Cooperative Spectrum Sharing Scheme for Enhancing Primary User Performance under Denial of Service Attack

Ahmed N. Elbattrawy[1,3] [a], Ahmed H. Abd El-Malek[2] [b], Sherif I. Rabia[1,4] [c] and
Waheed K. Zahra[1,5] [d]

[1]*Basic and Applied Sciences Institute, Egypt-Japan University of Science and Technology, Alexandria, Egypt*

[2]*School of Electronics, Communications and Computer Engineering,*
Egypt-Japan University of Science and Technology, Alexandria, Egypt

[3]*Department of Engineering Mathematics and Physics, Faculty of Engineering, Mansoura University, Dakahlia, Egypt*

[4]*Department of Engineering Mathematics and Physics, Faculty of Engineering, Alexandria University, Alexandria, Egypt*

[5]*Department of Engineering Physics and Mathematics, Faculty of Engineering, Tanta University, Tanta, Egypt*

Keywords: Cognitive Radio, Hybrid Channel Access, Cooperative Communication, DoS Attack, Jamming Attack, POMDP.

Abstract: In this paper, we investigate the performance of the primary user (PU) in a cognitive radio network under denial of service (DoS) attack. A cooperation scheme between the PU and an energy-constrained secondary user (SU) is proposed to mitigate the DoS attack effect on the PU, meanwhile, efficiently utilize the spectrum by allowing the SU to access the PU's spectrum in a hybrid underlay/overlay mode. Hence, a mutual benefit is achieved. To maintain its sustainability, the SU harvests energy from ambient sources. The location of the SU is optimized to maximize the PU performance. In addition, the PU sequential decision to cooperate with the SU or not is formulated as a multi-objective mixed-observable Markov decision process (MOMDP) to consider the performance of both the PU and the SU on the long run. Then, an optimal decision policy is obtained by solving the decision problem using a point-based value iteration (PBVI) algorithm with predetermined scalarization weights. The simulation results show the efficiency of the proposed scheme in enhancing the PU performance at different jamming levels.

1 INTRODUCTION

Cognitive radio (CR) technology emerged as a solution for the problem of spectrum scarcity and spectrum underutilization, by allowing unlicensed secondary users (SUs) to dynamically access the licensed spectrum of the primary users (PUs) considering no harmful interference affects the PUs (El Tanab and Hamouda, 2016). New wireless services are developed exploiting the CR technology, such as providing broadband access to rural areas using TV white space and disaster response networks (AlAqad et al., 2020).

Like other wireless networks, the wireless medium makes the CR networks (CRNs) vulnerable to security attacks. However, the situation is more challenging with CRNs, because in addition to the conventional security attacks like jamming and eavesdropping new types of attacks were introduced to CRNs such as primary user emulation attack (PUEA) and spectrum sensing data falsification (SSDF). Moreover, the security of both the PUs and the SUs must be ensured (Shu et al., 2013).

Jamming attacks are the most common and harmful attacks. Besides their destructive effects on communication links, they can be easily launched (Pirayesh and Zeng, 2021). In jamming attack the adversarial node floods the network with high-power interference reducing the legitimate nodes opportunity to access the network service, therefore, they are a type of denial of service (DoS) attacks. In CRNs, jamming may have two principal goals: avoiding all the communications of PUs and SUs, or preventing only SUs from accessing the free spectrum bands (Di Pietro and Oligeri, 2013).

Many research efforts were performed to defeat

[a] https://orcid.org/0000-0002-2730-4384
[b] https://orcid.org/0000-0002-7906-815X
[c] https://orcid.org/0000-0003-1471-8841
[d] https://orcid.org/0000-0002-6448-6877

the jamming attack. Classical spread spectrum (SS) modulation techniques have been used in several real-world wireless systems such as 3G cellular, ZigBee, and 802.11b (Pirayesh and Zeng, 2021). In direct sequence spread spectrum (DSSS) the original signal is spread across wider frequency band using a spreading code. While in frequency hopping spread spectrum (FHSS) the signal hops between frequencies within a fixed bandwidth and a pseudo-random code is used to avoid the disclosure of the hopping sequence, which requires synchronization between the transmitter and receiver nodes (Hasan et al., 2016). The shortcomings of these techniques are the need for wider bandwidth, in addition, they will be inefficient if the jammer can interfere multiple channels at once. Another approach is spatial retreat: when the jamming signal is detected, the user moves from the jammed location to another where no jamming exists. The shortcoming of spatial retreat is that the user may lose the current connection, and it is suitable only for portable devices (Shu et al., 2013).

Another investigated direction is to exploit cooperative communication for enhancing system security. Originally cooperative communication was used to increase system reliability (Liang et al., 2017; Su et al., 2012). But authors in (Thanh et al., 2018) considered enhancing SUs performance under jamming attack by cooperating with energy-harvesting relays for data transmission, and they proposed relay selection schemes to achieve that. Enhancing PU performance under jamming attack was rarely investigated, but rather under eavesdropping attack such as in (Zhang et al., 2014) and (Qin et al., 2018) where the SUs can act as a relay or a friendly jammer to improve the PU's secrecy. In this paper we will investigate the efficiency of cooperation between PUs and SUs in enhancing PUs performance under jamming attack. Therefore, we consider an underlay cognitive radio network (Liang et al., 2017), where a single PU and a single SU can transmit on the shared spectrum at the same time, provided that the SU interference upper bound is predefined by the PU to maintain its quality of service. The PU is suffering from random jamming attack (Pirayesh and Zeng, 2021) that cannot be perfectly detected due to the realistic sensing errors. The SU is energy-constrained, which is a common situation, so it depends on energy harvesting from ambient resources (solar, wind, temperature, etc.) for battery recharging without manual intervention (Bhowmick et al., 2015). Our contribution in this paper can be summarized as follows:

- Proposing a PU-SU cooperation scheme to mitigate the jamming effect on the PU. The scheme starts by optimizing the SU location to maximize
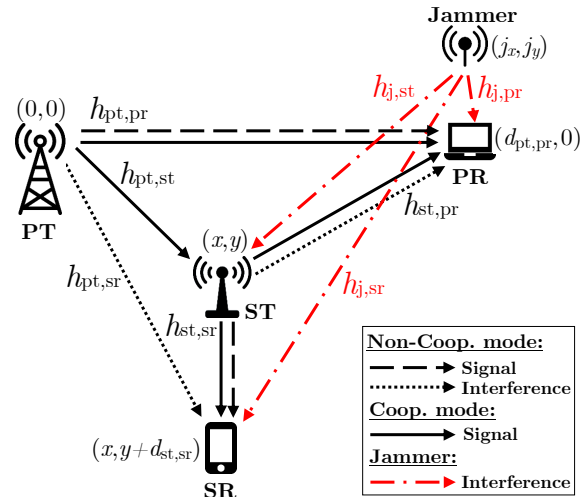

Figure 1: System model.

the PU performance, then the SU is exploited as a relay for the PU's data. The PU will have the choice either to cooperate with the SU or not.

- Due to the stochastic nature of the wireless environment (the jammer activity and the SU energy), the PU decision (either to cooperate with the SU or not) is modeled as a sequential decision making problem under uncertainty. To consider the performance of both the PU and the SU on the long run, the decision problem is formulated as a multi-objective mixed observable Markov decision process (MOMDP) where the PU imperfect sensing is considered. Then, the decision problem is solved for the optimal decision policy using a point-based value iteration (PBVI) algorithm.

Finally, simulation experiments are performed and the results show that, with the proposed scheme, the performance of the PU under jamming attack can be enhanced via exploiting cooperation with the SU.

The rest of the paper is organized as follows. System model and assumptions are presented in Section 2. The proposed scheme is explained in Section 3. The decision problem formulation and solution are presented in Section 4. The numerical results are shown in Section 5. Finally, the conclusion and future work are given in Section 6.

2 SYSTEM MODEL

2.1 Network Structure

We consider an underlay cognitive radio network consisting of a single PU (transmitter PT/receiver PR), a single SU (transmitter ST/receiver SR) and a ran-

dom jammer. The network components and their relative locations are depicted in Fig. 1. The jammer always has the enough energy to transmit jamming signals with power P_{jm}, and its activity follows a two-state Markov chain model in Fig. 2, where "J"/"NJ" stands for jamming/no jamming state. PT and ST always have data to transmit. All the channels between the PT, PR, ST, SR and the jammer are additive white Gaussian noise (AWGN) channels with channel coefficients $h_{pt,pr}, h_{pt,st}, h_{pt,sr}, h_{st,sr}, h_{st,pr}, h_{j,st}, h_{j,sr}$ and $h_{j,pr}$, such that

$$h_{tx,rx} = g_{tx,rx} \sqrt{\left(\frac{\lambda}{4\pi d_{tx,rx}}\right)^{\varepsilon}}, \qquad (1)$$

where $g_{tx,rx}$ is the channel fading coefficient between any two arbitrary nodes tx and rx, $d_{tx,rx}$ is the distance of the channel link, λ is the carrier wavelength and ε is the path loss coefficient assuming antennas' gains is 1.
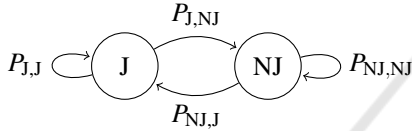

Figure 2: The jammer activity model.

2.2 Network Operating Modes

Our network has two operating modes: the non-cooperation (non-coop) mode and the cooperation (coop) mode.

In the non-cooperation mode, as depicted in Fig. 3a, the PU directly transmits (DT) its data through the direct channel between the PT and the PR for a period of T_t and allows the SU to concurrently transmit in underlay mode for the same period T_t on condition that ST interference power does not exceed the allowable interference threshold I_{th}.

While in the cooperation mode, as depicted in Fig. 3b, the ST operates as an amplify and forward (AF) relay for the PU data. In the first hop, the PT broadcasts its data to the PR and the ST for $\alpha\beta T_t$ period, then the ST amplifies the received signals (data and noise) and forwards the amplified signals to the PR in the second hop for $(1-\alpha)\beta T_t$ period. At the PR, signals received directly from PT in the first hop and indirectly from the ST in the second hop are combined using maximum-ratio combining (MRC) technique (Su et al., 2012). Finally, ST is allowed to transmit its own data for $(1-\beta)T_t$ with full power as an intensive for helping the PU.

At the beginning of each time slot T, the PU senses the direct channel (PT-PR) for jamming signals for T_s period, then chooses the proper operating mode.

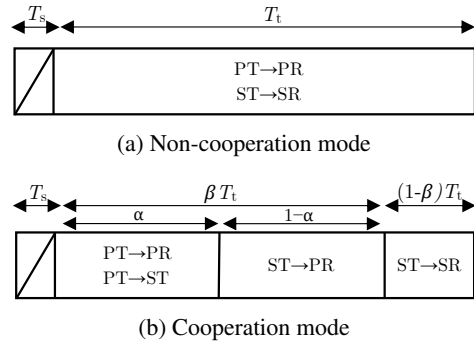(a) Non-cooperation mode

(b) Cooperation mode

Figure 3: Network operating modes.

The optimization of mode-selection is discussed in Section 4 taking into account the PU imperfect (realistic) sensing capabilities which is captured in the probabilities of detection P_d and false alarm P_f of the sensing mechanism. At the end of each time slot, the PT receives a feedback from the PR indicating the actual jamming status in that time slot. PR can detect the jammer activity from the received signal strength (RSS).

In the next subsections, we will compute the achievable throughput of the PU and the SU under the two specified network operating modes. Thus, we define two variables m and j indicating the network operating mode and jammer activity, respectively, as follows:

$$m = \begin{cases} 0 & , \text{non-cooperation} \\ 1 & , \text{cooperation,} \end{cases} \qquad (2)$$

$$j = \begin{cases} 0 & , \text{inactive} \\ 1 & , \text{active.} \end{cases} \qquad (3)$$

Normalized bandwidth is assumed in channels' capacities calculations, and any AWGN n_{rx} is with variance σ_{rx}^2.

2.3 Secondary User

ST operation relies on a battery with a limited capacity E_{max}. For battery recharging, ST harvests energy from ambient non-RF sources every time slot. The number of the harvested energy packets $e_{ha} \in \{1, 2, \ldots, n_a\}$ is random and follows a truncated Poisson distribution with rate μ_a. The maximum transmission power of ST is P_s^{max} which is a device constraint for better battery life.

2.3.1 Non-cooperation Mode

In this mode the SU access the licensed spectrum concurrently with the PU in underlay transmission (UT) mode. To avoid harmful interference to the

PU, ST transmits with maximum constrained power $P_{\mathrm{u}}^{\max} = \min\left(P_{\mathrm{s}}^{\max}, \frac{I_{\mathrm{th}}}{h_{\mathrm{st,pr}}}\right)$. Thus, the maximum consumed energy is $e_{\mathrm{u}}^{\max} = \min\left(P_{\mathrm{u}}^{\max} T_{\mathrm{t}}, E_{\max}\right)$ and the consumed energy at any time slot $e_{\mathrm{u}} = \min\left(e, e_{\mathrm{u}}^{\max}\right)$, where e is the SU available energy, and $P_{\mathrm{u}} = e_{\mathrm{u}}/T_{\mathrm{t}}$ is the transmission power. Transmission time fraction $\tau_{\mathrm{u}} = T_{\mathrm{t}}/T$.

2.3.2 Cooperation Mode

In this mode the SU access the licensed spectrum in overlay transmission (OT) mode. The ST is allowed to transmit its own data with full power $P_{\mathrm{s}}^{\max}$ for $(1 - \beta)T_{\mathrm{t}}$ period. We assume that the ST relays PT data and transmits its own data using the same power level $P_{\mathrm{o}}^{\mathrm{relay}} = P_{\mathrm{o}}^{\mathrm{own}}$. Thus, the consumed energy $e_{\mathrm{o}} = \min\left(e, e_{\mathrm{o}}^{\max}\right)$ where $e_{\mathrm{o}}^{\max} = \min\left(P_{\mathrm{s}}^{\max}\left(1 - \alpha\beta\right)T_{\mathrm{t}}, E_{\max}\right)$ is the maximum energy consumed in overlay mode. The transmission time fraction $\tau_{\mathrm{o}} = \frac{(1-\beta)T_{\mathrm{t}}}{T}$.

Thus, For the SU, the received signal at SR, channel capacity and achievable throughput are, respectively, given by

$$y_{\mathrm{st,sr}} = \sqrt{P_{\mathrm{su}}}h_{\mathrm{st,sr}}x_s + (1 - m)\sqrt{P_{\mathrm{p}}}h_{\mathrm{pt,sr}}x_p + j\sqrt{P_{\mathrm{jm}}}h_{\mathrm{j,sr}}x_{\mathrm{j}} + n_{\mathrm{sr}}, \quad (4)$$

$$C_{\mathrm{su}} = \log_2\left(1 + \frac{P_{\mathrm{su}}|h_{\mathrm{st,sr}}|^2}{(1-m)P_{\mathrm{p}}|h_{\mathrm{pt,sr}}|^2 + jP_{\mathrm{jm}}|h_{\mathrm{j,sr}}|^2 + \sigma_{\mathrm{sr}}^2}\right), \quad (5)$$

$$R_{\mathrm{su}} = \tau_{\mathrm{su}}C_{\mathrm{su}}, \quad (6)$$

where P_{p} is PT transmission power. The SU transmission power P_{su} and time fraction τ_{su} depends on the mode such that,

$$(P_{\mathrm{su}}, \tau_{\mathrm{su}}) = \begin{cases} (P_{\mathrm{u}}, \tau_{\mathrm{u}}) & , m = 0 \\ (P_{\mathrm{o}}, \tau_{\mathrm{o}}) & , m = 1. \end{cases} \quad (7)$$

2.4 Primary User

The PT has a fixed energy source that is sufficient for sensing jamming signals and transmitting data with power P_{p}. At the beginning of each time slot T, the PT has knowledge about the available energy e at the ST which will be exploited in optimizing PU decisions.

2.4.1 Non-cooperation Mode

In this mode the PU directly transmits (DT) through the direct channel (PT-PR). The received signal at PR,

channel capacity and achievable throughput are, respectively, given by

$$y_{\mathrm{pt,pr}} = \sqrt{P_{\mathrm{p}}}h_{\mathrm{pt,pr}}x_p + \sqrt{P_{\mathrm{u}}}h_{\mathrm{st,pr}}x_s + j\sqrt{P_{\mathrm{jm}}}h_{\mathrm{j,pr}}x_j + n_{\mathrm{pr}}, \quad (8)$$

$$C_{\mathrm{pu}}^{\mathrm{non\text{-}coop}} = \log_2\left(1 + \frac{P_{\mathrm{p}}|h_{\mathrm{pt,pr}}|^2}{P_{\mathrm{u}}|h_{\mathrm{st,pr}}|^2 + jP_{\mathrm{jm}}|h_{\mathrm{j,pr}}|^2 + \sigma_{\mathrm{pr}}^2}\right), \quad (9)$$

$$R_{\mathrm{pu}}^{\mathrm{non\text{-}coop}} = \tau_{\mathrm{pu}}^{\mathrm{non\text{-}coop}} C_{\mathrm{pu}}^{\mathrm{non\text{-}coop}}, \quad (10)$$

where $\tau_{\mathrm{pu}}^{\mathrm{non\text{-}coop}} = T_{\mathrm{t}}/T$ and the SU interference threshold was considered in computing P_{u}.

2.4.2 Cooperation Mode

In this mode the PR uses MRC to combine the received signals from both the direct channel (PT-PR) and the relayed channel (PT-ST-PR). For the relayed channel, the received signals $y_{\mathrm{pt,st}}$ at the ST and the relayed signal $y_{\mathrm{st,pr}}$ at PR are, respectively, given by

$$y_{\mathrm{pt,st}} = \sqrt{P_{\mathrm{p}}}h_{\mathrm{pt,st}}x_p + j\sqrt{P_{\mathrm{jm}}}h_{\mathrm{j,st}}x_j + n_{\mathrm{st}}, \quad (11)$$

$$y_{\mathrm{st,pr}} = \Lambda h_{\mathrm{st,pr}}y_{\mathrm{pt,st}} + j\sqrt{P_{\mathrm{jm}}}h_{\mathrm{j,pr}}x_j + n_{\mathrm{pr}}, \quad (12)$$

where Λ is the AF relay signal amplification factor that must satisify the condition

$$\Lambda \leq \sqrt{\frac{P_{\mathrm{o}}}{P_{\mathrm{p}}|h_{\mathrm{pt,st}}|^2 + jP_{\mathrm{jm}}|h_{\mathrm{j,st}}|^2 + \sigma_{\mathrm{st}}^2}}, \quad (13)$$

where P_{o} is the ST (AF relay) output power in the cooperation mode. We can assume that this constraint is met with equality (Laneman et al., 2004; Su et al., 2012). The SINR at the ST and PR are, respectively, given by

$$\mathrm{SINR}_{\mathrm{st}} = \frac{P_{\mathrm{p}}|h_{\mathrm{pt,st}}|^2}{jP_{\mathrm{jm}}|h_{\mathrm{j,st}}|^2 + \sigma_{\mathrm{st}}^2}, \quad (14)$$

$$\mathrm{SINR}_{\mathrm{pr}} = \frac{P_{\mathrm{o}}|h_{\mathrm{st,pr}}|^2}{jP_{\mathrm{jm}}|h_{\mathrm{j,pr}}|^2 + \sigma_{\mathrm{pr}}^2}. \quad (15)$$

While for the direct channel, the SINR at PR is given by

$$\mathrm{SINR}_{\mathrm{d}} = \frac{P_{\mathrm{p}}|h_{\mathrm{pt,pr}}|^2}{jP_{\mathrm{jm}}|h_{\mathrm{j,pr}}|^2 + \sigma_{\mathrm{pr}}^2}. \quad (16)$$

From (Laneman et al., 2004), the capacity of the combined channel and achievable throughput are, respectively, given by

$$C_{\mathrm{pu}}^{\mathrm{coop}} = \alpha\log_2\left(1 + \mathrm{SINR}_{\mathrm{d}} + f\left(\mathrm{SINR}_{\mathrm{st}}, \mathrm{SINR}_{\mathrm{pr}}\right)\right), \quad (17)$$

$$R_{\mathrm{pu}}^{\mathrm{coop}} = \tau_{\mathrm{pu}}^{\mathrm{coop}} C_{\mathrm{pu}}^{\mathrm{coop}}, \quad (18)$$

where $f(a,b) \doteq \frac{ab}{a+b+1}$ and $\tau_{\mathrm{pu}}^{\mathrm{coop}} = \frac{\beta T_{\mathrm{t}}}{T}$.
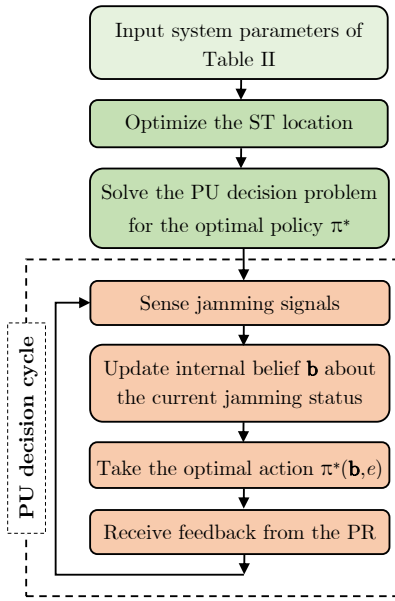
Figure 4: Proposed scheme steps.

3 PROPOSED SCHEME

Our proposed scheme is summarized in Fig. 4. After initializing the system parameters, the ST location is optimized to achieve the best PU performance in the cooperation mode when the jammer is active. Using the PU achievable throughput as a performance measure, we maximize (18) with respect to ST location (x,y) assuming $j = 1$ and $e = \mu_a$. R_{pu}^{coop} in (18) is related to ST location (x,y) though channels' coefficients $h_{pt,st}, h_{j,st}$ and $h_{st,pr}$ through the corresponding distances from (1) and Fig. 1,

$$d_{pt,st} = \sqrt{x^2 + y^2},$$

$$d_{j,st} = \sqrt{(x - j_x)^2 + (y - j_y)^2},$$

$$d_{st,pr} = \sqrt{(x - d_{pt,pr})^2 + y^2}.$$

In addition, we impose two constraints on the optimal solution: 1) $d_{pt,st} \leq d_{pt,pr}$, to ensure that the ST remains in the coverage area of the PT, 2) $d_{st,pr} \geq D$, where D is the radius of a prohibited area around the PR, allowing ST to operate in underlay mode with reasonable energy and not causing much interference to the PU. The ST location optimization problem can be formulated as

$$\begin{aligned} \max_{x,y} \quad & R_{pu}^{coop}\left(d_{pt,st}, d_{j,st}, d_{st,pr}\right) \\ \text{s.t.} \quad & d_{pt,st} \leq d_{pt,pr}, \\ & d_{st,pr} \geq D, \end{aligned} \qquad (19)$$

and the optimal solution can be obtained using the iterative interior point algorithm.

After optimizing the SU location, the PU solves the mode-selection decision problem and obtains the optimal policy π^* that maps between the system state (jammer status and the ST energy) and the corresponding optimal operating mode. After that, the system starts its normal operation by entering the PU decision cycle. This cycle starts by collecting the sensing data and the PR feedback about the previous transmission. This information is accumulated over time to infer a better estimate of the jammer current status as a probabilistic belief **b**. Then, the PU exploits π^* and takes the optimal action from the current state $(\mathbf{b}, e)$. The detailed formulation and solution of the decision problem are discussed in the next section.

4 DECISION PROBLEM

The PU is the decision maker who determines, at each time slot, the network operation mode. To optimize its decisions, the PU needs to solve a sequential decision making problem under uncertainty. We will consider only two sources of uncertainty, which are the jammer status j and the ST available energy e; they together constitute the system state (j,e) upon which the PU takes its decision.

Markov decision process (MDP) and its extensions are powerful mathematical tools for solving such decision problems (Abu Alsheikh et al., 2015). Due to the PU imperfect sensing, the jammer status j is a partially observable state-component, thus, the PU will maintain an internal probabilistic belief distribution $\mathbf{b} = [P_J \ P_{NJ}]$ over j possible values, where P_J/P_{NJ} is the PU belief that the jammer is active/inactive at the current time slot. Hence, system belief state becomes $(\mathbf{b}, e)$ or (P_J, e) for short.

Partially observable Markov decision process (POMDP) is the suitable MDP extension for modeling the PU decision problem. We can make use of the full observability of e to solve the POMDP problem in a more computationally efficient way known as mixed-observable Markov decision process (MOMDP) (Ong et al., 2009) by factoring the mixed-observable state space into union of disjoint subspaces each is corresponding to a value of the observable state-component e.

4.1 Problem Formulation: (POMDP Model)

The POMDP model is defined by a 6-tuple $< S,A,O,\mathcal{T},Z,R >$ where S is the state space of the underlying MDP. In our problem, $S = S_j \times S_e = \{0,1\} \times \{0,1,\ldots,E_{max}\}$.

$A = \{a_i\}_{i=1}^3$ is the action space, where a_1 represents the network's non-cooperation mode and a_2 represents the network's cooperation mode. While in a_3, both the PU and the SU stay silent to account for the case when the PU has a strong belief of jamming existence and $e = 0$. However, in all of these actions, the SU harvests energy from ambient sources.

Being in a mixed-observable state (j,e) at time slot $(t-1)$ and taking action a, the system transits to a next state (j',e') at time slot t where the PU will not be able to exactly observe the value of j' due to imperfect sensing. So the PU will use all available data (known as observations in POMDP context) to infer the jamming status as explained in subsection 4.3.1.

Observation space $O = O_{\text{feedback}} \times O_{\text{sensing}} = \{-, NJ, J\} \times \{SJ, SNJ\}$ where O_{feedback} contains the possible feedback values from PR about jamming status at $(t-1)$ with (-) indicates no feedback in case of taking a_3 where the PT is silent. O_{sensing} contains the possible sensing results at (t) which is either sensed jamming (SJ) or sensed no jamming (SNJ). We can enumerate the observation space elements, $\{o_i\}_{i=1}^6 = \{(-, SJ^t), (-, SNJ^t), (NJ^{t-1}, SJ^t), (NJ^{t-1}, SNJ^t), (J^{t-1}, SJ^t), (J^{t-1}, SNJ^t)\}$.

$\mathcal{T}(j',e'|j,e,a)$ is the state-transition probability function which assigns, for each action a, the probability of transition form state (j,e) to a next state (j',e'). Due to the probabilistic independence between j' and e', and between j' and a, $\mathcal{T}$ can be decomposed into $\Pr(e'|e,a)\Pr(j'|j)$ where $\Pr(j'|j)$ is available from the Markov chain model and $\Pr(e'|e,a)$ can be computed for each $a_i \in A$ as follows:

$$e' = \min(e - \zeta + e_{\text{ha}}, E_{\max}),\tag{20}$$

$\Pr(e'|e,a_i) =$

$$\begin{cases} \Pr(e_{\text{ha}} = e' - e + \zeta), & e - \zeta + 1 \leq e' < E_{\max} \\ \sum\limits_{k=E_{\max}-e+\zeta}^{n_{\text{a}}} \Pr(e_{\text{ha}} = k), & e' = E_{\max}, \end{cases}$$

$$\tag{21}$$

where $\zeta = e_{\text{u}}, e_{\text{o}}$ and 0 for $i = 1, 2$ and 3 respectively.

$Z(o|j',a,\mathbf{b})$ is the observation probability function which gives the probability of observing o as a result of system transition to j' after taking action a and having a belief $\mathbf{b}$. For transmitting actions $a \in \{a_1, a_2\}$, and since observations o_1 and o_2 do not include feedback from the PR, then $\Pr(o_1|j',a,\mathbf{b}^{t-1}) = \Pr(o_2|j',a,\mathbf{b}^{t-1}) = 0$. Moreover, $\Pr(o_3|J^{t-1},a,\mathbf{b}^{t-1}) = \Pr(NJ^{t-1}, SJ^t|J^t,a,\mathbf{b}^{t-1})$ and from the conditional independence of NJ^{t-1} and SJ^t given J^t, it can be ex-

pressed as

$$\begin{aligned} \Pr(o_3|J^t,a,\mathbf{b}^{t-1}) &= \Pr(NJ^{t-1}|J^t,a,\mathbf{b}^{t-1}) \\ &\times \Pr(SJ^t|J^t,a,\mathbf{b}^{t-1}) \\ &= P_{\text{NJ}}^{t-1}P_{\text{d}}. \end{aligned}\tag{22}$$

Similarly, we can compute

$$\Pr(o_i|J^t,a,\mathbf{b}^{t-1}) = \begin{cases} P_{\text{NJ}}^{t-1}(1-P_{\text{d}}) & ,i=4 \\ P_{\text{J}}^{t-1}P_{\text{d}} & ,i=5 \\ P_{\text{J}}^{t-1}(1-P_{\text{d}}) & ,i=6. \end{cases}\tag{23}$$

While for silent action a_3, $\Pr(o_i|j',a_3,\mathbf{b}^{t-1}) = 0$ for $i = 3, 4, 5, 6$ for any j', because the silent action does not expect feedback from the PR, and

$$\Pr(o_i|J^t,a_3,\mathbf{b}^{t-1}) = \begin{cases} P_{\text{d}} & ,i=1 \\ 1-P_{\text{d}} & ,i=2. \end{cases}\tag{24}$$

For $j' = NJ^t$, we only replace P_{d} by P_{f} in (22), (23) and (24).

Finally, $R(j,e,a)$ is the immediate reward function that assigns a scalar value for taking an action a from a mixed-observable state (j,e). Because j is hidden from the decision maker (agent), it uses the expected value $R(\mathbf{b},e,a) = \sum_j b(j)R(j,e,a)$ during solving the problem.

4.2 Objective Function and $R(j,e,a)$

In POMDP optimization problems, the agent aims to maximize the accumulation of the immediate rewards received from the environment. Solving the problem means finding the optimal policy $\pi^* : (\mathbf{b},e) \rightarrow a$ that agent follows to accomplish its goal. Using the discounted sum criterion for reward accumulation, the optimization problem can be written as $\pi^*(\mathbf{b},e) = \arg\max_a \sum_{t=0}^{\infty} \gamma^t E[R_{t+1}(\mathbf{b},e,a)]$, where γ is the discount factor.

$R(j,e,a)$ is determined in the light of the system objective. As our system is a CRN, it aims to maximize the throughputs of both the PU and the SU for better spectrum utilization. But according to our system structure, maximizing the throughputs of both the PU and the SU may be contradicting objectives in some situations. For example, at lower jamming levels, the PU may prefer to work in non-cooperation mode (a_1) where it can transmit during the whole T_{t}, while SU may prefer to operate in the cooperation mode (a_2) where it transmits with its full power for $(1 - \beta)T_{\text{t}}$. Thus, it is reasonable to take these contradiction situations into consideration during optimizing network performance.

Therefore, for each state (j,e) we consider the point $\left(R_{\text{pu}}(j,e,a), R_{\text{su}}(j,e,a)\right)$ that represents the

Table 1: PU and SU individual rewards.

-	a_1	a_2	a_3
$R_{\mathrm{pu}}(j,e,a)$	(10)	(18)	0
$R_{\mathrm{su}}(j,e,a)$	(6) ; $m=0$	(6) ; $m=1$	0

PU and the SU throughputs from Table 1, while the point $\left(R_{\mathrm{pu}}^*(j,e), R_{\mathrm{su}}^*(j,e)\right)$ is the optimal point that will be infeasible in contradiction situations, where $R_{\mathrm{pu}}^*(j,e) = \max_a R_{pu}(j,e,a)$ and $R_{\mathrm{su}}^*(j,e) = \max_a R_{su}(j,e,a)$. We will consider the negative of the normalized weighted distance between the previously specified two points as the immediate reward function or actually the penalty function here, given by

$$R(j,e,a) = -\max\left(w_{\mathrm{pu}} \left| 1 - \frac{R_{\mathrm{pu}}(j,e,a)}{R_{\mathrm{pu}}^*(j,e)} \right|, \right.$$
$$\left. w_{\mathrm{su}} \left| 1 - \frac{R_{\mathrm{su}}(j,e,a)}{R_{\mathrm{su}}^*(j,e)} \right| \right), \quad (25)$$

where the distance is measured in L_∞ norm and the weights w_{pu} and w_{su} sum to 1 as a usual practice, and reflect the importance of one objective over the other from the decision maker perspective. This formulation is known as the weighted Tchebycheff metric (Deb, 2001).

4.3 Problem Solution

POMDP problem can be decomposed into two sub-problems: Hidden state estimation and the optimal policy determination.
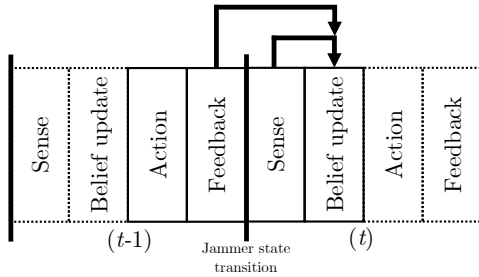
4.3.1 Hidden State Estimation

Figure 5: Time slots and the decision cycle.

The agent utilizes the history of actions and observations summarized in the previous belief $\mathbf{b}^{t-1}$, along with the most recent action-observation pair to update its current belief $\mathbf{b}^t$. Having a compound observation structure (feedback from $t-1$, sensing result at t), we want to exploit its two parts in updating the belief as depicted in Fig. 5. The first part (feedback from $t-1$) is exploited to get a prior belief

$P_{\mathrm{J}}^{t(\mathrm{prior})} = \Pr(\mathrm{J}^t | o_{i(\mathrm{feedback})}, a, \mathbf{b}^{t-1})$ about jamming existence at t, before considering the sensing result, this prior belief is given by

$$P_{\mathrm{J}}^{t(\mathrm{prior})} = \begin{cases} P_{\mathrm{J}}^{t-1} P_{\mathrm{J,J}} + P_{\mathrm{NJ}}^{t-1} P_{\mathrm{NJ,J}} & , i = 1,2 \\ P_{\mathrm{NJ,J}} & , i = 3,4 \\ P_{\mathrm{J,J}} & , i = 5,6. \end{cases} \quad (26)$$

To incorporate the sensing result at t (the second part of the observation), we want to go from the effect (sensing result) to the cause (jammer status). Bayes' rule from probability theory is typical for this situation. Thus, the updated belief is given by

$$P_{\mathrm{J}}^t = \Pr(\mathrm{J}^t | o_i, a, \mathbf{b}^{t-1}) = \begin{cases} \frac{P_{\mathrm{d}} P_{\mathrm{J}}^{t(\mathrm{prior})}}{\Pr(\mathrm{SJ}^t)} & , i = 1,3,5 \\ \frac{(1-P_{\mathrm{d}}) P_{\mathrm{J}}^{t(\mathrm{prior})}}{\Pr(\mathrm{SNJ}^t)} & , i = 2,4,6, \end{cases} \quad (27)$$

where $\Pr(\mathrm{SJ}^t)$ and $\Pr(\mathrm{SNJ}^t)$ are the total probabilities of sensing results at time slot t, given by

$$\Pr(\mathrm{SJ}^t) = P_{\mathrm{d}} P_{\mathrm{J}}^{t(\mathrm{prior})} + P_{\mathrm{f}} P_{\mathrm{NJ}}^{t(\mathrm{prior})}, \quad (28)$$

$$\Pr(\mathrm{SNJ}^t) = (1-P_{\mathrm{d}}) P_{\mathrm{J}}^{t(\mathrm{prior})} + (1-P_{\mathrm{f}}) P_{\mathrm{NJ}}^{t(\mathrm{prior})}. \quad (29)$$

4.3.2 Finding the Optimal Policy $\pi^*(\mathbf{b}, e)$

Exact solutions for POMDPs is an intractable problem (Spaan and Vlassis, 2005). So, instead of computing the optimal policy from all points $\mathbf{b}$ in a continues belief sub-space associated to a specific e value, we just sample a set B of reachable belief points under a random policy as suggested by the PBVI algorithm in (Spaan and Vlassis, 2005) to reach approximate solutions. At the nth iteration and a specific e, the optimal value function $V_n^{e,*}$ is proved to be piece-wise linear and convex (PWLC). So, it can be represented by a set $\Gamma_n^{e,*} = \{\upsilon_n^{e,a,i}\}_{i=1}^m$ of vectors $\upsilon_n^{e,a}$, where m is the number of the useful policies (conditional plans) at the nth iteration and e. These vectors can be used to evaluate $V_n^{e,*}(\mathbf{b})$ at any $\mathbf{b}$ using (30), which will be exact for all $\mathbf{b} \in B$ and approximate otherwise,

$$V_n^{e,*}(\mathbf{b}) = \max_{\upsilon_n^{e,a}} \mathbf{b}.\upsilon_n^{e,a}. \quad (30)$$

Also, each vector $\upsilon_n^{e,a} \in \Gamma_n^{e,*}$ is associated with an action a which is the optimal action over a partition of the belief sub-space. Therefore, we can get the optimal action $\pi_n^{e,*}(\mathbf{b})$ at any $\mathbf{b}$ similar to (30) but by using $\arg\max$ instead.

$\upsilon_n^{e,a}$ vectors can be backed up recursively between successive iterations using

$$\upsilon_n^a(j,e) = R(j,e,a) + \gamma \sum_{j'} \sum_o \sum_{e'} \Pr(j'|j) \Pr(o|j',a,\mathbf{b})$$
$$\times \Pr(e'|e,a) \upsilon_{n-1}^{*,o}(j',e'), \quad (31)$$

Table 2: Simulation parameters.

Parameter	Value	Parameter	Value
f	650 MHz	I_{th}	10^{-15}W
T	100 ms	$|g_{j,st}|^2 = |g_{j,sr}|^2$	10^{-3}
T_{s}	5 ms	Other $|g_{tx,rx}|^2$	0.5
β	0.9	All σ_x^2	4×10^{-18}W/Hz
α	0.5	$d_{pt,pr}$	3 km
P_{d}	0.99	$d_{st,sr}$	30 m
P_{f}	0.01	(j_x, j_y)	(2500, -200)
P_{jm}	100 mW	D	7
P_{p}	1000 mW	ε	3
$P_{\text{s}}^{\max}$	500 mW	Initial P_{J}	0.7
$E_{\max}$	30 mJ	Initial e	1 mJ
n_{a}	20	w_{pu}	1
μ_{a}	15	γ	0.95

where $\upsilon_{n-1}^{*,o}$ is the vector corresponding to the optimal action at the reachable belief in the e' belief sub-space in the $(n-1)^{\text{th}}$ iteration. This belief is reachable from **b** after taking action a and observing o.

5 NUMERICAL RESULTS

In this section, we evaluate the effect of applying our proposed scheme on the performance of both the PU and the SU. The simulation parameters are given in Table 2. P_{J}^{ss} is the steady state probability of the jammer's two-state Markov chain model, which can be achieved by setting $P_{\text{J,J}} = P_{\text{NJ,J}} = P_{\text{J}}^{\text{ss}}$. P_{J}^{ss} is used as a measure of the jamming level, i.e., how often the jammer is active. The performance of the PU and the SU is measured by their average throughputs. The performance of our proposed scheme is compared to the baseline (always operating in the non-cooperation mode after optimizing the ST location) and the random policy (randomly choosing between the cooperation and non-cooperation modes after optimizing the ST location).

Fig. 6 illustrates the PU performance under different schemes at different jamming levels. Generally, the performance of the PU degrades with the increase of the jamming level as expected. However, our proposed scheme achieves the best performance over the baseline and the random behavior. The superiority of our proposed scheme is due to exploiting the cooperation with the SU and optimizing the choice either to collaborate or not depending on the situation. For instance, at no jamming at all ($P_{\text{J}}^{\text{ss}} = 0$), our proposed policy agrees with the baseline in always choosing the non-cooperation mode because this achieves the best performance for the PU. With the increase of the jamming level, the baseline scheme performance degrades with a steeper slope. In order to relieve this steepness, our proposed policy relays on the cooperation-mode, and with the increase
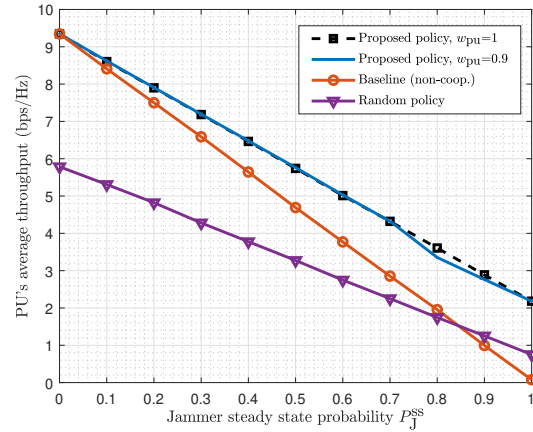

Figure 6: The PU's average throughput under different schemes at different jamming levels.

of the jamming level this reliance increases, until we reach $P_{\text{J}}^{\text{ss}} = 1$, cooperation mode is exploited all the time. The detailed behavior of our scheme through $P_{\text{J}}^{\text{ss}} \in [0, 1]$ is illustrated more in Fig. 7.

Also, from Fig. 6 we can notice the contribution of the preference weights $(w_{\text{pu}}, w_{\text{su}})$ at contradiction situations. Under the current simulation parameters, both the PU and the SU prefer cooperation when the jammer is active, while their goals contradict when the jammer is inactive; the PU prefers non-cooperation while the SU prefers cooperation. Therefore, we can observe that when $0.7 < P_{\text{J}}^{\text{ss}} < 1$ the PU performance under our proposed policy with $w_{\text{pu}} = 0.9$ is less than when $w_{\text{pu}} = 1$. This is because the jammer is active and cooperation is preferred for more than 70% of time, and the effect of $(w_{\text{pu}}, w_{\text{su}})$ appears when the jammer is inactive (30% of time or less). When $w_{\text{pu}} = 1$, non-cooperation is always chosen when the jammer is inactive. While when $w_{\text{pu}} = 0.9$, the SU desires are taken into consideration by also choosing cooperation when the jammer is inactive, which results in decreasing the PU performance at $w_{\text{pu}} = 0.9$ and $0.7 < P_{\text{J}}^{\text{ss}} < 1$.

Fig. 7 shows the percentage of choosing the cooperation mode under our proposed scheme at different jamming levels for $w_{\text{pu}} = 1$ and 0.9. This figure is complementary to Fig. 6, showing the effect of $w_{\text{pu}} = 1$ and 0.9 when $P_{\text{J}}^{\text{ss}} > 0.7$. It is noted that silent mode (a_3) is never chosen. a_3 was expected to be chosen when the jammer is active and the ST does not have enough energy to relay the PU data, but because at this situation non-cooperation (a_1) still achieves non-zero throughput, it is preferred over a_3. Hence, a_3 can be removed from the POMDP action space to speedup policy computation.

Regarding the SU performance, Fig. 8 shows the SU performance under our proposed scheme at
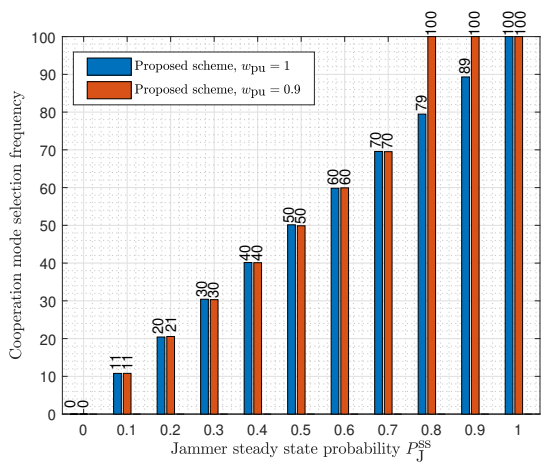
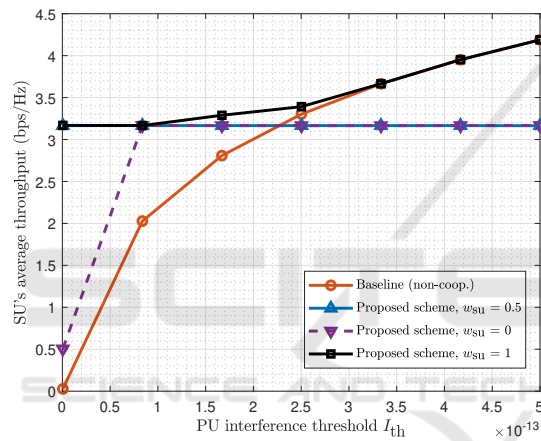Figure 7: Cooperation mode selection frequency under the proposed scheme at different jamming levels.

Figure 8: The SU's average throughput against the PU's allowable interference threshold I_{th}.

$w_{su} = 1, 0.5$ and 0 against the PU interference threshold I_{th} at $P_J^{ss} = 0.2$. It is obvious that our proposed scheme behavior is affected by the preference weights (w_{pu}, w_{su}), making its performance ranges from superiority to middle levels from the SU perspective. In underlay CRNs the I_{th} controls the performance of both the PU and the SU. At lower I_{th} values, the SU is allowed to transmit with lower power, causing less interference to the PU and resulting in a SU's lower performance, while as I_{th} increases, the reverse occurs; this is clear from the baseline behavior. This effect of I_{th} causes the PU and the SU preferences to change with I_{th}. At lower I_{th} values, the SU prefers cooperation which is more chosen when $w_{su} = 1$ and 0.5, while when $w_{su} = 0$, cooperation is only chosen when the jammer is active, i.e., for 20% of time ($P_J^{ss} = 0.2$). Thus, at lower I_{th}, the performance of the SU under our proposed scheme is better than the baseline for all w_{su}, and increasing the value of w_{su} increases the
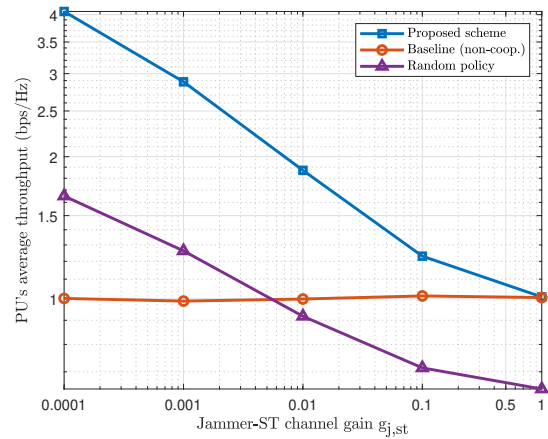

Figure 9: The PU's average throughput under different schemes against $g_{j,st}$.

SU's throughout.

At higher values of I_{th}, the SU prefers non-cooperation, which is achieved when $w_{su} = 1$, and the SU performance under the proposed scheme converges to the baseline. While at $w_{su} = 0.5$ and 0, the SU performance is below the baseline because the PU prefers cooperation at higher I_{th} values and its preference is taken into consideration.

Fig. 9 shows the PU performance against the fading coefficient $g_{j,st}$ of the channel (J-ST) between the jammer and the ST. It is clear that the performance of our proposed scheme is superior to the random and baseline behavior. However, it decreases with the increase of $g_{j,st}$, (i.e. better J-SR channel conditions and more jamming effect on the ST), until it meets with the baseline behavior at $g_{j,st} = 1$, (indicating no fading at all). This is reasonable because our proposed scheme depends on the cooperation between the PU and the SU as a main component, which will be more effective if the SU is less affected by the jammer interference (lower $g_{j,st}$) than the PU. Thus, along $g_{j,st}$ values range, our proposed scheme keeps optimizing the balance between cooperation and non-cooperation mode selection, until it always selects the non-cooperation mode at $g_{j,st} = 1$.

6 CONCLUSION

In this paper a PU-SU cooperation scheme was proposed for enhancing the PU performance under jamming attack in a cognitive radio network. The proposed scheme considers the PU imperfect sensing and the SU energy constraint. The SU is exploited as a relay for the PU's data, and its location is optimized to maximize the PU performance. The PU decision to cooperate with the SU or not, is formulated as a multi-

objective MOMDP problem to optimize the performance of both the PU and the SU on the long run. The decision problem is solved using a PBVI algorithm and an optimal decision policy was obtained. Compared to other schemes, the simulation results show the superiority of the proposed scheme in enhancing the PU performance at different jamming levels. In addition, the effectiveness of the proposed scheme will be higher if the SU is less affected by the jammer interference. As a future direction, cooperation with multiple SUs can be considered. Also, adaptive time allocation between the PU and the SU in cooperation mode can be studied. Finally, reinforcement learning techniques will be investigated to solve the decision problem, especially if the system's probabilistic models are not available.

ACKNOWLEDGMENT

REFERENCES

Abu Alsheikh, M., Hoang, D. T., Niyato, D., Tan, H., and Lin, S. (2015). Markov decision processes with applications in wireless sensor networks: A survey. *IEEE Commun. Surveys Tut.*, 17(3):1239–1267.

AlAqad, K. F., Burhanuddin, M., and Harum, N. B. (Jerusalem, Palestine, 16-17 Dec., 2020). A comprehensive survey on routing protocols for cognitive radio-based disaster response networks. In *2020 Int. Conf. Promising Electron. Technol. (ICPET)*, pages 133–139.

Bhowmick, A., Roy, S. D., and Kundu, S. (Kolkata, India, 15-18 Dec., 2015). Performance of secondary user with combined RF and non-RF based energy-harvesting in cognitive radio network. In *Proc. IEEE Int. Conf. Advanced Netw. Telecommun. Syst. (ANTS2015)*, pages 1–3.

Deb, K. (2001). Multi-objective optimization using evolutionary algorithms. John Wiley & Sons.

Di Pietro, R. and Oligeri, G. (2013). Jamming mitigation in cognitive radio networks. *IEEE Netw.*, 27(3):10–15.

El Tanab, M. and Hamouda, W. (2016). Resource allocation for underlay cognitive radio networks: A survey. *IEEE Commun. Surv. Tuto.*, 19(2):1249–1276.

Hasan, M., Thakur, J. M., and Podder, P. (2016). Design and implementation of FHSS and DSSS for secure data transmission. *Int. J. Sig. Process. Syst.*, 4(2):144–149.

Laneman, J. N., Tse, D. N. C., and Wornell, G. W. (2004). Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Trans. Inf. Theory*, 50(12):3062–3080.

Liang, W., Ng, S. X., and Hanzo, L. (2017). Cooperative overlay spectrum access in cognitive radio networks. *IEEE Commun. Surveys Tut.*, 19(3):1924–1944.

Ong, S. C., Png, S. W., Hsu, D., and Lee, W. S. (Seattle, USA, 28 Jun.- 1 Jul., 2009). POMDPs for robotic tasks with mixed observability. In *Proc. Robotics: Sci. Syst. Conf.*, volume 5, page 4.

Pirayesh, H. and Zeng, H. (2021). Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey. *arXiv preprint arXiv:2101.00292*.

Qin, M., Yang, S., Deng, H., and Lee, M. H. (2018). Enhancing security of primary user in underlay cognitive radio networks with secondary user selection. *IEEE Access*, 6:32624–32636.

Shu, Z., Qian, Y., and Ci, S. (2013). On physical layer security for cognitive radio networks. *IEEE Netw.*, 27(3):28–33.

Spaan, M. T. and Vlassis, N. (2005). Perseus: Randomized point-based value iteration for POMDPs. *J. Artif. Intell. Res.*, 24:195–220.

Su, W., Matyjas, J. D., and Batalama, S. (2012). Active cooperation between primary users and cognitive radio users in heterogeneous ad-hoc networks. *IEEE Trans. Signal Process.*, 60(4):1796–1805.

Thanh, P.-D., Vu-Van, H., and Koo, I. (2018). Efficient channel selection and routing algorithm for multihop, multichannel cognitive radio networks with energy harvesting under jamming attacks. *Secur. Commun. Netw.*, 2018.

Zhang, N., Cheng, N., Lu, N., Zhou, H., Mark, J. W., and Shen, X. S. (2014). Risk-aware cooperative spectrum access for multi-channel cognitive radio networks. *IEEE J. Sel. Areas in Commun.*, 32(3):516–527.