www.ietdl.org

Published in IET Communications
Received on 11th April 2010
Revised on 17th June 2011
doi: 10.1049/iet-com.2010.0285

ISSN 1751-8628

United nodes: cluster-based routing protocol for mobile cognitive radio networks

A.C. Talay D.T. Altilar

Department of Computer Engineering, Istanbul Technical University, 34469 Istanbul, Turkey
E-mail: ctalay@itu.edu.tr

Abstract: Advancement of cognitive radio (CR) technology can overcome the problems encountered from bandwidth and spectrum access limitations because of tremendous potential to improve the utilisation of the radio spectrum by efficiently reusing and sharing the licensed spectrum bands, as long as the interference power inflicted on the primary users of the band remains below a predefined threshold level. In mobile CR *ad hoc* networks, routing is one of the most important issues to be addressed and desires deep investigation. In this study, a distributed and efficient cluster-based spectrum and interference aware routing protocol is proposed. The protocol incorporates the spectrum availability cost and interference metrics into the routing algorithm to find better routes. A route preservation method is also implemented to repair the route when it is defective because of primary user activity. Extensive experimental evaluations are performed in the ns2 simulator. Results of the simulations illustrate that, the proposed algorithm can well fit into the mobile CR *ad hoc* networks and improve the network performance. The results indicate that the proposed protocol provides better adaptability to the environment than the existing ones. It also increases throughput and reduces data delivery latency in a number of realistic scenarios and outperforms recently proposed routing protocols for CR networks.

1 Introduction

Radio spectrum is among the most heavily used and expensive natural resource around the world. Although almost all the spectrum suitable for wireless communications has been allocated, recent studies and observations indicate that many portions of the radio spectrum are not used for a significant amount of time or in certain geographical areas while unlicensed spectrum bands are always crowded [1]. As a promising solution to scarce spectrum resource, cognitive radio (CR) [2] was proposed to enable unlicensed (secondary) users to sense and intelligently access the spectrum portions that are not used by the licensed (primary) users at that specific time and location. The essential components of CR networks, as shown in Fig. 1, can be classified into two groups: the licensed network and the CR network. The licensed network is referred to as an existing network, where the primary users have licenses issued by the government licensing authorities to operate in certain spectrum bands. Since the relevant spectrum portion is assigned to primary users, ideally their operations must not be affected by unlicensed users. In contrast, CR networks do not have and require a license to operate in a desired band. As illustrated in Fig. 1, the CR users have the opportunity to access their own CR base stations if they have, on both licensed and unlicensed spectrum bands. Communication and spectrum sharing policy within the CR network can be independent of that of primary networks. However, there have been some examples in which licensed and unlicensed networks can be managed via single spectrum sharing policy. The CR users can also access one of the licensed network's base stations through the licensed band. Such an approach can be considered as roaming to the licensed channel.

Another access type for a CR user is communicating with other CR users through *ad hoc* connections on both licensed and unlicensed spectrum bands therefore a multi-hop mobile CR network is established. In such networks, CR nodes sense spectrum and identify available frequency bands, (spectrum opportunities – SOP or white holes [3, 4]), and select an appropriate SOP with regard to a predetermined policy. Based on the sensed information, CR users access the licensed band opportunistically when no primary users are using that band and vacate the band immediately upon primary user activity detection in order not to cause harmful interference to the licensed user.

Using SOPs increases the utilisation of network. However, using spectrum bands in an opportunistic way comes with the expense of some important challenges such as routing problem in such a dynamic environment. Facing these challenges and research issues, several approaches have been introduced. One of the proposed approaches is partitioning the network into smaller segments, that is, clustering. Clustering in mobile *ad hoc* networks provides with some important benefits including optimising bandwidth usage, balanced distribution of resources and resolving scalability issues in combination with routing schemes.

Clustering schemes can be classified into two groups, clusters with or without cluster heads (CHs). Clusters

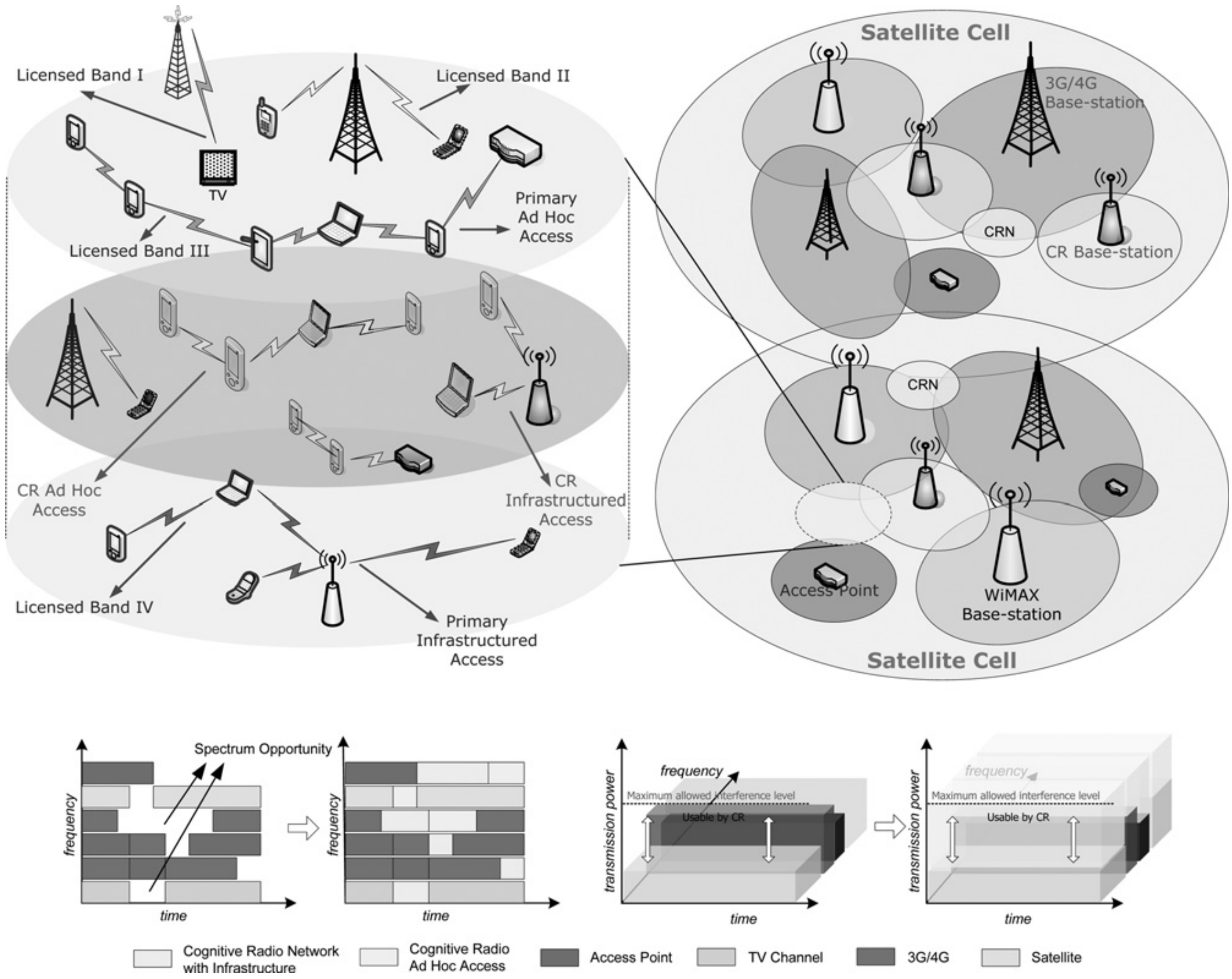

Fig. 1 *Illustration of a CR network: the use of SOPs and of varying transmission power*

without CHs provide a fair share of total communication load. However, CHs may serve for many purposes within a cluster, such as the allocation of resources to member nodes and coordinating transmission events for nodes in the cluster to avoid re-transmissions by reducing packet collisions [5]. Clusters controlled by CHs can be organised as either one-hop or *k*-hop (multi-hop) clusters. In one-hop clustering schemes all of the cluster members (CMs) are within transmission range of the CH. In *k*-hop clustering schemes, the maximum distance between the CH and any CM is at most *k* hops. Note that a CM may reside outside the communication range of the CH, where intermediate CMs relay messages to and from CHs.

A number of clustering algorithms have been proposed for wireless *ad hoc* networks [5–8], as well as for CR networks [9–13]. To the best of our knowledge, all of these algorithms considered other issues than routing, except for one [14]. In that paper [14], the users of different SOPs are gathered under different clusters.

In this paper, a mobile CR *ad hoc* network environment with a number of primary and secondary nodes, where all nodes communicate with each other in their own networks, is considered (Fig. 4). It is assumed that there is no communication (i.e. no cooperation) between primary and secondary networks. A novel algorithm, united nodes (UNITED), is proposed for maximising the network throughput and minimising the end-to-end delay. The UNITED operates autonomously in a distributed manner at every node. Initially, nodes organise themselves into several clusters by the clustering algorithm that is based on location, communication efficiency, network connectivity and spectrum availability. Following completion of cluster formation, routing is done according to the spectrum usage and interference metrics. Clusters adapt themselves dynamically with respect to spectrum availability, and the high mobility of the nodes.

The rest of the paper is organised as follows. Details of clustering are given in Section 2. The proposed routing protocol is explained in Section 3. The simulation and performance analysis of the UNITED is presented in Section 4. Finally, conclusions and future research directions are provided in Section 5.

2 Clustering algorithm

The clustering algorithm must operate even in conditions where mobility of the nodes and fluctuation in the available spectrum is very high. The clusters should be capable of adapting to cope with dynamic conditions imposed by mobility and more importantly the primary user activity. Even these challenges indicate that both mobile *ad hoc* network and CR issues have to be addressed in a clustering algorithm concurrently.

The proposed scalable clustering algorithm for mobile CR *ad hoc* networks makes autonomous decisions in a distributed manner. It is based on a combined weight metric that takes

doi: 10.1049/iet-com.2010.0285

into account several system parameters such as distance, transmission power, mobility, remaining power of nodes and sensed information about available spectrum. Depending on both the application and the environment, the contribution of these parameters to the final metric value may vary. The set of nodes are partitioned into clusters and each node is allowed to join only a single cluster. The CH selection procedure is invoked at the time of the system activation for all nodes. Initially, each node constitutes a cluster itself. Whenever a node receives data from a new neighbour node, it starts establishing a new cluster. Depending on the metrics, that are discussed in the further sections, either a new cluster is established by merging or the original status is kept. Once a new cluster is established, further cluster merge operations are controlled by the CH. The number of nodes d_n within a single cluster could not exceed a preset value of δ. For details, the CH selection algorithm is given in Fig. 2.

A node goes into an unclustered state if all of its links to other nodes within the cluster fail. Also, all nodes within the range of the primary transmission activity return to the unclustered state whenever a primary user activity is detected. When a node falls into an unclustered state, it runs cluster formation process as described in Fig. 3. Nodes in the network have to keep local information about the neighbouring nodes such as ID, speed, location, direction, cluster size and cluster membership. Such records have to be time-stamped in order to be expired after a predetermined time threshold, Δt_i. A node that falls into an unclustered state because of primary user activity usually has valid neighbour information in its table. If the set of valid neighbours is not empty, the algorithm starts with checking for these nodes to join a cluster. Otherwise, which is the rare case, the algorithm proceeds with searching for new neighbourhood nodes. The set of one-hop neighbours, S, is produced by collecting replies from CHs to periodic HELLO packets. Nodes receiving the HELLO packet utilises the node ID, the spectrum and the mobility information, that is, SOPs, location, speed and direction of travel, included in the packet to decide whether to reply or to ignore. Packets, either originating from the nodes that are moving away or having little spectrum access opportunity because of heavy primary user activity in the area, are ignored. Otherwise, the receiving node checks whether the maximum number of connections δ is exceeded. If not, it responds with a unicast response RESP packet. Upon receiving the first RESP packet, the unclustered node sets a timer to wait for all responses, that is, to let all neighbouring nodes to have an opportunity and respond.

After an unclustered node produces the set of its one-hop neighbour CHs, the set is sorted with respect to their weighted metrics in descending order and a search for actual connection is started. The unclustered node starts with sending a join request JOIN_REQ packet to the first CH in the list. This process is repeated for each node in the sorted set of S in order until a cluster is formed or $S = \emptyset$. If cluster is not formed the node seeks for a new set of S. If such a set cannot be established, the unclustered node creates its own cluster by declaring itself a CH and terminates the algorithm. An illustration is given in Fig. 4.

Algorithm 1

```
begin
    repeat
        ∀ node n, compute :
            the node degree Δ_n = |d_n − δ|;
            the mobility measure M_n(t);
            spectrum availability Sp_n(t);
            weighted value W_n = αΔ_n + βM_n(t) + γSp_n(t);
        choose the node with the highest weighted value as the Cluster Head;
        if n_j is CH for n_i and n_k is CH for n_j then
            n_i reselects CH excluding n_j;
        remove neighbour nodes of the chosen CH from set of nodes
    until G = ∅;
```

Fig. 2 *CH selection*

Algorithm 2

```
Input: S;
Output: id of CH;
begin
    NewList = false;
    while !NewList do
        if S = ∅ then
            NewList = true;
            repeat
                send hello packet with reply request;
            until S ≠ ∅;
        order(S);
        while S ≠ ∅ do
            i ← max(∀j ∈ S);
            send Join_Request(i);
            if Join_Response received then
                if connection accepted then
                    return clustered(i);
                else
                    S ← S − {i};
        return clustered(self);
```

Fig. 3 *Cluster formation*

3 Routing algorithm

Primary user activity prediction has been researched for nearly a decade. Recent reports [15, 16] indicate that prediction plays an important role in the system performance. These papers evaluate the issue from the perspective of the primary user. However, this paper, different from earlier work, takes CR perspective into consideration since the CR traffic essentially relies on the primary traffic. Considering the course of an individual flow in the CR network, many disruptions by primary network activities are expected. Therefore CR users are likely to access spectrum in a predictable fashion since the target node(s) would not change throughout such a disrupted course. In other words, if a node has gained opportunity to access a specific unused spectrum band and/or has a number of connections to a specific node recently (most probably within the same course), it is likely that it will establish the same connection again. The proposed routing algorithm makes use of these observations and information to improve routing performance by defining a novel spectrum availability cost metric and doing routing accordingly.

It is assumed that the usage pattern of the primary users which affects an arbitrary CR user i follows an independent two-stage ON/OFF random process. An ON period $T_{\text{on},j}$ represents the time that the primary users are active, that is, CR user i could not be active on the channel. An OFF period $T_{\text{off},j}$ represents the time that the primary users are inactive and CR user i has free access to the spectrum. To simplify our analysis, it is assumed that both $T_{\text{on},j}$ and $T_{\text{off},j}$ are exponentially distributed with means equal to $1/\mu_i$ and $1/\lambda_i$ seconds, respectively. The ON/OFF random processes of the primary users activity pattern affecting different CR users are assumed to be independent. There are two essential criteria to establish a route. From the perspective of the CR system, routes with the best end-to-end performance should be selected and from the perspective of system coexistence, routes have to be selected with the minimum interference to the primary systems. Note that communication overlap between primary and secondary users is called inter-system interference. While considering end-to-end throughput of a route for a CR network, interference from other CR users along the route should also be taken into account. This interference is called intra-system interference. We thought that if the metric is interference aware, routing performance would be improved. Therefore the defined metric comprises of the effects of variation in link loss ratio, differences in link transmission rate as well as inter-system and intra-system interferences.

3.1 Routing metrics

The proposed routing algorithm relies on the combined routing metric that comprises of two essential components: spectrum avaiability cost and interference cost.

3.1.1 Spectrum availability cost: Spectrum availability is an inherent characteristic in mobile CR *ad hoc* networks. Nodes usually get disconnected owing to characteristics of *ad hoc* networks and CR technology such as mobility, temporary obstructions, and ability to find a non-interfering communication channel with the primary users. In the UNITED, connectivity behaviour of each link is monitored in a 1 s long sliding window. Sensing period is found to be 10–160 ms in the literature and IEEE 802.22 standard. Window size values lower than a second would not allow us to observe sufficient number of transitions especially when longer frames are being transmitted. Window size values higher than a second result in both skipping the effect of fluctuating behaviour of spectrum usage and a slower response time. Initial experiments to decide on the window size indicate that a second-wide window provides a better performance and a balance between the above-mentioned effects.

The cost of the link $C_{i,j}$ is set to a value proportional to the total disconnection time within the recent window. Therefore a route can be kept between a source and a destination even if there is end-to-end connectivity failures that may occur temporally. On the other hand, multiple links may have similar average unavailable spectrum duration. The link with a history of more frequent disconnection-to-connection transition is assigned with a lower cost. It is assumed that for a given unavailable spectrum duration, a link with more frequent transitions is a better link since it has more opportunity to forward a packet to the other nodes. Note that, the cost of the link for a channel can be computed adaptively depending on the nature of the secondary users' traffic and the decision with regard to the transition count could be switched to less frequent number of transitions. With these guiding principles, $C_{i,j}$, the cost of a directional link $L_{i,j}$ is defined as

$$C_{i,j} = 1 + \frac{(T_{\text{cost\_window}} - \sum_{k=1}^{N_{i,j}^{\text{transition}}} T_{i,j}^{k})}{1 + N_{i,j}^{\text{transition}}} \quad (1)$$

$C_{i,j}$ is dynamically computed by node i based on its spectrum usage history over a discrete sliding window of length $T_{\text{cost\_window}}$. The number of transitions of the link status from transmission opportunity state to sensed primary user activity state within $T_{\text{cost\_window}}$ is represented by the parameter $N_{i,j}^{\text{transition}}$, and the duration of the kth transmission opportunity is represented by $T_{i,j}^{k}$. The term $\sum_{k=1}^{N_{i,j}^{\text{transition}}} T_{i,j}^{k}$ represents the total cumulative transmission opportunity duration within the last measurement window. For a non-disturbed secondary link, this term equals the duration of the measurement window itself, and $N_{i,j}^{\text{transition}}$ equals to zero hence the cost reduces to unity. Since the numerator of the expression for $C_{i,j}$ is dominated by the cumulative link primary user activity time, links with longer interfering times will have higher cost and thus will be avoided by the proposed algorithm as well as other the least cost algorithms. However, among multiple links with similar cumulative disconnectivity durations, the ones with higher transition counts $N_{i,j}^{\text{transition}}$ will have lower costs. This would also contribute to reduce data delivery latency. Note that the upper bound of the link cost will be determined by the parameter $T_{\text{cost\_window}}$.

As claimed before one of the novel approaches in the proposed algorithm is keeping a recently established route against temporal end-to-end disconnectivity. The details of the approach is given explained via two scenarios given in Fig. 5. Assume that node S starts to establish a path to send data to destination node D, whereas node S_1 is under the interference of primary user communication. Node S broadcasts an RREQ (route request) packet and finally a route reply packet containing the route and spectrum information arrives at node S from node D in return. The received packet provides with S, S_9, S_2, S_3 and S_4, D as route according to the costs scripted on the links in Fig. 5*a*. However, when node S checks its neighbour table, it would be seen that S_2 is a neighbour of its neighbour S_1 and a route to S_2 over S_1 costs low compared to S_5, S_9, S_2, sub-route even though the costs of S, S_1, S_1 and S_2 links have been increased because of the recent primary network activity. Whenever node S_1 have SOP again (which would be indicated by a HELLO beacon), node S sends an RREQ packet to node S_1 while feeding the established route. If S, S_1, S_2, S_3 and S_4, D route is able to convey the RREQ packet, then the new route will be used to transfer the next packet and the current route will be cancelled. As another case (Fig. 5*b*), assume that not only S_1 but also S_2 are affected from the primary communication. The route reply will suggest S, S_5, S_9, S_6, S_7, S_8 and D route. The communication starts flowing over that route. As soon as the primary communication is over, node S is informed by node S_1. Since S_1 resides at the direction of the destination and has a lower-cost value even though it is increased because of the recent primary activity, an RREQ for destination node D is sent to node S_1 with the current route information. Since node S_3 possesses the cost of S_3, S_4, D route, it sends a route enhancement reply when it receives the RREQ. Node S checks whether it is feasible to use the new route instead of the current one. If it is feasible the new route is selected and the current route is invalidated.

3.1.2 Interference cost: The impact of interference on the network performance is difficult to estimate. In order to have an accurate view of the current channel (link) state, it is necessary to factor in the indicators of the channel quality such as nominal throughput or packet loss. It is also critical to estimate the transmission delay resulting from concurrent data transmissions. Consequently, a routing metric properly tailored for CR networks that accounts for these different factors can improve the overall network performance by avoiding lossy links and congested zones.

doi: 10.1049/iet-com.2010.0285

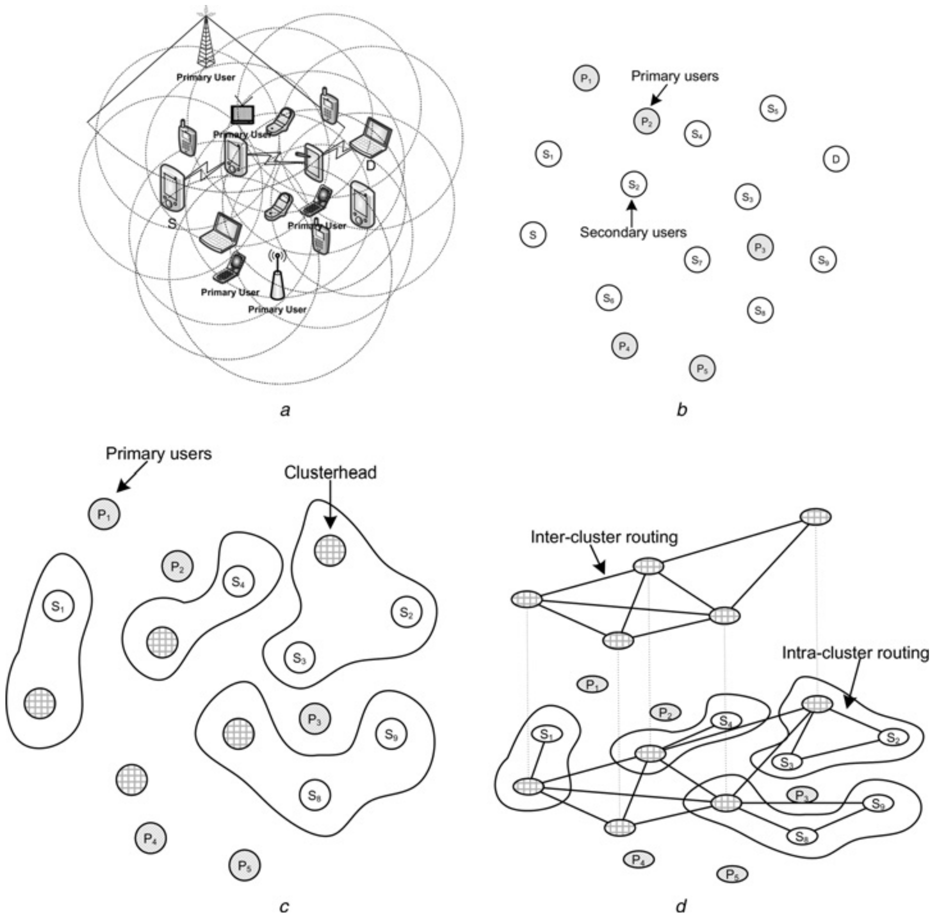

Fig. 4 *Illustration of a mobile CR ad hoc network*

a CR network illustrated
b Two-dimensional mapping of the CR network
c Clusterheads selected and clusters formed
d Connectivity in illustrated CR network for routing

A previously proposed routing metric for *ad hoc* networks, which is called expected transmission time (ETT) [17], gives an idea about the quality of the link quite well as links with less ETT give better throughput by neglecting the interference. ETT was defined as

$$\mathrm{ETT} = \mathrm{ETX}\frac{S}{B}, \quad \mathrm{ETX} = \frac{1}{P_\mathrm{f} \times P_\mathrm{r}} \tag{2}$$

where S is the packet size, and B is the bandwidth of the link. The expected number of transmissions required to successfully deliver a packet, that is, ETX, is computed by the underlying packet loss probability in both the forward and reverse directions, which are denoted by P_f, P_r, respectively.

Although it has been shown that ETT works for a limited number of interfering flows, it does not perform as expected for higher number of interfering flows in the network. We believe that interference imposed on the link, which fluctuates throughout a course of a connection, has to be modelled appropriately and factored into the routing metric to find better quality paths.

In a physical interference model [18], used to capture the interference experienced by links in the network, a communication between nodes m and n is successful if the SINR (signal to interference and noise ratio) at the receiver n is above a certain threshold, β, which has to be set depending on the desired transmission characteristics. SINR provides useful information on how strong the desired signal is compared with the interferer as well as the imposed noise in the network. Denoting the signal strength of a packet from node m at node n by $P_n(m)$, a packet on the link $L_{m,n}$ from node m to node n is correctly received if

$$\frac{P_n(m)}{N + \Sigma_{k\in V'} P_n(k)} \geq \beta \tag{3}$$

where N is the background noise, V' is the set of nodes simultaneously transmitting and β is a constant indicating the threshold. Considering all interfering nodes, SINR(m) can be defined as

$$\mathrm{SINR}\,(m) = \frac{P_m(n)}{N + \Sigma_{k\in\vartheta(m)\setminus\{m,n\}} \Gamma_k P_m(k)} \tag{4}$$

where the received interfering signal from node k is weighted using node k's transmission rate Γ_k, which is the normalised rate averaged over a period of time. It gives the fraction of time node k occupies the spectrum. The set of nodes that node m can hear or sense is associated with the set $\vartheta(m)$. We define interference ratio $I_i(m)$ for a node m in a link $L_{m,n}$ as the ratio of interference to the maximum interference $P_{\text{int}}^{\max}$ that a node can still communicate properly, and denoted as

$$I_i(m) = \frac{\sum_{k \in \vartheta(m)\setminus\{m,n\}} \Gamma_k P_m(k)}{P_{\text{int}}^{\max}} \tag{5}$$

where ($0 \leq I_i(m) \leq 1$). When considering a bidirectional link $L_{m,n}$, I_l is

$$I_l = \max(I_i(m), I_i(n)) \tag{6}$$

We define the interference metric of a link l as

$$\text{int}_l = \text{ETT}_l * \varphi(I_l) \tag{7}$$

where $\varphi(x)$ is the scaling function defined as $1 + (1/2)(x/\sqrt{x^2 + a})$, whereas a is the smoothing constant which is experimentally found and set to 100. Note that the scaling function $\varphi(x)$ provides a new range between 1.00 and 1.05 for a given I_l which initially ranges between 0 and 1. ETT_l is weighted with I_l to capture the interference experienced by the link from all of its neighbours including primary users (7). The contribution of I_l is limited to 5% as a result of initial experiments. Thus, int_l becomes the new cost value comprising of both ETT and interference.

3.1.3 Combined routing metric: We can combine the desirable properties of the two metrics described in (1) and (7) by taking their weighted average

$$\text{cost} = \alpha * C_{i,j} + (1 - \alpha) * \text{int}_l \tag{8}$$

where α is a tunable parameter subject to $0 \leq \alpha \leq 1$. The weighted average can be viewed as an attempt to balance between the spectrum availability and interference cost metrics. With respect to the simulation results, the impact of α on both throughput and end-to-end delay is shown in Fig. 6). The peak performance is achieved for $\alpha = 0.6$, where the throughput is 920 Kbps. Also, the shortest delay (40.6 ms) is achieved for $\alpha = 0.6$. Therefore α is set to 0.6 for the conducted experiments.

3.2 Design issues

In our design, a unique message identifier, a hop count, and an optional ACK request are associated with each message for routing. This identifier is a concatenation of the host's ID and a locally generated message ID. In our implementation, the hosts in the network are assigned with static ID's.

The hop count field determines the maximum number of exchanges that a particular message is subject to. Although the hop count is similar to the time-to-live field in IP packets, messages with a hop count of one will be delivered if and only if one of the neighbourhood nodes matches with the final destination. The packet will be dropped otherwise. Larger values for hop count will yield to distribute a message through the network wider and for some cases quicker. This will typically reduce average delivery time, but will also increase total resource consumption for message delivery. In order to minimise resource consumption, high priority messages might be marked with a high-hop count, whereas most messages can be marked with a value close to the expected number of hops for a given network configuration.

Certain applications may require acknowledgement to a delivered message. The ACK request field signals the destination of a message to provide an acknowledgement of message delivery. These acknowledgements are modelled as short return messages from the receiver back to the sender. Of course, the acknowledgement can also be piggybacked with any other message destined to the sender.

Each host sets a maximum buffer size to allocate for message distribution. The buffer size limits the amount of memory and network resources consumed by the UNITED. As known, hosts drop older messages in favour of newer ones when their buffers become full. As expected, there is an inherent trade-off between aggregate resource consumption and message delivery rate/latency. In order to ensure delivery of messages, at least a subset of nodes should have sufficiently large buffers. Otherwise, it is possible for older messages to be flushed from buffers before having a chance of delivery. A number of management strategies are possible for individual message buffers. The simplest policy is first-in-first-out (FIFO). As long as the buffer size on all hosts is larger than the expected number of messages in transit at any given time, FIFO is a very reasonable policy. However, if available buffer size is limited relative to the number of messages, FIFO is sub-optimal with respect to fairness and quality of service (QoS). For example, a host's aggregate buffer utilisation is directly proportional to the number of messages it sends, which may not be fair to other hosts. Furthermore, FIFO does not provide any mechanisms for preferentially delivering or storing high priority messages. Fair queueing algorithms, including weighted fair queueing, logically distribute available buffer space among competing hosts, providing differentiated QoS on a per-message granularity. The experiments are done by using FIFO implementation. Other algorithms are considered for a future study.

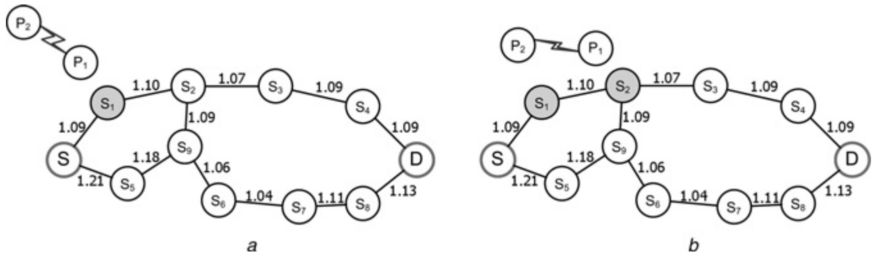

Fig. 5 *Keeping a recently established route against temporal end-to-end disconnectivity*

a Single-node disconnection
b Alternative sub-route disconnection

doi: 10.1049/iet-com.2010.0285

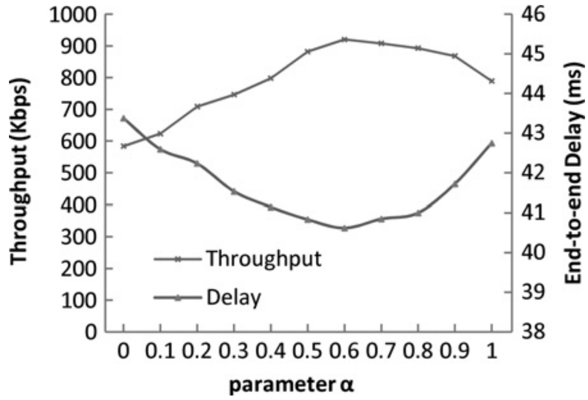

Fig. 6 *Throughput and end-to-end delay with respect to parameter α*

Algorithm 3

```
//For any node n_i and packet p;
begin
    if dest(p)=n_i then
        receive p;
    else
        next_hop ← search_rtable(dest(p));
        if next_hop≠ ∅ then
            forward(p, next_hop);
        else
            if is_ClusterHead(n_i) then
                drop p;
            else
                forward(p, R_default(n_i));
```

Fig. 7 *Forwarding strategy*

3.3 Forwarding strategy

Each node in a cluster can populate its routing table for intra-cluster routing based on the topology using shortest-path algorithms considering the proposed metric mentioned above. Each computed route is associated with a lifetime. The route will be removed when it is expired. If a node cannot find a route for the destination of a packet, the data packets will be forwarded to its default route R_{default} and then finally to the CH. This is common when the destination node is in a different cluster. In such a case, the CH will forward the packet to the CH of the destination cluster and in turn the packet is forwarded to the destination node. If the node is the CH and there are no available nodes to transmit the packet, the packet will be dropped. The data forwarding algorithm is shown in Fig. 7.

3.4 Route preservation and adaptation (local repair)

It is necessary to have a maintenance scheme in the routing protocol, where each node has to corroborate the area it belongs to, and update information for the mobility factor and spectrum availability. Whenever a new CH is selected, all neighbouring nodes receive a notification message to set the new address to transmit packets. Route maintenance can be done by: (a) skipping the broken node if the next hop in the path is reachable; (b) choosing another reachable node(s), which is out of the transmission range of the active primary use, to be the next hop that is reachable by the previous node and the next node in the path.

4 Simulation and performance analysis

Through simulations constructed in ns2, the performance and functional correctness of the UNITED and its relative performance compared with that of DORP [19], STOD-RP [20], SEARCH [21] and SCRP [14] are investigated. Simulations run with the following parameters. Two-ray ground propagation model is used at the radio layer. The bit rate for each channel is set to 2 Mbps. Varying number of mobile nodes up to 100 moving in a rectangular area 1800 m × 1800 m in dimension is modelled. Each node picks a random destination in the rectangle and moves there with a speed uniformly distributed between 0 and 20 m/s. Upon reaching the destination, the node picks a new destination and repeats the process. The activities of the primary user are modelled by using the exponential ON/OFF processes. The coverage range of the primary user on its operation channel is taken as 250 m. These values are taken from previous studies on a number of protocols to provide a mean for comparison. The communication pattern used in the simulation is a combination of both constant bit rate (CBR) and voice-over-IP (VoIP) traffic to make a more realistic scenario. Each source node generates and transmits CBR traffic and each message is 1 KB in length. The transmission interval for each node is set to 100 ms. The injected VoIP traffic is modelled by a two-state ON/OFF model with exponentially distributed duration of talk spurts and silence periods. A total of six VoIP CR users are randomly distributed over 8–128 Kbit/s with random arrival rates (including packetisation intervals according to the codec G.711, G.726 and G.729 recommendations). Fifty experiments are performed in random multi-hop network topologies, for each different parameter settings.

As noted earlier, DORP [19], a modified version of AODV, and spectrum-tree based on demand routing STOD-RP [20] and recently proposed a geographic forwarding-based spectrum aware routing protocol for cognitive *ad hoc* networks SEARCH [21], and spectrum-aware cluster-based routing protocol (SCRP) [14] that forms spectrum-cluster for each spectrum band are used for comparisons. The characteristics of the UNITED are explored under a number of different scenarios. Packet delivery ratio, throughput and end-to-end delay have been computed for DORP, STOD-RP, SEARCH, SCRP and UNITED. The comparisons show that the UNITED can fit well the multi-flow multi-channel environment and effectively exploit the potential large communication capacity in CR networks. In the simulations, the rate of flows is varied from 100 to 1800 kbps along with the injected VoIP traffic. The nodes are randomly placed in the area, and eight flows having the same traffic generation rate are initiated with VoIP traffic.

Fig. 8*a* illustrates how packet delivery ratio changes with flow rate. The measurements of the packet delivery ratio over all traffic flows are averaged and plotted for each protocol. Contrary to the other protocols, UNITED has a smooth trend especially for lower flow rates. Other protocols suffer from the congestion and unable to deliver more packets for higher flow rates. As seen clearly, the UNITED significantly improves the packet delivery ratio especially for higher-flow rates. The UNITED has a better packet delivery ratio because of the utilisation of proposed metrics and the route preservation and adaptation scheme.

As a second performance criterion, throughput analysis is conducted (Fig. 8*b*). When the traffic load is low (i.e. less than 400 kbps), all schemes perform with similar aggregate throughput. As the flow rate increases, the throughput of

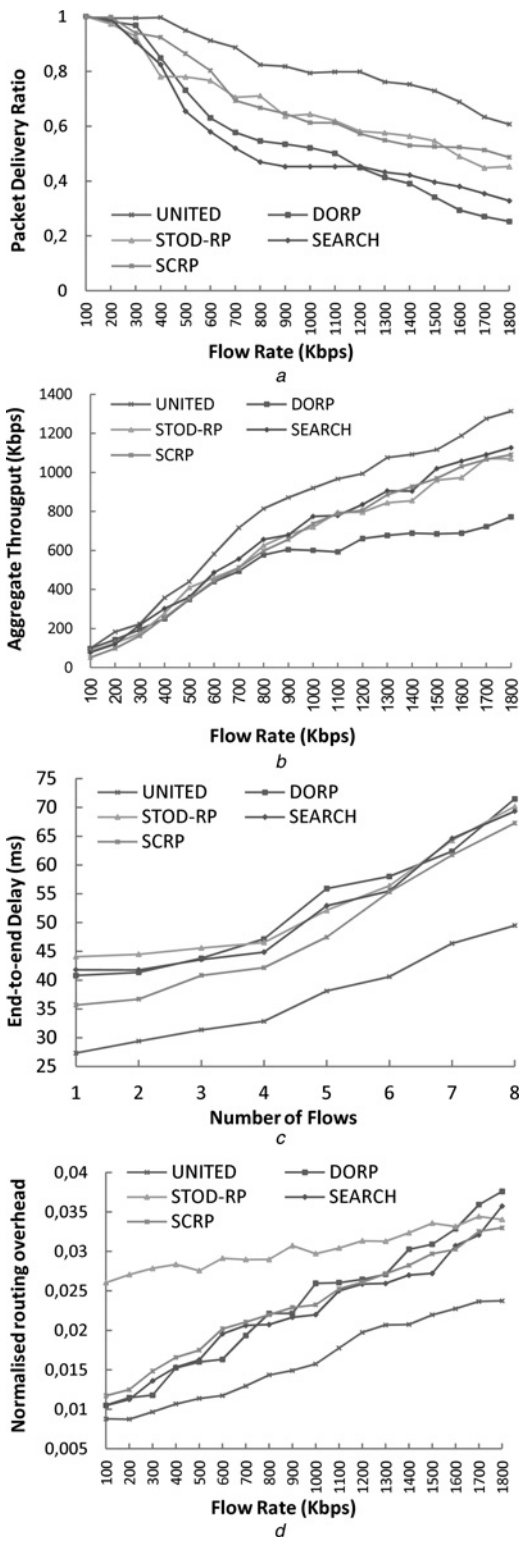

Fig. 8 *Overall performance analysis of different routing protocols for different performance criteria*

a Packet delivery ratio as the flow rate increases
b Throughput when two hosts, *S* and *D* communicates
c Average end-to-end delay performance against number of flows
d Normalised routing overhead as the flow rate increases

DORP increases slowly towards the limit of the established path. Also in DORP, nodes become disconnected because of primary user activity, and no packets are forwarded. STOD-RP, SEARCH and SCRP perform better than DORP since STOD-RP and SCRP employ route recovery algorithms and SEACRH has a route maintenance algorithm. As the traffic load increases, the performance improvement of UNITED compared to the other schemes becomes more significant because of used routing decision metric, effective spectrum usage and route preservation and adaptation scheme. In a dynamic environment (i.e. the network topology changes frequently), UNITED adapts itself to the environment to retain the secondary nodes communication path by making use of the clustering or to perform a local repair when the distance to the destination is not reachable without interfering primary user. Moreover, the established route in the UNITED is better than the route built in the other schemes in a frequently changing environment since we use interference as a routing metric.

The end-to-end delay performance of the UNITED with other schemes is also compared. The number of intersecting flows is adjusted between 1 and 8 to evaluate the performances on intersecting flows. The simulation result is shown in Fig. 8*c*. When the number of flows increases, the UNITED seeks a balance between assigning new frequency bands to allow simultaneous transmission and accommodating some nodes on one band to avoid switching delay. Also the re-route establishment time is low in the UNITED upon a primary user activity detection than the others, since those techniques have to reconstruct a path from source to destination upon primary user activity detection. Consequently, the UNITED achieves an overall minimum delay considering the other schemes as the number of intersecting flows grows as shown in Fig. 8*c*.

To assess the effectiveness of the UNITED, the normalised routing overhead is also used as a performance metric. Normalised routing overhead can be defined as the total number of control (N_{control}) and data packets sent (N_{data}) normalised by the total number of packets successfully delivered in the CR network, also considering the number of flows. The normalised routing overhead of the UNITED is illustrated in Fig. 8*d*.

5 Conclusion

In this paper, CR *ad hoc* networks are investigated and, the UNITED, a cluster-based routing protocol for mobile CR networks is proposed for maximising the network throughput and minimising the end-to-end delay. The UNITED considers the requirements of primary and secondary users simultaneously. It utilises spectrum availability cost and interference metrics to find better routes. These two metrics contributes to the final metric value by a weight factor α. The experiments show that the best performance is provided with $\alpha = 0.6$ (i.e. the 40% of interference metric and 60% of spectrum availability cost metric). Through an implementation in the ns2 simulator, it has been shown that the UNITED achieves significant improvement on the throughput and the end-to-end delay. The adaptability and efficiency of the scheme is shown in simulations. Performance comparisons indicate that the UNITED outperforms recently proposed algorithms which are DORP, STOD-RP, SCRP and SEARCH. The UNITED may be enhanced further by introducing dynamic parameter optimisation.

6 References

1 Peha, J.M.: 'Approaches to spectrum sharing', *IEEE Commun. Mag.*, 2005, **43**, (2), pp. 10–12

doi: 10.1049/iet-com.2010.0285

2 Mitola, J.: 'Cognitive radio: an integrated agent architecture for software defined radio'. PhD dissertation, Department of Teleinformatics, Royal Institute of Technology (KTH) Stockholm, Sweden, May 2000

3 Akyildiz, I.F., Lee, Y., Vuran, M.C., Mohanty, S.: 'A survey on spectrum management in cognitive radio networks', *IEEE Commun. Mag.*, 2008, **46**, (4), pp. 40–48

4 Shen, J., Liu, S., Wang, Y., Xie, G., Rashvand, H., Liu, Y.: 'Robust energy detection in cognitive radio', *IET Commun.*, 2009, **3**, (6), pp. 1016–1023

5 Yu, J.Y., Chong, P.H.J.: 'A survey of clustering schemes for mobile ad hoc networks', *IEEE Commun. Surv. Tuts.*, 2005, **7**, (1), pp. 32–48, First Quarter 2005

6 Chatterjee, M., Das, S.K., Turgut, D.: 'WCA: a weighted clustering algorithm for mobile ad hoc networks', *Cluster Comput.*, 2002, **5**, (2), pp. 193–204

7 Kawadia, V., Kumar, P.R.: 'Power control and clustering in ad hoc networks'. IEEE INFOCOM 2003. 22nd Annu. Joint Conf. IEEE Computer and Communications Societies, 2003, vol. 1, pp. 459–469

8 Younis, O., Fahmy, S.: 'HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks', *IEEE Trans. Mob. Comput.*, 2004, **3**, (4), pp. 366–379

9 Sun, C., Zhang, W., Letaief, K.B.: 'Cluster-based cooperative spectrum sensing in cognitive radio systems'. Proc. IEEE Int. Conf. on Communications (ICC'07), 24–28 June 2007, pp. 2511–2515

10 Qi, C., Wang, J., Li, S.: 'Weighted-clustering cooperative spectrum sensing in cognitive radio context'. Proc. WRI Int. Conf. on Communications and Mobile Computing (CMC'09), 6–8 January 2009, pp. 102–106

11 Chen, T., Zhang, H., Zhou, X., Maggio, G.M., Chlamtac, I.: 'CogMesh: a cluster based cognitive radio mesh network' (Springer, The Netherlands, 2007), Ch. 34

12 Alsarhan, A., Agarwal, A.: 'Cluster-based spectrum management using cognitive radios in wireless mesh network'. Proc. 18th Int. Conf. on Computer Communications and Networks, 2009 (IEEE ICCCN'09), 3–6 August 2009, pp. 1–6

13 Lazos, L., Liu, S., Krunz, M.: 'Spectrum opportunity-based control channel assignment in cognitive radio networks'. Proc. Sixth Annual IEEE Commun. Society Conf. on Sensor, Mesh and Ad Hoc Communications and Networks, (SECON'09), 22–26 June 2009, pp. 1–9

14 Jiang, W., Cui, H., Chen, J.: 'Spectrum-aware cluster-based routing protocol for multiple-hop cognitive wireless network'. Proc. IEEE Int. Conf. on Communications Technology and Applications (ICCTA'09), 16–18 October 2009, pp. 288–294

15 Kulkarni, P., Lewis, T., Fan, Z.: 'Simple traffic prediction mechanism and its applications in wireless networks', *Wirel. Pers. Commun.*, 2001, **49**, (1), pp. 1–14

16 Matinmikko, M., Höyhtyä, M., Mustonen, M., *et al.*: 'Cognitive radio: an intelligent wireless communication system, channel state estimation and spectrum management for cognitive radios'. Technical report VTTT-R-02219-08, VTT Technical Research Centre of Finland, 2008

17 Draves, R., Padhye, J., Zill, B.: 'Routing in multi-radio, multi-hop wireless mesh networks'. Proc. Tenth Annual Int. Conf. on Mobile Computing and Networking (MobiCom'04), 26 September–1 October 2004, pp. 114–128

18 Gupta, P., Kumar, P.R.: 'The capacity of wireless networks', *IEEE Trans. Inf. Theory*, 2000, **46**, (2), pp. 388–404

19 Cheng, G., Liu, W., Li, Y., Cheng, W.: 'Joint on-demand routing and spectrum assignment in cognitive radio networks'. Proc. IEEE Int. Conf. on Communications (ICC'07), 24–28 June 2007, pp. 6499–6503

20 Zhu, G.-M., Akyildiz, I.F., Kuo, G.S.: 'Stod-rp: a spectrum-tree based on-demand routing protocol for multi-hop cognitive radio networks'. Proc. IEEE Global Telecommunications Conf. (IEEE GLOBECOM 2008), 30 November–4 December 2008

21 Chowdhury, K., Felice, M.: 'SEARCH: a routing protocol for mobile cognitive radio ad-hoc networks', *Comput. Commun.*, 2009, **32**, (18), pp. 1983–1997