AFRL-RV-PS-
TR-2017-0132

COGNITIVE COMMUNICATIONS PROTOCOLS FOR SATCOM

Sudharman Jayaweera

Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131

20 Oct 2017

Final Report

AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
3550 Aberdeen Ave SE
AIR FORCE MATERIEL COMMAND
KIRTLAND AIR FORCE BASE, NM 87117-5776

DTIC COPY

NOTICE AND SIGNATURE PAGE

This report is the result of contracted fundamental research which is exempt from public affairs security and policy review in accordance with AFI 61-201, paragraph 2.3.5.1. This report is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RV-PS-TR-2017-0132 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//

JAMES LYKE
Program Manager

//SIGNED//

DAVID CARDIMONA
Technical Advisor, Space Based Advanced Sensing and Protection

//SIGNED//

JOHN BEAUCHEMIN
Chief Engineer, Spacecraft Technology Division
Space Vehicles Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
20-10-2017	Final Report	26 Apr 2016 to 26 Oct 2017

4. TITLE AND SUBTITLE

Cognitive Communications Protocols for SATCOM

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9453-16-1-0052

5c. PROGRAM ELEMENT NUMBER
62601F

6. AUTHOR(S)

Sudharman Jayaweera

5d. PROJECT NUMBER
8809

5e. TASK NUMBER
PPM00021771

5f. WORK UNIT NUMBER
EF128752

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Space Vehicles Directorate
3550 Aberdeen Ave SE
Kirtland AFB, NM 87117-5776

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RVSW

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RV-PS-TR-2017-0132

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The objective of this project was to conduct fundamental research that will lead to the development of cognitive communications protocols for satellite and space communications with possible broad applications in defense, homeland-security as well as consumer telecommunications. Such cognitive communications protocols are to be implemented on wideband autonomous cognitive radios (WACRs) that will have the ability to sense state of the radio frequency (RF) spectrum and the network and self-optimize their operating modes in response to this sensed state. The project developed three types of cognitive communications protocols: (a) machine learning aided cognitive anti-jamming and interference avoidance communications protocols, (b) cognitive cooperative beamforming, and (c) cognitive multi-mode communications protocols. Together, the advances made in this project form a foundation for achieving effective spectrum coexistence, interoperability and improved reliability in satellite transceivers in the presence of both inadvertent interference and deliberate jammers. Given the demonstrated promise, it is recommended that further research be pursued to advance the proposed WACR as the basis for future space communication systems that will offer significant benefits to national war-fighting and peacekeeping capabilities.

15. SUBJECT TERMS
satellite communication, cognitive radios, machine learning, anti-jamming, cognitive communications, cooperative communications, multi-mode communications

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			James Lyke
Unclassified	Unclassified	Unclassified	Unlimited	54	**19b. TELEPHONE NUMBER** *(include area code)*

(This page intentionally left blank)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This material is based on research sponsored by Air Force Research Laboratory under agreement number FA9453-16-1-0052. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

DISCLAIMER

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory or the U.S. Government.

1 SUMMARY

The objective of this project was to conduct fundamental research that will lead to the development of cognitive communications protocols for satellite and space communications with possible broad applications in defense, homeland-security, and civilian as well as consumer telecommunications. Such cognitive communications protocols are to be implemented on wideband autonomous cognitive radios (WACRs) that will have the ability to sense state of the radio frequency (RF) spectrum and the network and self-optimize their operating modes in response to this sensed state. The project focused on three types of cognitive communications protocols: First, machine learning aided cognitive anti-jamming and interference avoidance communications protocols were designed. These novel cognitive anti-jamming protocols were then implemented in software and the performance was evaluated to demonstrate their effectiveness in challenging multi-agent spectrum environments. Second, cognitive cooperative beamforming was considered for clusters of small satellites. In contrast to many previous theoretical studies, this project focused on practical protocols to achieve distributed transmit beamforming using node cooperation and cognition. Data synchronization that is necessary for cooperative beamforming was formulated as an optimization problem and a novel low-complexity practical solution algorithm was developed. Third, the project considered cognitive multi-mode communications that is identified as one of the key attributes of WACRs. A machine-learning aided cognitive radio access technology (RAT) selection algorithm was developed in the specific context of emerging 5th Generation (5G) cellular networks and its performance was demonstrated in computer simulations. The proposed approach can also be adapted to autonomous military and satellite communications networks based on WACR technology. Together, the advances made in this project form a foundation for achieving effective spectrum coexistence, interoperability and improved reliability in satellite transceivers in the presence of both inadvertent Radio Frequency Interference (RFI) and deliberate jammers. Cognitive satellite and space communications strategies based on the proposed WACR technology could be the foundation for future space communication systems that offer significant benefits to national war-fighting and peacekeeping capabilities through improved reliability and robustness in tactical and other communications networks.

2 INTRODUCTION

The wideband autonomous cognitive radios (WACRs) that we have been focused in this project present a potential future technology to realize autonomous radio communications over non-contiguous wide spectrum bands in the presence of adverse conditions with far-reaching implications for the future of telecommunications [1]. These wideband cognitive radios are radios that are capable of autonomously achieving arbitrary performance objectives of a user by utilizing built-in intelligence and self-awareness. This makes these radios to have much broader relevance and public purpose in that they have the potential to transform today's wireless telecommunications. In fact, already there are various cognitive algorithms that have been proposed to be incorporated in to the emerging 5G wireless cellular systems. As cognitive radio technology matures, we may expect it to play a significant role in 5G systems.

In military and satellite communications, the adverse conditions encountered may be both deliberate as well as inadvertent. Moreover, encroachment on previously-allocated spectrum resources by commercial and unlicensed/unauthorized users can only be expected to grow in the coming years. Combination of these traditional as well as evolving spectrum demands requires future telecommunications technologies to be intelligent, self-aware and spectrally agile. Wideband autonomous cognitive radios, first proposed by this project group in [1-6], and pursued in this project are radios with these defining characteristics that can lead to autonomous radio communications over non-contiguous wide spectrum bands in the presence of such adverse conditions.

In our previous work, we developed spectrum knowledge acquisition techniques that are suitable for wideband cognitive radios [4-6,7]. This project was focused on utilizing such acquired spectrum knowledge to help a WACR self-optimize its mode of operation in response to the state of the RF environment. Such communications approaches are termed cognitive communications protocols. During this project, we focused on developing cognitive communications protocols for three specific communications problems that are common to both military and civilian communications, but can be critically important for future space, satellite and tactical communications networks:

1. Cognitive anti-jamming and interference avoidance communications: Machine learning aided cognitive anti-jamming and interference avoidance protocols were

designed to achieve reliable communications in challenging multi-agent spectrum environments.

2. Cognitive cooperative communications: Novel, low-complexity algorithms for data synchronization among distributed mobile nodes were developed to facilitate cooperative beamforming in clusters of nodes formed by, for instance, small satellites.
3. Cognitive multi-mode communications: Machine-learning aided cognitive RAT selection algorithms were developed to help enable multi-mode communications in WACR-based communications networks.

One of the most important situations in which WACRs can be a great asset is against jamming attacks. When several radios attempt to operate in the presence of one or more malicious jamming signal transmissions aimed at disrupting the reliable communications, each radio will attempt to avoid the jammers as well as each other's transmissions, leading to a distributed decision-making problem in a challenging multi-agent environment. In this project, we have leveraged the cognitive and learning abilities of WACRs to develop effective and efficient machine-learning aided cognitive anti-jamming and interference-avoidance communications protocols. The first class of our cognitive anti-jamming algorithms are concerned with intelligently-switching the transmission frequency once the current transmission band is either jammed or interfered with [8-10]. However, such protocols can be vulnerable against smart jammers that may attempt to learn the cognitive radios own behavior. In response, our second class of proposed algorithms allows the radios to switch the operating frequency just before getting jammed or interfered with. These novel and first-in-its-kind algorithms [11, 12] can be extremely robust against smart jammers and a great asset in tactical communications networks.

Cognitive capabilities of WACRs can also facilitate implementation of advanced cooperative communications strategies in future wireless networks. With networks formed by clusters of small satellites such as cubesats in mind, we explored the cognitive cooperative beamforming to gain the advantages of distributed transmit beamforming [13,14]. In contrast to previous literature on theoretical treatment of the subject, during this project we proposed practical cognitive cooperative beamforming approaches. The proposed framework is made of two steps: data sharing among cooperative nodes and distributed transmit beamforming. During this project, a specific, low-complexity practical algorithm was developed to achieve to solve the first stage

problem of data synchronization among nodes and its effectiveness was demonstrated through examples.

One of the key attributes of the WACRs proposed by our team [1] is that they must support interoperability among systems through multi-mode communications over wide spectrum ranges. In recent years, multi-mode operation has also been found great interest in emerging 5G communications systems in general. As a result, we also considered the problem of multi-mode communications in the general setting of 5G networks and proposed a cognitive radio access technology selection algorithm that can also be adapted to future satellite and other military communications based on WACR technology in future.

The conventional mechanism for user association in cellular networks is based on the (Signal-to-Interference-plus-Noise-Ratio) max-SINR rule. While theoretically this choice should provide the highest channel capacity option to the user, in practice it has proven to be insufficient in the context of HetNets (Heterogeneous Networks) [15,16]. As a result, multi-parametric optimization solutions have been proposed in the literature to overcome the limitations of the max-SINR RAT selection approach by combining it with other criteria (e.g., cell load). However, these approaches suffer from their high computational complexity [16]. Moreover, the associated problem of efficient and practical multi-parametric modeling (representation of the network state) has not yet been completely solved [17,18]. On the other hand, the need for integrating multiple parameters in the optimization of network functionalities seems to be a design requirement for future networks. There is increasing agreement to look at these parameters under the umbrella of context-awareness [15,18-20].

Thus, during this project we proposed and developed a distributed cognitive framework for RAT selection in the specific context of 5G HetNets. Our proposed solution implements machine learning algorithms in order to satisfy the desired user association perspectives while providing a meaningful approach for context-aware multi-parametric modeling that echoes the principles of [21]. Our proposal was shown to be able to learn simple state representations out of the terminal experience and user behavior, reducing the complexity of the core network design requirements. Moreover, it allows multi-objective optimization of the association decisions while incurring minimal network overhead. Network simulation results shows that the proposed machine-learning

based approach can indeed be the basis for highly adaptable multi-mode cognitive communications protocols suitable for autonomous satellite and military communications networks

The long-term objective of this project is to develop WACR technology for future space and satellite radio systems that can be autonomous, self-aware and intelligent. The above outcomes from this project has advanced this objective and the project team expect to continue this work in the coming years to develop prototype WACRs.

3 METHODS, ASSUMPTIONS AND PROCEDURES

3.1 Cognitive anti-jamming and interference avoidance communications

3.1.1 Protocols to switch operating sub-band after getting jammed/interfered

First assume that a single WACR attempts to avoid a dynamic jammer as shown in figure 1. The wideband spectrum of interest can be considered as made of N_b sub-bands. Each sub-band may include a different number of communication channels. Following [8], we define the state of a sub-band based on the availability of bandwidth satisfying a minimum required bandwidth for transmission. Let β denotes the minimum required bandwidth for transmission. At any given time, if the available bandwidth in the sub-band is greater or equal to β then the sub-band is considered available (state 1) otherwise it's considered not-available (state 0). Thus, let us denote the state of the i-th sub-band at time n by $v_i[n] \in \{0,1\}$. We may assume that this state of a sub-band evolves according to a two-state first order Markov chain.

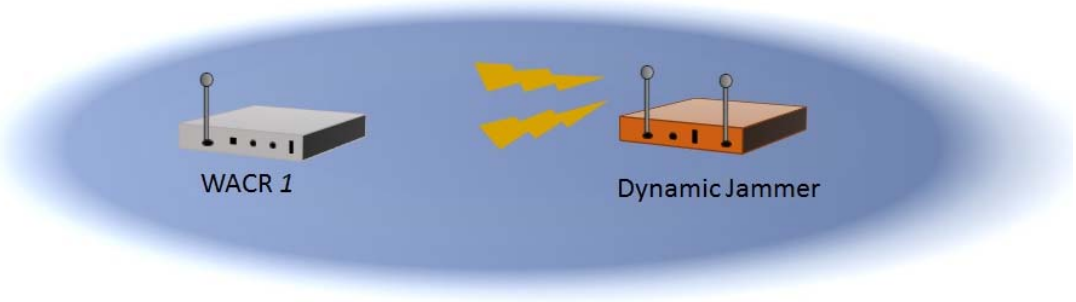

Figure 1. A single WACR challenged by a dynamic jammer.

We assume that the state of a sub-band is determined based on Neyman-Pearson (NP) criterion as discussed in [8]. This forms the *spectral activity detection* step that is an essential functionality of all WACRs [1]. The output of the NP detector is used to compute the maximum available contiguous bandwidth in a sub-band. The state of the sub-band is determined by comparing this to the minimum required bandwidth β for transmission. As an example, let us consider a sub-band made of 8 channels with equal bandwidth (5 MHz) as shown in figure 2. Only the last four channels are idle in this situation. If we assume that the minimum required bandwidth β=20 MHz, then this sub-band will be considered in state 1 (available).

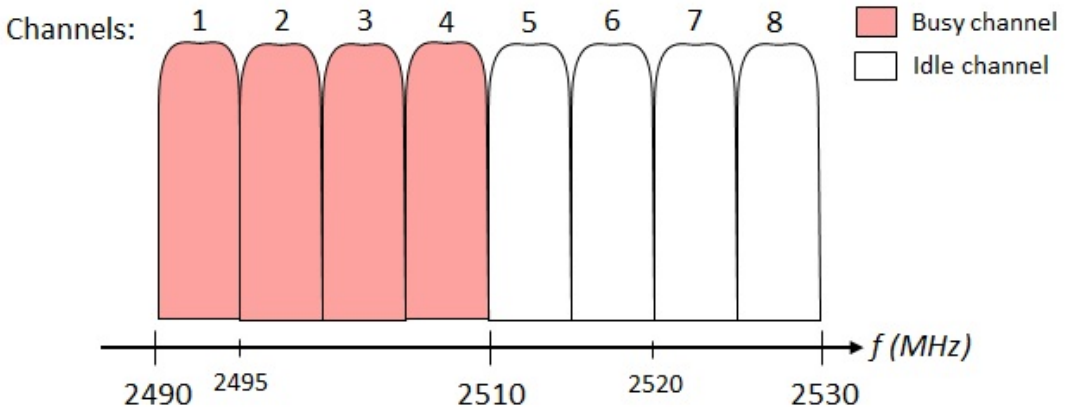

Figure 2. An example of a sub-band made of 8 channels in which last 4 channels are idle.

A WACR will select a new operating sub-band when the state of the current sub-band changes to not-available state (available bandwidth is less than β). For efficient operation with effective anti-jamming, the selected new sub-band should have low interference with high probability for the longest time.

Finding an optimal policy for the sub-band selection in an anti-jamming model can be difficult for many reasons. First, this may require high computational complexity. Second, a policy needs to be computed in real-time. Moreover, the model parameters may vary with time due to the dynamic nature of the wireless environment. All these make any attempt to directly compute an optimal policy almost impractical. An alternative is to let a WACR to learn an optimal policy through machine-learning instead of computing one. A special machine learning approach, called reinforcement learning (RL), could especially be suited when underlying state dynamics are Markov [1,22].

One of the most widely used reinforcement learning approaches is the Q-learning whose basic idea is to maintain a Q-table that contains what is called Q-values representing a measure of goodness resulting from taking an action a when in state s [1,23]. The number of possible states and possible actions determines the number of rows and columns of the Q-table, respectively. In our system model, the state s represents the index of the current operating sub-band. On the other hand, the action a refers to the index of the newly selected sub-band, with $a \in \{1,2,\cdots,N_b\}$.

For every iteration, the WACR identify whether the current operating sub-band s is available or not. If the sub-band is available, no further action is required. If the sub-band is not-available, the WACR updates the Q-table, based on a certain observed reward $r(s,a)$ that depends on the amount of time it takes for the jammer to interfere with the WACR transmission once it has switched to the a-th sub-band.:

$$Q(s[n-1],a[n-1]) \leftarrow Q(s[n-1],a[n-1]) + \alpha\left[r(s[n-1],a[n-1]) + \gamma \max_{a} \mathrm{Q}(s[\mathrm{n}],\mathrm{a}) - Q(s[n-1],a[n-1])\right] \quad (1)$$

where $\alpha \in (0,1)$ is the learning rate and $\gamma \in [0,1)$ is a discount factor. In Q-learning based policies, the actions are selected such that Q-value is maximized [1]:

$$a^* = \arg\max_{a \in A} \mathrm{Q}(s,\mathrm{a}) \quad (2)$$

An efficient frequency-switching can be achieved if the WACR can learn the pattern of the jammer's behavior. Indeed, following the above proposed cognitive approach, a single WACR may be able learn an optimal policy that enables it to switch the operating sub-band each time it gets jammed to a new one that is free of jamming signals for the longest possible time.

In practice, however, there could be multiple WACRs simultaneously operating over the same spectrum band of interest, producing a complicated multi-agent environment as shown in figure 3. In this case, each WACR needs to avoid both the malicious jammer as well as the transmissions of other radios. This can be modeled as a stochastic game, an extension of Markov Decision Processes (MDPs), in which interactions among different agents is considered [10].

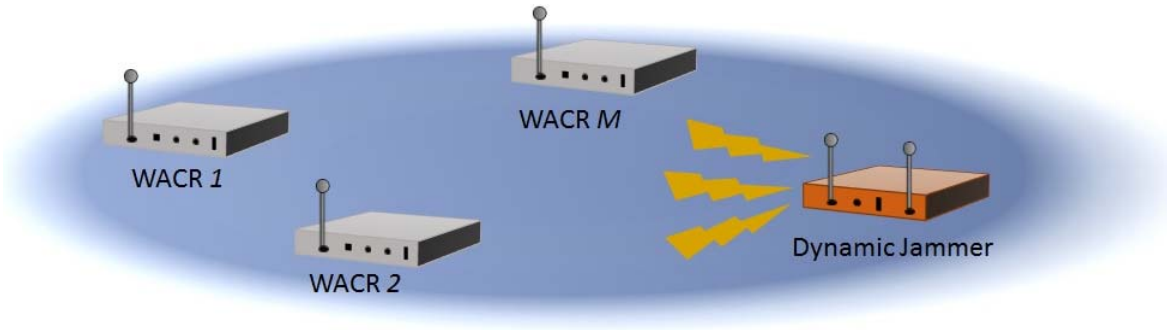

Figure 3. An Example of M WACRs operating in the same frequency range challenged by a dynamic jammer.

The stochastic game formulation includes a set of states and a collection of action sets denoted by $\mathcal{S}$ and $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_M$, respectively. The players of this game are the WACRs. The game moves from its current state to a new random state with transition probability determined by the current state and one action from each player $T: \mathcal{S} \times \mathcal{A}_1 \times \cdots \times \mathcal{A}_M \rightarrow PD(\mathcal{S})$. In this scenario, each WACR will use Q-learning algorithm to learn an optimal, or a near-optimal, policy for anti-jamming and interference-avoidance.

The state of m-th WACR at time n is defined as the index of the current operating sub-band denoted as $s_m[n] \in \mathcal{S}$. The space of game states is then given by $\mathcal{S} = \{1,2,\cdots,N_b\}$. The action taken by m-th WACR at time n, denoted by $a_m[n] \in \mathcal{A}_m$, represents the new operating sub-band selected by the m-th WACR at time n. At any time instant, if the current operating sub-band was declared to be not-available (the amount of available bandwidth is less than β) the WACR will choose a new operating sub-band according to (2) that will be most likely not facing any interference or jamming for the longest time. The m-th WACR's immediate reward for selecting action $a_m[n]$while in state $s_m[n]$ is denoted by $r_m(s_m[n], a_m[n])$.

3.1.2 Protocols to switch operating sub-band just before getting jammed/interfered

The cognitive anti-jamming protocols of the previous section, as with most of the existing literature, allows a WACR to switch its operating frequency only after getting jammed. As mentioned above, such protocols can be vulnerable against smart jammers. In addition, a smooth transition to a new frequency band may not be achievable. To address these concerns, we introduce a new cognitive anti-jamming communications protocol that allows the WACR to switch its

operating sub-band just before getting jammed [11,12]. The proposed novel cognitive anti-jamming communications framework is made of two operations: sensing and transmission. Each operation has its own learning algorithm based on Q-learning with different targets.

The goal of sensing is to learn the pattern of jammer's behavior and transmissions of other radios. On the other hand, the cognitive objective of the transmission operation is to determine when to switch the operating sub-band and to where so that jamming and interference is avoided for the longest possible duration. Thus, at any stage of the game, there are two sub-bands associated with a given WACR: one for sensing and one for transmission. If the sensing operation were to learn an optimal policy, the WACR would be able to accurately predict the jammed or interfered sub-bands. This will help the transmission operation since if the current transmission sub-band is predicted to be jammed during the next time instant, the WACR will switch to another sub-band and may avoid getting jammed.

Each sub-band can only be in one of two possible states: state "0" and state "1", as mentioned earlier. At any given time, if the sub-band gets jammed or faces interference, it is in state "0" (not-available). Otherwise, it is in state "1" (available). The set of sub-band states can be then given by $\mathcal{V} = \{0,1\}$. For both sensing and transmission operations, the game state $s_s[n] \in \mathcal{S}$ and $s_t[n] \in \mathcal{S}$ represent the index of selected sub-bands for sensing and transmission, respectively, at time n. Hence, the number of possible states for each operation is N_b, where N_b is the total number of sub-bands. While the sub-band state refers to the availability of the sub-band, the game state refers only to the index of the operating sub-band apart from whether it is available or not.

At any time instant, first the state of operating sub-bands for both sensing and transmission (the value of $v \in V$ for sub-band indices $s_s \in \mathcal{S}$ and $s_t \in \mathcal{S}$) are identified. After determining the states of both operating sub-bands, the WACR executes actions for both operations. We define actions $a_s[n] \in \mathcal{A}$ and $a_t[n] \in \mathcal{A}$ as the new operating sub-bands for sensing and transmission, respectively at time n.

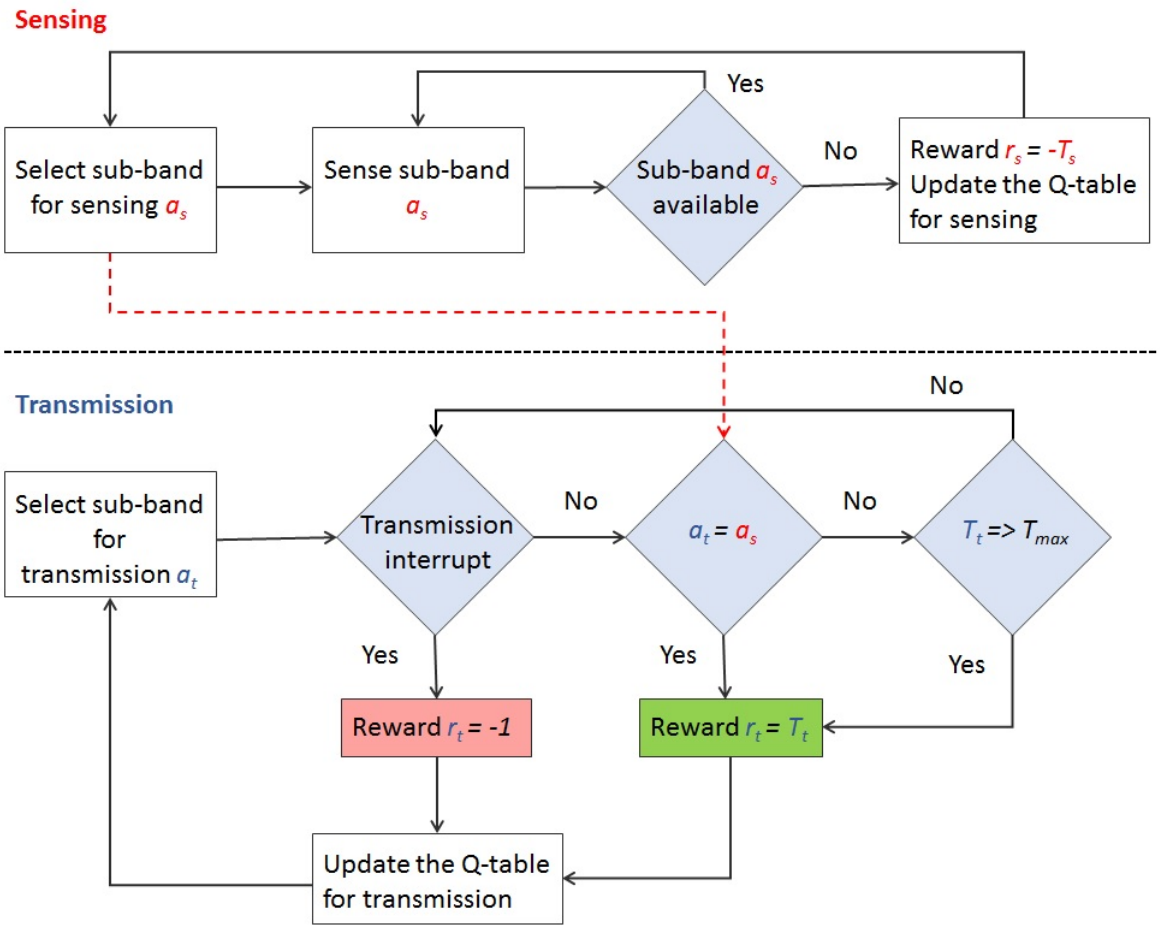

Figure 4. Cognitive anti-jamming communications for avoiding getting jammed/interfered.

Figure 4 shows both sensing and transmission operations for a given WACR [11,12]. For sensing operation, the WACR always attempts to sense the sub-band where the jammer or interference signal is located. For every new selected action (new sub-band), the WACR computes the time it takes until the jammer or interference signal arrives, denoted by T_s. The reward corresponding to a sensing action is defined as [11]:

$$r_s = -T_s \tag{3}$$

Note that, the WACR will select a new sub-band for sensing if and only if the current sensing sub-band gets jammed or faces interference. The actions $a_s \in \mathcal{A}$ for the sensing operation are selected such that rewards are maximized.

As mentioned earlier, the objective of transmission protocol is to switch the operating sub-band before getting jammed or facing an interference. As can be seen from figure 4, the operating sub-band for transmission may be changed under two scenarios 11]: First is if the transmission is interrupted, implying that the current operating sub-band is facing either too much interference or a jamming attack. The second condition is when the sensing operation predicts that the current transmission sub-band to be the one that will get jammed/interfered in the next time instant. Since the objective is to switch the operating sub-band before getting jammed, effective learning must lead to switching always due to the second condition while avoiding the first. In order to achieve this objective, the transmission reward function is defined as [11]:

$$r_t = \begin{cases} -1 & \text{if sub} - \text{band } a_t \text{ gets jammed} \\ T_t & \text{otherwise} \end{cases}, \quad (4)$$

where T_t is the transmission duration in sub-band at before switching to a new one.

3.2 Cognitive cooperative communications

Distributed transmit beamforming is a cooperative communications technique in which two, or more, spatially separated cooperative nodes act as elements of an antenna array to beamform common data to a destination node [13,24]. In this project, we explored practical distributed transmit beamforming approaches that may specifically be well-suited for networks of mobile cooperating nodes such as clusters of satellites.

In contrast to many existing theoretical contributions on this topic, we specifically considered practical approaches to achieve distributed transmit beamforming. To that end, our proposed approach divided the distributed transmit beamforming into two stages: data sharing and beamforming [24]. During the data sharing stage, cooperative nodes share their data with others to achieve data synchronization. During beamforming stage, cooperative nodes beamform the common data to the destination. Note that the presumption is that if distance among the cooperative nodes are sufficiently shorter than the distance between cooperative nodes and the destination node, energy required for data sharing may relatively be small compared to the energy required for data transmission to the destination node.

The simplest approach to data synchronization is uniform time sharing in which each node uses an equal duration of time to broadcast its data to all other nodes that can receive its data.

However, this may not be optimal when there is only a limited time available for data sharing. Indeed, allocated times do not necessarily need to be equal if the performance objective is to maximize the data transmitted to the destination node during the beamforming stage.

The focus of this project was to develop efficient cognitive cooperative protocols to solve the first stage problem identified above: data synchronization for distributed transmit beamforming. In particular, the motivation is to optimize data synchronization among nodes to achieve maximum throughput (transmitted data) during distributed transmit beamforming.

We may, for example, assume that cooperative nodes are satellites in specific orbits with known velocities and locations and the destination node is a satellite ground station. This can be an important application scenario for distributed transmit beamforming given the increasing popularity of using clusters of smaller satellites as an alternative to expensive large satellites. Our objective is to find the optimum number of packets that each cooperative node should send to others during each data sharing time interval so that the total throughput during the distributed transmit beamforming stage is maximized. We consider a general timing scenario for data sharing and beamforming, as shown in figure 5, and formulate the problem as an optimization problem subject to a set of constraints.

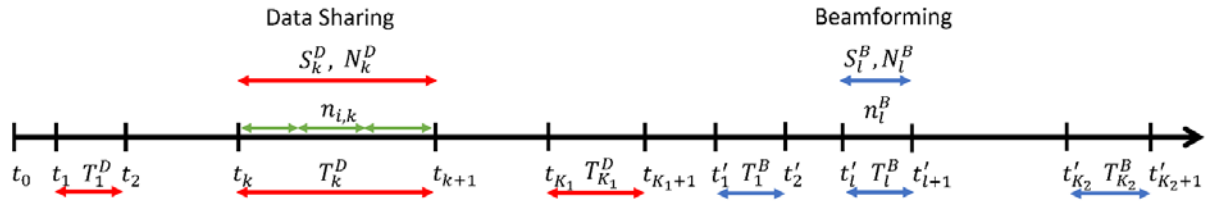

Figure 5. A general timing scenario during data sharing and beamforming stages in a distributed transmit beamforming application.

As figure 5 shows, assume that there are K_1 number of time intervals in data sharing stage and K_2 number of time intervals in beamforming stage. T_k^D is the k-th data-sharing time interval where $k = 1, \cdots, K_1$ and T_l^B is the l-th distributed transmit beamforming time interval where $l = 1, \cdots, K_2$. Set S_k^D denotes the indices of available cooperative nodes during the k-th data sharing time interval T_k^D. Set S_l^B denotes the indices of the cooperative nodes during beamforming interval T_l^B whose cooperative transmissions can provide enough beamforming gain to satisfy quality of

service requirements at the destination node (e.g., maximum tolerable bit error rate or minimum received Signal-to-Noise Ratio (SNR)). Without loss of generality, we may assume that, S_k^D is necessarily distinct from S_{k+1}^D, since otherwise we may combine them to create a single interval. The same is true with sets S_l^B.

We assume that each cooperative node i (for $i = 1, \cdots, M$) has enough packets (or theoretically unlimited number of packets) during each data sharing time interval. This is easily be justified since in most practical applications, there is a large amount of data but limited resources. Let N_k^D denote the total number of packets that the cooperative nodes in set S_k^D can share with each other during the k-th data sharing time interval T_k^D. We assume that each cooperative node i has N_i number of packets at the beginning of the data sharing stage which is enough to share during whole this stage. Let $N_{i,k}$ show the number of packets that node i has at the beginning of the k-th data sharing time interval T_k^D with $N_{i,1} = N_i$ and $N_{i,k} \geq N_k^D$. Let us denote the number of packets that cooperative node i sends to others during the k-th data sharing interval by $n_{i,k}$ where $0 \leq n_{i,k} \leq N_k^D$ and $\sum_i n_{i,k} = N_k^D$ for $k = 1, \cdots, K_1$, $i \in S_k^D$, and $n_{i,k}$ is an integer. Note that, if $i \notin S_k^D$ then $n_{i,k} = 0$. Then, $N_{i,k+1}$ must be updated as follows: $N_{i,k+1} = N_{i,k} - n_{i,k}$ where $k = 1, \cdots, K_1 - 1$. Our goal is to find the optimum values of $n_{i,k}$ so that the total number of transmitted packets during beamforming stage will be maximized.

The message bandwidth (B), packet rate (R), and maximum transmitted power (P_T) of all cooperative nodes are assumed to be the same. Also, we assume that the message bandwidth (B) is much less than the coherence bandwidth of the channel (B_c) between cooperative nodes and the destination node, i.e, $B \ll B_c$ so that the channel can assumed to be frequency-flat. The channels exhibit slow fading, i.e., the complex channel gain vector $\boldsymbol{h} = [h_1, \cdots, h_M]^T$is constant during several symbol periods (corresponding to one or more packets) where $h_i = |h_i| e^{j \measuredangle h_i}$is the complex channel gain between cooperative node i and the destination node.

The total number of packets that cooperative node i sent to cooperative node j during the data sharing stage can be written as $n_{i \to j} = \sum_{k=1}^{K_1} n_{i,k} I_k(j)$ where $i, j = 1, \cdots, M$, $i \neq j$, and

$$I_k(j) = \begin{cases} 1 & if\ j \in S_k^D \\ 0 & if\ j \notin S_k^D \end{cases} \tag{5}$$

We are now in a position to formulate the data synchronization problem for distributed transmit beamforming as the following optimization problem:

$$maximize \sum_{l=1}^{K_2} n_l^B, \tag{6}$$

where

$$n_l^B = \min(N_l^B, n_l^c), l = 1, \cdots, K_2, \tag{7}$$

$$n_l^c = \sum_{i \in S_l^B} \min_{j \in S_l^B} n_{i,j,l}^S, l = 1, \cdots, K_2, \tag{8}$$

$$n_{i,j,l}^S = \begin{cases} n_{i \to j} & if\ l = 1 \\ n_{i,j,l-1}^S - n_{l-1}^B & if\ l = 2, \cdots, K_2 \end{cases}$$

subject to

$$0 \le n_{i,k} \le N_k^D, \sum_{i \in S_l^B} n_{i,k} = N_k^D, k = 1, \cdots, K_1. \tag{9}$$

The optimization problem (6) - (9) is a nonlinear optimization problem for which there is no apparent direct closed-form solution. Therefore, in the following we first resort to exhaustive search.

3.2.1 Exhaustive search for data sharing in distributed transmit beamforming

The exhaustive search can be implemented by checking, for $k = 1, \cdots, K_1$, all possible integer values of $n\_\{i,k\}$ from 0 to $N\_k\hat{}D$ subject to $\sum_i n_{i,k} = N_k^D$ and finding their optimum values so that (6) would be maximized. For example, if $S_k^D = \{1, 2\}$ and $N_k^D = 5$, then for $n_{1,k}$ and $n_{2,k}$ we must check all possible integer values from 0 to 5 such that $n_{1,k} + n_{2,k} = 5$. In this case, the possible values of $n_{1,k}$ and $n_{2,k}$ are $(0,5), (1,4), (2,3), (3,2), (4,1)$, and $(5,0)$. In each beamforming time interval T_l^B, the exhaustive search uses (8) to find the maximum shared data between each node and other cooperative nodes and then uses (7) to calculate n_l^B [24].

In general, $N = N_k^D$ and $L = |S_k^D|$ be the cardinality of S_k^D. In each time interval T_k^D, exhaustive search has to check all possible integer combinations of $n_{i,k}$'s such that $\sum_i n_{i,k} = N$ for $i \in S_k^D$. For example, if $S_k^D = \{1,2\}$, $L = 2$, and $N = 10$, exhaustive search has to check all possible integer values (from 0 to 10) for $n_{1,k}$ and $n_{2,k}$ such that $(n_{1,k} + n_{2,k}) = 10$ which is 11 distinct pairs of values, i.e., $(0,10), (1,9)$, …, $(10,0)$. If $N_L(N)$ indicates the total number of distinct L-tuples of integer values for L number of $n_{i,k}$ whose sum is N, then, for $L > 4$, it is hard to find a closed form expression for $N_L(N)$. We can, however, show that approximately in the time interval T_k^D, computational complexity of exhaustive search is in the order of $O(N^{L-1})$. Hence, if we have K_1 number of time intervals and in each time interval T_k^D, we have $N_{i,k} \geq N_k^D$, then, computational complexity of exhaustive search will be on the order of $O(\prod_{k=1}^{K_1} N_k^{L-1})$ where $N_k = N_k^D$ and $L = |S_k^D|$.

The use of exhaustive search to find the optimum solution may not be desirable (due to its very high computational complexity) unless the number of cooperative nodes is relatively small. We may instead use meta-heuristic algorithms such as genetic algorithm or a heuristic method. We can, however, use the exhaustive search results as a reference for assessing the performance of the proposed heuristic method to solve the problem in (6) - (9).

3.2.2 A novel heuristic method for data sharing in distributed transmit beamforming

In our proposed heuristic method [24], first, we use the following two steps to remove those unnecessary sets S_k^D and S_l^B which don't influence the problem formulation, data sharing optimization, and throughput maximization:

1. Drop any S_l^B which is not the subset of any set or sets in the data sharing time intervals since cooperative nodes in that S_l^B will not be able to have common data to beamform based on the sets in the data sharing time intervals.
2. Drop any S_k^D none of whose subsets is a given set in then beamforming stage since the shared data between cooperative nodes in S_k^D cannot be transmitted to the destination node during beamforming stage.

After removing these redundant sets, the data sharing optimization and throughout maximization can be divided in to two steps:

1. Since cooperative nodes can in general be mobile, we start from T_1^D and then go to T_2^D and so on. In each data sharing time interval T_k^D (starting from T_1^D), we must select a subset or subsets of cooperative nodes (from set S_k^D) based on related sets S_l^B and N_l^B. We divide each time interval T_k^D to several disjoint subintervals based on the number of selected subsets. Now, assume that we have K_1' number of subintervals during data sharing stage ($T_{k'}^D$ for $k' = 1, \dots, K_1'$). Correspondingly, we define $S_{k'}^D$, $N_{k'}^D$, $N_{i,k'}$, and $n_{i,k'}$. For example, if $S_k^D = \{1, 2, 3\}$, $S_i^B = \{1,2\}$, and $S_j^B = \{1,2,3\}$, then we divide the k-th data sharing time interval to two subintervals $S_{k'}^D = \{1,2\}$ and $S_{k'+1}^D = \{1,2,3\}$.

   The values of $N_{k'}^D$ and $N_{k'+1}^D$ have to satisfy the following constraints:

$$N_{k'}^D + N_{k'+1}^D \leq N_k^D \tag{10}$$

$$N_{k'}^D \leq N_i^B \tag{11}$$

$$N_{k'+1}^D \leq N_j^B \tag{12}$$

   We impose the fairness criterion, $N_{k'}^D / N_{k'+1}^D = \mathrm{N_i^B}/\mathrm{N_j^B}$ \$. Therefore, $N_{k'}^D$ and $N_{k'+1}^D$ can be calculated based on the values of N_k^D, N_i^B, and N_j^B as follows:

   a. If $N_i^B + N_j^B \leq N_k^D$, we have $N_{k'}^D = N_i^B$ and $N_{k'+1}^D = N_j^B$.
   b. If $N_i^B + N_j^B > N_k^D$, we have

$$N_{k'}^D = \left\lceil N_k^D \left(\frac{N_i^B}{N_i^B + N_j^B} \right) \right\rceil, \tag{13}$$

$$N_{k'+1}^D = \left\lfloor N_k^D \left(\frac{N_j^B}{N_i^B + N_j^B} \right) \right\rfloor, \tag{14}$$

   where $\lceil . \rceil$ and $\lfloor . \rfloor$ functions map a real number to the greatest preceding or the least succeeding integer number, respectively.

   We follow a similar procedure as above when we have more than two subsets to find the values of $N_{k'}^D$'s. After finding $N_{k'}^D$ for each subinterval, we update corresponding

N_i^B by subtracting $N_{k'}^D$ from that N_i^B, i.e., $N_i^B = N_i^B - N_{k'}^D$ so that in the next subintervals, we try to provide as many packets as possible for beamforming time intervals based on the remaining N_i^B's.

2. The optimum values of $n_{i,k'}$ are computed based on $S_{k'}^D$, $N_{i,k'}$, and N_k^D such that a fairness criterion over cooperative nodes is maintained (i.e., the number of transmitted packets by source nodes should be proportional to their available data) [24]. Since all cooperative nodes have the same E_P and they have enough packets to share in each data sharing time interval, it turns out that $n_{i,k'} = (N_{i,k'} N_k^D) / \sum_{i \in S_{k'}^D} N_{i,k'}$.

3.3 Cognitive multi-mode communications

One of the key attributes of the WACRs proposed by our team [1] is that they must support multi-mode communications over wide spectrum ranges. In recent years, multi-mode operation has also been found great interest in emerging 5G communications systems in general. As a result, during this project, we considered the problem of multi-mode communications in the general setting of 5G networks and proposed a cognitive RAT selection algorithm that can also be adapted to satellite and other military communications in future.

In 5G HetNets, the variedness and availability of network attachment points through diversified RATs are expected to continue to mature and evolve. This development will turn the classic, and extremely important, user association problem into a complex decision process in order to guarantee efficacy and efficiency of the HetNet architecture. Thus, in the following we propose a framework for tackling the problem of determining which RAT standard and spectrum to utilize and which base stations (BS) or users to associate within the context of 5G HetNets.

The conventional mechanism for user association in cellular networks is based on the max-SINR rule. While theoretically this choice should provide the highest channel capacity option to the user, in practice it has proven to be insufficient in the context of HetNets [15,16]. Indeed, when the architecture of a network includes macro and small-cells (i.e., micro- pico-, femto-cells), the reduced coverage of the latter makes them less attractive to the terminal devices regardless of how loaded the macro-cells might be, particularly when both network tiers are using the same spectrum.

Multi-parametric optimization solutions have been proposed in the literature to overcome the limitations of the max-SINR RAT selection approach by combining it with other criteria (e.g., cell load). However, the computational complexity of these approaches has made this task very difficult [16]. Furthermore, the associated problem of efficient and practical multi-parametric modeling (representation of the network state) has not yet been completely solved [17,18]. On the other hand, the need for integrating multiple parameters in the optimization of network functionalities seems to be a design requirement for future networks. There is increasing agreement to look at these parameters under the umbrella of context-awareness [15,18-20].

Thus, we propose a distributed cognitive framework for RAT selection for 5G HetNets [25]. Each RAT is considered independently by our framework even when several of them are concentrated at a single physical node. Our proposed solution implements machine learning algorithms in order to satisfy the desired user association perspectives while providing a meaningful approach for context-aware multi-parametric modeling that echoes the principles of [21]. In particular, our solution advocates the use of cognition at the device-level in order to learn optimal, or at least reasonably well-performing, decision policies based on the experience of the device itself. Note that, cognition is important since there is no "one-size-fits-all" rule for association.

Figure 6. Machine-learning based cognitive RAT selection framework.

As is shown in figure 6, the proposed distributed cognitive framework for RAT selection can be segmented into three modules/stages:

1. **Defining the user/network state model:** Formulating a system model that is context aware can be a daunting task, especially for our case of interest. The state definitions must be reasonable across the envisioned heterogeneity and diversity of architectures, communication standards, international regulations, applications and protocols and deployment needs of future 5G networks. Furthermore, representing these states such that the overall processing and memory requirements as well as network overhead remain reasonable is important for efficiency. A centralized solution that involves bookkeeping all possible states in the core network nodes for all the users and making decisions for them, would demand a considerable computational capacity, even for a relatively small HetNet.

   Thus, it is desirable, to adopt a distributed design and build meaningful user/network states at the terminal level. As illustrated in figure 6, in the first stage of our framework, we propose clustering of appropriately defined feature vectors as a method for autonomously constructing suitable system models for the user association task. The choice of the standardized fields (or parameters) for formulating the feature vectors is extremely important and requires considerable domain knowledge in order to reduce the dimensionality. These features should contain both network-centric and user-centric information. This approach allows the system models to be custom-made for each node because the created cognitive states will depend on the specific situations it has experienced. Note also that handling all this information at the device-level dramatically decreases the overhead requirements and leads to low-complexity algorithms that try to optimize the gains of individual user according to its particular needs.

   As a result, the user state at any given time n is defined in terms of a collection of parameters that characterizes the mobile device and RAT situation in the network at that particular time. Such parameters can be collected and grouped as a tuple in order to formulate a vector $\boldsymbol{x}$ of d descriptors or features. Thus, we may represent an

observation of the user/network state as a point in a *d*-dimensional feature space. As mentioned earlier, the mobile device needs to collect relevant statistics that reflect both network state information and user needs/behavior information. The network state information could be obtained through Application Programming Interface (API) calls to the available network interfaces and broadcast messages, while estimating the user behavior might require sensing and packet-level analysis with the corresponding data post-processing.

From a machine learning perspective, the feature vectors are simply patterns. These patterns can be divided into a set $S = \{1, \ldots, M\}$ of M groups of similar characteristics or clusters, based on some measure of similarity. In our framework, these resulting clusters represent different cognitive states the terminal device can be in [25]. Note that these states are derived from the past data observed by the terminal device. Thus, the proposed approach builds a system model relying on the multi-parametric context information contained in the feature vectors. We refer to these states as cognitive states, since they are to be learned cognitively by each terminal device.

In our proposed approach, these cognitive states are generated by using the *X*-means algorithm [25]. Note that, the *X*-means is a clustering algorithm with relatively low complexity that, in addition to classifying data into a set of clusters, attempts to estimate the number of clusters M from data itself [1].

2. **Detecting the cognitive states:** The set of all possible system states S is the set of clusters from the first stage, obtained using the **X**-means algorithm (see figure 6). The current state s_t is the mapping of a current feature vector observed by the terminal device to one of those clusters. This mapping is done using the *k*-Nearest Neighbors algorithm.

   We define the set of actions to be the set of available RATs. At time t, each device must determine the best RAT association based on an observed feature vector $\boldsymbol{x}_t$. Although

there exist many supervised classification techniques to classify a new feature vector $\boldsymbol{x}_t$ into one of the cognitive states represented by clusters created by the X-means algorithm in the previous section, we advocate for the use of possibly the simplest approach, the k-Nearest Neighbors (kNN) algorithm [1]. The kNN rule assigns the new feature vector $\boldsymbol{x}_t$ to the m-th class to which the majority of k closest training feature vector(s) belong.

As depicted in figure 6, in our proposed approach, the training vectors and classes are supplied by the clustering stage preceding the kNN classifier. Let $\boldsymbol{y}_t$ be a tuple of feature vectors collected at decision time t by the terminal device. Then, the classification rule is a deterministic function that maps each feature vector in $\boldsymbol{y}_t$ to a cognitive state (i.e., the clusters). Hence, the output of the classifier is essentially the set of possible cognitive states the device may reach at the next time instant.

3. **Learning an effective user association policy:** The final goal is to obtain a decision policy to pick the best action (i.e., to associate with a RAT) given the state of the terminal [25]. We may assume that the terminal device is able to recognize the cognitive state that corresponds to the currently active RAT association and denote it by s_t. Let A_t denote the set of actions available to the device at time t, where actions a_t are identified as the available RATs. Choosing an action implies enforcing standard-specific procedures (i.e., network entry, handover, etc.) necessary for the new association. Notice that in our framework, the set of states reachable by the device at time $t+1$, depends on the chosen action at time t.

   As shown in figure 6, the third stage of our proposed framework relies on the model generated in the first stage to learn a good association policy. Specifically, we propose the use of the Q-learning algorithm for this reinforcement learning stage [25]. The Q-learning algorithm maintains a Q-table of values that represent a quantification of the goodness of taking a particular action when in a given state [1,22,23]. Each table entry, $Q(s_t, a_t)$, is associated with a state-action pair s_t and a_t. In our case, each Q-value is a measure of the "quality" of switching the currently active RAT association to either a

different RAT or keeping it unchanged. Note that, provided we defined the states as multi-parametric representations, the *Q*-Learning algorithm decisions are inherently context-aware.

Since the user association problem requires multi-objective optimization that jointly maximizes the user perceived average throughput and Quality of Service (QoS) while minimizing service interruptions due to mobility conditions, we define the reward function as [25]

$$R_t(s_{t-1}, s_t, a_{t-1}) = r_t \cdot U(r_t), \tag{15}$$

where

$$\begin{aligned} r_t = \ & \beta \cdot g(\text{Avg\_Measured\_Throughput}) \\ & - \lambda \cdot h(\text{HTTP\_RTP\_RTT\_Delay}) \\ & - \zeta \cdot c(\text{Handovers\_Calldrops}) \end{aligned} \tag{16}$$

with

1. R_t is the delayed reward function computed at instant t that evaluates the consequences of the action a_t taken at instant t while in state s_{t-1} that led to the current state a_t.
2. $U(.)$ is the Heaviside step function to ensure that $R_t(.)$ is non-negative.
3. β, λ and ζ are arbitrary coefficients defined on the interval $[0,1]$.
4. $g(.)$, $h(.)$ and $c(.)$ are suitably defined non-decreasing reward and cost functions of network performance metrics. $g(.)$ and $c(.)$ are also restricted to be non-negative.

The *Q*-values are then updated as

$$\begin{aligned} Q(s_{t-1}, a_{t-1}) \leftarrow & (1-\alpha) Q(s_{t-1}, a_{t-1}) \\ & + \alpha [R_t(s_{t-1}, s_t, a_{t-1}) \\ & + \gamma \max_{a_t} Q(s_t, a_t)] \end{aligned} \tag{17}$$

Once the *Q*-table is learned, the actions are selected according to:

$$a_t^* = \begin{cases} \underset{a_t}{\text{argmax}}\ Q(s_t, a_t) & \text{, with probability } 1-\epsilon \\ P(\mathcal{A}_t) & \text{, with probability } \epsilon \end{cases} \tag{18}$$

The novelty in the proposed approach is that the set of states are learned by the device itself and the number of states is also learned rather than pre-specified. This implies that a terminal, after collecting data during a training period, may formulate a system characterization and optimize its own association decisions without any external intervention.

4 RESULTS AND DISCUSSION

Cognitive Anti-Jamming and Interference Avoidance Communications

Protocols to switch operating sub-band after getting jammed/interfered

We start with the simple case in which a single WACR attempts to avoid a single sweeping jammer. Assume the WACR operates over a spectrum range of 200MHz in real-time and scans 40MHz-wide sub-bands at a time. In this case, there are 5 states and 5 actions. Note that, the action is the sub-band it selects for sensing and transmission during the next time instant to escape the jammer. The jammer sweeps the 200MHz-wide spectrum from lower frequency to higher frequency. The proposed cognitive anti-jamming technique was tested in two spectrum ranges: 2GHz-2.2GHz band and 3GHz-3.2GHz band.

In the presence of the sweeping jammer and the absence of any other interference the optimal sub-band selection policy to avoid the jammer is intuitive: The WACR should cyclically shift to the sub-band that is adjacent to the current sub-band from the lower frequency side. For example, if the WACR is currently sensing sub-band 5, it should choose sub-band 4 in order to avoid the jammer for the longest amount of time possible as shown on Table 1.

Table 1. Q-table with optimal anti-jamming policy.

s \ a	1	2	3	4	5
1	0	0	0	0	max Q-value
2	max Q-value	0	0	0	0
3	0	max Q-value	0	0	0
4	0	0	max Q-value	0	0
5	0	0	0	max Q-value	0

Results from our tests show that our WACR can indeed learn the above optimal sub-band selection policy to avoid deliberate jamming. Tables 2 and 3 show the Q-tables learned by the WACR, while operating in the 3GHz-3.2GHz band and the 2GHz-2.2GHz band, respectively [8]. In these experiments, user defined minimum required bandwidth in a sub-band is 30MHz.

Table 2. Learned Q-table in the 3 GHz to 3.2 GHz band.

s \ a	1	2	3	4	5
1	0.0461	0.0956	0.2907	0.4676	4.6945
2	4.8770	0.0830	0.2008	0.2872	0.9495
3	0.8342	4.6882	0.1628	0.2097	0.2882
4	0.3272	0.7844	4.5411	0.0645	0.2087
5	0.2048	0.7756	0.7705	4.5520	0.0851

Table 3. Learned Q-table in the 2 GHz to 2.2 GHz band.

s \ a	1	2	3	4	5
1	0.0971	0.3677	0.4801	0.4254	1.0584
2	1.5785	0.2964	0.1780	0.3003	0.6007
3	0.4680	1.4561	0.0940	0.1792	0.30792
4	0.3332	0.2704	1.4148	0.1881	0.1898
5	0.3323	0.5728	0.4249	1.2130	0.1328

Clearly, these Q-tables show that our proposed reinforcement learning based sub-band selection algorithm can indeed learn the sweeping jammer's behavior and perform as an effective cognitive anti-jamming and interference avoidance protocol. The Q-tables in Tables 2 and 3 show that if the system were to exploit (choose the actions resulting in the greatest reward), it will indeed choose the optimal sub-band that follows our intuition as previously mentioned and as shown in Table 1.

In a multi-agent scenario, the goal for each WACR is to learn the pattern of behavior of the jammer as well as other WACRs using the Q-learning algorithm. Each time the WACR gets jammed or faces an interference, it will select a new operating sub-band that has an idle bandwidth that is at least equal to β. The selected new sub-band must have low interference for the longest amount of time with high probability. Once the desired idle bandwidth condition is violated in the current sub-band due to an interferer or a jammer, the WACR will select another sub-band according to the decision policy in (2).

To evaluate its performance, the proposed cognitive anti-jamming and interference-avoidance policy in multi-agent environment is compared with a random sub-band selection scheme in which all sub-bands are selected with equal probabilities. As our performance metric, we use the normalized accumulated reward, defined as [8-11]

$$R_N = \frac{1}{N}\sum_{n=1}^{N} r(s[n], a[n]), \tag{19}$$

where $r(s[n], a[n])$ represents the immediate reward of taking action $a[n]$ when in state $s[n]$ and N is the number of iterations. Note that, the rewards in (15) are those that achieved after the convergence of the Q-table.

In our simulations for the multi-agent scenario we considered 2 test cases [10]. The first case assumes two WACRs and a sweeping jammer. The operating frequency band is taken to be from 2.0 to 2.2 GHz. This gives a total of 5 sub-bands each with a bandwidth of 40 MHz. In the second case, we assume three WACRs besides the sweeping jammer. The spectrum of interest in this case is taken to be from 2.0 to 2.4 GHz. This gives 10 sub-bands each with a bandwidth of 40 MHz. For any 2 units, having a short distance in-between, implies that the transmission of one will be received by the other with a high signal strength causing high interference impact if both are

operating on the same sub-band. We have used a continuous signal that sweeps the spectrum of interest from the lower to the higher frequency as the jammer.

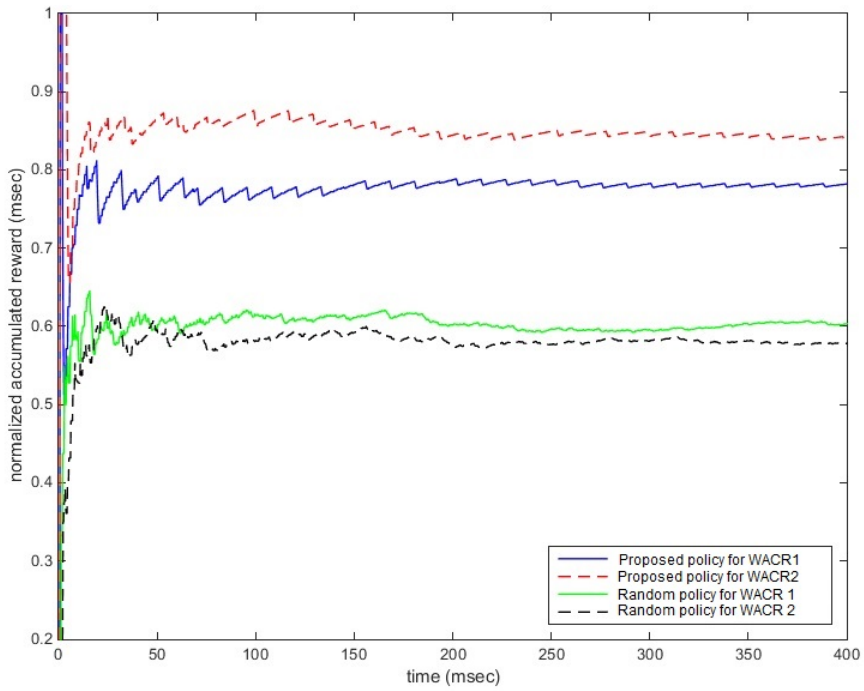

Figure 7. Normalized accumulated reward of WACR1 and WACR2: Test case 1.

Figure 7 shows the normalized accumulated reward achieved by the first and second WACR (WACR1 and WACR2) with the proposed policy and random action policy in test case 1 [10]. Note that, since there are 5 available sub-bands, the maximum immediate reward possible in this case is 1 msec. However, if we consider the interference caused by the transmission from other WACRs, it could affect the above maximum possible reward. From figure 7, the performance of the proposed policy lies somewhere between 75% to 90% of the above maximum possible reward of 1 msec. On the other hand, the random selection policy achieves only about 60% of the above maximum possible performance. These results show that the proposed policy can indeed provide noticeably better performance than simply selecting random sub-bands.

Next, we applied our proposed anti-jamming algorithm to the second test case in which there are 3 WACRs operating over 10 sub-bands. In this case, the maximum possible reward for a single WACR should be 2.25 msec since there are 10 sub-bands in the system. Figure 8 compares the

performance of the proposed policy for the first WACR (WACR1) with the random selection policy in test case 2. From figure 8 we observe that the proposed policy can achieve about 73% of the above mentioned maximum possible performance while the random selection policy can achieve only about 48%. Clearly, these results show that the proposed Q-learning based sub-band selection policy can be an effective cognitive anti-jamming and interference avoidance protocol [10].

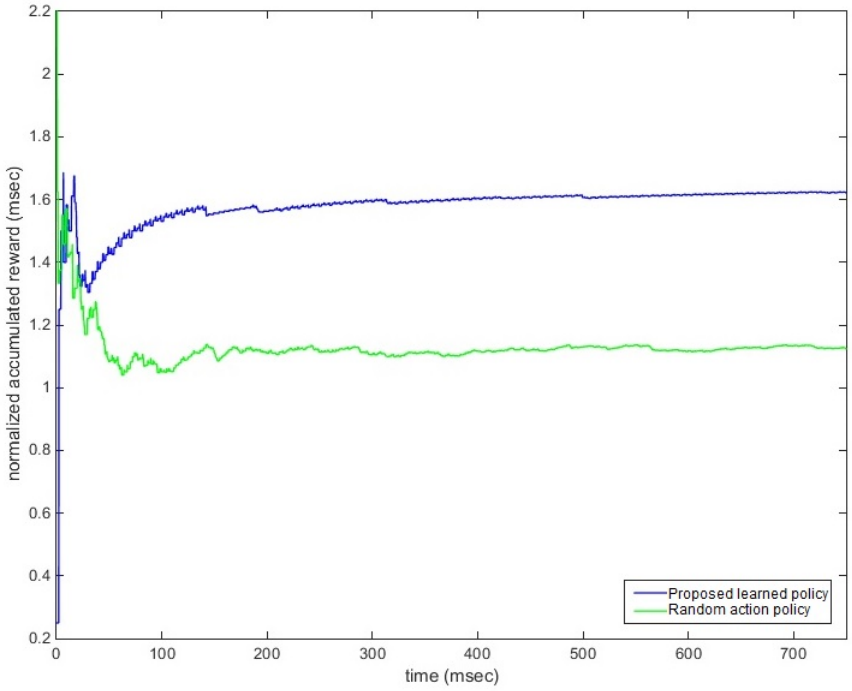

Figure 8. Normalized accumulated reward of WACR1 and WACR2: Test case 2.

Protocols to switch operating sub-band just before getting jammed/interfered

In evaluating the performance of cognitive anti-jamming and interference-avoidance stochastic game that enable the WACR to switch the operating sub-band before gets jammed/ interfered, the same performance metric in (15) is used, where $r(s[n], a[n])$ represents the immediate reward of taking action $a[n]$ when in state $s[n]$ for the transmission operation as shown in figure 4.

Table 4. Normalized accumulated reward values for different simulation scenarios (possible max reward).

Test Case	Scenario	Possible max. reward	WACR 1	WACR 2	WACR 3	WACR 4
1	1 WACR and 5 Sub-bands	4	Proposed: 3.9 Random: 2.5			
2	2 WACRs and 6 Sub-bands	4-5	Proposed: 2.9 Random: 0.5	Proposed: 3.5 Random: 0.5		
3	4 WACRs and 16 Sub-bands	11-15	Proposed: 7.8 Random: 2.1	Proposed: 9.5 Random: 1.4	Proposed: 9.5 Random: 1.4	Proposed: 8.7 Random: 1.4

Table 5. Normalized accumulated reward values for different simulation scenarios (probability of getting jammed).

Test Case	Scenario	WACR 1	WACR 2	WACR 3	WACR 4
1	1 WACR and 5 Sub-bands	Proposed: 0.92% Random: 1.79%			
2	2 WACRs and 6 Sub-bands	Proposed: 3.59% Random: 56.19%	Proposed: 5.64% Random: 56.23%		
3	4 WACRs and 16 Sub-bands	Proposed: 4.52% Random: 61.28%	Proposed: 6.12% Random: 69.37%	Proposed: 4.32% Random: 69.37%	Proposed: 8.16% Random: 67.06%

In all experiments a sweeping jammer is considered that sweeps the spectrum of interest from the lower to the higher frequency. Tables 4 and 5 summarize the obtained values of normalized accumulated reward and probability of getting jammed for all test cases, respectively [11].

The first experiment assumed a single WACR that operates over 5 sub-bands. The maximum possible reward for a single WACR is 4 time steps since there are 5 sub-bands in the system. Figure 9 shows that the proposed policy achieves about 97% of the maximum possible reward, while the random policy can achieve only about 61%.

Figure 10 shows the normalized accumulated reward for the second experiment, where 2 WACRs and 6 sub-bands were considered. The achieved accumulated reward of the proposed policy for both WACRs lies between 64% to 70% of the maximum possible reward. On the other hand, the random selection policy achieves only about 10% of the maximum possible performance.

In addition, as can be seen from Table 5, the proposed algorithm has a very low probability of getting jammed, while the random policy has a 56% probability of getting jammed.

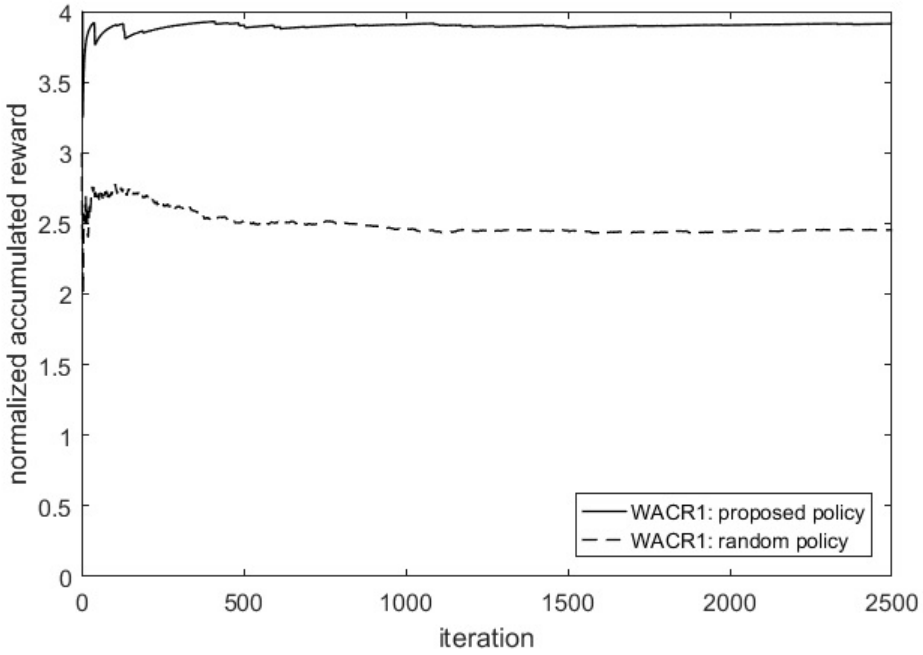

Figure 9. Normalized accumulated reward values for test case 1.

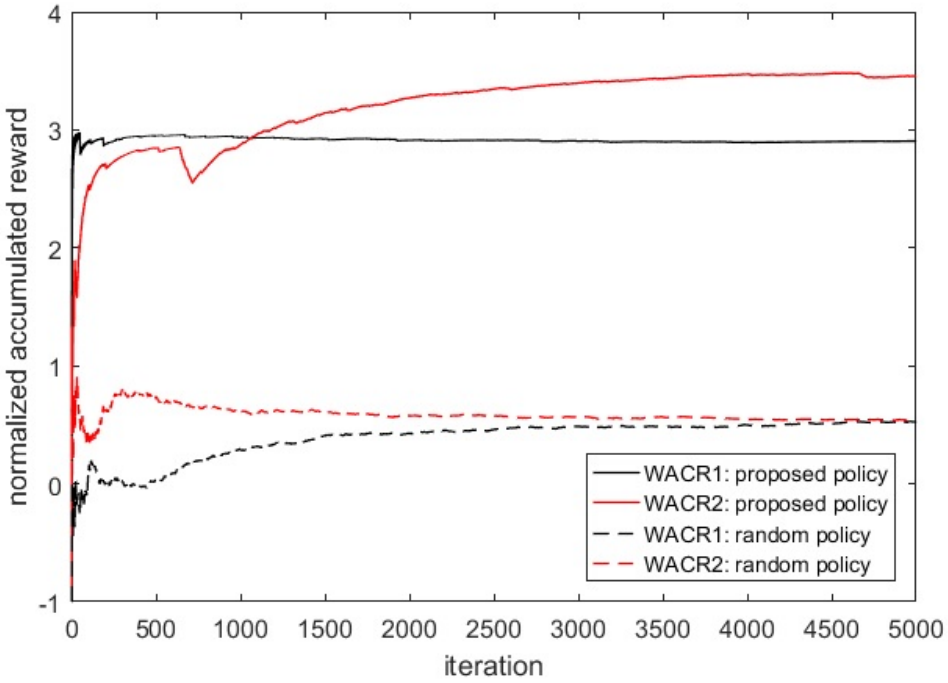

Figure 10. Normalized accumulated reward values for test case 2.

Finally, the third experiment assumed 4 WACRs operating over 16 sub-bands. As shown in Figure 11 and Table 4, the proposed algorithm achieves an acceptable normalized accumulated reward value between 58% to 73% of the maximum possible reward. The random policy on the other hand, achieves only 15% of the maximum possible reward. Moreover, the proposed policy significantly outperforms the random policy in terms of the probability of getting jammed as shown in Table 5.

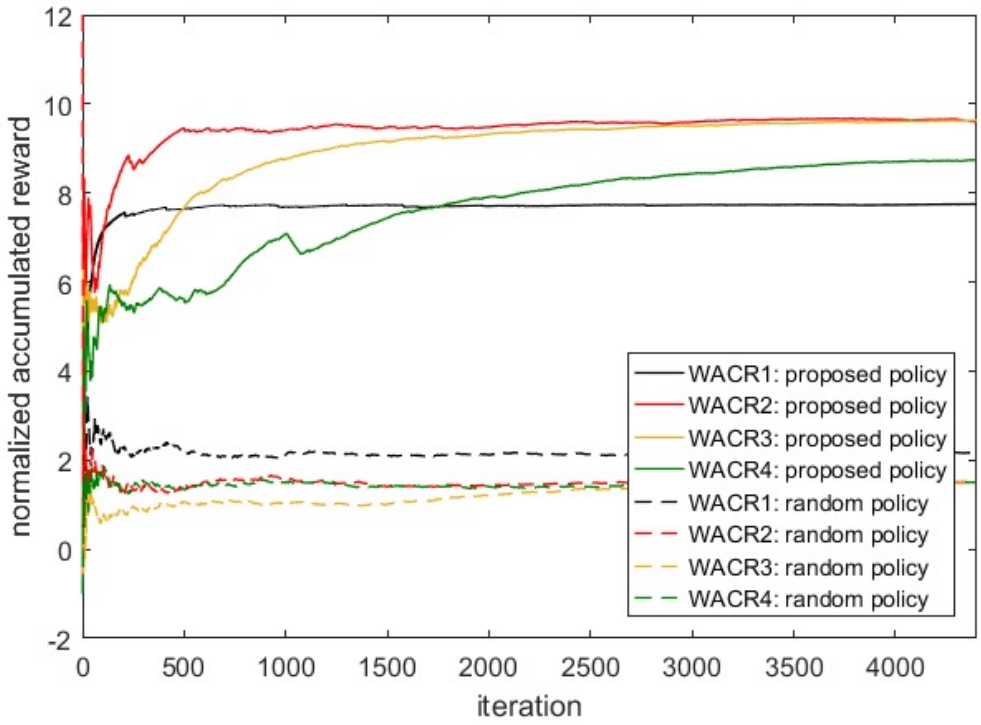

Figure 11. Normalized accumulated reward values for test case 3.

Cognitive cooperative communications:

To evaluate the performance of the exhaustive search and the proposed heuristic method to solve the data synchronization problem (6) – (9) for distributed beamforming, we considered 2 example scenarios which include multiple cooperating nodes and one destination node.

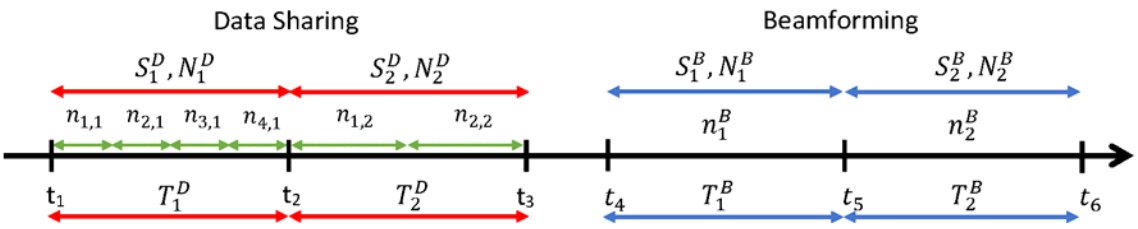

Figure 12. Timing schedule for 4 nodes in the 1st scenario.

In scenario 1, we assumed that there are 4 cooperative nodes (with $N_i = 40$ packets for $i = 1, 2, 3, 4$) and a single destination node timing schedule is shown in figure 12. The values of parameters are given in Table 6.

Table 6. The parameter values for the two simulation scenarios.

Parameter	*Value of scenario 1*	*Value of scenario 2*
M	4	3
K_1, K_2	2,2	4,4
N_1^D, N_2^D	22, 10 Packets	10, 20 Packets
N_3^D, N_4^D	N/A, N/A	20, 40 Packets
S_1^D, S_2^D	{1,2,3,4}, {1,2}	{1,2}, {1,3}
S_3^D, S_4^D	N/A, N/A	{2,3}, {1,2,3}
N_1^B, N_2^B	10, 100 Packets	10, 30 Packets
N_3^B, N_4^B	N/A, N/A	10, 30 Packets
S_1^B, S_2^B	{1,2}, {3,4}	{1,2}, {1,3}
S_3^B, S_4^B	N/A, N/A	{2,3}, {1,2,3}

Exhaustive search shows that scenario 1 in figure 12 has one optimum and fair solution with the maximum throughput of 32 packets. The optimal non-zero values of $n_{i,k}$ (in packets) are as follows: $n_{3,1} = 11, n_{4,1} = 11, n_{1,2} = 5$, and $n_{2,2} = 5$.

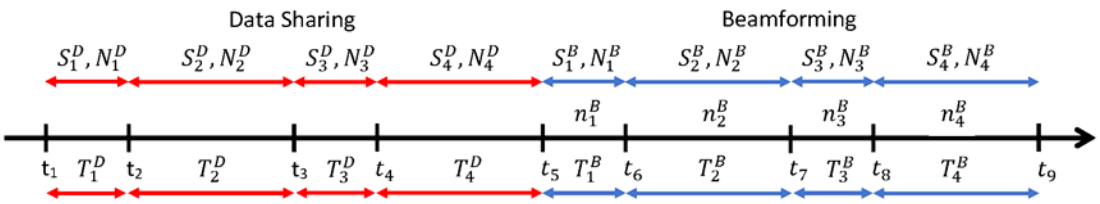

Figure 13. Timing schedule for 3 nodes in the 2nd scenario.

In scenario 2, we assume that there are 3 cooperative nodes (with $N_i = 100$ packets for for $i = 1, 2, 34$) and a single destination node. The timing schedule for the scenario 2 is shown in figure 13. The values of assumed simulation parameters are given in Table 6. This scenario has one fair and optimum solution with the maximum throughput of 80 packets. The optimum non-zero values of $n_{i,k}$ (in packets) corresponding to this solution are as follows: $n_{1,1} = 5$, $n_{2,1} = 5$, $n_{1,2} = 10$, $n_{3,2} = 10, n_{2,3} = 5, n_{3,3} = 5$, $n_{1,4} = 15$, $n_{2,4} = 10$, and $n_{3,4} = 15$.

Table 7. The total number of transmitted packets in two scenarios by the exhaustive search and the proposed heuristic method.

	Scenario 1	***Scenario 2***
Exhaustive	32	80
Heuristic	30	80

Table 7 shows the total number of transmitted packets during distributed transmit beamforming in the above two scenarios by exhaustive search and the proposed heuristic method. As it is seen, exhaustive search has the best performance (i.e., highest number of transmitted packets) with the highest computational complexity (since it checks all valid integer values of $n_{i,k}$). The heuristic method has a very low computational complexity while it has an excellent performance compared to the exhaustive search.

Cognitive multi-mode communications

To evaluate the proposed framework for user association in a multi-agent environment, we conducted several MATLAB simulations. The cognitive system models for each client node in our simulations were created using 3-dimensional feature vectors formed by the descriptors Peer identification (ID), i.e., base station ID (BSID) index, downlink SINR, and the downlink Cell Load (i.e., an average of radio resource utilization). An example is presented in figure 14, where each cluster represents regions of combinations of downlink (DL) SINR and Cell Load across two different network attachment points.

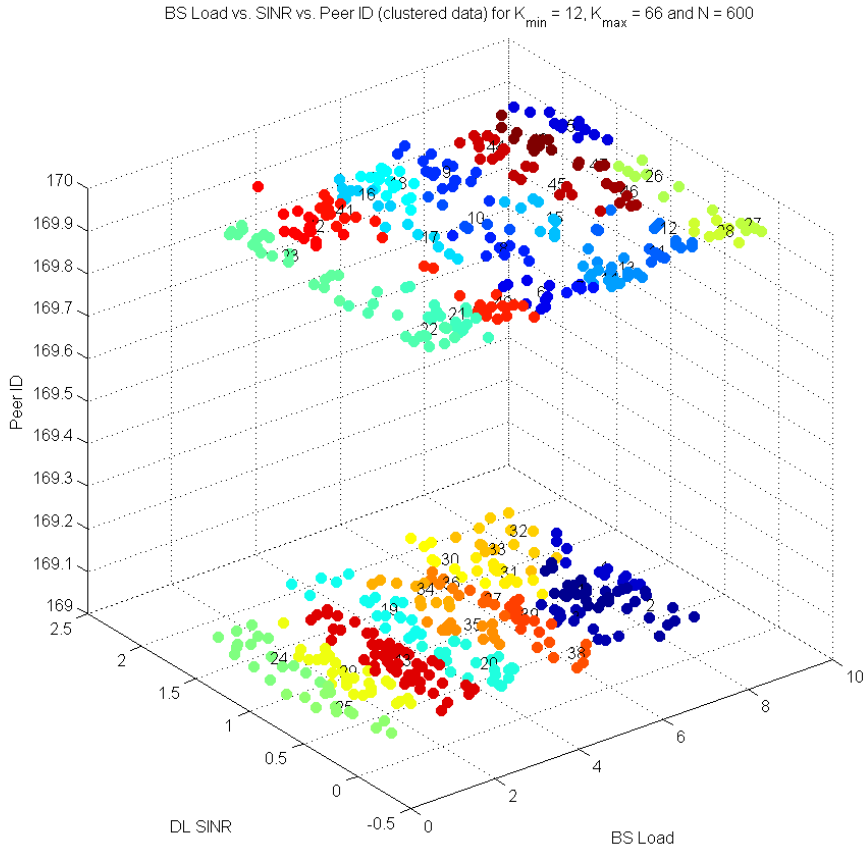

Figure 14. X-means clustering of feature vectors formed by Peer ID, SINR and BS Load. Both the SINR and BS Load values have been scaled by a factor of 0.1.

Our simulations compared the network behavior of different user association mechanisms under common RF and node mobility conditions. We restricted our implementation to the DL transmission and have assumed that there is no interference.

At each simulation time step, the SNRs of the client nodes relative to each serving node are recomputed according to (16), using the propagation parameters of table 8:

$$SNR_{dB} = P_{dBm} - PL(d)_{dB} - N_{dBm} \quad (20)$$

where, P_{dBm} is the serving node transmit power in dBm, $PL(d)_{dB}$ is the path loss of the wireless link assuming log-normal fading [26], and N_{dB} denotes the client node Noise Floor in dBm.

Table 8. Log-normal fading propagation model parameters.

Parameters	Macrocell (LTE-FDD 10 MHz)
Center Frequency (GHz)	1.8 (Band 3)
d_0 (Km.)	1.0
n	5.1
σ (dB)	8.0
P_{dBm}	40.0
N_{dBm}	-94

Let I and J denote the total number of client nodes and serving nodes in the simulation, respectively. Note that, the decisions of the i-th client node, for $i \in \{1, \ldots I\}$, associated to the j-th serving node, for $j \in \{1, \ldots J\}$, will affect its load. Let T_i^{req} be the required throughput of the i-th client node. Let $T_{ij}^{\max}$ be the mapping of SNR values to the maximum theoretical throughput according to the LTE (Long-term Evolution) standard [27]. Then, assume a time-sharing scheduling mechanism that allocates to each connected client node a time-slice according to the following proportion:

$$w_{ij} = \frac{T_{ij}^{max}}{\sum\limits_{i^{(j)}} T_{ij}^{max}} \tag{21}$$

where $\sum_{i^{(j)}}$ denotes summation over all the client nodes currently associated to the *j*-th cell. The average cell load L_j of the *j*-th cell is defined as

$$L_j = \sum_{i^{(j)}} \frac{T_i^{req}}{T_{ij}^{max}}, \tag{22}$$

In our simulations, the individual rewards obtained by the client node *i* after each decision instant *t* were computed using the following simplified form of the reward function presented above:

$$R_t^i(s_{t-1}, s_t, a_{t-1}) = \beta \cdot g(T_i^{req}, L_j) \tag{23}$$

where β = 0:001 is chosen for convenience, and

$$g(T_i^{req}, L_j) = \begin{cases} T_i^{req}, & \text{if } L_j \leq 1.0 \\ T_i^{req}/L_j, & \text{if } L_j > 1.0 \end{cases} \tag{24}$$

Figures 15 and 16 present results corresponding to one of the network simulations [25]. We have considered 2 LTE serving nodes and 3 client nodes. The client nodes were configured with high traffic rate requirements (i.e., $T_i^{req} = 40$ Mbps), and a random walk-based mobility model was assumed. In figure 15, the aggregated rewards computed by *Q*-Learning are compared with the ones obtained by a random decision mechanism and with the max-SINR rule. Figure 16 shows the comparison for the smoothed cell load is presented for each case. The smoothing operation is necessary because the cell load can be very noisy and difficult to interpret when a large number of simulation time steps are used. To reduce the effects of these fluctuations, we smoothed the original cell load values using an averaging rectangular sliding window of size *W*.

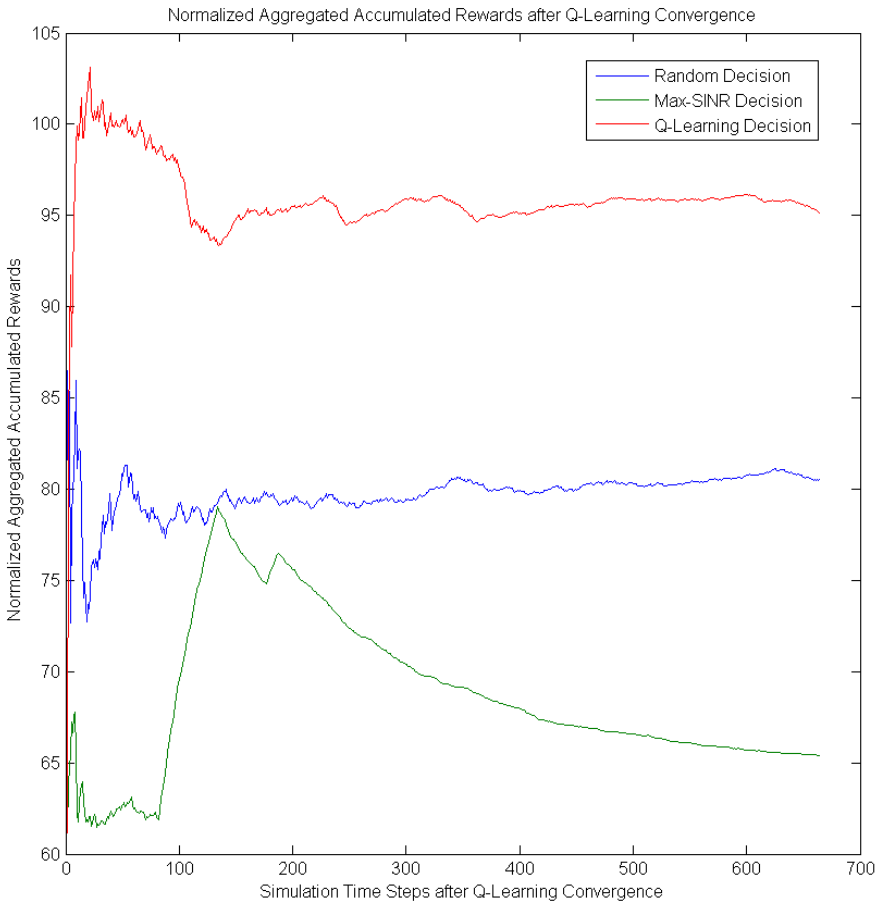

Figure 15. Simulation results for 3-client 2-serving node scenario in term of the normalized network rewards after convergence.

These results show that the Q-learning outperforms both other decision mechanisms. Q-learning obtained, on average, ~30% higher rewards than the max-SINR algorithm and ~10% higher rewards than the random decision mechanism. This is, evidently, a direct consequence of achieving a better network load balancing. Moreover, on average, the random decision mechanism generated more than 3 times the overhead of the Q-learning mechanism.

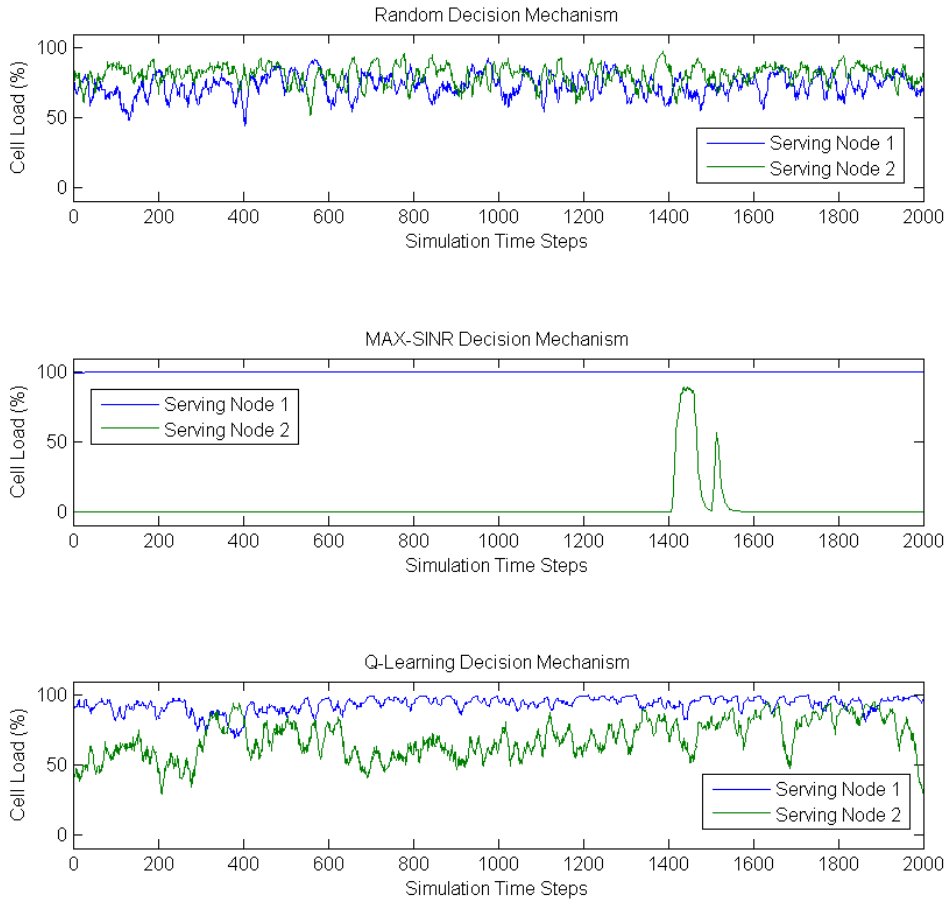

Figure 16. Simulation results for 3-client 2-serving node scenario in term of the smoothed cell load per decision mechanism using a smoothing window size of *W = 21*.

5 CONCLUSIONS

During this project, we developed two different types of cognitive anti-jamming and interference-avoidance communications protocols based on Q-learning that are suitable for multi-agent environments. In the first policy, the WACR selects a new operating sub-band that will lead to the longest possible uninterrupted transmission at any time it gets jammed or faces an interference. The simulation showed that, with the first approach, the WACR was able to successfully infer the jamming pattern and learn the optimal sub-band selection policy for jamming/interference avoidance. Moreover, it was shown through analysis and simulations that the first policy has low computational complexity and significantly outperforms the random sub-band selection policy.

The second cognitive anti-jamming and interference avoidance communications protocol allows each WACR to predict and avoid jamming attacks as well as interference from other radios. In this framework, each WACR has to perform two operations: sensing and transmission. The objective of the sensing operation is to track the jammed sub-bands. On the other hand, the transmission operation determines when to switch the operating sub-band and to where. Each operation uses *Q*-learning algorithm to learn optimal or near optimal policy. When compared with random selection policy, simulation results showed that the second policy has a very low probability of getting jammed and acceptable value for accumulated reward.

In this project, we also initiated the study of cognitive cooperative communications protocols suitable to achieve distributed transmit beamforming. As a practical approach, we divided the distributed transmit beamforming in to two stages and considered the problem of data synchronization among cooperative nodes in order to maximize the transmitted data throughput during distributed transmit beamforming. We formulated this as an optimization problem whose exhaustive search solution appears to be computationally too demanding. As an alternative, we proposed a novel heuristic method. Simulation results showed that the proposed heuristic method has a very low computational complexity while it has an excellent performance compared to the exhaustive search.

Motivated by the key attribute of WACRs as defined in [1], we also investigated cognitive multi-mode communications protocols that can be applicable in a wide range of applications including 5G networks as well as autonomous military communications systems. We developed a distributed cognitive framework based on machine learning for the RAT association problem in

the specific context of 5G HetNets. Our proposal was shown to be able to learn simple state representations out of the terminal experience and user behavior, reducing the complexity of the core network design requirements. Moreover, it allows multi-objective optimization of the association decisions while incurring minimal network overhead. Our network simulation results showed the benefits of the proposed framework compared to alternative decision methods in a multi-agent environment. These results show that the proposed machine-learning based approach can indeed be the basis for highly adaptable multi-mode cognitive communications protocols suitable for autonomous satellite and military communications networks.

6 RECOMMENDATIONS

The machine-learning aided cognitive anti-jamming communications protocols developed and evaluated during this project showed that WACR may use multi-agent reinforcement learning (MARL) to obtain effective and efficient decision policies to the underlying stochastic games. These promising results justifies further research on the wideband autonomous cognitive radio technology proposed by the University of New Mexico as a viable candidate for future space and satellite communications of interest to the Air Force [1]. Thus, it is recommended that further basic research on WACR technology and cognitive communications protocols as well as applied Research & Development (R&D) efforts in translating this technology to develop prototypes be continued.

The cognitive cooperative communications protocols, whose study was initiated during this project, are some of the most promising future research directions. We have laid out a roadmap for practical implementation of distributed transmit beamforming and provided an example solution approach for data sharing problem that is at the first stage of this approach. It is highly recommended that this be pursued to develop practical distributed transmit beamforming protocols that can be implemented on WACRs proposed by UNM in order to meet the needs of future Air Force and Department of Defense (DoD)satellite communications systems.

Finally, we demonstrated that machine-learning based approaches can be the basis for highly adaptable multi-mode cognitive communications protocols suitable for autonomous satellite and military communications networks. Development of the proposed WACR technology

can thus provide a viable and robust solution to the long-standing problem of interoperability of communications systems among the DoD community.

As more and more nations develop advanced communications and radar technologies, the combination of traditional as well as evolving spectrum challenges requires future communications technologies to be intelligent, self-aware and spectrally agile. The wideband autonomous cognitive radios proposed in [1] and pursued in this project are radios with these defining characteristics that will allow autonomous radio communications over non-contiguous wide spectrum bands in the presence of adverse conditions. As a result, our proposed WACR technology has the potential to revolutionize the future communications systems. Thus, it is recommended that further research on all aspects of WACR technology proposed by University of New Mexico (UNM) be pursued in order to bring this promising communications technology to maturity.

REFERENCES

1. S. K. Jayaweera, Signal Processing for Cognitive Radios, 1st ed. New York, NY, John Wiley & Sons Inc., 2014.

2. S. K. Jayaweera and C. G. Christodoulou, "Radiobots: Architecture, Algorithms and Realtime Reconfigurable Antenna Designs for Autonomous, Self-learning Future Cognitive Radios," University of New Mexico, Albuquerque, NM, EECE-TR-11-0001, Mar. 2011.

3. S. K. Jayaweera, Y. Li, M. Bkassiny, C. G. Christodoulou and K. A. Avery, "Radiobots: The autonomous, self-learning future cognitive radios," IEEE Intelligent Sig. Proc. and Commun. Systems (ISPACS'2011), Chiangmai, Thailand, pp. 1- 5, Dec. 2011.

4. M. Bkassiny, S. K. Jayaweera, Y. Li, and K. A. Avery, "Wideband spectrum sensing and non-parametric signal classification for autonomous self-learning cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2596–2606, July 2012.

5. Y. Li, S. K. Jayaweera, M. Bkassiny and C. Ghosh, "Learning-aided Sub-band Selection Algorithms for Spectrum Sensing in Wide-band Cognitive Radios," *IEEE Trans. in Wireless Commun.,* vol. 13, no. 4, pp. 2012 – 2024, April 2014.

6. M. Bkassiny, S. K. Jayaweera, and Y. Li, "Multidimensional Dirichlet Process-based Non-Parametric Signal Classification for Autonomous Self-Learning Cognitive Radios," *IEEE Trans. in Wireless Commun*., vol. 12, no. 11, pp. 5413-5423, Nov. 2013.

7. P. Das and S. K. Jayaweera, "Robust wideband spectrum sensing with compressive sampling in cognitive radios," IEEE Vehicular Technology Conference (VTC Fall), Boston, MA, pp. 1- 5, Sep. 2015.

8. S. Machuzak and S. K. Jayaweera, "Reinforcement learning based anti-jamming with wideband autonomous cognitive radios," IEEE/CIC International Conference on Communications in China (ICCC), Chengdu, China, pp. 1- 5, July 2016.

9. M. A. Aref, S. Machuzak, S. K. Jayaweera and S. Lane, "Replicated Q-learning based sub-band selection for wideband spectrum sensing in cognitive radios," IEEE/CIC International Conference on Communications in China (ICCC), Chengdu, China, pp. 1-6, July 2016.

10. M. A. Aref, S. K. Jayaweera and S. Machuzak, "Multi-agent Reinforcement Learning Based Cognitive Anti-jamming," IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, CA, pp. 1- 6, Mar. 2017.

11. M. A. Aref, and S. K. Jayaweera, "A Novel Cognitive Anti-jamming Stochastic Game," IEEE Cognitive Communications for Aerospace Applications Workshop (CCAA), Cleveland, OH, pp. 1- 4, June 2017.

12. M. A. Aref and S. K. Jayaweera, "A Cognitive Anti-jamming and Interference-Avoidance Stochastic Game," 16th IEEE International Conference on Cognitive Informatics & Cognitive Computing, Oxford, UK, July 2017.

13. R. Mudumbai, D. R. Brown III, U. Madhow and H. V. Poor, "Distributed transmit beamforming: challenges and recent progress," *IEEE Commun. Mag*., vol. 47, no. 2, pp. 102-110, Feb. 2009.

14. Y.-S. Tu and G. J. Pottie, "Coherent cooperative transmission from multiple adjacent antennas to a distant stationary antenna through AWGN channels," in Proc. IEEE 55th Veh. Technol. Conf. (VTC-Spring), Birmingham, AL, pp. 130-134, 2002.

15. H. ElSawy, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Virtualized cognitive network architecture for 5G cellular networks," *IEEE Communications Magazine*, vol. 53, no. 7, pp. 78–85, July 2015.

16. J. G. Andews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in HetNets: Old myths and open problems," *IEEE Communications Magazine*, vol. 21, no. 2, pp. 18–25, April 2014.

17. V. Yazici, U. C. Kozat, and M. O. Sunay, "A new control plane for 5G Network architecture with a case study on unified handoff, mobility and routing management," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 76–85, November 2014.

18. M. Zorzi, A. Zanella, A. Testolin, M. de Filippo de Grazia, and M. Zorzi, "Cognition-Based Networks: A new perspective on Network Optimization using Learning and Distributed Intelligence," *IEEE Access*, vol. 3, pp. 1512–1530, September 2015.

19. C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the Internet of Things: A survey," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 1, pp. 414–454, First Quarter 2014.

20. B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5G Era," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 90–96, February 2014.

21. F. Gao and K. Zhang, "Enhanced multi-parameter cognitive architecture for Future Wireless Communications," *IEEE Communications Magazine*, vol. 53, no. 7, pp. 86–92, July 2015.

22. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

23. C. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992.

24. A. Ghasempour, and S. K. Jayaweera, "Data Synchronization for Throughput Maximization in Distributed Transmit Beamforming," IEEE Cognitive Communications for Aerospace Applications Workshop (CCAA), Cleveland, OH, June 2017.

25. J. A. Perez, S. K. Jayaweera, and S. Lane, "Machine Learning Aided Cognitive RAT Selection for 5G Heterogeneous Networks," IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Istanbul, Turkey, June 2017.

26. B. Sklar, The Mobile Communications Handbook, 2nd ed., New York: CRC Press and IEEE Press, 1999, ch. 18.

27. 3GPP TS 36.213, Physical layer procedures (Release 9), Rev. V9.3.0, September 2010.

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

5G	5th Generation
API	Application Programming Interface
B	Bandwidth
BS	Base Station
BSID	Base Station ID
DL	Downlink
DoD	Department of Defense
HetNets	Heterogeneous Networks
ID	Identification
kNN	*k*-Nearest Neighbors
LTE	Long-term Evolution
MARL	Multi-agent Reinforcement Learning
MDP	Markov Decision Process
NP	Neyman-Pearson
P_T	Transmitted Power
QoS	Quality of Service
R	Packet Rate
RAT	Radio Access Technology
R&D	Research & Development
RF	Radio Frequency
RFI	Radio Frequency Interference
RL	Reinforcement Learning
SINR	Signal-to-Interference-plus-Noise-Ratio
SNR	Signal-to-Noise-Ratio
UNM	University of New Mexico
WACR	Wideband Autonomous Cognitive Radio

DISTRIBUTION LIST

DTIC/OCP
8725 John J. Kingman Rd, Suite 0944
Ft Belvoir, VA 22060-6218 1 cy

AFRL/RVIL
Kirtland AFB, NM 87117-5776 2 cys

Official Record Copy
AFRL/RVSW/James Lyke 1 cy

(This page intentionally left blank)