Abstract- Cognitive radio (CR) technology is rapidly developing these days due to its capability of adaptive learning and reconfiguration. Thus, using Cognitive Radio Networks (CRNs) spectrum efficiency can be increased by allowing the secondary users (SUs) to access the licensed band dynamically and opportunistically without interfering the primary users (PUs). Daniel H. and Ryan W. Thomas, define the CRNs in the context of machine learning as the network which improves its performance through experience gained over a period of time without complete information about the environment in which it operates. Thus, the dynamism and opportunism can be learnt by reinforcement learning, which is concerned with how software agents or learning agents ought to take actions in an environment so as to maximize some notion of cumulative reward. The paper proposes a routing scheme that uses Q-learning, which is the most widely used RL approach in wireless networks. In Q-learning, the learnt action value or Q-value, Q (state, event, action) is updated using the reward and is recorded. For each state-event pair, an appropriate action is rewarded and its Q-value is increased. Hence, the Q-value indicates the appropriateness of an action selection in a state-event pair. At any time instant, an action is chosen by the agent in such a way that it maximizes its Q-value. The reward corresponds to performance metric such as throughput.

Keywords- Opportunistic Routing, Reinforcement Learning, Reward.

I. INTRODUCTION

Cognitive Radio (CR) technology is a promising technology that allows unlicensed users to access licensed spectrum bands opportunistically in a dynamic and non-interfering manner. Cognitive Radio Network (CRN) users are distinguished into two types:

- Primary User: Licensed user having highest priority for spectrum utilization.
- Secondary User: Unlicensed user that can access spectrum in non-intrusive manner.

Due to these distinguished characteristics routing is an issue in CRNs as data should be routed via stable and reliable path to avoid rerouting and thus congestion in network, which in turn degrade the performance of network such as throughput and delay. Thus, routing in CRN is a challenging task due to the diversity in the available channels and data rates. The proposed routing scheme utilizes a reinforcement learning framework to opportunistically route the packets even if reliable knowledge about channel statistics and network model is not available. It jointly addresses the issues of learning and routing in an opportunistic context, where the network structure is characterized by the transmission success probabilities. In particular, this learning framework leads to a stochastic routing scheme that optimally explores and exploits the opportunities in the network.

II. LITERATURE SURVEY

Conventional infrastructure networks and mobile ad hoc networks have been studied since last decade and many routing protocols, e.g. proactive, reactive, hierarchical and multicast are available for such networks. Now it’s time to investigate the routing protocols for CRNs, as routing is a challenging task in such networks, particularly in multi hop CRNs due to the diversity in channel availability and data rates. Multi-interface enabled CR user can avail multiple available channels simultaneously, thus can help in increasing overall network performance as well as reduce the interference on the primary users. Due to this vital feature of CRNs conventional routing metrics such as hop count, congestion, etc. are not sufficient for routing decision in CRNs. Some conventional routing schemes fail to take advantage of broadcast nature and opportunities provided by the wireless medium, which result in unnecessary packet retransmissions. On the other hand, the opportunistic routing decisions are made in an online manner by choosing the next relay based on the actual transmission outcomes as well as a rank ordering of neighboring nodes.

Authors in propose opportunistic algorithms that depend on a precise probabilistic model of wireless connections and local topology of the network. It is shown that for the optimal routing decision, the next relay node is selected on the basis of a distance-vector summarizing the expected-cost-to-forward from the neighbors to the destination. In, Geographic Random Forwarding (GeRaF) focuses on the multi-hop performance of such a solution, in terms of average number of hops to reach a destination as a function of the distance and of the average number of available neighbors. It uses the smallest geographical distance from the destination as a criterion for selecting the next-hop. In the context of ExOR, the
optimal route is computed so as to minimize the expected number of transmissions (ETX). In, routing decisions are made based on immediate feedback from each transmission. The protocol detects and reacts to the stochastic result of each node’s local broadcast transmission, and an optimal route is constructed based on this immediate feedback.

However, these algorithms require an integrated approach to the probability estimation issue i.e., these probabilistic models need to be learned and maintained. So the question of jointly learning the channel statistics and routing opportunistically remains unexplored.

The main contribution of the routing scheme proposed in this paper is to provide an opportunistic routing algorithm that: 1) assumes zero knowledge about the channel statistics and network topology, but 2) uses a reinforcement learning framework in order to enable the nodes to adapt their routing strategies, and 3) optimally exploits the statistical opportunities. The proposed routing scheme utilizes a reinforcement learning framework to opportunistically route the packets even in the absence of reliable knowledge about channel statistics and network model. It jointly addresses the issues of learning and routing in an opportunistic context, where the network structure is characterized by the transmission success probabilities. In particular, this learning framework leads to a stochastic routing scheme that optimally “explores” and “exploits” the opportunities in the network.

III. IMPLEMENTATION DETAILS

Proposed system is divided into modules and these modules are integrated together for the execution of the system. The proposed system includes following modules:

[1] Relay discovery process
   a) RREQ (Relay Request) Packet
   b) RREP (Relay Reply) Packet
   [2] Route maintenance
   Here is brief description about each module.
   [1] Relay discovery process
      a) RREQ (Relay Request) Packet
      The sender or the source node frames and transmits a Relay request (RREQ) packet at time n if it has a packet to send. This packet contains message type, source address, current sequence number of source, destination address, broadcast ID and the hop count. The broadcast ID is incremented every time when the source node initiates a RREQ. In this way, the broadcast ID and the address of the source node form a unique identifier for the RREQ.
      b) RREP (Relay Reply) Packet
      Let S denote the (random) set of nodes that have received the packet transmitted by node i. In the reception and acknowledgment stage, successful reception of the packet transmitted by node i is acknowledged to it by all the nodes in S. For all nodes k in S, the ACK packet of node k to node i include the EBS message \( \uparrow \), where \( \uparrow \) represents estimated best score (EBS) for node i.

[2] Data transmission
   Node i selects the next forwarder and a channel (out of the top m available channels) using softmax action selection policy. Node i informs all the set of nodes who responded it, about the routing decision and additionally sends data packet to the selected next forwarder. If the action selected is not T that is, termination, then node prepares for forwarding in the next time slot, while other nodes expunge the packets. If termination action is chosen, all nodes in S expunge the packet. Upon selection of routing action, the counting variable \( \nu \) which is number of times up to n, nodes S have received packet from i and decision a is taken, is updated.

\[
\nu_n(i, S, a) = \begin{cases} 
\nu_{n-1}(i, S, a) + 1, & \text{if } (S, a) = (S^i, a_n^i) \\
\nu_{n-1}(i, S, a), & \text{if } (S, a) \neq (S^i, a_n^i).
\end{cases}
\]

After transmission and relaying is complete, the score vector \( \Lambda \) is updated by node i as

- For set of S nodes who responded if action a is chosen,

\[
\Lambda_{n+1}(i, S, a) = \Lambda_n(i, S, a) + \alpha \gamma_i(i, S, a)
\]

\[
x \left( -\Lambda_n(i, S, a) + g(S, a) + \Lambda_n^\alpha \right)
\]

Where \( g(S,a) \) is the reward obtained by taking decision a when set S of nodes receive a packet and \( \alpha \) is the learning rate [1].

- Otherwise,

\[
\Lambda_{n+1}(i, S, a) = \Lambda_n(i, S, a)
\]

Also, node i need to update its EBS for future acknowledgement. It does so as follows

\[
\Lambda_{\max}^i = \max_j A(S^i_n) \Lambda_{n+1}(i, S^i_n, j)
\]

Where A(S) is set of available actions when nodes in S receive a packet.

[3] Route maintenance

While transmitting the data packet to the selected next forwarder and on the selected available channel, if there is an arrival of primary user to use that channel then, reinforcement learning is employed to select another available channel either for the same relay or for other nodes.
At initial stage, all the nodes periodically shares information about channel availability as shown in figure 1, with each other so that all of them can estimate channel availability metric of all the respective available channels between two nodes. As shown in fig. 2(a), the source S whenever has data to be sent to destination D, frames a relay request message and send it to all its neighbors. The set of neighbors interested in forwarding the data of S reply back with an acknowledgement (shown in fig. 2(b)).

Node S uses reinforcement learning approach to select the next forwarder of data packet and sends data to it, apart from which it also informs the remaining nodes about the decision taken so that they can participate in some other data transmission. This has been shown in fig.1(c). This process is repeated by the selected node to find the next forwarder of the data packet, until destination is reached. For every successful data packet transmission, there is a reward with positive constant value.

![Fig. 1 Nodes sharing beacon messages](image)

![Fig. 2 (a) Source sends request to its neighbors (b) Interested neighbors reply back with an acknowledgement (c) Source sends data packet to the relay it selected as next forwarder and also informs other nodes about the decision taken (d) Selected relay node repeats the same process to find the next forwarder node.](image)

**CONCLUSION AND FUTURE SCOPE**

The opportunistic routing scheme discussed maximizes the expected average per packet reward from a source to a destination in the absence of any knowledge regarding network topology and link qualities. The use of reinforcement learning (RL) has been advocated to achieve context awareness and intelligence in wireless networks as it optimally “explores” and “exploits” the spectrum opportunities in the network in order to route the data [3]. Incorporating congestion control in opportunistic routing algorithms to minimize expected delay without the topology and the channel statistics knowledge is an area of future research.

**REFERENCES**


