows:

$$i_j^* = \max_{i_j} (log (W_{i_j} \times N_{i_j}^{Ch_{i_j}})). \quad (5)$$

This maximization problem can be simply solved by the well known descending sorting algorithm [10]. Therefore, a node with the highest log(CHEV) value forms the cluster and becomes the cluster head (CH). For example, when node $a$ and $d$ have 4 shared channels, while node $d$ has a higher number of neighbors compared to node $a$ under the conditions that they are static and have the same amount of power, node $d$ will be the CH. If the log(CHEV) value of a node $CR_i$ is smaller than its neighbor, $CR_j$ joins the neighbor, which has the highest log(CHEV) value, as cluster member (CM). Once the clusters are formed, CHs prioritize other cluster members based on log(CHEV) for the reserved cluster head (RCH) selection. CM with the highest log(CHEV) becomes the RCH for the cluster. The RCH takes charge of the cluster if current CH moves out, which reduce the possibility of re-clustering.

IV. ROUTING

By using the clustering mechanism, the proposed routing protocol uses both a proactive and a reactive routing in an adjustable hybrid manner. Each node is aware of the topology of the cluster that it belongs to. Therefore, the packet transmission inside each cluster occurs in a proactive manner. However, no matter how big is the network size, nodes does not require to know the topology of the whole network and the updates are propagated only locally inside each cluster. Meanwhile, routing between clusters happens in a reactive manner using CHs and gateways.

Source node sends the route request (RReq) to its cluster head. If the destination node is in the same cluster then cluster head informs the source node about the route to
the destination. Otherwise, CH broadcasts the route request (CHRReq) to the adjacent clusters using the GW nodes. Each CH that receives the CHRReq packet, it checks that whether the destination node is within its cluster members or not. If a CH finds the destination node is within its members, it replies the CHRReq by CHRRep. If it does not find the destination among its members, it adds its id, hop-count, cluster channel and cluster routing weight (CRW) to the packet and forwards the message to the adjacent clusters. Discussion about the function of CRW will be given in the following section.

Since in the proposed routing mechanism, routing queries propagate only among the CHs, thus it relatively uses small number of query messages.

A. Path optimization

Considering that improvement of different routing metrics may result in different routing paths, it is necessary to combine different individual routing metrics to form a global metric to achieve a performance trade-off among different routing metrics. In this paper, as we consider to improve multiple factors in routing such as delay, bandwidth and hop-count, path optimization is defined as multi-objective optimization as follows:

$$\min_x F(x) = (f_1(x), f_2(x), ..., f_n(x))^T$$

s.t. $n \geq 2, f_i(x) \rightarrow \mathbb{R}^+$

where $n$ is the number of objective functions, $F(x)$ is a vector of objective functions $f_i(x)$. It is needed to be noted that there is no single global solution to multi-objective optimization problems and it is more of a concept than a definition. To solve the above multi-objective optimization, we use very common method called weighted sum method [11].

$$M(r) = \sum_{i=1}^{n} w_i f_i(x)$$

where $M(r)$ corresponds to the global objective function $F$ and $r$ defines the chosen route. $W \triangleq w_i, w_i > 0, 1 \leq i \leq n$ and $\sum_{i=1}^{n} w_i = 1$ is a weight vector defined by each node and reflect the relative importance of that particular object.

In this paper, we aim to minimize the total delay in routing which is the main objective. The delay at each node is defined as follows:

$$D_{\text{total}} = D_t + D_s + D_b + D_q$$

where $D_t$ is transmission delay and is defined as the time taken by each node to transfer the packet to the next hop. $D_s$ is the switching delay and defined as the delay time caused by frequency band switching. Since in this paper we assume that the clustering scheme and the intra-cluster nodes are using the same frequency band, we ignore $D_s$ and consider it to be zero. $D_b$ is the backoff delay time that occurs when carrier sense fails due to collision or detecting another transmission. $D_q$ is the queueing delay time that highly depends on the traffic load, and it is defined as the total time a packet spends in a queue before transmitting.

Suppose we have $N$ hops between source to destination, so the backoff delay at hop $n$ denoted by $D_b,n$ and queuing delay is denoted by $D_q,n$. If the packet length is fixed, transmission delay would be fixed in all hops and it is denoted as $D_t$ as before. Therefore, the total delay over $N$ hop is

$$D(N) = \sum_{n=1}^{N} (D_t + D_b,n + D_q,n)$$

And average delay over $N$ hop is

$$E[D(N)] = N(D_t + \frac{\sum_{n=1}^{N} D_b,n}{N} + \frac{\sum_{n=1}^{N} D_q,n}{N})$$

Equation 10 shows that multihop delay linearly increases with the increasing number of hops.

Higher bandwidth assures faster transmission, where values of $D_q$ and $D_t$ are low. Meanwhile, chances of collision are increased in a highly dense network, where increased chance for collision increases $D_q$. This is because the packet should remain in the queue longer till the node find available free channel to transmit. Also it increases the $D_b$, as nodes have to perform backoff procedure more often due to collision. Therefore, $D_q$ and $D_b$ have direct relationship with node density. We consider node density on any area as the number of neighbors that the node has. Higher number of neighbors means higher node density in that area. Since all traffic flow through CHs, we only consider node density around CHs where number of neighbors for each CH represents node density.

$$\frac{D_q}{D_t} \propto \text{Node density}$$

According to [12], equation 11 is used for calculating the 1-hop network density:

$$\mu(r) = \frac{N \pi r^2}{A}$$

where $N$ is the total number of nodes in the network, $r$ is the transmission range of each node and $A$ is the size of the area where the network is deployed. Bandwidth of cluster is calculated similar to [13]. Defining $R_s(t)$ as the achievable data rate while PU correctly detected as idle without any false alarm, and $R_f(t)$ as achievable data rate during the falsely sensed idle channel, we have:

$$R_s(t) = (P_{off} - P_f) \frac{T - T_s - T_q}{T} \beta \log_2(1 + \frac{S^e_r}{n(t)})$$

$$R_f(t) = (P_{on} - P_d) \frac{T - T_s - T_q}{T} \beta \log_2(1 + \frac{S^e_r}{n(t) + S^e_r})$$

where $P_{on}$ and $P_{off}$ are defined as the probability of a channel being in occupied state and channel being in idle state, respectively. $P_d$ is defined as probability of detection and $P_f$
as false alarm. $T$ is the SU maximum frame period, while $T_s$ is defined as the sensing period and and $T_o$ is defined as overhead of negotiating the traffic channel between the pair of transmitter and receiver. $S_P$ is the received signal of PU by SU during spectrum detection. $S_P$ is the PU signal waveform, and $n(t)$ is a zero-mean additive white Gaussian noise (AWGN).

Having (12) and (13), considering a network having $C$ channels and $M$ PUs, the bandwidth of each SU can be calculated as follows:

$$R(t) = e^{-\theta_1 \frac{\tau_{on}}{\tau_t}} R_u(t) + (1 - e^{-\theta_1 \frac{\tau_{on}}{\tau_t}}) R_f(t) \quad (14)$$

where $\theta_1$ is the scaling factor and $\tau_{on}$ is the time that PUs are ON state.

Considering (11) and (14), each CH calculates a weight for its own cluster and includes it in the CHRReq. Defining $CH_{BW}$ as a bandwidth of cluster $i$, the cluster routing weight (CRW) is formulated as follows:

$$CRW_i = (\omega) \frac{1}{\mu} + (1 - \omega) CH_{BW} \quad (15)$$

where $\omega$ is the assigned weights (priority) given to the $\mu$ or $CH_{BW}$. From (7) and (15) we can have

$$M(r) = \sum_{i=1}^{N} (\omega) \frac{1}{\mu} + (1 - \omega) CH_{BW} \quad (16)$$

REFERENCES