This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2019.2933778, IEEE Journal on Selected Areas in Communications

Network Association in Machine-Learning Aided Cognitive Radar and Communication Co-Design

Jingjing Wang, *Member, IEEE*, Sanghai Guan, Chunxiao Jiang, *Senior Member, IEEE*,
Dimitrios Alanis, *Member, IEEE*, Yong Ren, *Senior Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—In order to beneficially exploit the scarce wireless spectral resources, spectrum sharing between communication and radar systems has become a promising research topic. However, traditional network association strategies may not result in efficient hybrid communication and radar systems. We circumvent this problem by formulating a partially observable Markov decision processes (POMDP) aided network association scheme, where the radar user acts as the primary user (PU), while the cognitive communication user is the secondary user (SU). For maximizing the network throughput, whilst minimizing the interference imposed on the radar user, the communication user is configured for adaptively selecting its underlay or overlay access mode. Moreover, a low-complexity near-optimal reinforcement learning algorithm is proposed for the co-design by considering both its complexity and feasibility. Finally, we quantify the performance of our proposed POMDP based network association scheme.**

***Index Terms*—Network association, POMDP, cognitive communication and radar co-design, underlay, overlay.**

I. Introduction

THE critical radar information infrastructure has been evolving towards environmentally aware adaptation, multi-functional implementations as well as towards big data awareness. In the spirit of holistic optimization, multiple single-function systems may collaborate with each other both for sharing the scarce spectral resources and for enhancing the network performance. As one of the critical information infrastructure components, radar systems are typically used for object-detection by analyzing the reflected radio waves to determine the range, angle or velocity of objects. By contrast, communication systems rely on the radio channels for transmitting information. In numerous civilian and military scenarios, the pair of systems co-exist and depend on each other. For example, the object-detection information emanating from the radar system should be promptly transmitted to the command center via the communication system at a high integrity. Furthermore, the frequency bands of next-generation communication system gradually extend to high-frequency microwave bands, some of which overlap the frequency bands of radar systems. Specific frequency bands have been invoked for spectrum sharing between the radar system and the communication system, such as the 3550–3650 MHz band. Hence, a well-designed network association scheme is beneficial in terms of mitigating the interference between the pair of systems, which requires cognitive communication and radar co-design [1].

However, traditional network association strategies face numerous challenges in hybrid communication and radar systems. Specifically, the frequency-hopping radar makes it difficult for communication users to accurately estimate the rapidly time-varying channel state information (CSI). Moreover, considering the bursty nature of traffic as well as the limited affordable power consumption, it is impractical for the communication system to incessantly sense the whole channel. Hence, the estimation of the system' channel state is one of its gravest challenges. Furthermore, conceiving a considerate spectrum sharing scheme is another open challenge, given the non-uniform sub-channel occupation and the presence of other ambient interferences. The aforementioned challenges require new network association methods, which are specifically designed for the communication and radar co-design.

Despite the above-mentioned challenges, some solutions have been proposed, which demonstrate that it is feasible to tackle the aforementioned problems. Game theory has been widely used in the literature for spectrum management [2]–[6]. To elaborate, in [4], Zhu *et al.* proposed a twin-level dynamic game model for spectrum sharing in two-tier cellular networks. Considering the users' dynamic decisions as well as the information delay, their proposed game yielded both an improved payoff and an increased convergence speed. In [5], a cooperative game was constructed by Liu *et al.* for maximizing the network utility of multi-user cognitive communications. Moreover, an efficient distributed algorithm was discussed, which led to rapid convergence to the optimal solution. Furthermore, Yi *et al.* [6] proposed a two-stage resource allocation scheme relying on combinatorial auction and on Stackelberg game aided spectrum management in the context of multiple heterogeneous spectrum sellers and buyers. However, all the aforementioned studies of spectrum management were designed for the co-existence of multiple communication users having different jurisdictions. Moreover,

This work is supported by the National Natural Science Foundation of China (91338203), Young Elite Scientists Sponsorship Program By CAST (2016QNRC001), the Pre-research Fund of Equipments of Ministry of Education of China (6141A02022615) the New Strategic Industries Development Projects of Shenzhen City (JCYJ20170816151922176). L. Hanzo would like to acknowledge the financial support of the EPSRC projects EP/Noo4558/1, EP/P34284 as well as of the ERC's Advanced Fellow Award QuantCom. (Corresponding author: Chunxiao Jiang)

J. Wang, S. Guan and Y. Ren are with the Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China. E-mail: chinaeephd@gmail.com, gsh17@mails.tsinghua.edu.cn, reny@tsinghua.edu.cn.

C. Jiang is with Tsinghua Space Center, Tsinghua University, Beijing, 100084, and also with the Research Institute of Tsinghua University in Shenzhen, Shenzhen 518057, China. E-mail: jchx@tsinghua.edu.cn.

D. Alanis and L. Hanzo are with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK. Email: d.alanis@soton.ac.uk, lh@ecs.soton.ac.uk.

emphasis has been predominantly on the utility of communication, such as the attainable throughput instead of focusing on the radar's performance.

As for the communication and radar co-design, a range of joint optimization schemes have been investigated [7]–[11]. Specifically, in [8], Turlapaty *et al.* proposed the joint design of the radar transmission waveform and the power spectral density of the multi-carrier communications system. This joint design was beneficial in terms both of enhancing the radar functions as well as of maintaining a substantial throughput for the communication system. Furthermore, the network association was formulated as a signal-to-interference-plus-noise ratio (SINR) maximization problem at the radar receiver, which was also subjected to the rate and power constraints of the communication system by Li *et al.* [9]. In [10], an adaptive radar beamforming technique was proposed by Geng *et al.* for eliminating the wireless interferences imposed by communication systems during spectrum sharing. Furthermore, sophisticated interference mitigation methods were designed in [12], [13].

However, joint optimization problems require accurate CSI and often suffer from a high computational complexity owing to their extended search-space. In reality, the bursty nature of traffic and the power constraint make it impossible for the communication system to estimate the CSI accurately. Reinforcement learning is a powerful decision-making tool, which maps situations to actions so as to maximize a numerical reward function [14]. As a popular member of the reinforcement learning family, both the Markov decision process (MDP) as well as the partially observable Markov decision process (POMDP) have been widely used for beneficial network association, such as access point selection in super-WiFi networks [15], rate and mode adaptation for WiFi/LTE-U hybrid networks [16], etc. In [17], Zhao *et al.* proposed a POMDP-based cognitive MAC protocol for opportunistic spectrum access (OSC) in wireless network, and provided a reduced-complexity suboptimal algorithm. Furthermore, in [18], Chen *et al.* separated the OSC into the sensing step and access step. A POMDP assisted joint optimal design was proposed, which was capable of considering the presence of sensing errors. By observing the strict energy constraint of cognitive radio networks, Hoang *et al.* [19] formulated a constrained POMDP framework for modeling the trade-off among energy consumption, delay and throughput. Moreover, a heuristic control policy was also proposed.

As evidenced by these contributions, POMDP has indeed been beneficially used for example in cognitive radio [17]–[19], because it is competent in constructing hypothesized states for estimating the partially unknown channel conditions, followed by exploiting them for decision-making. In this contribution, we go beyond the state-of-the-art by specifically designing the POMPD technique for the new amalgamated radar and communication system. Apart from a few exceptions, radar and communication co-design aims for guaranteeing the detection and estimation performance of the radar/communication systems by designing specific waveforms or beamformers for the sake of improving only a tolerable level of interference on the communication/radar system [20]. Although the POMDP technique constitutes an efficient tool of network association in communication and radar co-design in the face of rapidly time-varying channel states and partial observability from an upper-layer perspective, this research area is in its infancy. To make progress, in this paper, we conceive a novel POMDP based network association scheme for communication and radar co-design[1]. Our original contributions are summarized as follows:

- We formulate a POMDP based network association scheme for communication and radar co-design, which is capable of nimble adaptation to dynamically fluctuating environments, whilst efficiently exploiting the scarce spectral resources.
- A sampling-aided low-complexity co-design technique is proposed relying on the piecewise linearity and convexity of the value function, which provides a near-optimal solution for our network association problem.
- Simulations are conducted, which verify the compelling features of our proposed learning algorithm in terms of improving the network's throughput in the face of interference.

The remainder of this article is outlined as follows. The system model is detailed in Section II. An iterative POMDP-based sensing and access decision-making strategy is conceived for the communication and radar co-design in Section III. In Section IV, a low-complexity near-optimal online learning algorithm is designed and its complexity is analysed. In Section V, simulation results are provided for characterizing the POMDP based network association algorithms, followed by our conclusions in Section VI.

II. System Model

In this context, we construct a communication and radar co-design for the primary user (PU) and the secondary user (SU), as shown in Fig. 1. More specifically, PUs are unaware of the existence of SUs and they require unhindered access to the wireless channel without an extra authorization. In contrast to the PUs, SUs firstly sense the state of the wireless channel at the beginning of each time slot and then select an appropriate access strategy relying on their sensing results.

A. The Primary User

Radar systems detect and track objects by receiving and processing the waves reflected by the objects. In this treatise, radar users are viewed as the PUs, which constitute the primary network (PN). The frequency-hopping technique can beneficially improve the radar system's capability of avoiding both interference as well as of frequency selective fading,

[1] As for our theoretical contribution in comparison to [11], this paper defines the observation function, estimated state, hypothesized state as well as the reward function. Moreover, we provide the complete derivation and proof of the hypothesized state transition function relying on probability theory in this version. Furthermore, based on Bellman's principle, we elaborate on how to formulated the iterative value function. As for our simulations, we define the SU's achievable rate and PU's SNR degradation as a pair of performance metrics characterizing the for evaluating our proposed POMDP algorithm and its reduced complexity version. In excess of ten figures are added for characterizing the feasibility and superiority of our proposed algorithm.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2019.2933778, IEEE Journal on Selected Areas in Communications

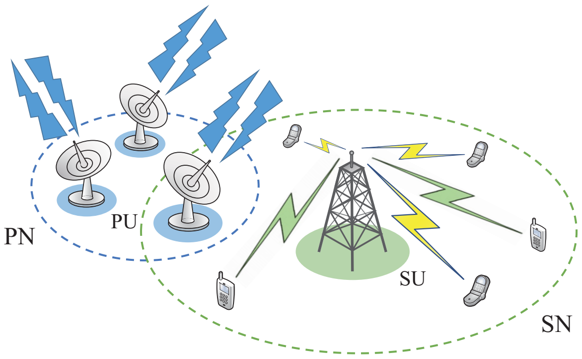


Fig. 1. The structure of radar and communication co-design, where PUs construct the primary network (PN) sharing the same frequency band with SUs in the secondary network (SN) in terms of an opportunistic access.

which is achieved by periodically hopping to a different frequency by retuning the frequency synthesizer. Provided that the system hops beyond the coherence frequency, independent fading is experienced, hence mitigating both the interference and fading effect. In our model, frequency-hopping radar systems are considered, which are characterized by the random scanning of the time-, frequency- and spatial resource slots. Thus, the spectrum holes created by the frequency hopping mechanism can be exploited by other communication systems for improving the join system's spectrum efficiency.

B. The Secondary User

Naturally, the SUs are radio communication users, who are served by a communication base station (BS). The BS is in charge of both sensing the channel and of formulating appropriate access strategies for the SUs. The SUs as well as the BS construct the secondary network (SN) utilizing the same frequency band as the PU. The SUs are capable of taking advantage of free channels and of sharing occupied channels, provided that the SINR constraints are not violated.

C. Co-Design Model

1) System State and its Transition Function: The total bandwidth of the co-design's channel is denoted as W, where N sub-channels can be sensed and accessed. The bandwidth of each sub-channel is represented by $W_1, W_2, \ldots, W_N$. These N sub-channels are assigned to the PUs, also termed as authorized users, which are capable of occupying any, or even all the sub-channels without any restriction. Thus, each sub-channel has two states at each time slot, i.e. the 'busy' state when the PUs are transmitting their signals and the 'idle' state, when the PUs are not using the sub-channel. Let $s_i(t)$ represent the state of the sub-channel i at the time slot t. Then we have $s_i(t) = 1$ if its state is busy, while $s_i(t) = 0$ if the sub-channel is idle. Therefore, the co-design's state at time slot t can be represented by a vector $\mathbf{S}(t) = [s_1(t), s_2(t), \ldots, s_N(t)], s_i(t) \in \{0, 1\}$. The co-design considered has a total of 2^N different states. Let $\mathbb{S}$ represent the co-design's state set, where we have $\mathbf{S} \in \mathbb{S}$ as well as $|\mathbb{S}| = 2^N$.

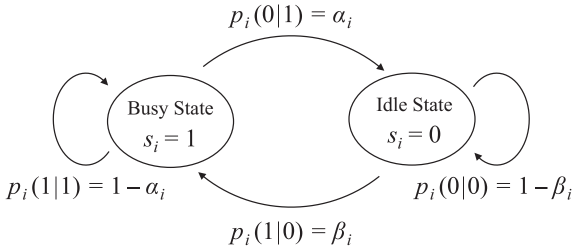


Fig. 2. The Markov process for the state transition of the sub-channel i, for example.

In our model, we assume that the state transitions of sub-channel i obey a Markov process, as shown in Fig. 2, where α_i represents the probability of the channel state traversing from busy to idle, while β_i denotes the probability of the state changing from idle to busy. Hence, the state transition probability of sub-channel i can be formulated as:

$$p_i(s'_i \mid s_i) = \Pr\{s_i(t+1) = s'_i \mid s_i(t) = s_i\}, \tag{1}$$

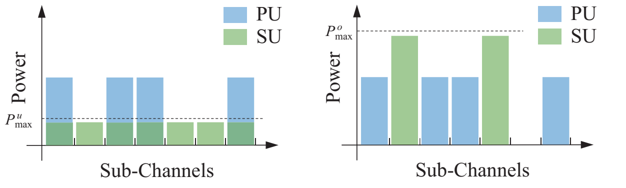
where s'_i represents the state of sub-channel i at the next time slot and $s_i, s'_i \in \{0, 1\}$. Let p_i^0 and p_i^1 represent the probability of the sub-channel i staying in the idle state and in the busy state, respectively, when the above-mentioned Markov process reaches its steady state. Hence, we have $p_i^0 = \alpha_i/(\alpha_i + \beta_i)$ and $p_i^1 = \beta_i/(\alpha_i + \beta_i)$. Relying on the independence of each sub-channel, the co-design's state transition function can be expressed by:

$$\begin{aligned} p(\mathbf{S}' \mid \mathbf{S}) &= \Pr\{\mathbf{S}(t+1) = \mathbf{S}' \mid \mathbf{S}(t) = \mathbf{S}\} \\ &= \prod_{i=1}^{N} \Pr\{s_i(t+1) = s'_i \mid s_i(t) = s_i\}, \end{aligned} \tag{2}$$

where $\mathbf{S}' = [s'_1, s'_2, \ldots, s'_N]$, $\mathbf{S} = [s_1, s_2, \ldots, s_N]$ and $\mathbf{S}', \mathbf{S} \in \mathbb{S}$.

2) Two Sequential Actions in SN: In our model, opportunistic spectrum management (OSM) is invoked for the SU's channel selection. The spectrum management decision-making can be divided into two stages, i.e. the sensing stage as well as the access stage.

Considering the energy constraint, the communication BS is capable of observing at most M sub-channels during the sensing stage at time slot t, where $M < N$. Relying on the previously observed results, the BS aims for selecting M of N sub-channels in order to better estimate the system's real state $\mathbf{S}(t)$ at time slot t, which can be viewed as the first action of the SN. In this paper, first action (Action 1) set is denoted by $\mathbb{A}_1 = \{\mathbf{A}_1\}$, where $\mathbf{A}_1 = [a_1^1, a_2^1, \ldots, a_N^1] \in \{0, 1\}^N$ and $|\mathbb{A}_1| = \binom{N}{M}$. To elaborate a little further, if SN decides to sense the ith sub-channel, we have $a_i^1 = 1$; otherwise we have $a_i^1 = 0$. Moreover, a maximum of M sub-channels sensing capacity of the SN yields $\sum_{i=1}^{N} a_i^1 \leq M$. Furthermore, the false-alarm rate and the missed-detection rate of the sensing stage are represented by ζ_f and ζ_m, respectively. Specifically, the false-alarm rate is the probability of falsely obtaining the sensing result that the sub-channel's state is busy but it is actually idle. By contrast, the missed-detection rate refers to



(a) Underlay access scheme.

(b) Overlay access scheme ($L = 2$, for example).

Fig. 3. The diagram of underlay and overlay access scheme.

the probability of mistakenly reckoning that the sub-channel is free but it is occupied in reality.

Relying on M sensed and observed sub-channels, we define the observation state vector of the co-design considered at time slot t, i.e. $\mathbf{O}(t) = [o_1(t), o_2(t), \ldots, o_N(t)]$, where $o_i(t) \in \{0, 1, \phi\}$. To elaborate, if the observed result of sub-channel i is idle, $o_i(t) = 0$, whilst if the observed result is busy, we have $o_i(t) = 1$. Moreover, $o_i(t) = \phi$ represents that the sub-channel i is not observed at time slot t. Let $\mathbb{O}$ represents the observation state set, i.e. $\mathbf{O} \in \mathbb{O}$.

In the access stage, based on the sensing result of the actual system's state, either the underlay or overlay access mode can be selected as the access scheme by SN. The second action (Action 2) set can be represented by $\mathbb{A}_2 = \{A_2\}$, where the variable $A_2 \in \{a_u^2, a_o^2\}$. Specifically, if the underlay access scheme is invoked as the second action, we have $A_2 = a_u^2$, while $A_2 = a_o^2$ if SN selects the overlay access scheme. Furthermore, the power allocated by the SN to each sub-channel is denoted by $\mathbf{P} = [P_1, P_2, \ldots, P_N]$. As shown in Fig. 3, the two aforementioned access schemes can be elaborated on a litter further as follows [21] [22].

- *Underlay Scheme*: SUs are capable of accessing the whole channel shared with PUs, who have a low and equally shared transmission power in each sub-channel. Hence, we have $P_1^u = P_2^u = \cdots = P_N^u$ and $P_i^u \leq P_{\max}^u$, where $P_{\max}^u$ represents the maximum allowable transmission power of the underlay scheme based on the interference constraint of PUs. Moreover, the interference constraint of PUs is parameterized by the frequency-hopping radar's performance in order to guarantee both its detection probability as well as false alarm probability specifications.
- *Overlay Scheme*: SUs can only access at most L sub-channels that are most likely to be idle relying on their observation of the channel state. Furthermore, the SUs are capable of using a higher transmission power than that in the underlay access mode. Let P_i^o represent the transmission power of the i-th assumed-to-be-idle sub-channel, and we have $P_i^o \leq P_{\max}^o$, where $P_{\max}^o$ represents the maximum affordable transmission power of the overlay scheme based on the transmitter's performance at the BS.

Therefore, the SN's decision-making pertaining to the spectrum management hinges on the above-mentioned two sequential actions, i.e. Action 1 as well as Action 2, which can be represented by the action vector $\mathbf{A} = [\mathbf{A}_1, A_2]$. Thus, the SN's action set of the whole spectrum management process can be formulated as $\mathbb{A} = \mathbb{A}_1 \times \mathbb{A}_2$, where × represents the Cartesian product and $|\mathbb{A}| = 2\binom{N}{M}$, while $\mathbf{A} \in \mathbb{A}$.

3) Reward: The reward of our proposed co-design, namely R, is defined as the total net reward that the SN acquires across all sub-channels, i.e.

$$R = \sum_{i=1}^{N} R_i. \tag{3}$$

Specifically, the net reward R_i consists of two parts, i.e. the *capacity gain* R_{ig} as well as the *interference penalty* R_{ip}. Hence, we have:

$$R_i = R_{ig} + R_{ip}. \tag{4}$$

To elaborate, if the SN successfully accesses an idle sub-channel, i.e. $s_i = 0$, the *capacity gain* R_{ig} can be formulated based on the Shannon formula, i.e.

$$R_{ig}(P_i, N_i) = \lambda_C W_i \log\left(1 + \frac{g_{sr} P_i}{(g_{pr} N_i + N_0) W_i}\right), \tag{5}$$

where λ_C is a weighting coefficient, while W_i represents the bandwidth of sub-channel i. Furthermore, N_i and N_0 denote the average power spectral density of the radar system and of the Gaussian white noise considered in sub-channel i, respectively. When sub-channel i is in the idle state, i.e. $s_i = 0$, we have $N_i = 0$. Moreover, g_{sr} and g_{pr} represent the receiver's power gain at the SUs and the transmitter's power gain at the PUs, respectively. Here, $R_{ig} \geq 0$.

However, as for the *interference penalty* mentioned above, if the SN mistakenly accesses a busy sub-channel occupied by the PN, i.e. $s_i = 1$, it will inevitably impose a serious interference on the PN, hence resulting in a substantial detection rate reduction for the radar system. Given the reckless access of the SN, the *interference penalty* R_{ip} can be formulated as:

$$R_{ip}(P_i, N_i) = -\lambda_I \cdot \frac{g_{sp}[P_i - P_{\max}^u]_+}{N_i W_i}, \tag{6}$$

where λ_I represents a weighting coefficient, while g_{sp} denotes the receiver's power gain of the PUs. Furthermore, we define the function $[\cdot]_+ = \max\{\cdot, 0\}$. We can find that $R_{ip} = 0$, when the SN selects the underlay scheme associated with $P_i \leq P_{\max}^u$, while there is a risk of a detrimental interference penalty quantified by $R_{ip} \leq 0$, when SN selects the overlay scheme, which may yield a higher capacity gain quantified by R_{ig}.

Furthermore, when the sub-channel i is idle, i.e. $s_i = 0$, regardless of which access scheme is selected, we have $R_{ip} = 0$. Finally, if the sub-channel i is not selected by the SN, we have $R_{ig} = R_{ip} = 0$.

In this paper, we focus our attention on the general cognitive radar and communication co-existence scenario, where the communication system is designed for spectrum sharing with frequency-agile radar. As for the radar's performance, different radar systems rely on different performance indices [23]–[25], such as the false alarm rate and miss detection rate, the estimation errors of the targets' range and velocity, its imaging performance, etc. However, these performance indices intrinsically rely on the radar receiver's SNR. Hence, in our manuscript, we use the receiver's SNR performance for quantifying the impact of communication users on the radar system.

The specific values of the associated weighting coefficients can be learned by comparing the receiver's SNR and the required target SNR performance.

III. The POMDP Approach

A. Observation Function

In this subsection, we define the observation function of $z_i(o_i|s_i, a_i^1)$, which refers to the probability of the observation state value of sub-channel i, i.e. the aforementioned o_i, under the condition of the first action a_i^1 at the system's state s_i, yielding:

$$z_i(o_i \mid s_i, a_i^1) = \Pr\left(o_i(t) = o_i \mid s_i(t) = s_i, a_i^1(t) = a_i^1\right). \tag{7}$$

Specifically, when we make the decision of the sensing stage as $a_i^1 = 1$, the observation function of the sub-channel i in Eq. (7) can be calculated as:

$$z_i(o_i \mid s_i, a_i^1) = \begin{cases} 1 - \zeta_f, & \text{if } o_i = 0, s_i = 0, a_i^1 = 1, \\ \zeta_m, & \text{if } o_i = 0, s_i = 1, a_i^1 = 1, \\ \zeta_f, & \text{if } o_i = 1, s_i = 0, a_i^1 = 1, \\ 1 - \zeta_m, & \text{if } o_i = 1, s_i = 1, a_i^1 = 1, \end{cases} \tag{8}$$

where ζ_f and ζ_m represent the above-mentioned false-alarm rate and missed-detection rate of the SN in the sensing stage, respectively. On the other hand, when $a_i^1 = 0$, the observation function is given by:

$$z_i(\phi \mid 1, 0) = 1, \tag{9}$$

as well as

$$z_i(\phi \mid 0, 0) = 1. \tag{10}$$

Considering that the state transition of each sub-channel is independent of that of the others, the co-design's observation function at time slot t can be formulated as:

$$\begin{aligned} z(\mathbf{O}|\mathbf{S}, \mathbf{A}_1) &= \Pr\big(\mathbf{O}(t) = \mathbf{O} \mid \mathbf{S}(t) = \mathbf{S}, \mathbf{A}_1(t) = \mathbf{A}_1\big) \\ &= \prod_{i=1}^{N} \Pr\big(o_i(t) = o_i \mid s_i(t) = s_i, a_i^1(t) = a_i^1\big). \end{aligned} \tag{11}$$

B. Estimated State

Given the energy and capacity constraint of the communication BS, the SN is unable to sense and estimate the accurate state of all sub-channels. Here, we define the estimated state in order to describe the system's state assumed after the sensing stage. The estimated state vector of N sub-channels in our co-design can be expressed by $\mathbf{\Theta}^{\mathbf{S}}(t) = [\theta_1^{s_1}(t), \theta_2^{s_2}(t), \ldots, \theta_N^{s_N}(t)]$, where $\theta_i^{s_i}(t)$ is the probability that sub-channel i is estimated to be at state s_i at time slot t. For the sake of simplification, $\theta_i^1(t)$ represents the probability that sub-channel i is estimated to be in the busy state at time slot t, i.e. $s_i \doteq 1$, while $\theta_i^0(t) = 1 - \theta_i^1(t)$ denotes the probability that sub-channel i is estimated to be in the idle state at time slot t. In our paper, we use $\doteq$ to represent the estimated value.

C. Hypothesized State and its Transition Function

The SN's hypothesized state of a certain legitimate system state $\mathbf{S}(t)$ at time slot t, namely $B_{\mathbf{S}}(t)$, refers to the conditional probability of the co-design's realistic state being $\mathbf{S}$, conditioned on the estimated state being $\mathbf{\Theta}^{\mathbf{S}}(t)$, i.e.

$$B_{\mathbf{S}}(t) = \Pr\big(\mathbf{S}(t) = \mathbf{S} \mid \mathbf{\Theta}^{\mathbf{S}}(t) = \mathbf{\Theta}^{\mathbf{S}}\big) = \prod_{i=1}^{N} \theta_i^{s_i}(t), \tag{12}$$

where $\mathbf{S} = [s_1, s_2, \ldots, s_N]$. Hence, we can represent the SN's hypothesized state vector at time slot t by $\mathbf{B}(t) = [B_{\mathbf{S}_1}(t), B_{\mathbf{S}_2}(t), \ldots, B_{\mathbf{S}_{2^N}}(t)] \in \mathbb{B}$, where $\mathbb{B}$ refers to the SN's hypothesized state set. More specifically, the elements of the vector $\mathbf{B}(t)$ can be viewed as a one-to-one mapping to the system's 2^N legitimate states, and we have $|\mathbf{B}(t)| = |\mathbb{S}| = 2^N$.

In the following, we define the hypothesis transition function $b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}_1)$ of our proposed co-design, which refers to the probability that the SN's hypothesized state traverses from $\mathbf{B}$ at time slot $(t-1)$ to $\mathbf{B}'$ under the sensing stage action $\mathbf{A}_1$ at time slot t. Then we have:

$$b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}_1) = \Pr\left(\mathbf{B}(t) = \mathbf{B}' \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\right), \tag{13}$$

where $\mathbf{B}', \mathbf{B} \in \mathbb{B}$.

After some derivations as shown in Appendix, the SN's hypothesis transition function of Eq. (13) can be expressed as:

$$\begin{aligned} b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}_1) = \sum_{\mathbf{O} \in \mathbb{O}} \Bigg(& I\left\{\mathbf{B}' = [B'_{\mathbf{S}_1}, B'_{\mathbf{S}_2}, \cdots, B'_{\mathbf{S}_{2^N}}]\right\} \\ & \cdot \sum_{\mathbf{S} \in \mathbb{S}} \Big(z(\mathbf{O} \mid \mathbf{S}, \mathbf{A}_1) \cdot \sum_{\mathbf{S}' \in \mathbb{S}} p(\mathbf{S} \mid \mathbf{S}') \cdot B_{\mathbf{S}'} \Big) \Bigg), \end{aligned} \tag{14}$$

where $I\{\cdot\}$ represents an indicator function, while $B'_{\mathbf{S}_1}, B'_{\mathbf{S}_2}, \cdots, B'_{\mathbf{S}_{2^N}}$ can be calculated from Eq. (33).

D. Access Scheme Selection

After the sensing stage action $\mathbf{A}_1$ at time slot t, the SN has to select either the underlay or overlay scheme as the access stage action, i.e. A_2, relying on the updated estimated state and hypothesized state. The access stage action A_2 aims for maximizing the expected reward received, which can be described as:

$$A_2^*(t) = \underset{A_2(t) \in \{a_u^2, a_o^2\}}{\arg\max} \; \mathbb{E}\left[R(t) \,|\, \mathbf{\Theta}^{\mathbf{S}}(t), A_2(t)\right]. \tag{15}$$

The SN may compare the expected reward obtained with both the underlay access scheme as well as with the overlay access scheme relying on the given estimated state. To elaborate a little further, if SN selects the underlay access scheme as the access stage action, i.e. $A_2 = a_u^2$, its expected reward can be expressed by:

$$\begin{aligned} \mathbb{E}\left[R \,|\, \mathbf{\Theta}^{\mathbf{S}}, A_2 = a_u^2\right] &= \sum_{i=1}^{N} \mathbb{E}\left[R_i \,|\, \theta_i^{s_i}, P_i = P_{\max}^u\right] \\ = \sum_{i=1}^{N} \Big(\theta_i^1 \, R_{ig}\left(P_{\max}^u, N_i\right) &+ \theta_i^0 \, R_{ig}\left(P_{\max}^u, 0\right) \Big). \end{aligned} \tag{16}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2019.2933778, IEEE Journal on Selected Areas in Communications

However, if SN selects the overlay access scheme as the access stage action, i.e. $A_2 = a_o^2$, it will access the L 'most-likely-to-be-idle' sub-channels, namely Ω. Hence, the SN, first of all, determines the transmission power P_i on L sub-channels for maximizing the expected reward, which can be formulated as:

$$\begin{aligned} \max_{P_i} \quad & \mathbb{E}\left[R \,|\, \mathbf{\Theta}^{\mathbf{S}}, A_2 = a_o^2\right], \\ \text{s.t.} \quad & P_{\max}^u \le P_i \le P_{\max}^o,\ i \in \Omega. \end{aligned} \tag{17}$$

Let P_i^{o*} represent the optimal power allocated to sub-channel i, where $i \in \Omega$. Thus, the reward expected for the overlay access scheme can be calculated as:

$$\begin{aligned} &\mathbb{E}\left[R \,|\, \mathbf{\Theta}^{\mathbf{S}}, A_2 = a_o^2\right] = \sum_{i\in\Omega} \mathbb{E}\left[R_i \,|\, \theta_i^{s_i}, P_i = P_i^{o*}\right] \\ &= \sum_{i\in\Omega} \Big(\theta_i^1 \left(R_{ig}\left(P_i^{o*}, N_i\right) + R_{ip}\left(P_i^{o*}, N_i\right)\right) + \theta_i^0 R_{ig}\left(P_i^{o*}, 0\right) \Big). \end{aligned} \tag{18}$$

Compared to the expected value of Eq. (16) and of Eq. (18), SN selects the better scheme as the access stage action A_2. In our paper, we assume that the access stage action A_2 does not influence the observed state $\Theta^{\mathbf{S}}$, for SN can only become informed of the total reward R after taking the action A_2, and it cannot acquire the accurate state information of each sub-channel.

E. A POMDP Framework

Based on the aforementioned assumptions and definitions, we can construct a POMDP framework of the network association for our proposed co-design, which can be formulated as a quintuple $\langle \mathbb{S}, \mathbb{B}, \mathbb{A}, b, r \rangle$. Specifically,

- *System's State Set:* $\mathbb{S} = \{\mathbf{S}\}$ is the set of all the possible system states $\mathbf{S}$, where $\mathbf{S} = [s_1, s_2, \cdots, s_N]$;
- *Hypothesized state Set:* $\mathbb{B} = \{\mathbf{B}\}$, where $\mathbf{B} = [B_{\mathbf{S}_1}, B_{\mathbf{S}_2}, \ldots, B_{\mathbf{S}_{2N}}]$ is the hypothesized state vector referring to the grade of similarity between each legitimate system state $\mathbf{S} \in \mathbb{S}$ and the estimated state $\Theta^{\mathbf{S}}$ relying on partial observation;
- *SN's Action Set:* $\mathbb{A}$ is the set of all the possible actions, i.e. $\mathbf{A} \in \mathbb{A}$, where $\mathbf{A} = [\mathbf{A}_1, A_2]$ represents a SN's specific action determining which M sub-channels to sense and which of the two available access mechanisms to select;
- *Hypothesis transition Function:* $b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}_1) \colon \mathbb{B} \times \mathbb{A}_1 \times \mathbb{B} \mapsto [0,1]$, where the operand '×' represents the Cartesian product, while $\mathbf{B}^{'} \in \mathbb{B}$ is the hypothesized state at the next time slot;
- *Reward Function:* $r(\mathbf{B}, \mathbf{A}_1, A_2) \colon \mathbb{B} \times \mathbb{A} \mapsto \mathbb{R}$, which indicates the immediate expected reward received as a benchmark of the pair of sequential actions $\mathbf{A}_1$ and A_2 relying on the hypothesized state $\mathbf{B}$. Moreover, we have $r(\mathbf{B}, \mathbf{A}_1, A_2) = \sum_{\mathbf{S}\in\mathbb{S}} B_{\mathbf{S}} \cdot R(\mathbf{S}, \mathbf{A}_1, A_2)$.

In order to better understand our proposed POMDP framework, important definitions and their internal relationships are illustrated in Fig. 4.

F. Optimal Policy

As we mentioned before, we have converted the discrete POMDP problem into a continuous MDP problem relying on the concept of hypothesized state as well as its state transition function. In order to search for the optimal action of the SN in each step, let $G(t)$ represent the discounted accumulated reward of the co-design commenced at time slot t, namely the *return*, which can be expressed by:

$$G(t) = \sum_{k=0}^{\infty} \gamma^k \cdot r\big[\mathbf{B}(t+k), \mathbf{A}_1(t+k), A_2(t+k)\big], \tag{19}$$

where γ $(0 \le \gamma \le 1)$ denotes the discount rate, which determines the weight of the future reward towards the co-design. Specifically, a large γ means that the co-design is 'farsighted' and focuses more attention on the future reward, and vice versa. The SN's goal is to maximize the return $G(t)$ by jointly considering the current hypothesized state $\mathbf{B}(t)$ and its appropriate action $\mathbf{A}(t)$. Here, we define the *policy* as a mapping $\pi \colon \mathbf{B} \mapsto \mathbf{A}$, where $\mathbf{B} \in \mathbb{B}$ and $\mathbf{A} \in \mathbb{A}$. As for a given value of $\pi(\mathbf{A} \mid \mathbf{B})$, it refers to the probability of taking action $\mathbf{A}$, when in the hypothesized state $\mathbf{B}$. In terms of different possible policies for a given hypothesized state, the *hypothesized value function* $V^{\pi}(\mathbf{B})$ is defined in order to characterize the expected return $G(t)$ of the hypothesized state $\mathbf{B}$, which can be formulated by:

$$V^{\pi}(\mathbf{B}) = \mathbb{E}_{\pi}\left[G(t) | \mathbf{B}(t) = \mathbf{B}\right]. \tag{20}$$

Given Bellman's principle [26], we can satisfy the Bellman formula of $V^{\pi}(\mathbf{B})$ as:

$$\begin{aligned} &V^{\pi}(\mathbf{B}) = \mathbb{E}_{\pi}\left[G(t) | \mathbf{B}(t) = \mathbf{B}\right] \\ &= \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^k \cdot r(t+k) | \mathbf{B}(t) = \mathbf{B}\right] \\ &= \mathbb{E}_{\pi}\left[r(t) + \gamma \sum_{k=0}^{\infty} \gamma^k \cdot r(t+k+1) | \mathbf{B}(t) = \mathbf{B}\right] \\ &= \sum_{\mathbf{A}\in\mathbb{A}} \pi(\mathbf{A} \mid \mathbf{B}) \sum_{\mathbf{B}'\in\mathbb{B}} b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}) \\ &\quad \cdot \left(r(t) + \gamma \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^k \cdot r(t+k+1) | \mathbf{B}(t+1) = \mathbf{B}'\right]\right) \\ &= \sum_{\mathbf{A}\in\mathbb{A}} \pi(\mathbf{A} \mid \mathbf{B}) \sum_{\mathbf{B}'\in\mathbb{B}} b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}) \left(r(t) + \gamma V_{\pi}(\mathbf{B}')\right). \end{aligned} \tag{21}$$

conditioned on a Markov state transition, where $r(t)$ is the abbreviation of $r[\mathbf{B}(t), \mathbf{A}_1(t), A_2(t)]$.

Hence, we can achieve the optimal mixed-policy π^* under the hypothesized state from:

$$\begin{aligned} &\pi^*(\mathbf{A} \mid \mathbf{B}) = \arg\max_{\pi} \mathbb{E}_{\pi}\left[G(t) \mid \mathbf{B}(t) = \mathbf{B}, \mathbf{A}(t) = \mathbf{A}\right] \\ &= \arg\max_{\pi} \sum_{\mathbf{A}\in\mathbb{A}} \pi(\mathbf{A} \mid \mathbf{B}) \sum_{\mathbf{B}'\in\mathbb{B}} b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}) \left[r(t) + \gamma V^*(\mathbf{B}')\right], \end{aligned} \tag{22}$$

which yields the best value function with the best mixed-probability aided actions for that hypothesized state, in the

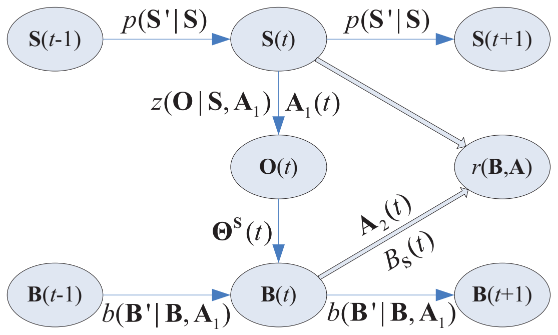


Fig. 4. The diagram of proposed POMDP framework illustrating the internal relations of some important definitions.

form of:

$$\begin{aligned} V^*(\mathbf{B}) &= \max_{\pi} \mathbb{E}_{\pi}\left[G(t) \mid \mathbf{B}(t)=\mathbf{B}, \mathbf{A}(t)=\mathbf{A}\right] \\ &= \max_{\pi} \sum_{\mathbf{A}\in\mathbb{A}} \pi(\mathbf{A} \mid \mathbf{B}) \sum_{\mathbf{B}'\in\mathbb{B}} b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A})\left[r(t) + \gamma V^*(\mathbf{B}')\right]. \end{aligned} \tag{23}$$

In reality, in each time slot, an action pair should be provided for the SUs by the BS, which also guarantees the convergence of the iterative algorithm considered. Then we have:

$$\pi^*(\mathbf{A} \mid \mathbf{B}) = \arg\max_{\mathbf{A}\in\mathbb{A}} \sum_{\mathbf{B}'\in\mathbb{B}} b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A})\left[r(t) + \gamma V^*(\mathbf{B}')\right]. \tag{24}$$

Thus, the value function of an infinite time slots is expressed as:

$$V^*(\mathbf{B}) = \max_{\mathbf{A}\in\mathbb{A}} \sum_{\mathbf{B}'\in\mathbb{B}} b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A})\left[r(t) + \gamma V^*(\mathbf{B}')\right]. \tag{25}$$

The aforementioned iterative POMDP based optimal technique is described in Algorithm 1, where we discretize the continuous hypothesized states in $\mathbb{B}$.

Algorithm 1: Iterative based POMDP Optimal Algorithm

```
discretize $\mathbf{B} \in \mathbb{B}$;
initialize $V^{(0)}(\mathbf{B}) \leftarrow 0$ and $\pi^{(0)}(\mathbf{A} \mid \mathbf{B})$ for all
  discretized $\mathbf{B} \in \mathbb{B}$;
while $\max_{\mathbf{B}\in\mathbb{B}} |V^{(k+1)}(\mathbf{B}) - V^{(k)}(\mathbf{B})| > \epsilon$ do
    for $\mathbf{B} \in \mathbb{B}$ do
        calculate $b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A})$ relying on Eq. (35);
        value iteration $V^{(k+1)}(\mathbf{B}) \leftarrow$
          $\max_{\mathbf{A}\in\mathbb{A}} \sum_{\mathbf{B}'\in\mathbb{B}} b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}) \left[r(k) + \gamma V^{(k)}(\mathbf{B}')\right]$;
        policy improvement
        $\pi^{(k+1)}(\mathbf{A} \mid \mathbf{B}) \leftarrow \arg\max_{\mathbf{A}\in\mathbb{A}} V^{(k+1)}(\mathbf{B})$;
    end
end
return $V^{(k+1)}(\mathbf{B})$ and $\pi^{(k+1)}(\mathbf{A} \mid \mathbf{B})$;
```

IV. A Low-Complexity Near-Optimal Algorithm

For the sake of efficiently searching for the feasible solution of our POMDP formulation, in this section, we propose a low-complexity near-optimal algorithm relying on the specific form of the value function of Eq. (25) [27]. Specifically, in our model, the optimal value function of Eq. (25) is piecewise linear and convex with respect to the hypothesis transition function of Eq. (35), which is closely related to the hypothesized state $\mathbf{B}$. It is reasonable to assume that a strong belief in the idle nature of a sub-channel contributes a high reward. As shown in Eq. (34), the hypothesized state $\mathbf{B}$ can be calculated with the aid of the estimated state vector $\boldsymbol{\Theta}$. Hereinafter, we use $\boldsymbol{\Theta}^1$ to denote the estimated state vector, which is mathematically equivalent to $\boldsymbol{\Theta}$.

Hence, we approximately reformulate the value function $V(\mathbf{B})$ in the form of a non-linear polynomial function with respect to $\boldsymbol{\Theta}^1$, which is expressed as:

$$\tilde{V}(\mathbf{B}) \triangleq f(\boldsymbol{\Theta}^1) = \boldsymbol{\mu}^{\mathrm{T}}\phi(\boldsymbol{\Theta}^1), \tag{26}$$

where $\boldsymbol{\mu} = [\mu_0, \mu_1, \mu_2, \ldots]^{\mathrm{T}}$ represents the regression coefficient vector, while $\phi(\boldsymbol{\Theta}^1)$ is an N-dimension expansion vector of $\boldsymbol{\Theta}^1$, i.e.

$$\phi(\boldsymbol{\Theta}^1) = \left[1, \theta_1^1, \ldots, \theta_N^1, \theta_1^1\theta_2^1, \ldots, \theta_{N-1}^1\theta_N^1, \ldots, \theta_1^1\theta_2^1\cdots\theta_N^1\right]^{\mathrm{T}}. \tag{27}$$

For an N sub-channel co-design, the length of the expansion vector is $|\phi(\boldsymbol{\Theta}^1)| = \sum_{i=0}^{N}\binom{N}{i}$, and we have $|\boldsymbol{\mu}| = \sum_{i=0}^{N}\binom{N}{i}$ in Eq. (26). In this paper, we assume that the BS can only sense M sub-channels in each time slot, i.e. $M < N$. Thus, the length of the expansion vector reduces to $|\phi(\boldsymbol{\Theta}^1)| = \sum_{i=0}^{M}\binom{N}{i}$.

In Algorithm 2, we propose a low-complexity sampling-aided value iteration algorithm for the POMDP formulation considered. In contrast to having discretized continuous-value hypothesized states and then calculating the optimal policy for each hypothesized state as shown in Algorithm 1, Algorithm 2 aims for iteratively optimizing the regression coefficient $\boldsymbol{\mu}$ by sampling a sufficiently large estimated state vector set $\boldsymbol{\Theta}^1$ by relying on the least square (LS) principle.

Then, we can arrive at a near-optimal approximated value function $\tilde{V}(\mathbf{B})$ from the regression coefficient vector $\boldsymbol{\mu}$ received relying on Eq. (26). Hence, for a given hypothesized state $\mathbf{B}$, the expected accumulated reward with respect to the action pair $\mathbf{A} = \{\mathbf{A}_1, A_2\}$ can be calculated as:

$$Q(\mathbf{A}_1, A_2|\mathbf{B}) = r(\mathbf{A}_1, A_2|\mathbf{B}) + \gamma \sum_{\mathbf{B}'\in\mathbb{B}} b(\mathbf{B}'|\mathbf{B}, \mathbf{A}_1)\cdot\tilde{V}(\mathbf{B}'). \tag{28}$$

Thus, we can obtain the near-optimal policy given by:

$$\pi^*(\mathbf{B}) = \arg\max_{\mathbf{A}\in\mathbb{A}} Q(\mathbf{A}_1, A_2|\mathbf{B}). \tag{29}$$

Note that our algorithm can also be extended to the scenario, where the BS needs no prior knowledge concerning to the sub-channels' states.

As for the computational complexity, if we discretize the continuous $\mathbf{B}$ values into discrete Y values in Algorithm 1, we have $|\mathbb{B}| = Y^{2^N}$. The computational complexity of

Algorithm 2: Low-Complexity Sampling-Aided Value Iteration Algorithm

```
generate $X$ estimated state vectors, such as $\boldsymbol{\Theta}^1_{(1)}, \ldots, \boldsymbol{\Theta}^1_{(X)}$ randomly;
calculate corresponding hypothesized states, i.e. $\mathbf{B}_{(1)}, \ldots, \mathbf{B}_{(X)}$ relying on Eq. (33);
initialize $\boldsymbol{\mu} \leftarrow 0$, $\boldsymbol{\mu}' \leftarrow \infty$ and $\overline{V}(\mathbf{B}_{(x)}) \leftarrow 0$ for all $x = 1, \ldots, X$;
while $\max |\boldsymbol{\mu} - \boldsymbol{\mu}'| > \epsilon$ do
    update $\boldsymbol{\mu}' \leftarrow \boldsymbol{\mu}$;
    for $x = 1, \ldots, X$ do
        $\overline{V}(\mathbf{B}_{(x)}) \leftarrow \max_{\mathbf{A} \in \mathbb{A}} \big( r(\mathbf{B}_{(x)}, \mathbf{A}_1, A_2) + \gamma \cdot \sum_{\mathbf{B}' \in \mathbb{B}_X} b(\mathbf{B}' \,|\, \mathbf{B}_{(x)}, \mathbf{A}_1) \cdot \overline{V}(\mathbf{B}') \big)$;
    end
    optimize based on LS principle
    $\boldsymbol{\mu} \leftarrow \arg\min_{\boldsymbol{\mu}} \sum_{x=1}^{X} \big( \boldsymbol{\mu}^{\mathrm{T}} \phi(\boldsymbol{\Theta}^1_{(x)}) - \overline{V}(\mathbf{B}_{(x)}) \big)^2$;
end
return $\boldsymbol{\mu}$;
```

calculating the hypothesized state transition function is on the order of $O\left(Y^{2^N} \cdot Y^{2^N} \cdot \binom{N}{M}\right) = O\left(Y^{2^{(N+1)}} \cdot \binom{N}{M}\right)$. Furthermore, the computational complexity of calculating the value function is $O\left(Y^{2^N} \cdot \binom{N}{M}\right)$ and the look-up table aided policy improvement imposes a computational complexity on the order of $O\left(Y^{2^N}\right)$. If the maximum number of iterations of the algorithm's external loop is set to T, the total computational complexity is given by $O\left(Y^{2^{(N+1)}} \cdot \binom{N}{M} + Y^{2^N} \cdot \binom{N}{M} + T \cdot Y^{2^N} \cdot Y^{2^N}\right) \doteq O\left(Y^{2^N} \cdot \binom{N}{M}\right)$, which exponentially increases with the number of sub-channels N. By contrast, as for the sampling-aided low complexity Algorithm 2, if we sample X hypothesized states, the computational complexity of calculating the hypothesized state transition function and the value function is $O\left(X^2 \cdot \binom{N}{M}\right)$ and $O\left(X \cdot \binom{N}{M}\right)$, respectively. Similarly, the computational complexity of the value improvement of X only entails table-look-up operations. Moreover, the computational complexity of solving the associated LS optimization problem is of order $O\left(|\phi(\boldsymbol{\Theta}^1)|^2 \cdot X\right)$. If the maximum number of iterations of the external loop is still T, we can obtain the total computational complexity of Algorithm 2 as $O\left(X^2 \cdot \binom{N}{M} + X \cdot \binom{N}{M} + T \cdot X \cdot X + T \cdot |\phi(\boldsymbol{\Theta}^1)|^2 \cdot X\right)$, which can be approximately viewed as $O\left(X^2 \cdot \binom{N}{M}\right)$ and is not related to the number of sub-channels N. Therefore, we can conclude that the sampling-aided POMDP solution algorithm substantially reduce the computational complexity in comparison to the original algorithm.

V. Simulation Results

In our simulations, we assume that the radar and communication co-design contains five sub-channels, i.e. $N = 5$. Moreover, all the five sub-channels have the same initial utilization rate of $p^1_i = 40\%$. The bandwidth of each sub-channel is $W_i = 10$ MHz in conjunction with $N_i = 5 \times 10^{-7}$ W/Hz as well as $N_0 = 1 \times 10^{-7}$ W/Hz. Let the maximum transmission power of SN be $P^u_{\max} = 2$ W for underlay access scheme, while $P^o_{\max} = 20$ W for the overlay access scheme. Furthermore, without any loss of generality, we set the power gain to $g_{sr} = g_{pr} = g_{sp} = 1$. Let the weighting coefficients $\lambda_C = 1.15 \times 10^{-7}$ (b/s)$^{-1}$ and $\lambda_I = 5$. Moreover, the discount factor is $\gamma = 0.8$. Hence, according to Eq. (4), if the SU accesses an idle sub-channel of the underlay scheme associated with a SN transmission power of $P^u_{\max}$, its reward will be 2, while its reward will be 5 in terms of the overlay scheme with transmission power $P^o_{\max}$. By contrast, upon accessing a busy sub-channel, its reward is set to 0.5 and -15 for the underlay scheme and the overlay scheme having the maximum transmission power, respectively.

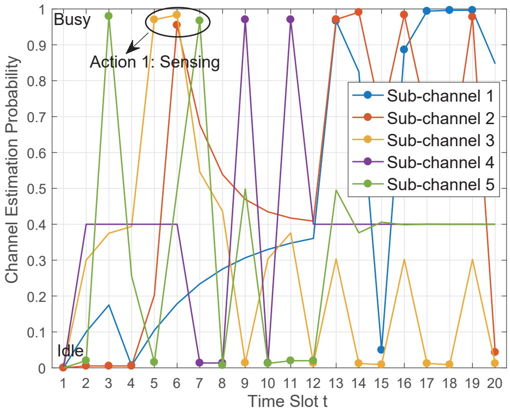


Fig. 5. Channel state estimation probability for the SU's first-step action decision during the first 20 time slots ($N = 5$, $M = 2$, $L = 2$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $\beta = [10\%, 20\%, 30\%, 40\%, 50\%]$, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

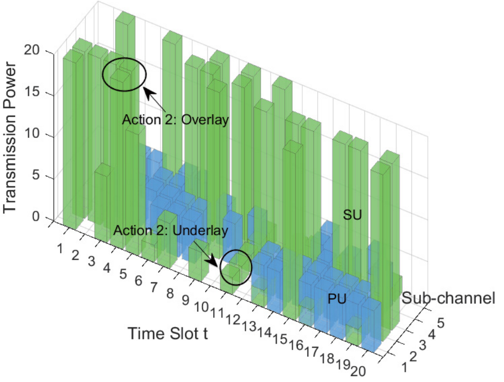


Fig. 6. Transmission power of the SU for its second-step action decision during the first 20 time slots ($N = 5$, $M = 2$, $L = 2$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $\beta = [10\%, 20\%, 30\%, 40\%, 50\%]$, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

First of all, we verify the feasibility of our proposed network association mechanism by conducting a numerical simulation spanning over 100 time slots relying on Algorithm 2. Specifically, for example, we assume that the SU is capable of sensing and accessing two channels in each time slot, and we have $M = 2$ as well as $L = 2$. Moreover, the subchannels' busy-to-idle transition probability is $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, respectively, while their idle-to-busy transition probability is $\beta = [10\%, 20\%, 30\%, 40\%, 50\%]$, respectively. In this context, we assume the SU's false-alarm rate to be $\zeta_f = 2\%$ and its miss-detection rate to be $\zeta_m = 2\%$, respectively. Here, we generate $X = 5000$ samples to optimize the coefficient μ in Algorithm 2.

Fig. 5 shows the result of the channel state estimation probability as well as the SU's first-step action decision during the first 20 time slots, where the dot mark represents the sub-channel being chosen for sensing by the SU at that time slot. The related channel estimation probability value represents the relative frequency of the sub-channel's temporal state. In other words, a high value of the channel state estimation probability indicates that the sub-channel is likely to be in the busy state, while a close-to-zero value suggests that the sub-channel is likely to be idle. We may conclude that the SU preferably chooses the specific sub-channels for sensing whose temporal states are estimated to be either busy or idle with a high confidence. Fig. 6 shows the result of the final transmission power of the SU as well as its second-step action decision during the first 20 time slots. As for the underlay access scheme, the SU accesses the whole channel shared with the PUs at an identical but low transmission power, while the SU can only access $L = 2$ sub-channels, when it selects the overlay access scheme. Our simulation results have verified the feasibility of our proposed POMDP scheme in cooperative channel sensing and access decision making in the context of our radar and communication co-design.

In the following, we carry out the performance analysis of our proposed POMDP scheme in comparison to the idealized optimal strategy having perfect knowledge of all the present channel states (termed as, Full info). For convenience, let us define the channel's occupancy rate p_i^{on} of sub-channel i as $\beta_i = \frac{p_i^{on}}{1-p_i^{on}}\alpha_i$. Furthermore, a pair of benchmark schemes are proposed for evaluating the performance of the POMDP algorithm. Specifically, SU's achievable rate Λ is given by summing the achievable rate of all the sub-channels that the SU has accessed, i.e. $\Lambda = \sum_{n=1}^{N} I(n) \log[1 + P_{SU}^n/(P_{PU}^n + N_0 W_n)]$ bits/s/Hz, where $I(n) = 1$ if the SU accesses sub-channel n; otherwise $I(n) = 0$. As for evaluating the influence of the SU on the PU, the SNR degradation is defined as $\Delta SNR = \Upsilon N_0 W_i/[\sum_{n=1}^{N} \delta(n) P_{SU}^n + \Upsilon N_0 W_i]$, where Υ represents the number of the sub-channels occupied by the PU, while W_i is the corresponding bandwidth of the sub-channel considered. Moreover, $\delta(n) = 1$ represents that the sub-channel n is also occupied by the PU, otherwise $\delta(n) = 0$.

Fig. 7 and Fig. 8 show the SU's achievable rate and the PU's SNR degradation versus the channel's occupancy rate p_i^{on} parameterized by the number of sub-channels being sensed and accessed for both the proposed POMDP and for the idealized full-information based algorithms. The busy-to-idle transition probability is $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$. We can conclude that the SU's achievable rate decreases upon increasing the channel's occupancy rate both in the context of POMDP and of the full-information based algorithm, since the SU has to select a more conservative access strategy, when the channel becomes busy. Moreover, the SNR degradation imposed on the PU becomes more severe in busy channel conditions. It is noted that when $M = L = 1$, the SU has a higher probability of opting for the underlay scheme than for the overlay scheme, because less CSI information is acquired, which results in the highest SNR degradation inflicted upon the PU and the lowest achievable rate for the SU.

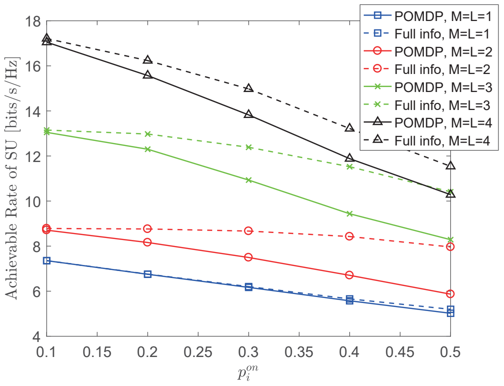


Fig. 7. The SU's achievable rate versus the channel's occupancy rate parameterized by the number of sub-channels being sensed and accessed for both proposed POMDP and full information algorithms ($N = 5$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P_{\max}^u = 2$ W, $P_{\max}^o = 20$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

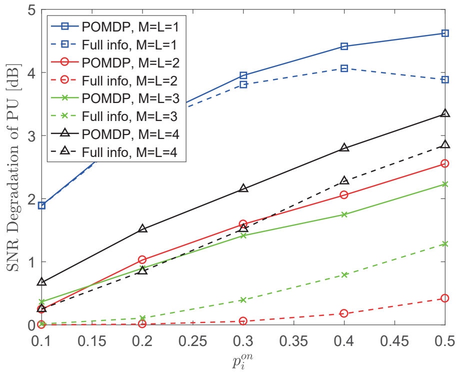


Fig. 8. The SNR degradation imposed on PU versus the channel's occupancy rate parameterized by the number of sub-channels being sensed and accessed for both proposed POMDP and full information algorithms ($N = 5$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P_{\max}^u = 2$ W, $P_{\max}^o = 20$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

Fig. 9 and Fig. 10 evaluate the impact of the number of sub-channels that are being sensed imposed both on the SU's achievable rate and on the PU's SNR degradation versus the channel's occupancy rate p_i^{on} both in the context of our proposed POMDP and for the idealized full-CSI based algorithms. The busy-to-idle transition probability is $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$. Moreover, we fix the number of sub-channels that are being accessed to $L = 3$. Since the idealized full-CSI based algorithm exploits the perfect CSI for its final decision-making, the performance of both the SU

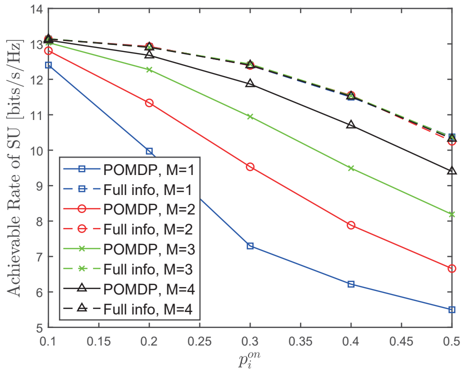


Fig. 9. The SU's achievable rate versus the channel's occupancy rate parameterized by the number of sub-channels being sensed for both proposed POMDP and full information algorithms ($N = 5$, $L = 3$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $P^o_{\max} = 20$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

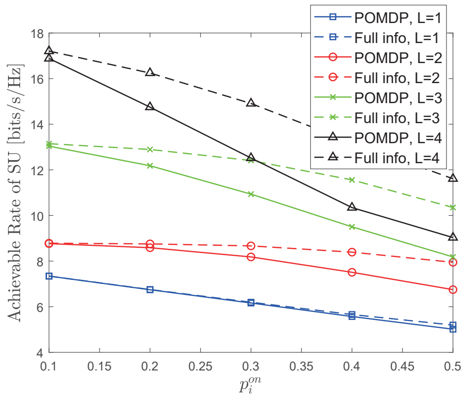


Fig. 11. The SU's achievable rate versus the channel's occupancy rate parameterized by the number of sub-channels being accessed for both proposed POMDP and full information algorithms ($N = 5$, $M = 3$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $P^o_{\max} = 20$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

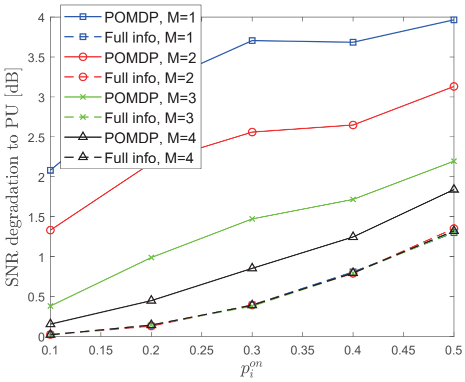


Fig. 10. The SNR degradation imposed on PU versus the channel's occupancy rate parameterized by the number of sub-channels being sensed for both proposed POMDP and full information algorithms ($N = 5$, $L = 3$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $P^o_{\max} = 20$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

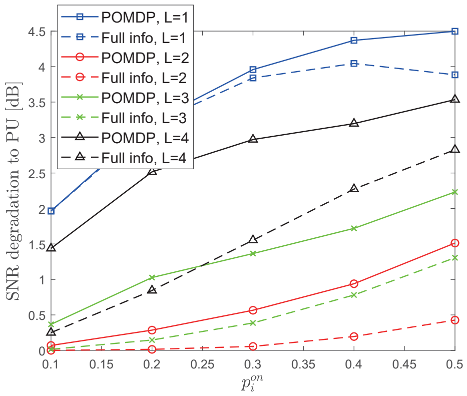


Fig. 12. The SNR degradation imposed on PU versus the channel's occupancy rate parameterized by the number of sub-channels being accessed for both proposed POMDP and full information algorithms ($N = 5$, $M = 3$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $P^o_{\max} = 20$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

and of the PU remains the same, regardless of the number of sub-channels being sensed. Furthermore, the SU's achievable rate can be substantially improved by increasing the number of sub-channels being sensed, while the PU's SNR degradation is actually reduced upon increasing the number of sub-channels being sensed. This trend prevails because if perfect CSI is used in support of the access-related decision-making, the sub-channels can be more efficiently shared without precipitating avalanche-like collision.

Fig. 11 and Fig. 12 show the impact of the number of sub-channels that are being accessed imposed on both the SU's achievable rate and on the PU's SNR degradation versus the channel's occupancy rate p_i^{on} both for our proposed POMDP and for the idealized full-CSI based algorithms. Similarly, the busy-to-idle transition probability is $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, and we fix the number of sub-channels being sensed to $M = 3$. We may conclude that with less sub-channels being accessed, the SU's achievable rate and the PU's SNR degradation associated approach that of the optimal solution relying on decision-making associated with full CSI, especially in the context of a low channel occupancy rate. When the number of sub-channels being accessed is higher than that of the sub-channels being sensed, the performance gap between the pair of algorithms considered is increased, because the SUs have to explore unknown sub-channels for their trial-and-error based access strategy in each decision-making round.

Fig. 13 and Fig. 14 evaluate the impact of weighting coefficients imposed on both the SU's achievable rate and the PU's SNR degradation versus the channel's occupancy rate p_i^{on}. Without loss of generality, here we focus our attention on the penalty weighting coefficient λ_I, for example. The busy-to-idle transition probability is $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, and the number of sub-channels being sensed and accessed is $M = L = 3$. Since the idealized full-information based algorithm does not rely on the reward-and-penalty incentive mechanism, it has the same SU achievable rate and the same PU SNR degradation, regardless of the value of λ_I. We can see that the weighting coefficients have different influence on SU's and PU's performance. Since λ_I is the penalty weighting coefficient, a large λ_I can beneficially reduce the PU's SNR degradation, gradually approaching the

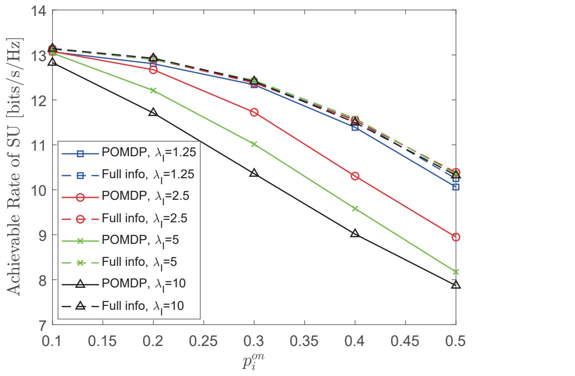


Fig. 13. The SU's achievable rate versus the channel's occupancy rate parameterized by the weighting coefficient λ_I for both proposed POMDP and full information algorithms ($N = 5$, $M = L = 3$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $P^o_{\max} = 20$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

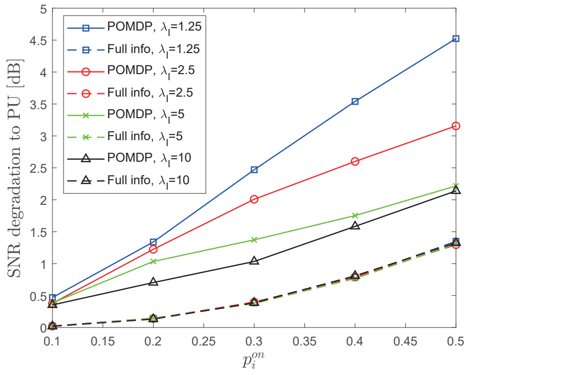


Fig. 14. The SNR degradation imposed on PU versus the channel's occupancy rate parameterized by the weighting coefficient λ_I for both proposed POMDP and full information algorithms ($N = 5$, $M = L = 3$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $P^o_{\max} = 20$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

optimal lower bound, while a small λ_I is capable of yielding a near-optimal SU rate. Hence, in oure cognitive radar and communication co-design, we should appropriately choose the values of weighting coefficients according to the particular specifications of the system. Specially, if we want to reduce the SNR degradation imposed by the SUs on the PU, we can choose a large penalty weighting coefficient. By contrast, if we want to maximize the achievable rate of SUs, we may increase the value of the reward weighting coefficient.

Fig. 15 to Fig. 18 show the influence of both the maximum tolerable transmission power $P^o_{\max}$ of the overlay mode as well as of the maximum tolerable transmission power $P^u_{\max}$ of the underlay mode imposed on the SU's achievable rate and the PU's SNR degradation versus the channel's occupancy rate. In this context, let $M = L = 3$, for example, and the busy-to-idle transition probability be $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$. We can conclude from Fig. 15 and Fig. 16 that a high $P^o_{\max}$ results in a high achievable rate for the SU, while the SNR degradation of the PU may not increase upon increasing $P^o_{\max}$, because the SU is capable of gaining access at the maximum tolerable transmission power of the underlay mode, when it detects a conflict with the PU.

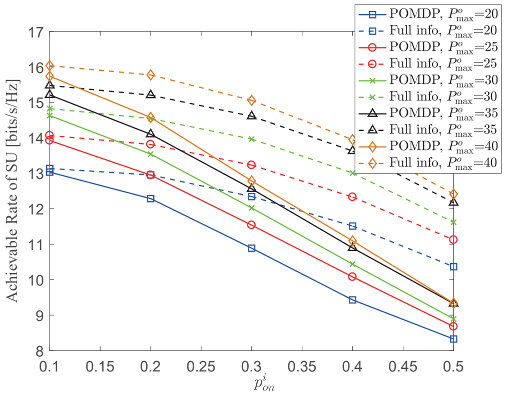


Fig. 15. The SU's achievable rate versus the channel's occupancy rate parameterized by the maximum tolerable transmission power $P^o_{\max}$ of the overlay mode in the context of $M = L = 3$, for example ($N = 5$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

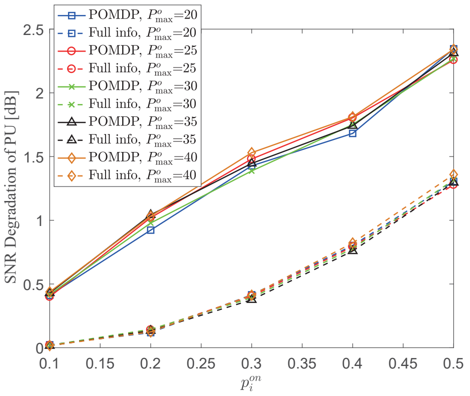


Fig. 16. The PU's SNR degradation versus the channel's occupancy rate parameterized by the maximum tolerable transmission power $P^o_{\max}$ of the overlay mode in the context of $M = L = 3$, for example ($N = 5$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

By contrast, as shown in Fig. 17 and Fig. 18, both the SU's achievable rate and the PU's SNR degradation are improved upon increasing the maximum tolerable transmission power $P^u_{\max}$ of the underlay mode. It is also noted that when the channel is less occupied, $P^o_{\max}$ plays a critical part in improving the SU's achievable rate, since the SU is more likely to opt for the overlay access scheme, while $P^u_{\max}$ becomes the dominant factor, when the channel is busy.

Fig. 19 to Fig. 22 highlight the influence of both the false-alarm rate ζ_f and the missed-detection rate ζ_m imposed on the SU's achievable rate and the PU's SNR degradation versus the channel's occupancy rate. Similarly, let us consider $M = L = 3$ for example, and again the busy-to-idle transition probability of $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$. Given the perfect full CSI of the optimal algorithm, its performance does not change with ζ_f and ζ_m. As for the proposed POMDP algorithm, it is plausible that a large value of ζ_f and ζ_m reduces the SU's achievable rate and simultaneously degrades the SNR of the PU.

$$\begin{aligned} & b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}_1) = \Pr\left(\mathbf{B}(t) = \mathbf{B}' \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\right) \\ &= \sum_{\mathbf{O}\in\mathbb{O}} \Big(\Pr(\mathbf{B}(t) = \mathbf{B}' \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1, \mathbf{O}(t) = \mathbf{O}) \cdot \Pr\left(\mathbf{O}(t) = \mathbf{O} \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\right)\Big). \end{aligned} \tag{30}$$

$$\begin{aligned} & \Pr\big(\mathbf{O}(t) = \mathbf{O} \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big) \\ &= \sum_{\mathbf{S}\in\mathbb{S}} \Big(\Pr\big(\mathbf{O}(t) = \mathbf{O} \mid \mathbf{S}(t) = \mathbf{S}, \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big) \cdot \Pr\big(\mathbf{S}(t) = \mathbf{S} \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big)\Big) \\ &= \sum_{\mathbf{S}\in\mathbb{S}} \Bigg(\Pr\big(\mathbf{O}(t) = \mathbf{O} \mid \mathbf{S}(t) = \mathbf{S}, \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big) \\ &\quad \times \sum_{\mathbf{S}'\in\mathbb{S}} \Big(\Pr\big(\mathbf{S}(t) = \mathbf{S} \mid \mathbf{S}'(t-1) = \mathbf{S}', \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big) \cdot \Pr\big(\mathbf{S}'(t-1) = \mathbf{S}' \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big)\Big)\Bigg) \\ &= \sum_{\mathbf{S}\in\mathbb{S}} \Big(\Pr\big(\mathbf{O}(t) = \mathbf{O} \mid \mathbf{S}(t) = \mathbf{S}, \mathbf{A}_1(t) = \mathbf{A}_1\big) \sum_{\mathbf{S}'\in\mathbb{S}} p(\mathbf{S} \mid \mathbf{S}') \cdot B_{\mathbf{S}'}\Big) \\ &= \sum_{\mathbf{S}\in\mathbb{S}} \Big(z(\mathbf{O} \mid \mathbf{S}, \mathbf{A}_1) \cdot \sum_{\mathbf{S}'\in\mathbb{S}} p(\mathbf{S} \mid \mathbf{S}') \cdot B_{\mathbf{S}'}\Big). \end{aligned} \tag{31}$$

$$\begin{aligned} & B'_{\mathbf{S}}(t) = \Pr\big(\mathbf{S}(t) = \mathbf{S} \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1, \mathbf{O}(t) = \mathbf{O}\big) = \frac{\Pr\big(\mathbf{S}(t) = \mathbf{S}, \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1, \mathbf{O}(t) = \mathbf{O}\big)}{\Pr\big(\mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1, \mathbf{O}(t) = \mathbf{O}\big)} \\ &= \frac{\Pr\big(\mathbf{S}(t) = \mathbf{S}, \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1, \mathbf{O}(t) = \mathbf{O}\big)}{\Pr\big(\mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1, \mathbf{S}(t) = \mathbf{S}\big)} \cdot \frac{\Pr\big(\mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1, \mathbf{S}(t) = \mathbf{S}\big)}{\Pr\big(\mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big)} \\ &\quad \times \frac{\Pr\big(\mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big)}{\Pr\big(\mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1, \mathbf{O}(t) = \mathbf{O}\big)} \\ &= \frac{\Pr\big(\mathbf{O}(t) = \mathbf{O} \mid \mathbf{S}(t) = \mathbf{S}, \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big)}{\Pr\big(\mathbf{O}(t) = \mathbf{O} \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big)} \cdot \Pr\big(\mathbf{S}(t) = \mathbf{S} \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big). \end{aligned} \tag{32}$$

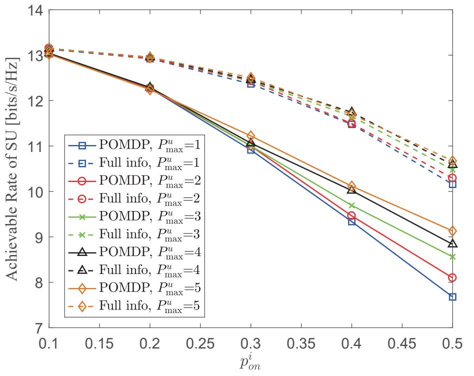


Fig. 17. The SU's achievable rate versus the channel's occupancy rate parameterized by the maximum tolerable transmission power $P^u_{\max}$ of the underlay mode in the context of $M = L = 3$, for example ($N = 5$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^o_{\max} = 20$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

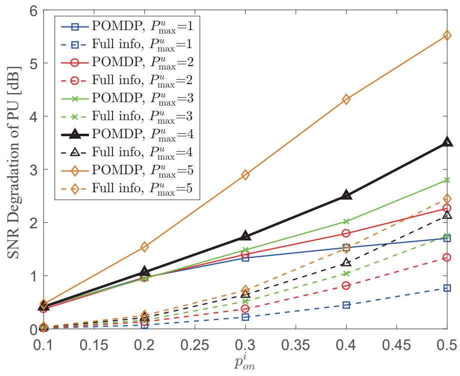


Fig. 18. The PU's SNR degradation versus the channel's occupancy rate parameterized by the maximum tolerable transmission power $P^u_{\max}$ of the underlay mode in the context of $M = L = 3$, for example ($N = 5$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^o_{\max} = 20$ W, $\zeta_f = 2\%$ and $\zeta_m = 2\%$).

VI. Conclusions

In this paper, we have constructed a learning assisted network association mechanism for communication and radar co-design. Firstly, we formulated the co-design as a POMDP problem and provided its solution, demonstrating that it is suitable for the scarce spectral resources even in partially

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2019.2933778, IEEE Journal on Selected Areas in Communications

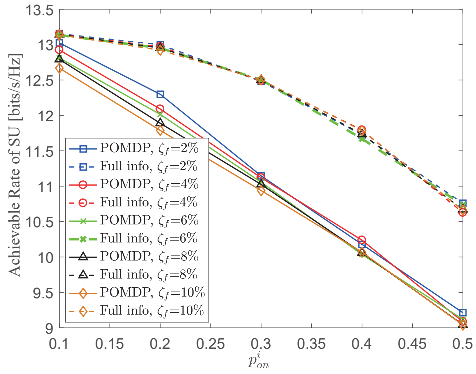


Fig. 19. The SU's achievable rate versus the channel's occupancy rate parameterized by the false-alarm rate ζ_f in the context of $M = L = 3$, for example ($N = 5$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $P^o_{\max} = 20$ W and $\zeta_m = 2\%$).

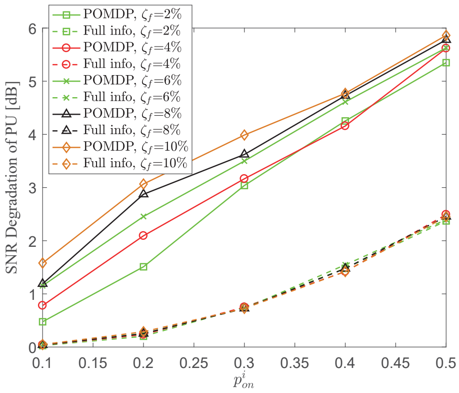


Fig. 20. The PU's SNR degradation versus the channel's occupancy rate parameterized by the false-alarm rate ζ_f in the context of $M = L = 3$, for example ($N = 5$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $P^o_{\max} = 20$ W and $\zeta_m = 2\%$).

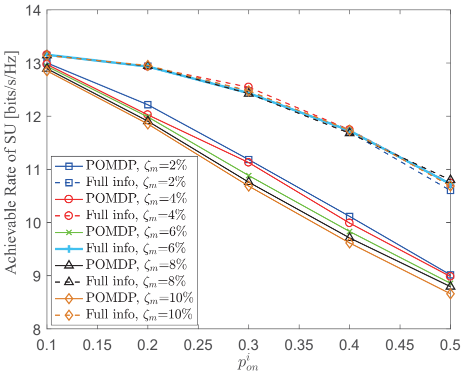


Fig. 21. The SU's achievable rate versus the channel's occupancy rate parameterized by the missed-detection rate ζ_m in the context of $M = L = 3$, for example ($N = 5$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $P^o_{\max} = 20$ W and $\zeta_f = 2\%$).

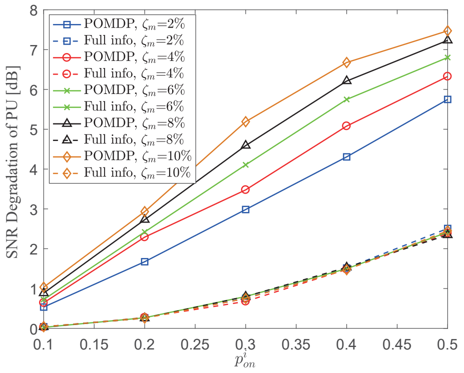


Fig. 22. The PU's SNR degradation versus the channel's occupancy rate parameterized by the missed-detection rate ζ_m in the context of $M = L = 3$, for example ($N = 5$, $\alpha = [15\%, 30\%, 45\%, 60\%, 75\%]$, $P^u_{\max} = 2$ W, $P^o_{\max} = 20$ W and $\zeta_f = 2\%$).

observed CSI scenarios. Moreover, we conceived a low-complexity algorithm for finding a beneficial near-optimal policy. Finally, our simulation results demonstrated that the proposed POMDP algorithm improved the achievable rate of the SU as well as the SNR of the PU even in comparison to the full-information based algorithm by relying on beneficially designing the number of sub-channels sensed and accessed, as well as by adjusting the maximum tolerable transmission power of both access modes.

Appendix

The derivation of SN's hypothesis transition function $b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}_1)$: Relying on the law of total probability, Eq. (13) can be rewritten as Eq. (30). In a similar way, upon relying on Eq. (11) and Eq. (12), we obtain Eq. (31). Note that vector $\mathbf{S}'$ in Eq. (31) represents the system's state at time slot $(t-1)$.

Furthermore, based on the hypothesized state vector $\mathbf{B}(t-1)$ at time slot $(t-1)$ and on the observation state vector $\mathbf{O}(t)$ at time slot t yielded by the sensing stage action $\mathbf{A}_1(t)$, we become capable of updating the hypothesized state vector $\mathbf{B}'(t)$. Specifically, without loss of generality, one of the 2^N elements of $\mathbf{B}'(t)$, considering $B'_{\mathbf{S}}$ for example, can be calculated as Eq. (32).

Relying on the intermediate results of Eq. (31), $B'_{\mathbf{S}}(t)$ can be further reformulated as Eq. (33), where $\mathbf{S}$ and $\mathbf{S}''$ independently denote the system's state at time slot t, while $\mathbf{S}'$ represents that at time slot $(t-1)$. Moreover, Eq. (33) underlines the update of the estimated state vector of $\mathbf{\Theta}^{\mathbf{S}}(t)$. Then, we arrive at:

$$\Pr\big(\mathbf{B}(t) = \mathbf{B}' \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1, \mathbf{O}(t) = \mathbf{O}\big) \\ = I\left\{\mathbf{B}' = [B'_{\mathbf{S}_1}, B'_{\mathbf{S}_2}, \cdots, B'_{\mathbf{S}_{2^N}}]\right\}, \tag{34}$$

where $I\{\cdot\}$ represents an indicator function, while $B'_{\mathbf{S}_1}, B'_{\mathbf{S}_2}, \cdots, B'_{\mathbf{S}_{2N}}$ can be calculated from Eq. (33). Hence, the SN's hypothesis transition function of Eq. (13)

$$
\begin{aligned}
B'_{\mathbf{S}}(t) &= \frac{z(\mathbf{O} \mid \mathbf{S}, \mathbf{A}_1) \cdot \sum_{\mathbf{S}' \in \mathbb{S}} p(\mathbf{S} \mid \mathbf{S}') \cdot B_{\mathbf{S}'}}{\Pr\big(\mathbf{O}(t) = \mathbf{O} \mid \mathbf{B}(t-1) = \mathbf{B}, \mathbf{A}_1(t) = \mathbf{A}_1\big)} = \frac{z(\mathbf{O} \mid \mathbf{S}, \mathbf{A}_1) \cdot \sum_{\mathbf{S}' \in \mathbb{S}} p(\mathbf{S} \mid \mathbf{S}') \cdot B_{\mathbf{S}'}}{\sum_{\mathbf{S}'' \in \mathbb{S}}\big(z(\mathbf{O} \mid \mathbf{S}'', \mathbf{A}_1) \cdot \sum_{\mathbf{S}' \in \mathbb{S}} p(\mathbf{S}'' \mid \mathbf{S}') \cdot B_{\mathbf{S}'}\big)} \\
&= \prod_{i=1}^{N} \theta_i^{s_i}(t) = \prod_{i=1}^{N} \frac{z_i(o_i \mid s_i, a_i^1) \cdot \sum_{s'_i \in \{0,1\}} p(s_i \mid s'_i) \cdot \theta_i^{s'_i}(t-1)}{\sum_{s''_i \in \{0,1\}} \Big(z_i(o_i \mid s''_i, a_i^1) \cdot \sum_{s'_i \in \{0,1\}} p(s''_i \mid s'_i) \cdot \theta_i^{s'_i}(t-1)\Big)} \\
&= \prod_{i=1}^{N} \frac{z_i(o_i \mid s_i, a_i^1) \cdot \Big(p\big(s_i \mid s'_i = 1\big) \cdot \theta_i^1(t-1) + p\big(s_i \mid s'_i = 0\big) \cdot \theta_i^0(t-1)\Big)}{\sum_{s''_i \in \{0,1\}} \Big(z_i(o_i \mid s''_i, a_i^1) \cdot \Big(p\big(s''_i \mid s'_i = 1\big) \cdot \theta_i^1(t-1) + p\big(s''_i \mid s'_i = 0\big) \cdot \theta_i^0(t-1)\Big)\Big)}.
\end{aligned} \tag{33}
$$

can be expressed as:

$$
\begin{aligned}
b(\mathbf{B}' \mid \mathbf{B}, \mathbf{A}_1) = \sum_{\mathbf{O} \in \mathbb{O}} \bigg(I\Big\{\mathbf{B}' = [B'_{\mathbf{S}_1}, B'_{\mathbf{S}_2}, \cdots, B'_{\mathbf{S}_{2^N}}]\Big\} \\
\cdot \sum_{\mathbf{S} \in \mathbb{S}} \Big(z(\mathbf{O} \mid \mathbf{S}, \mathbf{A}_1) \cdot \sum_{\mathbf{S}' \in \mathbb{S}} p(\mathbf{S} \mid \mathbf{S}') \cdot B_{\mathbf{S}'} \Big) \bigg).
\end{aligned} \tag{35}
$$

Relying on the observation function seen in Eq. (8) and on the system's Markov transition probability shown in Fig. 2, the numerator of the fraction in Eq. (33) can be expressed as:

$$
\begin{aligned}
&z_i(o_i \mid s_i, a_i^1) \cdot \sum_{s'_i \in \{0,1\}} p\big(s_i \mid s'_i\big) \cdot \theta_i^{s'_i}(t-1) \\
&= \begin{cases}
(1-\zeta_m) \cdot \vartheta_i^1, & \text{if } o_i = 1, s_i = 1, a_i^1 = 1, \\
\zeta_f \cdot \vartheta_i^0, & \text{if } o_i = 1, s_i = 0, a_i^1 = 1, \\
\zeta_m \cdot \vartheta_i^1, & \text{if } o_i = 0, s_i = 1, a_i^1 = 1, \\
(1-\zeta_f) \cdot \vartheta_i^0, & \text{if } o_i = 0, s_i = 0, a_i^1 = 1, \\
\vartheta_i^1, & \text{if } o_i = \phi, s_i = 1, a_i^1 = 0, \\
\vartheta_i^0, & \text{if } o_i = \phi, s_i = 0, a_i^1 = 0, \\
0, & \text{otherwise},
\end{cases}
\end{aligned} \tag{36}
$$

where we have

$$
\vartheta_i^1 = (1-\alpha_i) \cdot \theta_i^1(t-1) + \beta_i \cdot \theta_i^0(t-1), \tag{37}
$$

as well as

$$
\vartheta_i^0 = \alpha_i \cdot \theta_i^1(t-1) + (1-\beta_i) \cdot \theta_i^0(t-1). \tag{38}
$$

References

[1] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.

[2] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 136–144, Jun. 2014.

[3] H. Zhang, X. Chu, W. Guo, and S. Wang, "Coexistence of Wi-Fi and heterogeneous small cell networks sharing unlicensed spectrum," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 158–164, Mar. 2015.

[4] K. Zhu, E. Hossain, and D. Niyato, "Pricing, spectrum sharing, and service selection in two-tier small cell networks: A hierarchical dynamic game approach," *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1843–1856, Jul. 2014.

[5] Y. Liu and L. Dong, "Spectrum sharing in MIMO cognitive radio networks based on cooperative game theory," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 4807–4820, 2014.

[6] C. Yi and J. Cai, "Two-stage spectrum sharing with combinatorial auction and Stackelberg game in recall-based cognitive radio networks," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3740–3752, Oct. 2014.

[7] H. Hayvaci and B. Tavli, "Spectrum sharing in radar and wireless communication systems: A review," in *IEEE International Conference on Electromagnetics in Advanced Applications (ICEAA)*, Palm Beach, Netherlands Antilles, Sep. 2014, pp. 810–813.

[8] A. Turlapaty and Y. Jin, "A joint design of transmit waveforms for radar and communications systems in coexistence," in *IEEE Radar Conference*, Cincinnati, OH, Aug. 2014, pp. 0315–0319.

[9] B. Li, H. Kumar, and A. P. Petropulu, "A joint design approach for spectrum sharing between radar and communication systems," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3306–3310.

[10] Z. Geng, H. Deng, and B. Himed, "Adaptive radar beamforming for interference mitigation in radar-wireless spectrum sharing," *IEEE Signal Processing Letters*, vol. 22, no. 4, pp. 484–488, Otc. 2015.

[11] J. Wang, S. Guan, C. Jiang, H. Zhang, Y. Ren, and L. Hanzo, "Network association for cognitive communication and radar co-systems: A POMDP formulation," in *IEEE International Communication Conference (ICC)*, Kansas City, MO, May 2018.

[12] H. Deng and B. Himed, "Interference mitigation processing for spectrum-sharing between radar and wireless communications systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 3, pp. 1911–1919, Jul. 2013.

[13] F. Hessar and S. Roy, "Spectrum sharing between a surveillance radar and secondary Wi-Fi networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 3, pp. 1434–1448, Jul. 2016.

[14] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. Cambridge: MIT press, 1998.

[15] J. Wang, C. Jiang, Z. Han, Y. Ren, and L. Hanzo, "Network association strategies for an energy harvesting aided super-WiFi network relying on measured solar activity," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3785–3797, Dec. 2016.

[16] M. Hirzallah, W. Afifi, and M. Krunz, "Full-duplex-based rate/mode adaptation strategies for Wi-Fi/LTE-U coexistence: A POMDP approach," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 1, pp. 20–29, Nov. 2017.

[17] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[18] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2053–2071, Apr. 2008.

[19] A. T. Hoang, Y. C. Liang, D. T. C. Wong, Y. Zeng, and R. Zhang, "Opportunistic spectrum access for energy-constrained cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1206–1211, Mar. 2009.

[20] L. Zheng, M. Lops, X. Wang, and E. Grossi, "Joint design of overlaid communication systems and pulsed radars," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 139–154, 2018.

[21] S. Senthuran, A. Anpalagan, and O. Das, "Throughput analysis of opportunistic access strategies in hybrid underlay-overlay cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2024–2035, Apr. 2012.

[22] J. Zou, H. Xiong, D. Wang, and C. W. Chen, "Optimal power allocation for hybrid overlay/underlay spectrum sharing in multiband cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1827–1837, Dec. 2013.

[23] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Radar detection of distributed targets in homogeneous interference whose inverse

covariance structure is defined via unitary invariant functions," *IEEE Transactions on Signal Processing*, vol. 61, no. 20, pp. 4949–4961, Oct. 2013.

[24] A. Al-Hourani, R. J. Evans, S. Kandeepan, B. Moran, and H. Eltom, "Stochastic geometry methods for modeling automotive radar interference," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 333–344, Feb. 2018.

[25] L. Liu, F. Zhou, M. Tao, P. Sun, and Z. Zhang, "Adaptive translational motion compensation method for ISAR imaging under low SNR based on particle swarm optimization," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 11, pp. 5146–5157, Nov. 2015.

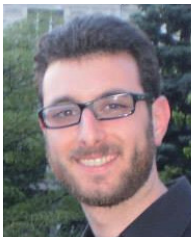
[26] R. Bellman, "On a routing problem," *Quarterly of Applied Mathematics*, vol. 16, no. 1, pp. 87–90, 1958.

[27] E. J. Sondik, "The optimal control of partially observable markov processes over the infinite horizon: Discounted costs," *Operations Research*, vol. 26, no. 2, pp. 282–304, Mar. 1978.

Jingjing Wang (S'14-M'19) received his B.S. degree in Electronic Information Engineering from Dalian University of Technology in 2014 with the highest honor. He currently works for his PhD degree in Department of Electronic Engineering, Tsinghua University, Beijing. From 2017 to 2018, he has been a joint PhD student in Next Generation Wireless Group chaired by Prof. Lajos Hanzo, University of Southampton, UK. His research interests include the resource allocation and network association, learning theory aided modeling, analysis and signal processing, as well as information diffusion theory for mobile wireless networks. He received China Postgraduate National Scholarship Award in 2017, Best Journal Paper Award of IEEE Technical Committee on Green Communications & Computing in 2018.

Sanghai Guan received the B.Eng. degree in electronic engineering from Dalian University of Technology, Dalian, Liaoning, China, in 2017, with the honor of excellent graduate of Liaoning Province. He is now pursuing the M.S. degree in information and electronic engineering at Tsinghua University, Beijing, China. His research interests include complex networks and systems, multi-agent networked systems, and network association.

Chunxiao Jiang (S'09-M'13-SM'15) received the B.S. degree from Beihang University, Beijing in 2008 and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing in 2013, both with the highest honors. From 2011 to 2012, he visited University of Maryland, College Park as a joint PhD supported by China Scholarship Council. From 2013 to 2016, he was a postdoc with Tsinghua University, during which he visited University of Maryland, College Park and University of Southampton. Since July 2016, he became an assistant professor in Tsinghua Space Center, Tsinghua University. His research interests include space networks and heterogeneous networks. Dr. Jiang is the recipient of the Best Paper Award from IEEE GLOBECOM in 2013, the Best Student Paper Award from IEEE GlobalSIP in 2015, IEEE Communications Society Young Author Best Paper Award in 2017, the Best Paper Award IWCMC in 2017, the Best Journal Paper Award of IEEE ComSoc Technical Committee on Communications Systems Integration and Modeling in 2018.



Dimitrios Alanis (S'13) received the M.Eng. degree in electrical and computer engineering from the Aristotle University of Thessaloniki in 2011 and the M.Sc. and Ph.D. degrees from the University of Southampton in 2012 and 2017, respectively. During 2017 and 2018, he was a Research Fellow in the Next Generation Wireless (NGW) group, School of Electronics and Computer Science, University of Southampton. He is currently working as an Algorithm Engineer for Viavi Solutions, Wireless.

His research interests include quantum computation and quantum information theory, quantum search algorithms, cooperative communications, resource allocation for self-organizing networks, bio-inspired optimization algorithms and classical and quantum game theory and machine learning.

Yong Ren (SM'16) received his B.S, M.S and Ph.D. degrees in electronic engineering from Harbin Institute of Technology, China, in 1984, 1987, and 1994, respectively. He worked as a post doctor at Department of Electronics Engineering, Tsinghua University, China from 1995 to 1997. Now he is a professor of Department of Electronics Engineering and the director of the Complexity Engineered Systems Lab in Tsinghua University. He holds 12 patents, and has authored or co-authored more than 100 technical papers in the behavior of computer network, P2P network and cognitive networks. He has serves as a reviewer of IEICE Transactions on Communications, Digital Signal Processing, Chinese Physics Letters, Chinese Journal of Electronics, Chinese Journal of Computer Science and Technology, Chinese Journal of Aeronautics and so on. His current research interests include complex systems theory and its applications to the optimization and information sharing of the Internet, Internet of Things and ubiquitous network, cognitive networks and Cyber-Physical Systems.

Lajos Hanzo (http://www-mobile.ecs.soton.ac.uk) FREng, F'04, FIET, Fellow of EURASIP, received his 5-year degree in electronics in 1976 and his doctorate in 1983 from the Technical University of Budapest. In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 119 PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1800+ research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE ComSoc and VTS. He is a former Editor-in-Chief of the IEEE Press and a former Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk