Cognitive Anti-jamming Satellite-to-Ground Communications on NASA's SCaN Testbed

Sudharman K. Jayaweera[1,2], Shuang Feng[1], Dale Mortensen[3], Abriel Holland[1], Marie Piasecki[3], Mike Evans[3], Christos Christodoulou[1,2]
[1]Bluecom Systems and Consulting LLC, Albuquerque, NM
[2]Communications and Information Sciences Lab (CISL), Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM
[3]NASA Glenn Research Center, Cleveland, OH

***Abstract*—Machine learning aided cognitive anti-jamming communications is designed, developed and demonstrated on a live satellite-to-ground link. A wideband autonomous cognitive radio (WACR) is designed and implemented as a hardware-in-the-loop (HITL) prototype. The cognitive engine (CE) of the WACR is implemented on a PC while the software-defined radio (SDR) platform utilized two different radios for spectrum sensing and actual communications. The cognitive engine performs spectrum knowledge acquisition over the complete spectrum range available for the SATCOM system operation and learns an anti-jamming communications protocol to avoid both intentional jammers and inadvertent interferers using reinforcement learning. When the current satellite-to-ground link is jammed, the cognitive engine of the ground receiver directs the satellite transmitter to switch to a new channel that is predicted to be jammer-free for the longest possible duration. The end-to-end, closed-loop system was tested on the NASA Space Communications and Navigation (SCaN) Testbed on the International Space Station (ISS). The experimental results demonstrated the feasibility of satellite-to-ground cognitive anti-jamming communications along with excellent anti-jamming capability of machine learning aided cognitive protocols against several different types of jammers.**

***Index Terms*—Cognitive anti-jamming communications, cognitive radios, machine learning, Q-learning, reinforcement learning, satellite communications, wideband autonomous cognitive radios.**

I. Introduction

In [1]–[3], wideband autonomous cognitive radios (WACRs) were proposed as radios that can self-configure the mode of operation in response to the given state of the overall system consisting of the radio, spectrum and the end-user. These inherent capabilities make this technology especially suited for various military, satellite, space and homeland security applications. Indeed, as proliferation of wireless telecommunications skyrocket, spectrum awareness and agility enabled by WACR technology can be essential to both coexist with friendly users as well as to counter or defeat malicious and hostile agents.

Among many potential applications of WACR technology is the cognitive anti-jamming communications. In [4]–[9], several machine learning aided cognitive anti-jamming communications protocols have been proposed earlier. Essentially, all these approaches rely on the spectrum knowledge acquisition capability of WACRs to scan the total spectrum of interest and learn an effective anti-jamming communications policy. The reinforcement learning based protocols of [4]–[7] allow these policies to be updated in realtime possibly leading to highly responsive anti-jamming systems. However, almost all previously reported performance of these cognitive anti-jamming communications protocols are based on simulations and simple laboratory tests of one end of a link. In particular, to the best of our knowledge, there has not been any effort to implement an end-to-end closed loop cognitive anti-jamming systems. In this paper, we report the design and development of a WACR system for satellite-to-ground communications and results of live closed-loop testing of the developed system on the NASA Space Communications and Navigation (SCaN) Testbed on the International Space Station (ISS).

The rest of the paper is organized as follows: Section II provides a brief introduction to the wideband autonomous cognitive radio technology along with the system architecture developed for the particular satellite-to-ground cognitive anti-jamming communications. Section III describes the cognitive engine design for the cognitive anti-jamming communications. Section IV details the live real-time testing of the developed cognitive anti-jamming communications system on the NASA SCaN Testbed and the experimental results. Section V provides an analysis of the results to demonstrate the feasibility and effectiveness of cognitive anti-jamming communications for SATCOM. Finally, the paper is concluded in Section VI.

II. Wideband Autonomous Cognitive Radio Technology

The overall concept of wideband autonomous cognitive radios, as envisioned in [1], [2], is shown on Fig. 1. It is made of a reconfigurable RF front-end, a software-defined radio (SDR) baseband module and a cognitive engine (CE). The cognitive engine acts as the brain of the WACR by managing the overall cognitive and intelligent operation of the radio. The defining feature of a cognitive radio is its spectrum knowledge acquisition functionality which is commonly referred to as spectrum sensing [1]. Note that the spectrum knowledge acquisition may cover a broader functionality than simple whitespace detection as meant in literature on spectrum sensing [10].

As can be seen from Fig. 1, cognitive processing performed within the cognitive engine can be divided in to two parts:

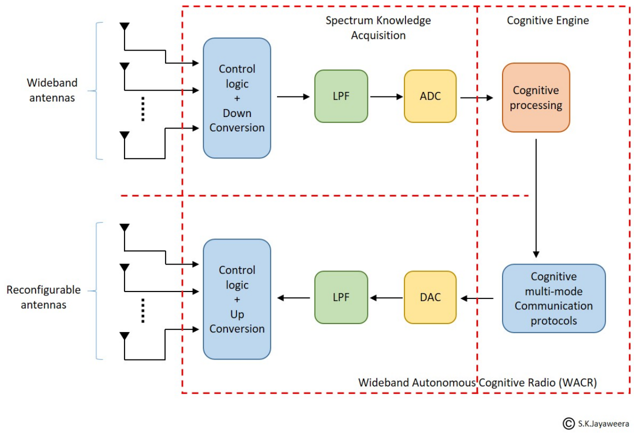

Fig. 1: The concept and basic architecture of a wideband autonomous cognitive radio.

spectrum knowledge acquisition and cognitive communications protocols. Spectrum knowledge acquisition deals with gaining knowledge and comprehension about the states of the RF environment, network, radio and the user [1]. The cognitive communications protocols utilize this knowledge to decide and act in order to achieve user communications objectives. This module is responsible for issuing instructions to both the SDR and the RF front-end on how to reconfigure their modes of operations and parameters in response to the interpreted states of the RF environment, radio network, WACR itself and user. Thus, the key to cognitive operation is the design of the cognitive engine.

In this work, a specific WACR system is designed for the purpose of cognitive anti-jamming communications. The system is comprised of an SDR that is controlled by the cognitive engine implemented on a host PC.

III. Cognitive Engine Design for Cognitive Anti-jamming Communications

The cognitive engine consists of two parts: the spectrum knowledge acquisition and cognitive anti-jamming communications protocol operation. It is assumed that the SDR may only sense a single channel out of the total set of channels available. The operation of the WACR is divided into two stages: training and live communications. During training mode, the cognitive engine performs only the spectrum sensing and cognitive policy learning. In the developed system, this involves the radio randomly picking a channel for sensing and awaiting till the channel is jammed or interfered with. Once the channel is jammed, the cognitive engine instructs the SDR to sense another randomly picked channel while it uses the time it took for the channel to get jammed as a measure of a reward to learn an effective cognitive anti-jamming policy.

In this work, the policy learning was based on a modified version of the Q-learning algorithm [11], [12]. As has been shown previously in [4]–[6], reinforcement learning approaches such as Q-learning can be highly effective in learning good anti-jamming communications policies. In the current cognitive engine design, the anti-jamming policy learning problem was defined as learning in an RF environment that undergoes state changes. For the purpose of learning, the index of the WACR's current channel is defined as the state. The set of actions available to the radio is the total set of available channels. During the training mode, once the current operating channel $s_t = s$ is jammed, the cognitive engine selects a random channel index $a_t = a$ as the new channel. If the reward associated with this choice is r by the time the new channel gets jammed, then the cognitive engine updates a Q-table using the Watkin's Q-learning algorithm as follows [1], [11]:

$$Q(s,a) \;=\; (1-\alpha)Q(s,a) + \alpha\left(r + \gamma \max_{a'} Q(a,a')\right) \quad (1)$$

where $\alpha \in [0,1]$ and $\gamma \in [0,1]$ are the learning rate and the forgetting factor. At the end of the training period, the cognitive engine extracts an anti-jamming policy from the learned Q-table:

$$\pi(s) \;=\; \arg\max_{a'} Q(s,a')$$

The advantage of this reinforcement learning approach is that it allows this policy to be continuously updated during the live communications. Indeed, when the satellite-to-ground communications link is determined to be jammed by the cognitive ground receiver, it updates the corresponding Q-table entry using the same Q-learning (1). In this case, if we denote the current Q-table at time t by $Q_t(.,.)$, then the most up to date cognitive anti-jamming policy at time t is given by

$$\pi_t(s) \;=\; \arg\max_{a'} Q_t(s,a').$$

In order to prevent reinforcement learning getting trapped in a local optima and facilitate the policy learning over all possible regions of the state-action space, the actual action selection also incorporates a possible exploration rate $\epsilon \in [0,1]$ [1], [4], [12]:

$$a_t(s) \;=\; \begin{cases} \pi_t(s) & \text{with probability } 1-\epsilon \\ \mathcal{U}(\mathcal{A}\backslash\{s\}) & \text{with probability } \epsilon \end{cases} \quad (2)$$

where state at time t is $s_t = s$ and $\mathcal{U}(\mathcal{A})$ denotes the uniform distribution over the action set $\mathcal{A}$.

IV. Live Cognitive Anti-jamming Communications on NASA's SCaN Testbed

The SCaN Testbed is an externally mounted payload onboard the ISS housing three software defined radios capable of reconfiguration to support a variety of experiments. As shown in Fig. 2, fixed and steerable antennas on the SCaN Testbed multiplex with the radios enclosed to support links with other satellites as well as ground stations. The Testbed operates at S- and Ka-band frequencies, but Ka-band is not available for direct-to-ground links [13]. One objective of SCaN Testbed is to demonstrate the Space Telecommunications Radio System (STRS) Architecture, which promotes the reuse of communications software [14]. There is a growing library of STRS radio

Fig. 2: NASA's SCaN Testbed on board the ISS.

waveform applications that have been developed for SCaN Testbed SDRs.

In addition to SCaN Testbed, NASA Glenn Research Center (GRC) also operates an S-band ground station, which can support Testbed experiments. For this WACR experiment a temporary antenna was set up for the jammer signal in proximity to the main tracking antenna as shown in Fig. 3. This jammer was configured to emulate a fixed terrestrial signal source that is uncooperative with the ground station receiving dish. Ground terminals can be particularly sensitive to this type of interference as signals from these sources are often stronger than ones from the space asset. Even when the antenna is not pointed directly at the jamming source, interference can still occur due to the antenna sidelobes or multipath. Each low-earth orbit (LEO) satellite pass is approximately eight minutes long. During these relatively short passes interference of this type can cause a significant decrease in the quality of service (QoS). Figure 4 shows the power received at the ground station from both SCaN Testbed and the jammer for the same gimbal motion. During the pass there are periods where the jammer is stronger than the desired signal and vice versa. This variation throughout the pass can be challenging for a human operator to mitigate as the interference would appear unpredictable. Application of a CE for this volatile environment will potentially improve satellite reception.

A live cognitive satellite-to-ground anti-jamming communications experiment was performed using an SDR that is currently available on the SCaN Testbed. The testing utilized 5 MHz of licensed spectrum in S-band that was divided into nine 500 kHz-wide channels with a small guard band on each end, as shown in Figure 5. This experiment required that the previously-existing adaptive coding and modulation system described in [15], which includes a Digital Video Broadcasting Second Generation (DVB-S2) compliant waveform, be modified to support rapid changes in downlink carrier frequency offset based on command parameters embedded within the uplink data stream. Throughout the experiment, the following

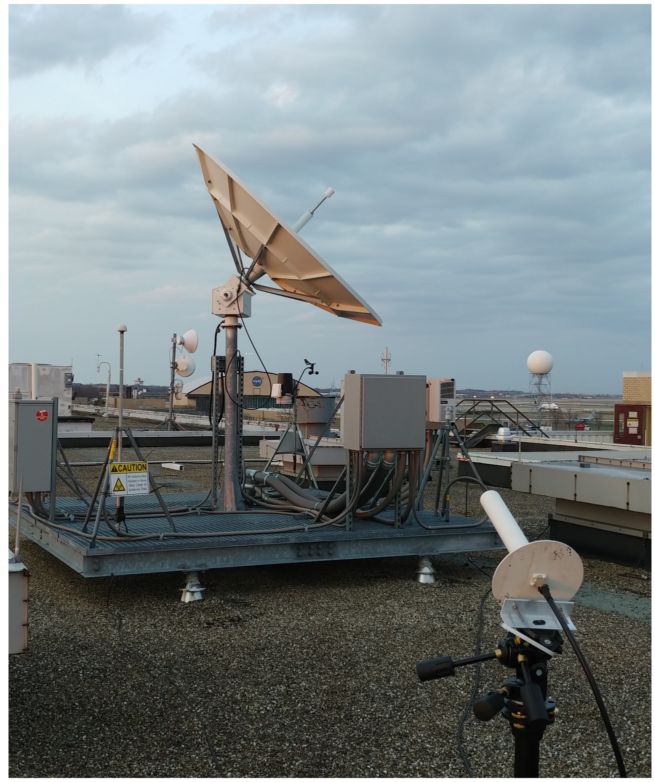

Fig. 3: NASA GRC Ground Station with jamming antenna (foreground).

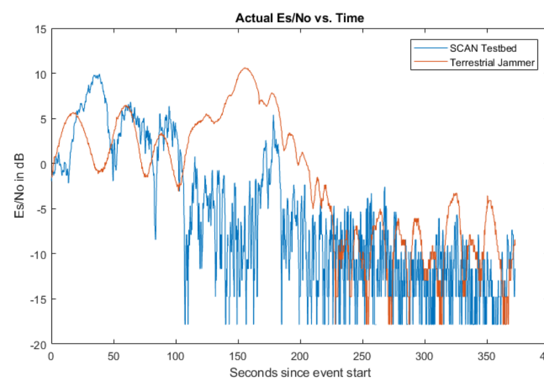

Fig. 4: Example plot showing satellite and jammer power versus time.

physical layer downlink parameters were held static: rate-1/4 QPSK with short frames (16200 bits), channel symbol rate of 300 kBaud, pilot symbol insertion enabled, and a square-root raised cosine filter rolloff of 0.20.

Figure 6 shows the closed-loop, end-to-end cognitive satellite-to-ground communications systems architecture. The key element of this architecture is the Bluecom Systems Cognitive Engine, Radiobot 1.0. For the purpose of this live SCaN Testbed experiment, the Cognitive Engine was implemented in software on a host PC. The WACR System was completed by combining the software-implemented Radiobot 1.0 CE with

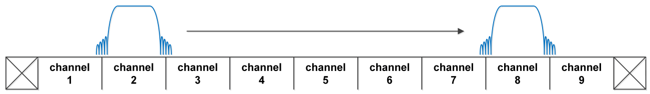

Fig. 5: Frequency plan for the SCaN Testbed cognitive satellite-to-ground communications experiment.

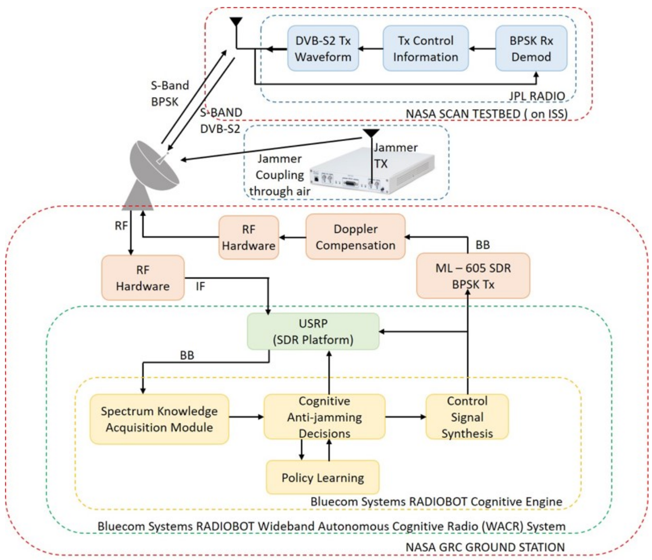

Fig. 6: Experimental Set up of the cognitive satellite-to-ground communications systems architecture.

the SDR platform. In the absence of a single commercial SDR that supports both spectrum sensing as well as cognitive satellite communications, the SDR platform of the WACR was achieved by using two SDRs: A National Instruments USRP 2953R was used as the SDR platform for the spectrum knowledge acquisition branch of the WACR, while a NASA GRC-developed radio based on the Xilinx ML605 evaluation board, was used as the SDR platform on the cognitive satellite communications branch, as shown in Fig. 7.

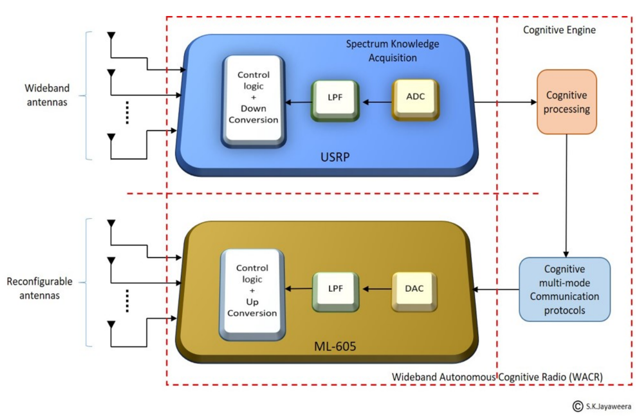

Fig. 7: WACR system made of two SDR modules for cognitive satellite-to-ground communications on the SCaN Testbed.

As can be seen from Fig. 6, the ML605 SDR is capable of transmitting a binary phase-shift keying (BPSK) waveform over the S-band frequency range. The ML605 provides an Ethernet interface to the Host PC running the Bluecom Cognitive Engine. Cognitive decisions made by the Radiobot CE are forwarded to the ML605 as a channel index, which then transmits these over the uplink to the SCaN Testbed SDR instructing it to switch to this new channel in order to help the ground receiver avoid getting jammed.

The spectrum sensing is performed with the aid of a USRP 2953R SDR. In the so-called switch-after-getting jammed configuration of the cognitive anti-jamming implementation, the Radiobot Cognitive Engine listens to the current channel on which the SCaN Testbed SDR is transmitting to the ground receiver. In order to allow the spectrum sensing to be achieved without knowing the exact details of the waveform, and in particular, without having to demodulate the signal, the Radiobot Cognitive Engine only assumed the knowledge of the communication's link(DVB-S2) modulation type and the order. Based on this information, the Radiobot Cognitive Engine was able to determine when the current channel is jammed or interfered with. It used this information to both learn a cognitive anti-jamming communications policy as well as to instruct the SCaN Testbed radio that it must switch its operating channel to a new channel. The cognitive anti-jamming communications protocol, implemented on Radiobot 1.0 cognitive engine, allows the ground station receiver to learn an effective anti-jamming communications policy so that this new channel will most likely stay jammer-free for the longest possible duration.

Live testing of the cognitive anti-jamming communications system was conducted at NASA GRC in Cleveland, Ohio during two weeks in April 2018 by a project team of Bluecom Systems' and NASA GRC's personnel. Two types of jammer signals were considered for these tests: a sweep jammer that sequentially jams the channels and a Markov jammer that jams channels according to a certain probability distribution over the channel indices. Both jammers were synthesized in software and transmitted using an independent SDR located near the satellite ground station antenna (Fig. 3). Note that the jammer antenna is a fixed antenna, whereas ground station satellite antenna is gimbaled. Indeed, the ground station's antenna orientation varies during each pass of the ISS in order to track the signal from the SCaN Testbed SDR.

Table I summarizes the anti-jamming performance data recorded during the above tests. All time data is scaled such that one time-unit equals one sensing period (SP) of the cognitive engine. The number of channel transitions indicates how many times the Radiobot switched channels during an event-pass due to signal quality falling below a certain specified quality which it interprets as channel being jammed. The total number of sensing periods with sufficient signal quality between channel transitions is a measure of the total time the WACR was able to receive the signal from the SCaN Testbed above this specified quality.

As can be seen from the Test Plan in Table I, the learning performance of the Radiobot was to be evaluated by comparing with a random channel selection anti-jamming policy. It is,

Test #	Description	Total Number of SPs During the Complete Event-pass	Total Number of SPs with Sufficient Signal Quality Between CH Transitions	Number of CH Transitions
1	Jammer type: sweep jammer. Policy type: a random policy.	214710	29380	21
2	Jammer type: sweep jammer. Policy type: a random policy.	218545	96337	77
3	Jammer type: sweep jammer. Policy type: a pre-learned CAJ policy with $\epsilon = 0.3$.	235192	132380	81
4	Jammer type: sweep jammer. Policy type: a pre-learned CAJ policy with $\epsilon = 0.3$.	120370	298	4
5	Jammer type: Markov jammer. Policy type: a random policy.	192751	51412	67
6	Jammer type: Markov jammer. Policy type: a pre-learned CAJ policy with $\epsilon = 0$.	229520	72908	79
7	Jammer type: Markov jammer. Policy type: a pre-learned CAJ policy with $\epsilon = 0.3$.	266661	112660	115

TABLE I: Raw test data from the SCaN Testbed experiment on cognitive anti-jamming (CAJ) communications.

however, important to point out that the random channel selection policy does not mean a traditional radio. Although the channel selection policy is random, still the radio is a WACR and there are some very important cognitive functions overseen by the Radiobot cognitive engine in order for the radio to achieve successful channel transitions even in this case. Specifically, channel transitions are triggered as a result of the WACRs spectrum sensing that allows it to detect the jammer existence. Without this cognitive operation, the radio may not have achieved channel transitions immediately when it is jammed. Hence, the performance comparison against the random channel selection policy must be understood only as a comparison of the effectiveness of the learning process rather than the effectiveness of the cognitive communications operation through the Radiobot. Both policies are enabled by the Radiobot cognitive engine, and thus the performance of all tests are a validation of the applicability of the WACR to achieve successful cognitive anti-jamming communications.

For all flight tests, a Bluecom software synthesized 64QAM signal at a rate of 200 kSymbols/sec was used as the jammer signal. In the case of the sweep jammer, the synthesized signal was made to sweep across the 9 channels in the order of channel 1 to 9 sequentially. Similarly, in the case of the Markov jammer, the synthesized signal was made to hop across the 9 channels according to a fixed Markov pattern. These jammer signals were transmitted from a separate USRP 2953R that was located at a different location on the NASA GRC from where the Radiobot cognitive engine and its associated USRP 2953R were located. The jammer signal from this secondary USRP 2953R was then transmitted over the air from an antenna located near the SCaN Testbed Ground Station Receiver.

V. Analysis of Satellite-to-Ground Cognitive Anti-jamming Communications Experiment Results

First of all, as can be observed from the Table I, Test #4 had corrupted results. This was due to the presence of an unexpected fixed interference in the frequency range that the cognitive anti-jamming experiment was performed during this particular event-pass. It was later determined that this interference was from another satellite transmitter on-board the SpaceX Dragon spacecraft docked to ISS. Since its presence could only be detected once the event-pass at GRC started and since these event-passes are very short, it did not allow sufficient time to request turning off the transmission, rendering Test #4 results useless. Hence, in the following discussion we will ignore the Test #4 data. However, it must be pointed out that, such static interference can easily be handled by the developed cognitive anti-jamming communications protocol with a small modification to the algorithm that was not implemented at the time of this experiment.

Table II uses the raw data in Table I to compute metrics that quantify the performance of the cognitive anti-jamming communications. Note that, the average time between two channel transitions is not a direct measure of the effectiveness of cognitive anti-jamming algorithm. The reason is that it does not show us exactly how much of that time the cognitive ground receiver was actually able to receive the signal from the SCaN Testbed. It is possible that the average time between two channel transitions is high in certain scenarios giving the impression that the cognitive ground receiver was selecting channels that did not get jammed for a long period. However, this interpretation is incorrect because the ground receiver will stay in a new channel until it receives the satellite signal correctly. Hence, it is possible that the average time between two channel transitions is high because the receiver had to wait in a new channel for a longer time before it first was able to receive the signal from the SCaN Testbed correctly. This may be an indication that actually the policy was so bad that the

Test #	Jammer Type	Policy Type	Avg Time Between Two CH Transitions	Avg Time in a CH Without Being Jammed	Avg Fraction of Time in a CH Without Being Jammed
1	Sweep.	Random policy.	10224	1399	0.1368
2	Sweep.	Random policy.	2838	1251	0.4408
1 & 2	Sweep	Random policy.	6531	1325	0.2029
3	Sweep.	Pre-learned and continuously updated through exploration with $\epsilon = 0.3$.	2904	1634	0.5627
5	Markov.	Random policy.	2877	767	0.2666
6	Markov.	Pre-learned and fixed.	2905	922	0.3174
7	Markov.	Pre-learned and continuously updated through exploration with $\epsilon = 0.3$	2319	980	0.4226

TABLE II: Performance evaluation of SCaN Testbed experiment on cognitive anti-jamming communications.

channels it selected were either already jammed or got jammed immediately when it starts to receive on it. Hence, in Table II, we have proposed two additional performance metrics to better quantify the cognitive anti-jamming performance: the first is the average time the WACR stays in a channel without its signal being jammed (average time in a channel without being jammed). The second metric is the average fraction of time the radio spends in a channel without being jammed with respect to the average time between two channel transitions.

Both these metrics are needed to understand the full impact of an anti-jamming policy. The reason that the average time in a channel without being jammed is not adequate to fully characterize the performance should also be clear from the above discussion regarding why the average time between two channel transitions is not adequate as a performance metric. Indeed, it is possible that when the radio switches to a channel, it may find that channel already jammed. In such a situation, the radio may spend a considerable amount of time waiting in that channel till the channel becomes free of jamming. Once this happens, perhaps the radio may be able to stay in that channel for a significant amount of time without again getting jammed. This may result in a large average time in a channel without being jammed. However, it does not tell anything about how much time it wasted waiting in the channel for the channel to be free of jamming to begin with. It is reasonable to expect that a good channel selection policy will make the radio to switch to a new channel that not only leads to a longer average time in a channel without being jammed, but also to a shorter wait-time for the channel to be actually jammer-free once the radio switches to it. Hence, in addition to the average time in a channel without being jammed, in Table II we also compute the average fraction of time the radio spends in a channel without being jammed with respect to the average length of time the radio spends in a channel (which is the time between two channel transitions).

In the case of a sweep jammer our original intent was to be able to repeat the same experiment twice so we may average the performance over two trials to reach better conclusions. Although in tests #1 and #2 we were able to achieve this for the random policy, since Test #4 had to be discarded we could not do this for test #3. In the third row of Table II we have shown the average performance of the random policy against a sweep jammer obtained by averaging the tests #1 and #2.

As we observe from Table II, the comparison between Tests #1 and #3 shows the excellent cognitive anti-jamming performance improvement achieved by using the Radiobots cognitive anti-jamming communications policy. With the random channel selection policy, the radio stays in a channel for an average time of 1399 time-units without being jammed. However, when it uses the cognitive anti-jamming communications policy with an exploration rate of $\epsilon = 0.3$, it can stay in a channel for an average time of 1634 time-units without being jammed. To put this in to context, assume an ideal system in which the jammer moves from one channel to the next within exactly one sensing period. In this case, it can be shown that the optimal cognitive anti-jamming policy against a sweep jammer will allow the radio to stay in a channel without getting jammed for 8 time-units. On the other hand, the average time in a channel without being jammed, that can be achieved by a random policy in this ideal system is 4.5 time-units. Hence, the best performance we can expect in such an ideal system is an improvement of about 78% of the average time in a channel without being jammed with respect to a random channel selection policy. The performance in Test #3 versus the performance of Test #1 shows that it has achieved about 22% of this best possible performance even with a policy learned under laboratory conditions along with completely un-optimized parameter settings. The full significance of the above performance improvement can be viewed when it is combined with the performance metric of average fraction of time in a channel without being jammed. Indeed, note from Table II that out of the total time the radio spends in a channel, only about 14% is jammer-free when it uses a random policy. On the other hand, when it follows the cognitive anti-jamming policy, about 56% of the time in a channel it will find itself be free of jamming. That is 300% of improvement.

Perhaps a better evaluation of the effectiveness of cognitive anti-jamming communications against a sweep jammer can be obtained by comparing the performance of test #3 to the average performance of tests #1 and #3 shown in Table II. Indeed, this shows that in following a random channel selection policy the radio may stay only about 20% of its time in a channel jammer-free compared to that of 56% with the cognitive anti-jamming policy. This is a considerable advantage since with a random policy the radio will stay idly in a channel for a long duration expecting it to become jammer-free. Moreover, it is about 30% of the best possible performance an ideal policy would have achieved under complete synchronous conditions.

We may draw similar conclusions regarding the performance against a Markov jammer from the results of tests #5, #6 and #7 shown on Table II. Comparison of tests #5 and #6, shows that the pre-learned policy against the Markov jammer under laboratory conditions still performs well enough even during the flight testing. Indeed, the average time in a channel without being jammed has increased by more than 20% compared to that with the random channel selection policy. The average fraction of jammer-free time in channel has also improved with the CAJ policy to about 32% as opposed to 27% achieved by random channel transitions. As can be seen from Table II, allowing exploration with $\epsilon = 0.3$ has further improved the performance of cognitive anti-jamming communications against a Markov jammer both in terms of the average time in a channel without being jammed as well as the average fraction of jammer-free time in channel.

Thus, from the results of these tests we may conclude that:

- WACR is an effective anti-jamming tool regardless of what type of jammer, learning and channel selection algorithms are used.
- Reinforcement learning aided cognitive anti-jamming communications policy significantly outperforms the random channel selection policy, both in terms of the average time in a channel without being jammed as well as the fraction of time in a channel without being jammed.
- Allowing learning-based policy update and policy exploration during actual cognitive communications will lead to better cognitive anti-jamming performance.
- Best possible performance improvements with the cognitive anti-jamming communications policy can expected to be higher than what is observed in these tests since these tests only allowed a very short learning period length and all parameters of the algorithms were un-optimized arbitrary values.

VI. Conclusion

A complete cognitive anti-jamming communications WACR system was designed, implemented and tested on an actual closed-loop satellite-to-ground communications link. The WACR system was made of a cognitive engine implemented in software run on a host PC and two SDRs controlled by this cognitive engine. The anti-jamming communications system was implemented to operative over a 5 MHz wide S-band spectrum that was divided into nine channels. The satellite transmitter was an S-band SDR that currently available on the SCaN Testbed on the ISS. The cognitive engine used a reinforcement learning approach to learn an effective anti-jamming communications policy. When the current satellite-to-ground link is jammed during live communications, the cognitive engine used this policy to instruct the SCaN Testbed SDR to switch the transmission to a new channel that will most likely stay jammer-free for the longest possible duration. Live testing was performed against two types of terrestrial-born jammers: a sweeping jammer and a Markov jammer. Results obtained from a sequence of live experiments over several satellite passes over the ground station showed that indeed machine learning aided cognitive anti-jamming communications provided by WACR technology is a feasible technology for future satellite and space communications. Moreover, the experimental results showed that even under highly dynamic challenging channel conditions, reinforcement learning based approaches can be very effective in learning good anti-jamming communications policies.

References

[1] S. K. Jayaweera, *Signal Processing for Cognitive Radios*, 1st ed. New York, NY, USA: 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, Technical Report 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," in *IEEE Intelligent Sig. Proc. and Commun. Systems (ISPACS 2011)*, Chiangmai, Thailand, Dec. 2011.

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

[5] M. A. Aref, S. K. Jayaweera, and S. Machuzak, "Multi-agent reinforcement learning based cognitive anti-jamming," in *IEEE Wireless Communications and Networking Conference (WCNC)*, San Francisco, CA, Mar. 2017.

[6] M. A. Aref and S. K. Jayaweera, "A novel cognitive anti-jamming stochastic game," in *IEEE Cognitive Communications for Aerospace Applications Workshop (CCAA)*, Cleveland, OH, June 2017.

[7] Y. Gwon, S. Dastangoo, C. Fossa, and H. T. Kung, "Competing mobile network game: Embracing anti-jamming and jamming strategies with reinforcement learning," in *IEEE Conference in Communications and Network Security (CNS'13)*, National Harbor, MD, Oct. 2013.

[8] ——, "Fast online learning of antijamming and jamming strategies," in *2015 IEEE Global Communications Conference (GLOBECOM'15)*, San Diego, CA, Dec. 2015.

[9] B. Wang, Y. Wu, K. Liu, and T. Clancy, "An anti-jamming stochastic game for cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, 2011.

[10] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE JSAC*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[11] C. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3, p. 279292, 1992.

[12] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[13] R. C. Reinhart and J. P. Lux, "Space-based reconfigurable software defined radio test bed aboard international space station," in *AIAA Space Ops Conference GRC-E-DAA-TN13717*, May 2014.

[14] "Space telecommunications radio system (strs) architecture standard release 1.02.1," NASA Technical Memorandum 2010-216809/REV1, Mar. 2012, available online at the NASA Technical Reports Server, or https://strs.grc.nasa.gov.

[15] J. Downey, D. Mortensen, M. Evans, J. Briones, and N. Tollis, "Adaptive coding and modulation experiment with nasa's space communication and navigation testbed," in *International Communications Satellite Systems Conference GRC-E-DAA-TN35246*, Oct. 2016.