A Survey on the Communication Protocols and Security in Cognitive Radio Networks

Natarajan Meghanathan

Department of Computer Science, Jackson State University,
Mail Box 18839, 1400 John R. Lynch Street, Jackson, MS 39217, USA
natarajan.meghanathan@jsums.edu

Abstract: A cognitive radio (CR) is a radio that can change its transmission parameters based on the perceived availability of the spectrum bands in its operating environment. CRs support dynamic spectrum access and can facilitate a secondary unlicensed user to efficiently utilize the available underutilized spectrum allocated to the primary licensed users. A cognitive radio network (CRN) is composed of both the secondary users with CR-enabled radios and the primary users whose radios need not be CR-enabled. Most of the active research conducted in the area of CRNs has been so far focused on spectrum sensing, allocation and sharing. There is no comprehensive review paper available on the strategies for medium access control (MAC), routing and transport layer protocols, and the appropriate representative solutions for CRNs. In this paper, we provide an exhaustive analysis of the various techniques/mechanisms that have been proposed in the literature for communication protocols (at the MAC, routing and transport layers), in the context of a CRN, as well as discuss in detail several security attacks that could be launched on CRNs and the countermeasure solutions that have been proposed to avoid or mitigate them. This paper would serve as a good comprehensive review and analysis of the strategies for routing and transport protocols and security issues for CRNs as well as would lay a strong foundation for someone to further delve onto any particular aspect in greater depth.

Keywords**:** Cognitive Radio, Secondary User, Medium Access Control, Routing Protocols, Transport Layer Protocols, Security Attacks and Solutions

1. Introduction

A cognitive radio is defined as a radio that can change its transmitter parameters based on the interaction with the environment in which it operates [1]. A cognitive radio (CR) has the ability (cognitive capability) to sense and gather information (such as the transmission frequency, bandwidth, power, modulation, etc) from the surrounding environment [2] as well as has the ability (reconfigurability) to swiftly adapt the operational parameters, for optimal performance, according to the information sensed [3]. With the above features, the cognitive radio technology is being perceived as the key enabling technology for the next generation dynamic spectrum access networks that can efficiently utilize the available underutilized spectrum allocated by the Federal Communications Commission (FCC) to licensed holders, known as *primary users*. Cognitive radios facilitate a more flexible and comprehensive use of the limited and underutilized spectrum [4] for the *secondary users*, who have no spectrum licenses.

Cognitive radios enable the usage of temporally unused spectrum, referred to as *spectrum hole* or *white space* [3], and if a primary user intends to use this band, then the secondary user should seamlessly move to another spectrum hole or stay in the same band, altering its transmission power level or modulation scheme to avoid interfering with the primary user. Traditional spectrum allocation schemes [5] and spectrum access protocols may no longer be applicable when secondary unlicensed users coexist with primary licensed users. If secondary users are allowed to transmit data along with primary users, the transmissions should not interfere with each other beyond a threshold. On the other hand, if secondary users can transmit only in the absence of primary users, then a secondary user transmitting data in the absence of a primary user should be able to detect the reappearance of the primary user and vacate the band. There is a significant amount of research currently being conducted and more need to be performed to develop new spectrum management approaches related to cognitive radio for both spectrum sensing and dynamic spectrum sharing.

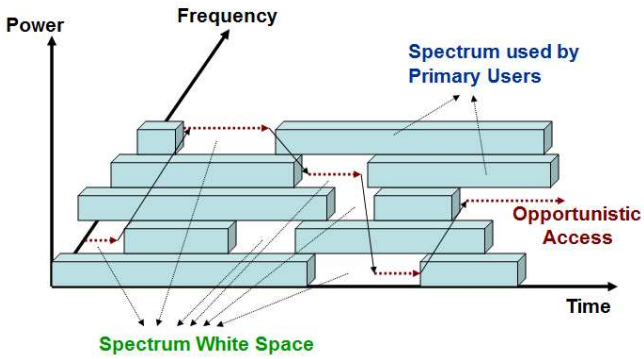

Figure 1. Spectrum Usage – Opportunistic Access of Spectrum White Space and Channel Switching by a Cognitive Radio User

A cognitive radio network architecture (Figure 2) includes components corresponding to both the secondary users (secondary network) and the primary users (primary network). The secondary network is composed of a set of secondary users with or without a secondary base station, all of which are equipped with CR functions. A secondary network with a base station is referred to as the infrastructure-based CR network; the base station acts as a hub collecting the observations and results of spectrum analysis performed by each CR secondary user and deciding

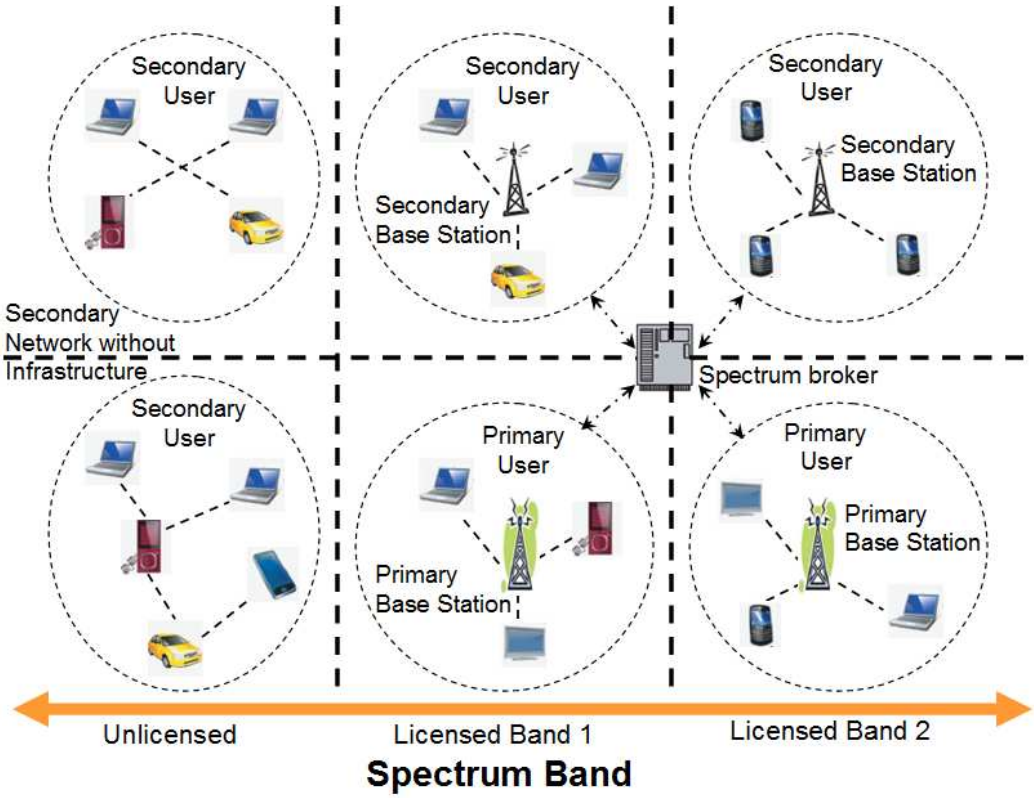

Figure 2. A Cognitive Radio Network Architecture with Primary and Secondary User Networks

on how to avoid interference with the primary networks. As per this decision, each CR secondary user reconfigures his communication parameters. A secondary network without a base station is referred to as the infrastructure less – cognitive radio ad hoc network (CRAHN). In a CRAHN, the CR secondary users employ cooperation schemes to exchange locally observed information among the devices to broaden their knowledge on the entire network, and decide on their actions based on this perceived global knowledge. A primary network comprises of primary users and one or more primary base stations, all of which are in general not equipped with CR functions. Hence, if a secondary network shares a licensed spectrum band with a primary network, the secondary network is required to be able detect the presence of a primary user and direct the secondary transmission to another available band that will not interfere with the primary transmission. Figure 1 illustrates the opportunistic access of the spectrum white space and switching of the frequency bands by a CR secondary user at the incidence of use by a primary user. Figure 2 illustrates cognitive radio network architecture with both the primary user network and the secondary user network (with and without infrastructure – base station support).

A cognitive radio network architecture (Figure 2) includes components corresponding to both the secondary users (secondary network) and the primary users (primary network). The secondary network is composed of a set of secondary users with or without a secondary base station, all of which are equipped with CR functions. A secondary network with a base station is referred to as the infrastructure-based CR network; the base station acts as a hub collecting the observations and results of spectrum analysis performed by each CR secondary user and deciding on how to avoid interference with the primary networks. As per this decision, each CR secondary user reconfigures his communication parameters. A secondary network without a base station is referred to as the infrastructure less – cognitive radio ad hoc network (CRAHN). In a CRAHN, the CR secondary users employ cooperation schemes to exchange locally observed information among the devices to broaden their knowledge on the entire network, and decide on their actions based on this perceived global knowledge. A primary network comprises of primary users and one or more primary base stations, all of which are in general not equipped with CR functions. Hence, if a secondary network shares a licensed spectrum band with a primary network, the secondary network is required to be able detect the presence of a primary user and direct the secondary transmission to another available band that will not interfere with the primary transmission. Figure 1 illustrates the opportunistic access of the spectrum white space and switching of the frequency bands by a CR secondary user at the incidence of use by a primary user. Figure 2 illustrates cognitive radio network architecture with both the primary user network and the secondary user network (with and without infrastructure – base station support).

The current spectrum allocation and sharing schemes according to three criteria: (1) Spectrum bands in use by a CR user; (2) Network architecture and (3) Access behavior of CR users.

- **Classification based on Spectrum Bands used by the CR User:** Based on the spectrum bands in use by a secondary user, the spectrum sharing scheme could be classified as open spectrum sharing and hierarchical spectrum access model. In the open spectrum sharing model, the secondary users access

the unlicensed spectrum band and no user owns any spectrum license; hence, all users have the same access rights in using the unlicensed spectrum. In the hierarchical spectrum access model [6], the secondary users share the licensed spectrum bands with the primary users. Since primary users need not be equipped with cognitive radio, they have all the priority to use the spectrum band. Hence, when a primary user reclaims a spectrum band for use, the secondary users currently using the spectrum band and the near by bands will have to adjust their operating parameters (such as power, frequency and bandwidth) to avoid interrupting the primary users. The hierarchical spectrum access model can be further divided into two categories, depending on the access restrictions on the secondary users:

- ***Spectrum underlay*:** With this model, the secondary CR users coexist along with the primary users, and use the licensed spectrum band without exceeding the interference temperature limit/threshold. If primary users transmit data all the time in a constant mode, there is no need for the secondary CR users to detect for available spectrum band; instead, they can just continue to use the spectrum (of course, only for short-range communication).
- ***Spectrum overlay*:** With this model, the secondary CR users can only use the licensed spectrum when the primary users are not transmitting. So, there is no need for the CR users to operate under an interference temperature limit; however, the tradeoff is that the CR users need to repeatedly sense the licensed frequency band and detect the spectrum white space, to avoid interfering with the primary users. If a primary user is detected, the CR users have to change to another spectrum.

- **Classification based on the Network Architecture:** Based on the network architecture, the spectrum sharing model can be divided into centralized and distributed architectures. Under the centralized model, a central entity controls and coordinates the spectrum allocation and access of secondary users. With the distributed spectrum sharing model, the users make their own decision regarding spectrum access based on their local observation of the spectrum dynamics. The centralized controller model is expensive and also not suitable for ad hoc emergency or military use. The distributed spectrum sharing model is relatively less expensive and can be used in infrastructure less mode.
- **Classification based on Access Behavior of Secondary CR Users**: Based on the access behavior of secondary users, the spectrum sharing model can be categorized as either cooperative or non cooperative. Under the cooperative model, the secondary users often belong to the same service provider and coordinate between themselves to collectively maximize the benefit to the entire group. On the other hand, under the non cooperative model, secondary users access the open spectrum band, and aim at maximizing their own benefit from using the spectrum resources.

The above introduction on the basics of CRNs lays the groundwork for understanding the rest of the paper, which is organized as follows:

- Section 2 reviews the medium access control protocols proposed for both infrastructure-based and ad hoc CRNs. MAC protocols are typically either time-slotted or random access based. We discuss the characteristics of random-access based and time-slot based MAC protocols (along with some representative solutions) for both infrastructure-based and ad hoc CRNs.
- Section 3 reviews the two broad categories of routing protocols (those based on full spectrum knowledge and those based on local spectrum knowledge) along with their representative solutions and analyzes their pros and cons.
- Section 4 presents state-of-the-art transport layer solutions for CRNs. With spectrum sensing and sharing being integral to the functioning of a CRN, cross-layer protocol design (for interaction and sharing of state information across layers) is considered a more logical approach for designing transport layer protocols for CRNs. Nevertheless, there also exist some transport layer solutions that preserve the layering approach. As can be seen, only a handful of end-to-end transport layer protocols have been proposed for CRNs from both the cross-layer and layer preserving perspectives and much work needs to be done in this area.
- Section 5 reviews the various security attacks possible on CRNs by exploiting the characteristics of these networks and the operating principles adopted by the various categories of communication protocols at the physical, MAC, routing and transport layers.
- Section 6 concludes the paper by drawing some general inferences from the survey on the communication protocols for the MAC, routing and transport layers, and the associated security issues.

2. Medium Access Control Protocols for Cognitive Radio Networks

In this section, we will focus on the spectrum access problem wherein multiple CR users share the spectrum and determine who gets access to the channel and when. In this context, we discuss the medium access control (MAC) protocols that have been proposed for both the infrastructure-based and decentralized/ ad hoc cognitive radio networks. The MAC protocols for both categories of CR networks can be either time-slotted, random access or both. The time-slotted MAC protocols require network-wide synchronization and operate by dividing time into discrete slots for both the control channel and data transmission. On the other hand, the random access protocols do not require time synchronization, and are based on the CSMA/CA (carrier sense multiple access/collision avoidance) principle wherein a CR user monitors the spectrum band to detect the presence of any transmission from peer CR users and if so, transmits after backing off for a random duration, to reduce collisions due to simultaneous transmissions. We begin this section with a description of the common control channel (CCC) – a

key component in the design of the MAC protocols for decentralized/ad hoc cognitive radio networks.

2.1 Common Control Channel

The CCC is used for neighbor discovery as well as for path discovery and establishment. Nodes share their neighbor information on different interfaces through the broadcast messages sent out on the CCC to all the potential neighbors, using a high transmission power, corresponding to the maximum transmission range of the CR nodes. The CCC could be either in-band or out-of-band with respect to the data channels. If in-band, the CCC may be one of the data channels to which all nodes can tune in; if a data channel common to all CR nodes is not possible to be found, then the network could employ more than one CCC, each of which having certain region of coverage. In the case of out-of-band CCC, a dedicated control channel, separate from the data channels, is used for control signaling, either network-wide or coverage-based. The CCC and the data channels could all be accessed through a single radio, in which case the routing solutions are prone to the channel deafening problem wherein the control message received on one channel is not received when the radio is tuned to a different data channel. If a dedicated radio is allotted for the CCC, one could avoid the channel deafening problem [6]; however it would be expensive to employ more than one radio per CR node, and also CR nodes employing more than one radio suffer from the cosite interference problem [6] according to which when two or more radios are located on the same device – signals transmitted and/or received on one radio interfere with signals transmitted and/or received on the other radio.

2.2 MAC Protocols for Infrastructure-based Cognitive Radio Networks

Random Access Protocols: In [7], a CSMA based random access protocol was proposed for an infrastructure-based cognitive radio network under the assumption of use of a single transceiver and in-band signaling. The protocol facilitates the coexistence of the primary and CR users by requiring the latter to adapt their transmission power to maintain the interference to the primary users within a pre-decided threshold. The primary users coordinate with a primary base station and the CR users coordinate with a CR base station, and establish a direct single-hop connection with their respective base stations. The primary network follows the classical CSMA protocol according to which a primary user senses the channel for a period (τ_p) before sending the Request to Send (RTS) packet to its base station for which the latter may reply with a Clear to Send (CTS) signal if available for the data transfer. The CR users have a relatively much longer carrier sensing time (τ_s, where $\tau_s >> \tau_p$) so that the primary users get the priority to access the spectrum. The CR base station decides on the transmission power and data rate for the transfer depending on the distance between itself and the CR users. A CR user is allowed to send just one packet in a round of negotiation to reduce or avoid interference and collisions with the transmissions of other primary users. The random access protocols require significant interaction between the primary and CR networks; otherwise, the CR users are oblivious of any failed transmissions of a primary user. Also, the transmission power of the CR users needs to be partitioned to several discrete levels (not just low and high levels) to reliably protect the primary users from interference as well as to maximize throughput by operating the CR devices at the appropriate level.

Time Slotted Protocols: The time slotted protocols follow the IEEE 802.22 centralized MAC standard [8] for cognitive radio networks. The 802.22 standard uses simple time division multiplexing in the downstream direction, and demand assigned TDMA (Time Division Multiple Access) in the upstream direction. The base station manages all the CR users in its cell. Time is slotted into multiple superframes, each comprising multiple MAC frames preceded by the frame preamble. A Superframe Control Header (SCH) is located at the start of each superframe to inform the CR users about the current available channels, different bandwidths supported, future spectrum access time, and etc. The MAC frame is composed of a DS subframe and a US subframe. The DS subframe consists of a preamble that deals with synchronization and channel estimation, a frame control header containing the sizes of the DS- and US-MAP fields with channel descriptors, and the DS/US-MAPs provide the scheduling information for user bursts. The US subframe consists of an Urgent Coexistence Situation (UCS) notification field that informs about the primary licensees that have just been detected; the other fields are used to derive the distance from the base station and the individual bandwidth requests. The main drawback with the time-slotted protocols is the use of heavy headers as part of the frames, leading to a reduced throughput.

2.3 MAC Protocols for Cognitive Radio Ad hoc Networks

The MAC protocols for infrastructure less cognitive radio ad hoc networks (decentralized CR networks) require increased cooperation among neighboring nodes to facilitate a scalable architecture that supports flexible deployment, distributed spectrum sensing, sharing and access. The main design issues include network-wide time synchronization and information exchange among neighboring nodes with minimum overhead.

Random Access Protocols: Random access protocols for CR ad hoc networks can be categorized depending on the number of transceivers required per CR and the requirement for a Common Control Channel (CCC): (1) Protocols requiring the support of multiple radio transceivers; (2) Protocols requiring the support of only a single radio transceiver; (3) Protocols that assume the existence of a CCC; and (4) Protocols that make use of a non-global CCC. In this section, we discuss a representative protocol from each category.

In [9], the authors proposed a distributed channel assignment (DCA) based MAC protocol that uses multiple transceivers, a dedicated out-of-band control channel for signaling, as well as spectrum pooling to reliably detect the activity of the primary network. Each node maintains a list of currently used channels of its neighbor nodes and a list of free channels derived from the former and the spectrum pool. During a RTS-CTS handshake, the sender and receiver match their list of free channels and agree on a

common channel to use. The RTS-CTS messages also facilitate the neighboring CR users to update their used channel and free channel lists. The main drawback of the DCA protocol is the requirement for a separate control channel to support the RTS-CTS exchange, and also there is no primary user-related adaptation for channel usage. In [10], a single radio transceiver version of the DCA protocol has been proposed with the idea of alternately monitoring the control channel and the data spectrum bands for signals. The Single Radio Adaptive Channel (SRAC) algorithm proposed in [10] uses a frequency division multiplexing like scheme wherein a CR user transmits packets on a larger spectrum but receives return acknowledgments over smaller spectrum bands for efficient spectrum utilization. A CR node maintains the list of receive bands of all its neighbor nodes. When a CR node senses its current transmission channel to be occupied by a primary user, it sends a notification packet in the receive bands of its neighbor nodes, and switches to the band that is confirmed to by all the neighbor nodes. In the meanwhile, the CR node transmits on the receive band of a neighboring node that is yet to acknowledge for the notification packet. The drawback is that the signaling traffic overhead associated with maintaining the updated receive spectrum bands of all the neighbor nodes. Also, control messages that are not sent on the receive bands of a node are not listened to, leading to longer *deaf* periods.

The CREAM-MAC (Cognitive Radio EnAbled Multichannel MAC) [11] and SCA-MAC (Statistical Channel Allocation MAC) protocol [12] are examples of MAC protocols that assume the existence of a global CCC (from the unlicensed 2.4 GHz band) that is agreed upon by all the CR nodes in their neighborhood. Under this assumption, the functioning of this category of MAC protocols mimics that of the CSMA-standard for centralized networks. While CREAM-MAC is designed based on a four-way dialog (RTS, CTS, Channel-State-Transmitter: CST and Channel-State-Receiver: CSR packets) on the GCCC, the SCA-MAC employs only a two-way handshake of the control frames (Channel Request to Send and Channel Clear to Send) on the GCCC to facilitate the sender and receiver to tune their transceivers to a mutually agreed upon data channel. While the four-way dialog of CREAM-MAC facilitates the prospective sender and receiver to exchange additional control information with regards to the availability, reliability and quality of the data channels, it could add considerable delay for real-time applications as well as increase energy consumption at the CR nodes. In this context, SCA-MAC is more suitable for delay-sensitive real-time applications involving energy-constrained CR nodes; where as CREAM-MAC is more suitable for QoS-sensitive and/or delay-tolerant applications that can operate at higher energy costs. SCA-MAC also explores the availability of a backup data channel to increase throughput. The main weakness of CREAM-MAC and SCA-MAC kind of protocols is their total dependence on an offline mechanism to facilitate the availability of a global CCC. The

The Opportunistic Cognitive-MAC (OC-MAC) [13] and the more recent Decentralized Non-Global MAC (DNG-MAC) [14] protocols are examples of MAC protocols that do not require the presence of a global CCC for deciding spectrum access among neighboring CR users. OC-MAC assumes that the CRN co-exists with a wireless local area network (WLAN) and use of the IEEE 802.11 DCF (Distributed Coordination function) mechanism at the CR nodes to compete with one another for data channel reservation. The assumption of co-existence with a WLAN is questionable because WLANs typically operate in the unlicensed ISM bands (e.g., 2.4 GHz) and cognitive radio networks operate in licensed spectrum bands. The DNG-MAC protocol uses the TDMA (Time division multiplexing mechanism) to fairly allocate the control channel to all the available CR nodes; the common control channel is one of the best available channels selected by the first CR node that initiates the data communication. The CCC is divided into time slots of fixed length; with each time slot comprising of a listening period (to which all CR nodes are synchronized to listen) and a transceiving period (during which the CR nodes exchange the list of freely available data channels). The premise of DNG-MAC is that since all CR nodes starve for a data channel to use, there will not be wastage of the resources with the assignment of a time slot of the control channel for every CR node. Though this assumption simplifies the design of DNG-MAC and avoids the complex synchronization overhead typically seen with time-slot based MAC protocols (see section 2.3.2), it is difficult to expect the data channels to be available for the same duration as that of the time slots of the control channel (the availability of the spectrum holes in a CRN is non-deterministic) and the time slot per CR node has to be re-calculated upon the inclusion/exclusion of a CR node in the network. This also implies that the MAC protocol to be also not flexible for changes in the network topology due to node mobility.

Time Slotted Protocols: For this category of CR ad hoc network protocols, we discuss the C-MAC (Cognitive MAC) protocol [15], based on synchronized time slots, and include the use of a rendezvous channel (RC) and a backup channel (BC). The RC is the channel that exists for the longest time for use for the CR users throughout the network and is used for node coordination, primary user detection, as well as multi-channel resource reservation. The BC is locally determined at each CR user, through out-of-band measurements, and is used as an alternate spectrum band in the case of appearance of a primary user. In C-MAC, each spectrum band comprises of recurring superframes, each composed of a beacon period (BP) and a data transmission period (DTP). Each BP is time slotted so that the individual CR users can transmit their beacons without interference. The RC is used to exchange the BP schedules of nodes to prevent simultaneous transmission over all the spectrum bands. A CR user announces the need for any new data spectrum band through the beacons, and also informs about any spectrum change over the RC. Periodic tuning to the RC allows a CR user to re-synchronize and obtain the recent neighborhood topology information. The time slotted nature of C-MAC also facilitates the use of a non-overlapping quiet period (QP) for each spectrum band, through which one could differentiate a primary user from a CR user. The main drawbacks of the C-MAC are that it requires the RC to be

a dedicated spectrum band that is not used by any primary user, which is difficult to guarantee in distributed networks. Also, due to the requirement to include the beacons with the load and channel usage information in the BP of a superframe, the protocol is not scalable for a larger number of CR users. It is difficult to enforce the non-overlapping nature of the BPs and the quiet periods, without the presence of a central entity. In [16], a distributed slotted protocol was proposed to circumvent the use of a RC by providing in-band signaling through a dedicated control window in addition to the beacon and data transfer periods.

2.4 Channel Hopping based Control Channel Identification for MAC Protocols

Recently, the research community has started to explore the use of channel hopping among the CR nodes as a potential alternative strategy for control channel identification in the design MAC protocols for CRNs. The idea here is to let the CR nodes to generate their own channel hopping sequences and when two CR nodes (sender and receiver) hop to a common channel, they can exchange control packets on the common channel and negotiate for data communication [17]. Because the CR nodes can rendezvous on every available channel, channel hopping-based control channel identification can overcome the control channel saturation problem seen with the previous approaches where only few channels are considered candidates for a common control channel. Also, channel hopping based control channel identification requires only a pair-wise rendezvous (between the sender and receiver) and does not need a globally available common control channel. However, the main drawback of the channel hopping-based approach is the channel access delay as a CR node has to keep switching one channel after another before it can initiate any communication with its neighbor. This would incur a significant amount of time (called the Time to Rendezvous, TTR) as the number of available channels increases [6]. If there are N available channels, it would require at most N^2 time slots for two CR nodes to identify a common channel [18][19]. While coordinator-based channel hopping [20] can reduce the TTR value, it is a centralized approach and is not scalable. Permutation-based [18] and Quorum-based [19] mechanisms are some of the examples for scalable distributed channel-hopping based control channel identification mechanisms proposed in the literature. A potential drawback with channel-hopping based control channel identification is that the technique is likely to suffer a significantly long delay when used for end-to-end communication in multi-hop CRNs. At the worst case, the end-to-end delay could be the sum of per-hop TTR values.

3. Routing in Cognitive Radio Networks

The problem of routing in multi-hop cognitive radio networks (CRNs) refers to the creation and maintenance of wireless multi-hop paths among the CR users (also called Secondary Users, SUs) by deciding the relay nodes and the spectrum to be used on each of the links in the path. Even though the above problem definition exhibits similarities with routing in multi-channel, multi-hop ad hoc networks and mesh networks, the challenge in the form of dynamic changes in the available spectrum bands due to simultaneous transmissions involving primary users needs to be handled. Any routing solution for multi-hop CRNs needs to be tightly coupled with spectrum management functionalities [16] so that the routing modules can take more accurate decisions based on the dynamic changes in the surrounding physical environment. As the topology of multi-hop CRNs is highly influenced by the behavior of the PUs, the route metrics should be embedded with measures on path stability, spectrum availability, PU presence, etc. For instance, if the PU activity is low-to-moderate, then the topology of the SUs is almost static, and classical routing metrics adopted for wireless mesh networks could be employed; on the other hand, if PUs become active very frequently, then the routing techniques employed for ad hoc networks could be more applicable [4]. Also, the routing protocols should be able to repair broken paths (in terms of nodes or used channels) due to the sudden reappearance of a PU.

With respect to the issue of spectrum-awareness, the routing solutions for CRNs could be classified as those based on the full spectrum knowledge and local spectrum knowledge. In the former case, the spectrum availability between any two nodes in the network is known to all the nodes (or to a central control entity). This is often facilitated through a centrally-maintained spectrum database to indicate channel availabilities over time and space. The routing solutions built on the top of the availability of full spectrum knowledge are mostly based on a graph abstraction of the CRN and, though not often practically feasible for implementation, are used to derive benchmarks for routing performance. The routing module is not tightly coupled with the spectrum management functionalities for centralized full spectrum knowledge-based solutions. On the other hand, for local spectrum knowledge based solutions, information about spectrum availability is exchanged among the network nodes along with traditional network state information (such as the routing metrics, node mobility, traffic and etc). On these lines, the local spectrum knowledge-based routing protocols could be further classified as those that aim to minimize the end-to-end delay, maximize the throughput and maximize the path stability. In addition to the above, we have also come across probabilistic approaches for routing (e.g., [23, 24]) in which CR users opportunistically transmit over any spectrum band available during the short idle periods of the surrounding primary users.

3.1 Routing Solutions based on Full Spectrum Knowledge

The general strategy under this approach is to first abstract the physical network as a graph with nodes and edges with weights, all capturing the network dynamicity and spectrum availability, and then run a route calculation algorithm on the graph to find a path/tree or any appropriate communication topology connecting the nodes (source-destination pairs for unicast routes; source-receiver nodes for multicast communication and etc). In [25], the authors propose a generic framework for modeling CRNs comprising of nodes with a single half-duplex cognitive radio transceiver, which can be tuned to the various available spectrum bands or channels. The framework is

based on creating a layered graph that features a number of layers equal to the number of available channels. Each CR device is represented in the layered graph with a node, *A*, and *M* additional sub nodes A_1, A_2, …, A_M, one for each available channel, and *M* is the total number of available channels. Three kinds of edges exist in this layered graph: The *access* edges connect a node with all its corresponding sub nodes; the *horizontal* edges connect the sub nodes of two different nodes on the same logical layer if the two nodes can be tune to the corresponding channel; the *vertical* edges connect sub nodes of different layers of a single CR device to switch from one channel to another. Figure 3 illustrates a layered graph with four nodes and two channels. The weights of the horizontal edges typically capture the cost involved in propagating data from one CR device to another node on the particular channel and the weights of the vertical edges typically capture the cost involved in switching from one channel to another at a particular CR device. Graph theoretic algorithms optimizing the overall cost of a path between every source-destination pair, or trees connecting a group of nodes (including all the nodes in the graph, in the case of spanning trees) could then be run on such a weighted layered graph. As an example of the application of the layered graph model, in [26], the authors represent the horizontal edge weights to be proportional to the traffic load and interference, and propose a centralized heuristic algorithm to calculate shortest paths. The main weaknesses of the layered graph model presented above are that it requires a network-wide signaling to generate such a global graph at each node and it may not scale well as the network dimensions increase.

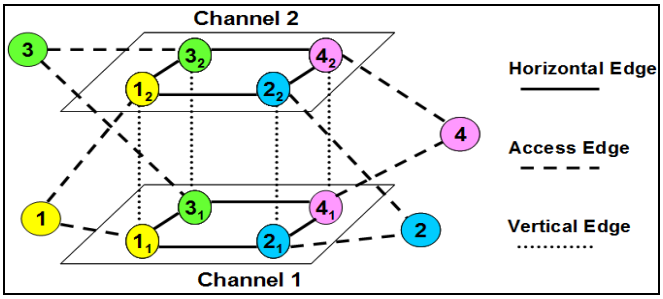

Figure 3. Example for Layered Graph Model

To circumvent the scalability problem, an edge coloring model was proposed in [27] that gets away with representing sub nodes of a node in multiple layers, and instead connects the nodes with edges of different colors, with each edge color indicating whether the nodes can communicate on a particular channel (i.e., one color per channel). The edge coloring model has also been extended to locally optimize the adjacent hop interference.

Another solution is to capture the network as a conflict graph [28] where each node in the conflict graph is actually an edge between two nodes in the network graph and there exists an edge in the conflict graph only if the edges corresponding to the two end nodes of the conflict graph cannot be active at the same time. One can then run a maximum independent set (or maximum clique) heuristic on the conflict graph to derive a conflict-free channel assignment for the original network graph. Nevertheless, all of the three graph theory models (layered, colored or conflict graphs) suffer from the weakness of being centralized in nature and requiring the full knowledge of the network topology and the available spectrum bands.

3.2 Routing Solutions based on Local Spectrum Knowledge

The routing solutions based on local spectrum knowledge (that varies both in time and space) are distributed in nature and differ depending on the specific metric used to assess the route quality. One class of routing distributed local spectrum knowledge-based routing protocols assume the availability of a Common Control Channel (CCC) [29] across all the CR nodes in the network. Route discovery is launched through a Route-Request-Reply (RREQ-RREP) cycle (similar to that of the classical ad hoc networks) run on the CCC at all the nodes. An AODV (ad hoc on-demand distance vector) [30]-style routing protocol for CRNs has been proposed in [31]: the RREQs are broadcast on the CCC; the intermediate forwarding nodes keep track of the cost accumulated on the path traversed by the RREQs; the destination selects the path with the minimum cost incurred and the RREP is propagated back on the reverse route setup during RREQ propagation. A major flaw with the use of CCC for RREQ propagation is that the availability of the data channels at the intermediate CR nodes cannot be easily/accurately captured.

An alternate strategy for route discovery without using the CCC is to broadcast the RREQ packets on all the available channels and let a flood of RREQ packets reach the destination, on multiple paths and on multiple channels. The destination processes these RREQ packets and selects the best path(s) that satisfies the route selection criteria. The RREQ messages are forwarded on all the available channels. The CAODV-BR [32] protocol, a cognitive adaptation of the AODV routing protocol, chooses backup routes in conjunction with a primary route and reverts to one of these backup routes when one or more hops/channels in the primary route is occupied by a primary user. In a similar vein, the authors in [33] propose to use a backup control channel, in addition to a principal control channel (both of which are locally selected) at a node to coordinate the route discovery and channel switching mechanisms. Nevertheless, broadcasting across all the spectrum bands for route discovery would be too much of an overhead compared to broadcasting the RREQ packets on a single CCC and including information about all the available channels at each node in these RREQ packets.

Minimum Power Routing: In the minimum power routing protocol proposed in [34], the weight of a link (for each interface) is modeled as the transmission power to be spent to reach the other end of the link within an appreciable received signal threshold. An energy loss is associated to switch from one frequency channel to another. An intermediate forwarding node includes in the RREQ the transmission power loss to be incurred for each of its outgoing channels. The destination receives the Route Request packets along all the paths and finds the path that minimizes the sum of the energy lost across all the links and their corresponding channels as well as the

switching energy loss, if any, is incurred. The Route-Reply packet containing information on the chosen route is sent through the CCC. The main weakness of the minimum power routing protocol is that it is oblivious to the presence of primary users and their impact on neighbor discovery among the CR users. In [35], the authors propose an energy-efficient quality-of-service aware routing (EQR) protocol, built on top of the Dynamic Source Routing (DSR) protocol [36] for MANETs: the idea is that the source estimates and specifies, in the RREQ packets, the number of time slots needed for an ongoing session with a destination node; only those intermediate nodes that can commit the requested number of time slots forward the RREQ packets. EQR has been extended as Spectrum and energy-aware routing (SER) protocol [37] for multi-path routing. Both EQR and SER are not suitable for dynamic CRNs in which the availability of the PU channels changes quite unpredictably.

Minimum Delay-based Routing: In [38], the authors propose routing protocols aimed at optimizing the various components of the delay incurred at a node, with the overall objective of minimizing the delay incurred on a path. The delay at a relay node is conceived as the sum of the delays incurred due to switching from channel to another; accessing the channel corresponding to the chosen spectrum band; and the queuing delay suffered by the packet before it is transmitted on the particular channel. The switching delay includes two components: the delay to switch the packet from one frequency band to another frequency band – a measure of the separation of the two frequency bands, and also the delay incurred due to the scheduling (the round-robin scheduling is often chosen for fairness) of the packet transmissions at the node across the spectrum bands in use. Note that the queuing delay suffered by a packet is also influenced by the channel scheduling component of the switching delay. While [38, 39] focused on minimizing the sum of the switching and access delays incurred at the relay nodes; [40] focused on minimizing the sum of the queuing delays at the relay nodes. In [41], the authors proposed a routing protocol that lets an intersecting node (a node that lies on more than one path from the source to the destination) to locally coordinate among the neighboring nodes to decide whether to accommodate an incoming new flow or to redirect it to one of its neighbors to obtain a relief to the workload on the node. If such a route redirection materializes, this would actually lead to a scenario wherein the route discovery RREQ-RREP packets and the data packets traverse different paths – the RREQ-RREP packets traverse through the intersecting node, and the data packets traverse through the neighbor node that took up the load from the intersecting node to provide relief to the latter's workload. In another related work [42], the authors propose to shift traffic to the edge of the network away from the high-density regions to effectively use the available capacity throughout the CRN. This strategy has been observed to maximize the utilization of channel capacity in CRNs, compared to shortest path routing.

Maximum Throughput-based Routing: In [43, 44, 45], throughput-based solutions for routing in CRNs have been proposed. The Spectrum Aware Mesh Routing (SAMER) protocol [43] first establishes paths based on the periodically collected global states, and at the time of packet transmissions, the packets are delivered opportunistically along the path with the highest value for a throughput metric, referred to as the Path Spectrum Availability (PSA). The PSA captures the number of available spectrum blocks at each node as well as their aggregated bandwidth and loss rate. Though throughput is the primary routing objective, SAMER imposes an upper bound on the number of intermediate nodes to be used on the path and for which the PSA values are calculated. In [44, 45], the authors propose a spectrum utility based routing algorithm called ROSA (Routing and Spectrum Allocation) to maximize the throughput. The spectrum utility of a link (i, j) is the product of the achievable capacity of the link and the maximum differential backlog of packets between nodes i and j. ROSA maximizes the weighted sum of the differential backlogs and thereby gives preference to high-capacity links without generating harmful interference to other users (the bit error rate is guaranteed to be within a threshold) – all of these leading to increase in the throughput of the end-to-end communication [45]. ROSA is one of the several cross-layer approach-based solutions for designing spectrum-aware routing protocols to maximize end-to-end throughput in multi-hop CRNs.

In [46], a bandwidth footprint (BFP) minimization-based maximum throughput routing protocol has been proposed to find an appropriate channel and capacity for a session with minimal impact (with respect to interference and throughput) on the ongoing sessions of the PU and SU users. The BFP for a node refers to the interference area of the node for a given transmission power. With a node switching from one band to another and each band incurring a certain footprint corresponding to its transmission power, the objective of the protocol is to minimize the network-wide BFP, which is the sum of the BFPs of all the nodes. The routing protocol goes through an iterative procedure to fit in an incoming session request with the existing sessions. First, the session is assigned to an available capacity on a channel; if this is not sufficient, the transmission power of the band is increased to increase the session rate (referred to as Conservative Iterative Procedure, CIP). However, if the increase in transmission power violates the interference constraints and significantly increases the BFP, the alternative channels are considered to migrate the session to achieve the targeted session rate. To do this, the capacity allocated for the existing sessions in the alternate channel need to be reduced (referred to as Aggressive Iterative Procedure, AIP). If the reduction impacts the quality-of-service guaranteed for these sessions beyond a limit, then the new session is accommodated; otherwise, it is allocated a capacity in the alternate channel. In [47], the above work has been extended to develop a distributed cross-layer optimization algorithm (encompassed with routing, scheduling and power control modules) to iteratively increase the data rates for user communication sessions based on the notions of the CIP and AIP.

In [48], the authors propose a weighted cumulative expected transmission time (WCETT)-based routing

protocol to determine high-throughput routing paths in multi-radio, multi-hop CRNs. The WCETT of a path is defined as the weighted average of (1) the sum of the expected transmission times of the individual links on the path and (2) the maximum value of sum of the expected transmission times of the bottleneck channel used across one or more links/hops on the path. The idea is to avoid the use of the same link over more than one hop on a multi-hop path to reduce co-channel interference along adjacent links. The hypothesis behind WCETT to maximize throughput is to choose a path that would incur channel switching along the links to minimize the delay incurred to wait for the same channel across several links. However, the protocol is only suitable for multi-radio environments, where channel switching is feasible.

In [49], the authors propose a routing metric called Cognitive Transport Throughput (CTT) to capture the potential relay gain over the next hop. The locally calculated CTT values of the links (based on the local channel usage statistics) form the basis for selecting the best relay node with the highest forwarding gain in the Opportunistic Cognitive Routing (OCR) protocol for multi-hop CRNs.

Geographic Routing: In [50], the authors proposed a routing protocol whose objective is to choose the next hop that would minimize the interference to the PUs operating in the vicinity of the transmission and satisfying the QoS parameters for the SUs to the maximum. In this context, they evaluated the use of Nearest Neighbor Routing (NNR) and Farthest Neighbor Routing (FNR) to decide the next hop neighbor for a CR node employing geometric/geographic forwarding. The tradeoff observed is that the FNR scheme achieves a better end to end channel utilization and reliability; whereas, the NNR scheme has a better energy efficiency.

In [51], a spectrum-aware beaconless (SABE) geographical routing protocol has been proposed. The strategy used to select the next hop forwarding node is described as follows: A source or intermediate CR node broadcasts a *request-to-forward* (RTF) packet in its neighborhood. The receiver CR nodes set their reply timer to be proportional to the distance to the destination node – i.e., the receiver node that lies closest to the destination responds the earliest with an *accept-to-forward* (ATF) packet. The RTF-ATF exchange happens on the common control channel and the two nodes negotiate on the data channel to use for the actual packet transfer. A weakness in the above strategy is the implicit assumption that all the nodes in the network know the exact location of the destination at any time; this assumption cannot hold true in the presence of node mobility. Besides, SABE suffers from the *dead end problem*, typical of geographic routing protocols. With the neighbor nodes not exchanging periodic beacon packets, they have to resort to a technique called Beaconless Forwarder Planarization (BFP) [52] to overcome the scenario wherein there are no neighbor nodes that are closer to the destination node vis-à-vis the source or the intermediate node trying to forward the message to the destination. BFP leads to identifying the neighbor node that is closest to the forwarder node to further relay the data towards the destination. However, this would result in significant waiting time for an ATF packet at the forwarder node. A forwarder node has to wait for a maximum timeout period expecting an ATF packet from its neighbor nodes and when only none of them respond within this period, the forwarder node can switch to BFP/nearest neighbor node forwarding.

Class-based Routing: In [53], the authors observed the Farthest Neighbor Routing (FNR) strategy to be more effective to offer better service differentiation for high-priority traffic in dynamic CRNs where the average duration of the availability of the communication channel can be much shorter than the communication time. This observation formed the basis for the development of the Opportunistic Service Differentiation Routing Protocol (OSDRP) for dynamic CRNs. At each node, OSDRP basically sets up multiple forwarding nodes for a destination node, depending on the priority of the traffic flowing to the destination: The larger the priority of the traffic, the farther is the next forwarding node (i.e., more closer to the destination). In another related work [54] on class-based routing for CRNs, the authors propose two routing classes: Class I for routes that require lower end-to-end delay while guaranteeing minimum PU interference avoidance; Class II for routes that prioritize PU protection at the expense of permissible performance degradation for the CR users. The spectrum and next hop forwarding node are selected simultaneously at the time of route search: the RREQs of Class I routes are given priority (for spectrum and next hop node selection) over Class II routes.

4. Transport Layer Issues and Solutions for Cognitive Radio Networks

Research on transport layer protocols for cognitive radio networks is very much in its nascent stages. We have come across relatively few proposals for transport layer protocols for CRNs and performance evaluation studies (e.g. [55]) of the traditional TCP protocols for CRAHNs. In this section, we first identify the reasons for possible packet drops in a mobile wireless CRN; analyze the potential performance degradation when the traditional TCP is run on a CRN; and then discuss the currently available solutions for transport layer protocols for CRNs in the literature.

4.1 Transport Layer Issues and Motivating Examples

Packet losses in a CRN involving mobile wireless nodes may occur due to one of the following factors: (i) Traditional network congestion that could be further aggravated due to reduced link capacity and loss of connection, (ii) Link error, (iii) Collision due to simultaneous transmissions, (iv) mobility of a node from one base station to another, (v) mobility of the intermediate forwarding nodes, (vi) the intermittent spectrum sensing undertaken by the CR users, (vii) the switching of a CR-node between transmitting and spectrum sensing states, (viii) the activity of the primary licensed users of the spectrum, and (ix) large-scale bandwidth variation due to spectrum availability. Factors (vi) through (ix) are characteristic of CRNs and these factors have not been considered in the design of the transport layer protocols for

other categories of wireless networks (e.g. wireless mobile ad hoc networks or sensor networks), motivating the need to design transport layer protocols exclusively for CRNs.

Figure 4. A Hybrid Cognitive Radio Network Layout for Example 1 (adapted from [56])

Example 1: Assume a CR-node is sending data to a destination node in the Internet through an intermediate router (refer Figure 4). Let the link between the router and the destination node be wired and the link between the source node and the router be a CR-link. As is characteristic of a CRN, the source-router CR-link alternates between spectrum sensing and transmission modes. When the nodes for the CR-link enter into the spectrum sensing mode, the source does not receive any acknowledgment packets and hence cannot estimate a RTT (round trip time) for the link. Once spectrum sensing is completed, the source node starts receiving the acknowledgment packets that were waiting at the router. The RTTs for these acknowledgment packets that waited in the network would be quite high as they correspond to the spectrum sensing duration and do not capture the congestion level in the network. Once the backlog of the acknowledgment packets is cleared, the source node starts receiving the acknowledgments for the packets sent at the end of the sensing mode and notices a sudden decrease in the RTT. However, the retransmission timeout (RTO) value for the congestion control algorithm gets unnecessarily increased to extraneous values because of the RTTs of the acknowledgment packets that waited at the router. These acknowledgment packets would have been received if the source node were not in the sensing mode. It takes awhile for the congestion control algorithm to lower its estimate for the RTO value even if the RTT value starts decreasing abruptly once the backlog of acknowledgment packets is cleared. Additive Increase and Multiplicative Decrease (AIMD) is a core principle of standard TCP congestion control algorithms. This contributes to a lower throughput and an under-utilization of available bandwidth.

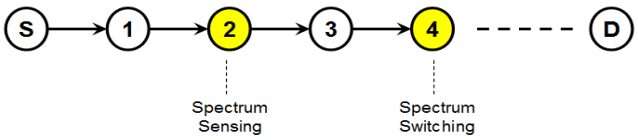

Figure 5. A Multi-hop Cognitive Radio Ad hoc Network Layout for Example 2 (adapted from [57])

Example 2: Consider a multi-hop CRAHN (cognitive radio ad hoc network) shown in Figure 5, with nodes *S* and *D* as the source and destination nodes respectively (adapted from [57]). As we can see from the figure, due to node 2 entering into the spectrum sensing mode, the *S*-*D* path is virtually broken into two connected segments: *S* – 1 and 3 – *D*. Source *S* may eventually timeout waiting for the acknowledgment packets for the transmitted data packets, and this could trigger a retransmission of the data packets, even in the absence of true congestion. If the source *S* does not limit its transmission rate during the spectrum sensing duration and continues to transmit/retransmit the data packets, the queue at the intermediate node 1 may soon be overwhelmed and will succumb to a buffer overflow. In addition, a proper balance has to be maintained between the sensing interval and the data packet transmission time so that the throughput of the connection can be maintained as well as the interference with the PU activities is minimized [58]. A longer sensing interval would correspond to the CR user spending most of the time monitoring the channel rather than transmitting the data packets; on the other hand, a shorter sensing interval could increase the risk of interfering with the activity of a PU [59].

Another factor that needs to be considered in the design of transport layer protocols for CRNs is the uncertain delay caused due to the need for a CR user to successfully search for an available channel once a PU activity is detected in the currently used channel. Unlike spectrum sensing, the time spent to hunt for an available channel is not deterministic and cannot be known in advance to the source on a multi-hop path. This necessitates the need for transport layer protocols to differentiate the spectrum switching state from other causes of route disconnections by requiring an explicit feedback from the nodes affected by the PU activity (node 4 in Figure 5).

4.2 Cross-Layer Approach for Transport Layer Solutions

The solutions to handle the RTO-increase problem and other related issues at the transport layer due to a CR-node entering the spectrum sensing/switching states can be effective only if at least the current status of the node (could be: normal, spectrum sensing, spectrum switching and route failure) is known at the higher end-to-end layers, starting from the network layer. A cross-layer approach to solve the above problems could involve the use of a cognitive knowledge module that is shared by all the layers (Figure 6.1). The physical layer could update a Boolean flag to indicate the current status of the node, and all the other layers could refer to this information to infer whether the node is in the normal transmission mode or spectrum sensing/switching/route failure modes. Below, we describe some of the cross-layer design based solutions to maximize TCP throughput for CRNs.

In [60], the authors propose the use of a finite-state Markov chain (FSMC) model to represent the time-varying behavior of the underlying spectrum and integrate the wireless channel and residual energy state transitions (captured through the FSMC model) with the modules at the

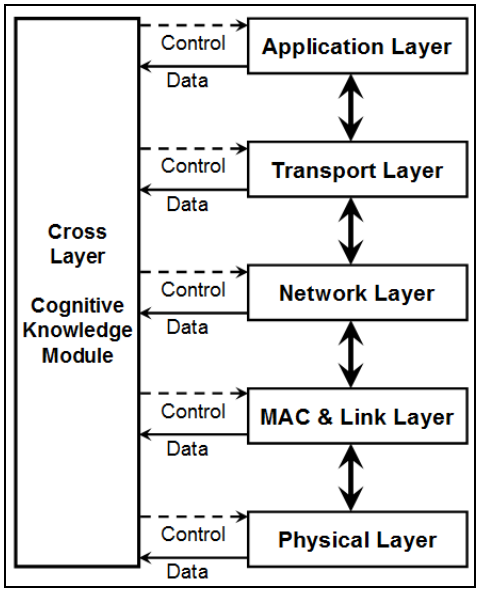

Figure 6.1. Cross-Layer Approach

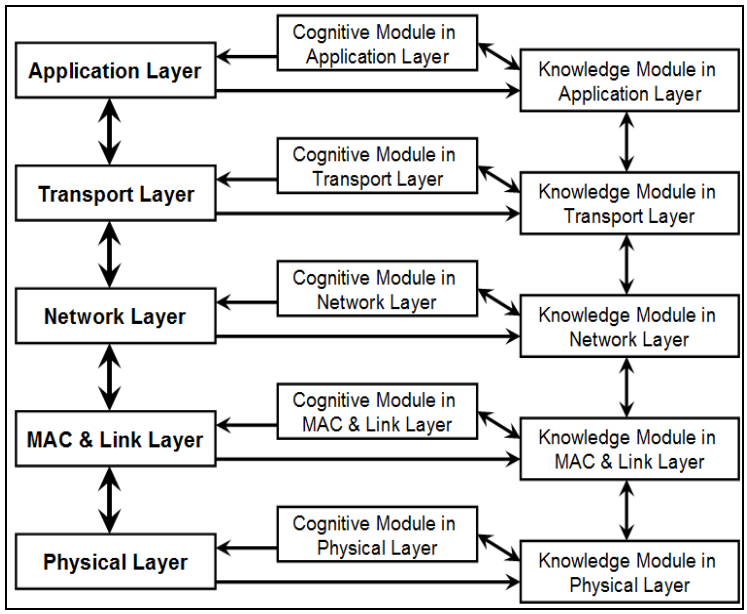

Figure 6.2. Layer-Preserving Approach

Figure 6. Approaches to Extend the Standard TCP/IP Layer Suite for Cognitive Radio Networks (adapted from [57])

various layers: (i) Adaptive modulation and coding, and power control modules at the physical layer; (ii) Adaptive frame size and Automatic repeat request modules at the data link layer and (iii) relay selection module at the network layer, implemented through a distributed primal-dual priority-index heuristic [61] run in the neighborhood of a forwarding node. In [62], the authors showed that the above FSMC-based cross-layer model is also energy-efficient. Though the above scheme contributed to an increase in TCP throughput and extended network lifetime (due to the energy savings), the scheme can be applied only to underlay CRNs and not for overlay CRNs, wherein the additional complexity of channel switching (including spectrum sensing and allocation to facilitate the switching) needs to be considered. In a related work [63], for mobile CRNs, the IP network layer at a node is enhanced with a Channel Assigning Agent (CAA) that forecasts the possibility of link failures due to node mobility and proactively informs the physical layer to sense for available channels and prepare for a channel switch. The CAA was later integrated with the TCP layer [64] to facilitate the congestion control algorithm running at the transport layer to reset the RTO timers and adapt for a prospective channel switch in the near future.

In [65], the authors propose a partially observable Markov decision process (POMDP)-based cross-layer optimization architecture wherein the TCP layer takes the decisions for the functioning of the lower layers based on the feedback (sensing decisions) passed on by the physical layer on the available channels. The TCP layer decides whether to access the channel, and if so, informs the corresponding modulation and coding schemes to be adopted at the physical layer and the frame size to be used at the data link layer. The POMDP model provides a relatively higher degree of freedom for the transport layer compared to the lower layers. The model lets a higher end-to-end layer to completely decide on the operating values for the parameters of the protocols running at the lower layers. Though this appears to be fine from cross-layer perspective, the transport layer might be oblivious to some of the operating constraints of the lower layers in the local neighborhood – as a result, the decision taken may not be appropriate and is bound to have some estimation errors. It would be rather more prudent if the lower layer protocols can have some control over the operating values of their parameters. In another related paper [66] from the same authors of [65], they formulated the channel access problem in a CR network as a stochastic optimization restless bandit problem [67] and developed an indexable version of the optimal channel access policy to facilitate the selection of channels with highest index for maximizing TCP throughput.

4.3 Layer-Preserving Approach for Transport Layer Solutions

In [56], the authors propose a layer-preserving approach to extend the standard TCP/IP layer suite for cognitive radio networks. The idea is to implement two modules – Knowledge module and Cognitive module – as part of each of the layers. The *Knowledge module* at a layer stores information about the application's need and status of local and global networks, all pertaining to the appropriate layer; the *Cognitive module* at a layer is responsible for the algorithms/heuristics to gather knowledge and to generate control signals for managing the operation of the layer based on the information in the Knowledge module. The separation of the knowledge and cognitive decision making modules from the standard modules for each layer preserves the modularity and abstraction concept of the TCP/IP protocol stack as well as reduces the development and maintenance time of new software that would need to be implemented for any of these layers in the context of cognitive radio networks. The layer-preserving architecture (shown in Figure 6.2) can serve as a generic architecture to deploy families of protocols to fulfill the requirements of individual applications, without affecting the core

functionality of the layers in the standard TCP/IP layer stack.

Two solutions, which adhere to the layer-preserving approach, have been proposed in the literature to avoid the abnormal increase in the RTO values because of spectrum sensing/switching. One solution [57] is not to consider the RTTs of the acknowledgment packets that are forced to wait in the network due to the source node or the network (i.e., the intermediate nodes on a multi-hop path) entering the spectrum sensing/switching states. The Knowledge module at the transport layer learns about the node or the network entering into the spectrum sensing/switching states through its interaction with the Knowledge module at the lower layers and updates the Cognitive module. Once the Cognitive module learns about the node or the network entering into the spectrum sensing/switching states, it marks every data packet (that were already sent) whose acknowledgment is yet to be received and updates the TCP process accordingly. When the source node or the network gets back to the transmission mode, the TCP process at the source node, in consultation with the Knowledge and Cognitive modules, decides not to consider (to estimate the RTO value) the RTTs of the acknowledgment packets received for the data packets marked during the node/network's sojourn in the spectrum sensing/switching states in the past.

Another related solution proposed in [56] is to mark an acknowledgment ACK packet (or a sliding window worth of ACK packets) as delayed due to the node/network entering the spectrum sensing/switching states if the RTT of the acknowledgment packet (or the window of packets) at the time of reception is greater than the RTT of the latest acknowledgment packet (or the latest window of packets) received plus 0.9*the spectrum sensing/switching duration. The value for the spectrum sensing/switching duration is estimated by the Cognitive module through the interactions of the Knowledge module with its counterparts in the lower layers. While the duration for spectrum sensing may be fixed per node, the duration for spectrum switching is a stochastic parameter that can be only best estimated, mostly based on the past history (including the statistics of the PU activities). Once an ACK packet (or a window of ACK packets) is perceived to be delayed because of the node/network entering into the spectrum sensing/switching states, the TCP process does not update the RTO and the estimated current estimated RTT. If the RTO timer expires while the node/network is in the spectrum sensing/switching states, the RTO timer is simply reset and no further action is taken.

5. Security Attacks on Cognitive Radio Networks and Countermeasures

In this section, we discuss several security attacks that could be launched on cognitive radio networks and the countermeasure solutions to thwart or mitigate them.

5.1 Attacks on the Common Control Channel (CCC) and Solutions

The centralized and cooperative CRNs are more vulnerable to masquerade and denial of service attacks. The CCC is a single point of failure and is vulnerable for a jamming attack that can effectively destroy the entire CRN. An attacker can inject a strong interference signal to the CCC and disable the reception of valid control messages at the CR receivers, essentially leading to a denial of service (DoS). It is a more energy-efficient and effective strategy for an attacker to just jam the CCC and bring the network down, rather than jamming the entire spectrum band [68, 69]. For centralized CRNs, one can avoid CCC saturation attacks by requiring the MAC control frames to be authenticated and stamped by the base station. The CCC anti-jamming solutions that are currently available for distributed/cooperative CRNs include: (1) Dynamic CCC allocation and (2) CCC key distribution. Dynamic CCC allocation can be accomplished using cross-channel communication [10] and frequency hopping [71]. The idea behind the cross-channel communication approach is to use a CCC currently under jamming attack to notify CR users about the new CCC for receiving control messages if the receiving nodes are free of jamming. Information about the new CCC can be conveyed through a unique frequency hopping sequence that is known only to the CR users. However, any CR user who is compromised by the jammer could receive the notification about the change of the CCC and be able to jam the new CCC. For increased robustness against CCC jamming attacks, the CCC key distribution method is preferable, though it involves significant overhead. The idea behind the CCC key distribution method is to use multiple CCC channels for transmitting control signals. A CR user is assigned the keys for only certain CCCs and not to all of them. This way, even if a CR user is compromised, he cannot extract information from the CCCs for which he does not know the key. The random key distribution approach [69, 72] has been observed to be the most effective approach for CCC key distribution.

5.2 Jamming Attack

The jamming attack is the most common mode of attack for triggering denial of service (DoS) to legitimate primary and secondary users in a CRN. Jamming attacks could be of four types [73]: Constant jammer, Deceptive jammer, Random jammer, and Reactive jammer. A constant jammer sends out data packets continuously without any regards to other users on that channel. A deceptive jammer tricks a legitimate user to switch to "receive" state as they detect a constant stream of incoming data packets. A random jammer takes random breaks while sending jamming signals; during its jamming phase, it may behave either as a constant or random jammer. A reactive jammer senses the channel all the time and transmits the jamming signals upon sensing a communication in the channel. Jamming driven DoS attacks at the physical layer requires an attacker to use a device that is capable of emitting energy at the same frequency used by other devices to communicate and interfere with their communication. In [74], the authors describe an attack scenario involving a single cognitive radio that can repeatedly switch back and forth between several channels after sending the jamming packets in each of them for a fixed period. Jamming driven DoS attacks at the link layer [75] involve the attacker sending out packets on a specific radio channel forcing all

the devices within the radio range to assume that the channel is not idle and postpone their data transmission.

Several detection techniques have been proposed for user at the devices to conclude whether they have been subjected to a jamming driven DoS attack. If a device never passes the carrier-sensing phase of the CSMA (Carrier Sense Multiple Access) medium access control protocol, then it could conclude that it is a victim of a DoS attack. At the physical layer, a strategy proposed [75] is to have a legitimate device collect enough data packets and build a statistical model to distinguish between normal and abnormal levels of noise on the channel. In [73], the authors propose a jamming detection technique that leverages the relationship between signal strength (SS) and packet delivery ratio (PDR): A node concludes itself to having been a victim of jamming attack if its SS is high and PDR is low, and none of its neighbors have a high SS as well as a high PDR. Another related technique, called the Location Consistency Checks technique, suggested in [73] is based on the idea that if all the nearby neighbors of a node have low PDR values, then either the node is being jammed or the quality of the links with its neighbors is poor. Given the above jamming detection techniques, two strategies for defense against jamming attacks have been suggested in the literature [75]: link-layer defense – frequency hopping (switch to a different channel) or network layer defense – spatial retreat (legitimate users change their location to escape from the interference range imposed by an attacker).

Recently, the authors in [76] propose machine learning/game theory-based approaches to defend from jamming attacks. For single channel CRNs (the SUs have access to only one channel at a time), the SUs first go through a Markov decision process [77]-based reinforcement learning phase to learn the useful operating parameters of a prospective jammer and then apply Q-learning [78] to learn and update any additional information needed to train the decision engine for maximum likelihood estimation. The decision engine is then used to derive the channel hopping sequence for the SUs so that they are highly unlikely to be exposed to the physical layer jamming attack. For multi-channel CRNs (each SU has simultaneous access to multiple channels, through multiple radios), the defense strategy is to randomly vary the transmission power to be used for the SU on these channels. The probability distribution of the allocated power on the channels is generated from a Colonel Blotto game [79]-based decision engine. For higher spectrum throughput and efficiency, it would be better to switch between the control channels and data channels, according to a stochastic pattern that will negate a successful jamming attack perpetrated through channel hopping [80]. On similar lines (i.e., from a game theoretic perspective), stochastic swarm intelligence-based optimization algorithms [81] and Markov decision process-based models [82] have been proposed to learn about the attackers' access patterns and detect jamming attacks simultaneously initiated from multiple sources. Though it is not easy for an attacker to locate the frequency channels that would be in use between a transmitter/receiver pair, collaborative jamming (one attacker per channel) is still possible. It would be more prudent for the regular CR users to coordinate among themselves (e.g., form tiers to exploit the temporal and spatial diversity to avoid jamming [83]) and thwart such collaborative jamming attacks. More active research needs to be conducted to develop collaborative defense strategies to thwart collaborative jamming attacks.

5.3 Primary User Emulation (PUE) Attacks

The primary user emulation (PUE) attack [84] is launched by a malicious or selfish secondary user emulating or masquerading as a primary user to obtain complete access to the spectrum bands of a given channel and not sharing it with other secondary users. While a selfish PUE attack could be intended to increase the attacker's share of the spectrum resources, the malicious PUE attack is typically targeted at preventing the legitimate secondary users from using the spectrum holes. More sophisticated PUE attacks could be performed if the attacker has knowledge about the CRN. For example [85], an attacker can transmit during the "quiet period" of a CRN and masquerade as a primary user to the rest of the nodes (secondary users). A quiet period is the time period during which all secondary users desist from transmitting to facilitate spectrum sensing.

PUE attacks can also be launched when the CRN makes a frequency handoff (i.e. switches from one channel to another), leading to degradation in the data throughput. For a while, the TCP (transmission control protocol) process at the transport layer of the sender side will be unaware of the frequency handoff at the physical layer and will keep creating logical connections/sending data packets without receiving acknowledgments. Perceiving this as an impending congestion, the TCP process backs off and doubles its retransmission timeout value, resulting in transient delays and packet losses. If an attacker is able to intercept the messages and predict the frequency bands used in a handoff, he can launch the PUE attack on both the old and new frequency bands, leading to total network starvation. Such a manifestation of the PUE attack disrupting TCP connections at the transport layer is called the Lion attack [86].

Several solutions have been proposed to defend against PUE attacks. As suggested in [87], one could employ the traditional option of a public-key infrastructure (PKI) for user authentication and require PUs to digitally sign their messages using public-key certificates. However, such a solution would require all PUs to own a public-key certificate and mandatorily use them to digitally sign their messages. Besides, it would require the SUs to be configured with the certificates of the PUs in the network and/or use a centralized certificate authority to authenticate the transmissions. An important criterion for these solutions is that they should not require any change in the operating mode or characteristics of the primary users. In [84], the authors suggest the use of a Distance Ratio Test (DRT) or a Distance Difference Test (DDT) to determine the location of a transmitting source and crosscheck the transmitter location with trusted location verifiers (LVs) that have a database of the coordinates of the primary users (e.g., TV broadcast towers) obtained through secure GPS systems. The LVs need to be tightly synchronized with each other and exchange information in encrypted form. Still, attackers could subvert the LV-DRT/DDT detection

mechanism by transmitting signals from the vicinity of a primary user. A solution [88] to detect such PUE attacks is to measure the energy level of the received signals and compare them with that expected from an authentic primary user. The assumption behind this strategy is that primary users (such as TV towers) have a fixed location and the energy level of their received signal would be much stronger than that of the signals received from secondary users who are also often mobile.

As an extension of the above idea of using the received signal strength, in [89], the authors use the notion of belief propagation, according to which the individual CR nodes mutually exchange their individual beliefs/perception of the received signal strength and then collectively decide whether or not the suspect is a malicious secondary user or a genuine primary user. The classification function used could be as simple as computing the average of the belief values received from each CR user and then checking whether it is below a threshold value (to classify the suspect as a PUE attacker). The CR nodes then mutually exchange their final decisions. The idea of using belief propagation is distributed in nature and does not require any additional hardware or sensor nodes to be part of the CR nodes. Belief propagation is also vouched for defense against the routing-towards primary-user (RPU) attack, wherein the malicious nodes route large amount of packets towards or around the primary users to cause interference around the neighborhood of the latter. The impending delay incurred in data transmission along the CR routes can be excessively high, especially if multiple honest CR nodes unknowingly end up forwarding the RPU packets. Belief propagation has been vouched for as a successful defense strategy against RPU attack: To start with, the source establishes a route to the destination by going through the regular route establishment process. Each intermediate node exchanges their knowledge about the belief values of every other node in the network. Once the source node compiles the belief values from all the nodes on the source-destination path, it decides whether or not to use the path. Potential malicious nodes are decided based on their belief values. Simulation results in [89, 90 and 91] have shown that the belief propagation mechanism to identify a PUE/RPU attacker can converge quickly. However, a major problem with the belief propagation mechanism to defend against PUE attacks is that if the PUE attacker colludes with one or more CR nodes in the neighborhood, then the calculations of the average beliefs could go awry and the benign CR nodes would not be able to effectively detect the presence of the PUE attacker. To defend against the RPU attack, each node is estimated to maintain accurate record of the perceived belief values for every other node in the network, which will incur too much of communication overhead and bookkeeping overload. Any error (could be also perpetrated by a colluding CR node) introduced in the belief values propagated by a node can poison the belief values for every other node in the network.

Apart from the above, another category of solutions (e.g., [92][93]) based on fingerprinting have been proposed for CRNs. These solutions are based on the idea of extracting unique distinctive patterns in the initial waveforms emitted by a transceiver and use these as an authentic means of identifying the transmitting source. For example, in [92], the authors propose to use the variance of the received signal power at the SU nodes to estimate the "noise power" channel parameter (σ^2, an invariant representing the channel transmission characteristics), used as the fingerprint to classify the suspected transmissions as those from a genuine primary user or from an attacker. The hypothesis is that it may be possible for an attacker to emulate the primary user by making the raw values of the received signal strength (and even the mean values within a time period) to fall within the threshold limits to classify suspected PUE transmissions; however, it will be extremely difficult to emulate the PU channel transmissions such that the variance of the received signal strength falls within a threshold. Though relatively more credible, the fingerprinting-based solutions have been observed to require large samples of training data as well as more storage and significant computation plus signal processing overhead. It has been observed [93] that a coordinated extraction/analysis of the channel parameters coupled with cooperative/joint decision making can increase the detection accuracy as well as reduce the overhead at the individual nodes.

5.4 Objective Function Attack

The cognitive engine of a cognitive radio is responsible for adjusting the radio parameters (such as center frequency, bandwidth, power, modulation type, coding rate, channel access protocol, etc) to meet specific requirements (such as low energy consumption, high data rate, high security and etc). An attacker could launch an attack to manipulate the values of the parameters that he has on control to tailor the results of the objective function to suit his interests. For example, consider a scenario (presented in [94]) of a cognitive engine attempting to maximize an objective function f composed of transmission rate (R) and security (S) given by: $f = w_1R + w_2S$, where w_1 and w_2 represent the weights for parameters R and S. If the attacker gets to know that the cognitive engine is attempting to maximize f by increasing S, he may launch a jamming attack on the radio and reduce R, so that the overall value of f could get lower. To prevent the value of f from getting lowered, the cognitive engine may choose to operate at a lower value for the security level. Though no concrete solutions have been proposed for the Objective function attacks, an idea proposed in [85] is to impose a threshold on the values for every updatabale radio parameter and stop the communication if the values of the parameters fall outside the thresholds.

5.5 Smaller Backoff Window Attack

The Smaller Backoff Window (SBW) attack [95] is launched at the MAC layer of cognitive radio networks that employ the IEEE 802.11 DCF (Distributed Coordination Function) as the channel coordination mechanism. Typically, a node backs off for a random time within a maximum duration window (determined based on the number of times the node has backed off so far and the maximum backoff duration when a node backs off for the first time) to be able to successfully gain access to the channel. The backoff time should typically increase with every failure to gain access to the channel in order to alleviate channel contention in the neighborhood and

provide fair chances for every node to gain access to the channel. However, a malicious node may not increase its backoff timer and may indeed operate with a smaller backoff window for successive (re)transmission attempts. This is a denial of service attack on the MAC layer. A cumulative distribution function (CDF)-based solution is available in the literature (first proposed in [96] and later enhanced in [97]) to detect and quarantine nodes launching SBW attacks. The idea is to have a node observe the transmission attempts of its neighbor node and compare the observed CDF of the backoff time window with that of a theoretical hypothetical CDF, expected of a non-malicious node in a certain neighborhood. If there is incidence of a series of mismatches between the observed CDF and the expected hypothetical CDF within a certain time period, then the neighbor node is flagged malicious and an alarm can be raised to notify the intrusion detection system (IDS) for the network. As a long as the IDS node is not compromised, there are good chances that the above scheme will work and incur fewer false negatives. However, the scheme is vulnerable for many false positives as its success depends on the perceived neighborhood density and knowledge about the transmission requirements/attempts of every node in the neighborhood to generate an accurate hypothetical CDF that can be expected for a neighbor node.

5.6 Spectrum Sensing Data Falsification (SSDF) Attack

The SSDF attack (a.k.a. Byzantine attack) [98, 99] happens when an attacker sends false local spectrum sensing results to its neighbors (for a distributed CRN) or a fusion center (for a centralized CRN) to make them take a wrong spectrum sensing decision. A Byzantine attack on distributed CRNs is hard to control because the false information can propagate quickly; whereas, in a centralized CRN, the fusion center (that collects all the sensed data and makes a decision on which frequency bands are occupied and which are free) can lessen the impact of false information by comparing the data received from all the CR nodes.

One category of data fusion techniques proposed to detect the Byzantine attack are based on the idea [100] of summing up the number of sensing terminals reporting "busy" and if the sum is greater than a fixed threshold, then the channel is considered to have been occupied. Distributed approaches (e.g., [101]) to arrive at a consensus among the sensing terminals have also been developed. While a threshold value of 1 (one attacker is sufficient to mislead the neighborhood) may trigger several false alarms; a larger value for the threshold could lead to detection misses (i.e., the presence of a primary user may not be detected) and could be still prone to Byzantine attack if multiple attackers collude.

Another category of data fusion techniques proposed to detect Byzantine attacks is based on the notion of trust factor/indicator, typically built up based on the past behaviors. The trust value for a node increases slowly with time due to good behavior but decreases quickly due to bad behavior [21]. In [22], the authors proposed a data fusion technique called the Weighted Sequential Ratio Test (WSRT) that takes into consideration both the actual status reported by a CR node as well as its reputation value (initialized to 0 and incremented by 1 for each correct local spectrum report). A similar trust-based scheme was proposed in [70] that (in addition to the regular nodes) also assigns a trust factor for permanently malicious nodes – "Always Yes" and "Always No" nodes – such that the reports from these malicious nodes help to identify the malicious nodes that are only sometimes faulty.

5.7 Cross-Layer Attacks and Solutions

In [97], the authors design cross-layer attack strategies (i.e., simultaneously launching attacks at more than one layer of the TCP/IP protocol stack) and propose appropriate defensive solutions to combat these attacks. A cross-layer attack is defined as a collection of attack activities conducted coordinately on multiple layers of the TCP/IP protocol stack to achieve specific attack goals. It is being argued (in [97]) that with a cross-layer attack, the attacker can increase the damage at a lower risk of being detected, relative to launching an attack at a single layer. For example, one can effectively reduce channel utilization by simultaneously launching physical layer attack (PUE attack, SSDF attack and etc) in coordination with a MAC layer attack (CCC Denial of Service attack, Small Backoff Window attack and etc) rather than launching the attack on just one layer. Nevertheless, there are some attacks that need to be launched across more than one layer in order to fructify the attack. For example, to cause interference at the PUs, the benign SU nodes should fail to detect the presence of a PU (this can be done through an SSDF attack at the physical layer) and the routing protocol should be attacked at the network layer to facilitate the malicious nodes to route the packets towards the benign SU nodes who are close to the PU.

Binary trust-based defensive solution has been proposed to mitigate the cross-layer attacks. The idea is to concurrently run the defense modules for each layer and classify the attack at a layer with a Yes/No (0/1) decision. The per-layer binary results reported from each node are gathered at a monitoring node (deployed for intrusion detection) and the results are weighted to calculate a cross-layer overall trust value for the node. A cumulative trust value for the entire neighborhood is then determined based on the individual cross-layer overall trust values. Nodes that consecutively report results that are abnormal and different from majority of the neighborhood are flagged as malicious. Since the trust value for a node is calculated based on the multi-layer response, the number of false positives resulting from the cross-layer defensive approach is likely to be lower than the single layer defensive strategies.

To mitigate the chances a TCP session from being intercepted and subjected to a PUE attack/Lion attack during frequency handoff, the authors in [86] suggest cross-layer data sharing between the physical and transport layers. This would facilitate the TCP session to freeze the connection parameters until the frequency handoff is completed and adapt them to the new network. In addition, a group key management mechanism could facilitate the CRN members to encrypt, decrypt and authenticate each other and prevent an attacker from intercepting the TCP session/frequency handoff to infer the control parameters.

6. Conclusions

In this paper, we have presented an exhaustive review and analysis of a host of issues and mechanisms that have been proposed in the literature for cognitive radio networks, with regards to the medium access control protocols (time slotted and random access protocols for both infrastructure-based and infrastructure less CRNs), routing protocols (protocol solutions based on full spectrum knowledge and local spectrum knowledge) and transport layer protocols (issues for effective design of new protocol solutions, and existing solutions based on cross-layered and layer-preserving approaches). In the later part of paper, we have analyzed in detail the various security attacks (control channel saturation attack, primary user emulation attack, small backoff window attack, jamming attack, objective function attack and spectrum sensing data falsification attack) and solutions that could be deployed to counter these attacks. The security attacks analyzed in this paper are characteristic of CRNs. In addition to these attacks, a CRN could be subjected to security attacks (for example, routing re-direction based sink-hole and Hello flood attacks on multi-hop topologies) that are characteristic of wireless networks in general.

From a design point of view, a common thread that should be prevalent in any proposed mechanism for CRNs is that the solution should not require the primary user to be capable of adapting its transmission parameters due to the presence of the secondary CR user. In fact, a licensed user need not be even aware of the presence of the unlicensed CR users, and there should be no appreciable degradation in the quality of service for the primary users. While the solutions proposed for centralized and/or infrastructure-based CRNs are typically construed to provide performance benchmarks for the appropriate paradigm, the solutions proposed for distributed/cooperative and/or infrastructure less ad hoc CRNs capture the practical difficulties and performance bottlenecks in real-time implementations. Most of the active research conducted in the area of CRNs has been so far focused on spectrum sensing, allocation and sharing, and medium control access. Recently, the research community has also started looking at development of end-to-end solutions, starting from the routing protocols and transport layer protocols, which are needed to fully realize the potential of cognitive radios from an application standpoint. Of course, security of the underlying CRN and the end users would also need to be a key ingredient/design criterion for any proposed solution.

References

[1] FCC, ET Docket No 03-222, Notice of Proposed Rule Making and Order, December 2003.

[2] FCC, ET Docket No 02-135, Spectrum Policy Task Force Report, November 2002.

[3] S. Haykin, “Cognitive radio: Brain-empowered wireless communications,” IEEE Journal of Selected Areas in Communications, vol. 23, no. 2, pp. 201–220, February 2005.

[4] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, “Next Generation/ Dynamic Spectrum Access/ Cognitive Radio Wireless Networks: A Survey,” Computer Networks, vol. 50, pp. 2127–2159, May 2006.

[5] Z. Han and K. J. R. Liu, Resource Allocation for Wireless Networks: Basics, Techniques and Applications, Cambridge University Press, Cambridge, UK, 2008.

[6] Q. Zhao and B. Sadler, “A Survey of Dynamic Spectrum Access,” IEEE Signal Processing Magazine, vol. 24, no. 3, pp. 79-89, May 2007.

[7] S-Y. Lien, C-C. Tseng, and K-C. Chen, “Carrier Sensing based Multiple Access Protocols for Cognitive Radio Networks,” Proceedings of IEEE International Conference on Communications, pp. 3208-3214, May 2008.

[8] C. Cordeiro, K. Challapali, and M. Ghosh, “Cognitive PHY and MAC Layers for Dynamic Spectrum Access and Sharing of TV Bands,” Proceedings of IEEE International Workshop on Technology and Policy for Accessing Spectrum, p. 222, August 2006.

[9] P. Pawelczak, R. Venkatesha Prasad, L. Xia, and I. G. M. M. Niemegeers, “Cognitive Radio Emergency Networks – Requirements and Design,” Proceedings of the IEEE Dynamic Spectrum Access Networks, pp. 601-606, November 2005.

[10] L. Ma, C-C. Shen and B. Ryu, “Single-Radio Adaptive Channel Algorithm for Spectrum Agile Wireless Ad hoc Networks,” Proceedings of the IEEE Dynamic Spectrum Access Networks, pp. 547-558, April 2007.

[11] H. Su, “CREAM-MAC: An Efficient Cognitive Radio-enabled Multi-Channel MAC Protocol for Wireless Networks,” Proceedings of the International Symposium on World of Wireless, Mobile and Multimedia Networks, June 2008.

[12] Y. R. Kondareddy and P. Agrawal, “Synchronized MAC Protocol for Multi-hop Cognitive Radio Networks,” Proceedings of the IEEE International Conference on Communications, pp. 3198-3202, 2008.

[13] S. Y. Hung, E. H. K. Wu and G. H. Chen, “An Opportunistic Cognitive MAC Protocol for Coexistence with WLAN,” Proceedings of the IEEE International Conference on Communication, pp. 4059-4063, 2008.

[14] M. A. Shah, S. Zhang and C. Maple, “An Analysis on Decentralized Adaptive MAC Protocols for Cognitive Radio Networks,” Proceedings of the 18th International Conference on Automation and Computing, 2012.

[15] C. Cordeiro and K. Challapali, “C-MAC: A Cognitive MAC Protocol for Multichannel Wireless Networks,” Proceedings of the IEEE Dynamic Spectrum and Access Networks, pp. 147-157, April 2007.

[16] J. Zhao, H. Zheng, and G-H. Yang, “Spectrum Sharing through Distributed Coordination in Dynamic Spectrum Access Networks,” Wireless Communications and Mobile Computing, vol. 7, no. 9, pp. 1061-1075, 2007.

[17] Z. Htike, J. Lee and C-S. Hong, “A MAC Protocol for Cognitive Radio Networks with Reliable Control

Channels Assignment," Proceedings of the International Conference on Information Networking, pp. 81-85, 2012.

[18] L. DaSilva and I. Guerreiro, "Sequence-based Rendezvous for Dynamic Spectrum Access," Proceedings of the IEEE Conference on Dynamic Spectrum Access Networks, 2008.

[19] K. Bian, J-M. Park and R. Chen, "A Quorum-based Framework for Establishing Control Channels in Dynamic Spectrum Access Networks," Proceedings of the ACM International Conference on Mobile Computing (MobiCom), pp. 25-36, 2009.

[20] Y-H. Lee and D. Kim, "A Slow Hopping MAC Protocol for Coordinator-based Cognitive Radio Network," Proceedings of the IEEE Consumer Communications and Networking Conference, pp. 854-858, 2012.

[21] W. Wang, H. Li, Y. Sun and Z. Han, "Attack-Proof Collaborative Spectrum Sensing in Cognitive Radio Networks," Proceedings of the 43rd Annual Conference on Information Sciences and Systems, pp. 130-134, Baltimore, MD, USA, March 2009.

[22] R. Chen, J-M. Park, Y. T. Hou and J. H. Reed, "Toward Secure Distributed Spectrum Sensing in Cognitive Radio Networks," IEEE Communications Magazine, vol. 46, no. 4, pp. 50-55, 2008.

[23] H. Khalife, N. Malouch, and S. Fdida, "Multihop Cognitive Radio Networks: To Route or not to Route," IEEE Network Magazine, vol. 23, no. 4, pp. 20-25, 2009.

[24] A. C. Talay and D. T. Altilar, "ROPCORN: Routing Protocol for Cognitive Radio Ad hoc Networks," Proceedings of the International Conference on Ultra Modern Telecommunications & Workshops, pp. 1-6, October 2009.

[25] C. Xin, B. Xie, and C-C. Shen, "A Novel Layered Graph Model for Topology Formation and Routing in Dynamic Spectrum Access Networks," Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 308-317, 2005.

[26] C. Xin, L. Ma, and C-C. Shen, "A Path-centric Channel Assignment Framework for Cognitive Radio Wireless Networks," Mobile Networks and Applications, vol. 13, no. 5m pp. 463-476, 2008.

[27] X. Zhou, L. Lin, J. Wang and X. Zhang, "Cross-layer Routing Design in Cognitive Radio Networks by Colored Multi graph Model," Wireless Personal Communications, vol. 49, no. 1, pp. 123-131, 2009.

[28] Q. Wang and H. Zheng, "Route and Spectrum Selection in Dynamic Spectrum Networks," Proceedings of the 3rd IEEE Consumer Communications and Networking Conference, vol. 1, pp. 625-629, 2006.

[29] B. F. Lo, "A Survey of Common Control Channel Design in Cognitive Radio Networks," Physical Communication, vol. 4, pp. 26-39, 2011.

[30] C. Perkins, E. Royer and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," IETF RFC 3561, July 2003.

[31] J-P. Sheu and I-L. Lao, "Cooperative Routing Protocol in Cognitive Radio Ad hoc Networks," Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), pp. 2916-2921, April 2012.

[32] H. Liu, Z. Ren and C. An, "Backup-Route-Based Routing for Cognitive Ad hoc Networks," Proceedings of the IEEE International Conference on Oxide Materials for Electronic Engineering, pp. 618-622, September 2012.

[33] M. Zeeshan, M. F. Manzoor and J. Qadir, "Backup Channel and Cooperative Channel Switching On-demand Routing Protocol for Multi-hop Cognitive Radio Ad hoc Networks (BCCCS)," Proceedings of the 6th International Conference on Emerging Technologies, pp. 394-399, October 2010.

[34] C. W. Pyo and M. Hasegawa, "Minimum Weight Routing based on a Common Link Control Radio for Cognitive Wireless Ad hoc Networks," Proceedings of the International Conference on Wireless Communications and Mobile Computing, pp. 399-404, 2007.

[35] S. M. Kamruzzaman, E. Kim and D. G. Jeong, "An Energy Efficient QoS Routing Protocol for Cognitive Radio Ad hoc Networks," Proceedings of the 13th International Conference on Advanced Communication Technology, pp. 344-349, February 2011.

[36] D. Johnson, D. Maltz and Y. Hu, "The Dynamic Source Routing Protocol for Mobile Ad hoc Networks," IETF RFC 4728, February 2007.

[37] S. M. Kamruzzaman, E. Kim, D. G. Jeong and W. S. Jeon, "Energy-aware Routing Protocol for Cognitive Radio Ad hoc Networks," IET Communications, vol. 6, no. 14, pp. 2159-2168, September 2012.

[38] H. Ma, L. Zheng, X. Ma and Y. Iuo, "Spectrum Aware Routing for Multi-hop Cognitive Radio Networks with a Single Transceiver," Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, pp. 1-6, 2008.

[39] G. Cheng, W. Liu, Y. Li and W. Cheng, "Spectrum Aware On-demand Routing in Cognitive Radio Networks," Proceedings of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 571-574, 2007.

[40] G. Cheng, W. Liu, Y. Li and W. Cheng, "Joint On-demand Routing and Spectrum Assignment in Cognitive Radio Networks," Proceedings of the IEEE International Conference on Communications, pp. 6499-6503, 2007.

[41] Z. Yang, G. Cheng, W. Liu, W. Yuan and W. Cheng, "Local Coordination based Routing and Spectrum Assignment in Multi-hop Cognitive Radio Networks," Mobile Networks and Applications, vol. 13, no. 1-2, pp. 67-81, 2008.

[42] Y. Liu and D. Grace, "Improving Capacity for Wireless Ad hoc Communications using Cognitive Routing," Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, May 2008.

[43] I. Pefkianakis, S. Wong and S. Lu, "SAMER: Spectrum Aware Mesh Routing in Cognitive Radio Networks," Proceedings of the 3rd IEEE Symposium

on New Frontiers in Dynamic Spectrum Access Networks, pp. 1-5, 2008.

[44] L. Ding, T. Melodia, S. Batalama and M. J. Medley, "ROSA: Distributed Joint Routing and Dynamic Spectrum Allocation in Cognitive Radio Ad hoc Networks," Proceedings of the 12th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, pp. 13-20, 2009.

[45] L. Ding, T. Melodia, S. N. Batalama, J. D. Matyjas, and M. J. Medley, "Cross-Layer Routing and Dynamic Spectrum Allocation in Cognitive Radio Ad hoc Networks," IEEE Transactions on Vehicular Technology, vol. 59, no. 4, pp. 1969-1979, May 2010.

[46] Y. Shi and Y. Hou, "A Distributed Optimization Algorithm for Multi-hop Cognitive Radio Networks," Proceedings of the 27th IEEE Conference on Computer Communications, pp. 1292-1300, 2008.

[47] Y. Shi, Y. T. Hou, H. Zhou and S. F. Midkiff, "Distributed Cross-layer Optimization for Cognitive Radio Networks," IEEE Transactions on Vehicular Technology, pp. 4058-4069, October 2010.

[48] A. Chehata, W. Ajib and H. Elbiaze, "An On-demand Routing Protocol for Multi-hop Multi-radio Multi-channel Cognitive Radio Networks," Proceedings of the IFIP Wireless Days Conference, October 2011.

[49] Y. Liu, L. X. Cai and X. Shen, "Spectrum-aware Opportunistic Routing in Multi-hop Cognitive Radio Networks," IEEE Journal on Selected Areas in Communications, vol. 30, no. 10, pp. 1958-1968, November 2012.

[50] M. Xie, W. Zhang and K-K. Wong, "A Geometric Approach to Improve Spectrum Efficiency for Cognitive Relay Networks," IEEE Transactions on Wireless Communications, vol. 9, no. 1, pp. 268-281, 2010.

[51] J. Kim and M. Krunz, "Spectrum-Aware Beaconless Geographical Routing Protocol for mobile Cognitive Radio Networks," Proceedings of the IEEE Global Telecommunications Conference, December 2011.

[52] S. Ruhrup, H. Kalosha, A. Nayak and I. Stojmenovic, "Message-Efficient Beaconless Georouting with Guaranteed Delivery in Wireless Sensor, Ad hoc and Actuator Networks," IEEE/ACM Transactions on Networking, vol. 18, no. 1, pp. 95-108, February 2010.

[53] K. C. How, M. Ma and Y. Qin, "An Opportunistic Service Differentiation Routing Protocol for Cognitive Radio Networks," Proceedings of the IEEE Global Telecommunications Conference, December 2010.

[54] K. R. Chowdhury and I. F. Akyildiz, "CRP: A Routing Protocol for Cognitive Radio Ad hoc Networks," IEEE Journal on Selected Areas in Communications, vol. 29, no. 4, pp. 794-804, April 2011.

[55] M. Di Felice, K. R. Chowdhury and L. Bononi, "Modeling and Performance Evaluation of Transmission Control Protocol over Cognitive Radio Ad hoc Networks," Proceedings of the 12th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, pp. 4-12, 2009.

[56] D. Sarkar and H. Narayan, "Transport Layer Protocols for Cognitive Networks," Proceedings of the IEEE INFOCOM Workshops, March 2010.

[57] K. R. Chowdhury, M. Di Felice and I. F. Akyildiz, "TP-CRAHN: A Transport Protocol for Cognitive Radio Ad-hoc Networks," Proceedings of the IEEE INFOCOM Conference, pp. 2482-2490, 2009.

[58] A. M. R. Slingerland, P. Pawelczak, R. V. Prasad, A. Lo and R. Hekmat, "Performance of Transport Control Protocol over Dynamic Spectrum Access Links," Proceedings of the IEEE International Conference on Dynamic Spectrum Access Networks, pp. 486-495, April 2007.

[59] W. Y. Lee and I. F. Akyildiz, "Optimal Spectrum Sensing Framework for Cognitive Radio Networks," IEEE Transactions on Wireless Communications, vol. 7, no. 10, pp. 3845-3857, Ocotber 2008.

[60] D. Chen, H. Ji and V. C. M. Leung, "Distributed Best-Relay Selection for Improving TCP Performance over Cognitive Radio Networks: A Cross-Layer Design Approach," IEEE Journal on Selected Areas in Communications, vol. 30, no. 2, pp. 315-322, February 2012.

[61] D. Chen, H. Ji and V. C. M. Leung, "Distributed Optimal Relay Selection for Improving TCP Throughput over Cognitive Radio Networks: A Cross-Layer Design Approach," Proceedings of the IEEE International Conference on Communications (ICC), June 2011.

[62] G. Li, Z. Hu, G. Zhang, L. Zhao, W. Li and H. Tian, "Cross-Layer Design for Energy-Efficiency of TCP Traffic in Cognitive Radio Networks," Proceedings of the IEEE Vehicular Technology Conference (VTC Fall), September 2011.

[63] A. Al-Dulaimi and J. Cosmas, "Mobility Management for Microcells based Cognitive Radio over Fiber Networks," Proceedings of the World Wireless Research Forum (WWRF#25) Meeting, London, November 2010.

[64] A. Al-Dulaimi, S. Al-Rubaye and J. Cosmas, "Adaptive Congestion Control for Mobility in Cognitive Radio Networks," Proceedings of the Wireless Advanced Conference, pp. 273-277, June 2011.

[65] C. Luo, F. R. Yu, H. Ji and V. C. M. Leung, "Cross-Layer Design for TCP Performance Improvement in Cognitive Radio Networks," IEEE Transactions on Vehicular Technology, vol. 59, no. 5, pp. 2485-2495, June 2010.

[66] C. Luo, F. R. Yu, H. Ji and V. C. M. Leung, "Optimal Channel Access for TCP Performance Improvement in Cognitive Radio Networks: A Cross-Layer Design Approach," Proceedings of the IEEE Global Telecommunications Conference, 2009.

[67] D. Berstimas and J. Nino-Mora, "Restless Bandits, Linear Programming Relaxations, and a Primal Dual Index Heuristic," Operations Research, vol. 48, no. 1, pp. 80-90, 2000.

[68] A. Chan, X. Liu, G. Noubir and B. Thapa, "Broadcast Control Channel Jamming: Resilience

and Identification of Traitors," Proceedings of the IEEE International Symposium on Information Theory, pp. 2496-2500, 2007.

[69] P. Tague, M. Li and R. Poovendran, "Probabilistic Mitigation of Control Channel Jamming via Random Key Distribution," Proceedings of the 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1-5, 2007.

[70] P. Kaligineedi, M. Khabbazian and V. K. Bhargava, "Secure Cooperative Sensing Techniques for Cognitive Radio Systems," IEEE International Conference on Communications, pp. 3406-3410, Beijing, China, May 2008.

[71] L. Lazos, S. Liu and M. Krunz, "Mitigating Control-Channel Jamming Attacks in Multi-Channel Ad hoc Networks," Proceedings of the 2nd ACM Conference on Wireless Network Security, pp. 169-180, 2009.

[72] P. Tague, M. Li and R. Poovendran, "Mitigation of Control Channel Jamming under Node Capture Attacks," IEEE Transactions on Mobile Computing, vol. 8, no. 9, pp. 1221-1234, 2009.

[73] W. Xu, W. Trappe, Y. Zhang and T. Wood, "The Feasibility of Launching and Detecting Jamming Attacks in Wireless Networks," Proceedings of the ACM International Symposium on Mobile Ad hoc Networking and Computing, pp. 46-57, Urbana, IL, USA, May 2005.

[74] A. Sampath, H. Dai, H. Zheng and B. Y. Zhao, "Multi-channel Jamming Attacks using Cognitive Radios," Proceedings of the 16th International Conference on Computer Communications and Networks (ICCCN), pp. 352-357, Honolulu, HI, USA, August 2007.

[75] W. Xu, T. Wood, W. Trappe and Y. Zhang, "Channel Surfing and Spatial Retreats: Defenses against Wireless Denial of Service," Proceedings of the 3rd ACM Workshop on Wireless Security, pp. 80-89, Philadelphia, PA, USA, January 2004.

[76] Y. Wu, B. Wang, K. J. R. Liu and T. C. Clancy, "Anti-Jamming Games in Multi-Channel Cognitive Radio Networks," IEEE Journal on Selected Areas in Communications, vol. 30, no. 1, pp. 4-15, January 2012.

[77] M. L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, 1994.

[78] Y. Shoham and K. Leyton-Brown, Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations, Cambridge University Press, 2008.

[79] B. Roberson, "The Colonel Blotto Game," Economic Theory, vol. 29, no. 1, pp. 1-24, September 2006.

[80] B. Wang, Y. Wu, K. J. R. Liu and T. C. Clancy, "An Anti-Jamming Stochastic Game for Cognitive Radio Networks," IEEE Journal on Selected Areas in Communications, vol. 29, no. 4, pp. 877-889, April 2011.

[81] H. Zhang, Z. Liu and Q. Hui, "Optimal Defense Synthesis for Jamming Attacks in Cognitive Radio Networks via Swarm Optimization," Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications, July 2012.

[82] Y. Wu, B. Wang and K. J. R. Liu, "Optimal Defense against Jamming Attacks in Cognitive Radio Networks using the Markov Decision Process Approach," Proceedings of the IEEE Global Telecommunications Conference, December 2010.

[83] W. Wang, M. Chatterjee and K. Kwiat, "Collaborative Jamming and Collaborative Defense in Cognitive Radio Networks," Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), June 2011.

[84] R. Chen and J-M. Park, "Ensuring Trustworthy Spectrum Sensing in Cognitive Radio Networks," Proceedings of the 1st IEEE Workshop on Networking Technologies for Software Defined Radio Networks, pp. 110-119, Reston, VA, September 2006.

[85] O. Leon, J. Hernandez-Serrano and M. Soriano, "Securing Cognitive Radio Networks," International Journal of Communication Systems, vol. 23, no. 5, pp. 633-652, 2010.

[86] J. Hernandez-Serrano, O. Leon and M. Soriano, "Modeling the Lion Attack in Cognitive Radio Networks," EURASIP Journal on Wireless Communications and Networking, vol. 2011, article id: 242304, 10 pages, 2011.

[87] C. Zou and C. Chigan, "Licensed Receiver Detection and Authentication in Simplex Licensed Networks," Proceedings of the IEEE International Workshop on Recent Advances in Cognitive Communications and Networking, pp. 924-929, December 2011.

[88] R. Chen, J-M. Park and J. H. Reed, "Defense against Primary User Emulation Attacks in Cognitive Radio Networks," IEEE Journal on Selected Areas in Communications, vol. 26, no. 1, pp. 25-37, 2008.

[89] Z. Yuan, D. Niyato, H. Li, J. B. Song and Z. Han, "Defeating Primary User Emulation Attacks using Belief Propagation in Cognitive Radio Networks," IEEE Journal on Selected Areas in Communications, vol. 30, no. 10, pp. 1850-1860, November 2012.

[90] Z. Yuan, D. Niyato, H. Li and Z. Han, "Defense against Primary User Emulation Attacks using Belief Propagation of Location Information in Cognitive Radio Networks," Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), pp. 599-604, March 2011.

[91] Z. Yuan, Z. Han, Y. Sun, H. Li and J. Song, "Routing-Toward-Primary-User Attack and Belief Propagation Based Defense in Cognitive Radio Networks," IEEE Transactions on Mobile Computing, DOI: 10.1109/TMC.2012.137, June 2012.

[92] Z. Chen, T. Cooklev, C. Chen and C. Pomalaza-Raez, "Modeling Primary User Emulation Attacks and Defenses in Cognitive Radio Networks," Proceedings of the 28th IEEE International Performance Computing and Communications Conference, pp. 208-215, December 2009.

[93] T. Yang, H. Chen and L. Xie, "Cooperative Primary User Emulation Attack and defense in Cognitive Radio Networks," Proceedings of the 7th International Conference on Wireless Communications, Networking and Mobile Computing, September 2011.

[94] T. C. Clancy and N. Goergen, "Security in Cognitive Radio Networks: Threats and Mitigation,"

Proceedings of the International Conference on Cognitive Radio Oriented Wireless Networks and Communications, pp. 1-8, Singapore, May 2008.

[95] K. Bian and J. M. Park, "MAC-layer Misbehaviors in Multi-hop Cognitive Radio Networks," Proceedings of the 2006 US-Korea Conference on Science, Technology and Entrepreneurship, August 2006.

[96] A. L. Toledo and X. Wang, "Robust Detection of Selfish Misbehavior in Wireless Networks," IEEE Journal of Selected Areas in Communications, vol. 25, no. 6, pp. 1124-1134, August 2007.

[97] W. Wang, Y. Sun, H. Li and Z. Han, "Cross-Layer Attack and Defense in Cognitive Radio Networks," Proceedings of the IEEE Global Telecommunications Conference, December 2010.

[98] C. Karlof and D. Wagner, "Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures," Proceedings of the 1st IEEE International Workshop on Sensor Network Protocols and Applications, pp. 113-127, Berkeley, CA, USA, May 2003.

[99] C. Mathur and K. Subbalakshmi, "Security Issues in Cognitive Radio Networks," Cognitive Networks: Towards Self-Aware Networks, pp. 284-293, Wiley, New York, 2007.

[100] A. Pandharipande et al., "IEEE P802.22 Wireless RANs: Technology Proposal Package for IEEE 802.22," IEEE 802.22 Working Group on WRANs, November 2005.

[101] F. R. Yu, H. Tang, M. Huang, Z. Li and P. C. Mason, "Defense against Spectrum Sensing Data Falsification Attacks in Mobile Ad hoc Networks with Cognitive Radios," Proceedings of the IEEE Military Communications Conference, October 2009.