AFRL-RV-PS-
TR-2014-0207

ROBUST SATELLITE COMMUNICATIONS UNDER HOSTILE INTERFERENCE

Marc Lichtman and Jeffrey Reed

Virginia Tech
1880 Pratt Drive, Ste. 2006
Blacksburg, VA 24060

8 Jan 2015

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 deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. 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-2014-0207 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//
STEVEN A. LANE
Program Manager

//SIGNED//
PAUL HAUSGEN
Technical Advisor, Spacecraft Component Technology Branch

//SIGNED//
BENJAMIN M. COOK, Lt Col, USAF
Deputy Chief, 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)*
08-01-2015	Final Report	03 Apr 2013 to 30 Nov 2014

4. TITLE AND SUBTITLE

Robust Satellite Communications Under Hostile Interference

5a. CONTRACT NUMBER
FA9453-13-1-0237

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
63401F

6. AUTHOR(S)

Marc Lichtman and Jeffrey Reed

5d. PROJECT NUMBER
2181

5e. TASK NUMBER
PPM00013284

5f. WORK UNIT NUMBER
EF008906

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Virginia Tech
1880 Pratt Drive, Ste. 2006
Blacksburg, VA 24060

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/RVSV

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RV-PS-TR-2014-0207

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In this report, we investigate a strategy to avoid or mitigate reactive forms of jamming using a reinforcement learning approach. The mitigation strategy focuses on finding an effective channel hopping and idling pattern to maximize link throughput. We also analyze the feasibility of reactive jamming (including repeater jamming) in a satellite communications scenario, and propose a countermeasure that takes advantage of the constraints associated with reactive jamming.

15. SUBJECT TERMS
Space communication, reactive jamming, reinforcement learning, assured communication

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Steven A. Lane
Unclassified	Unclassified	Unclassified	Unlimited	50	**19b. TELEPHONE NUMBER** *(include area code)*

(This page intentionally left blank)

TABLE OF CONTENTS

LIST OF FIGURES ii
LIST OF TABLES iii
1 SUMMARY 1
2 INTRODUCTION 2
2.1 Reinforcement Learning 2
2.2 Reactive Jammer 3
3 METHODS, ASSUMPTIONS, AND PROCEDURES 5
3.1 Reactive Jamming Mitigation Using Reinforcement Learning 5
3.1.1 System Model and Problem Formulation 5
3.1.2 Strategy for Mitigation of Reactive Jamming 7
3.1.3 Reinforcement Learning Background 8
3.1.4 Markov Decision Process Formulation 10
3.1.5 Knowledge Decay 14
3.1.6 Comparison with Traditional Parameter Optimization 15
3.2 Analysis of Reactive Jamming Against Satellite Communications 15
3.2.1 Received Signal-to-Noise Ratio at the Jammer 15
3.2.2 SNR Threshold for Repeating the Signal 16
3.2.3 SNR During Uplink Jamming 16
3.2.4 SNR During Downlink Jamming 21
3.2.5 Jammer-to-Signal Ratio Component 24
3.2.6 Uplink Jamming JSR 25
3.2.7 Downlink Jamming JSR 26
3.2.8 Geometric Component 27
3.2.9 Fraction of Each Hop That Must Be Jammed 28
4 RESULTS AND DISCUSSION 30
4.1 Recursive Learning Mitigation Strategy 30
4.2 Simulation Using Systems Tool Kit 32
4.3 Coding and Interleaving 34
5 CONCLUSIONS 36
6 RECOMMENDATIONS 37
REFERENCES 38
LIST OF SYMBOLS, ACRONYMS, AND ABBREVIATIONS 40

LIST OF FIGURES

Figure 1. A Satellite Communication Jamming Scenario Involving an Inter-Satellite Link 4
Figure 2. System Model of a Transmitter, Receiver and Reactive Jammer 6
Figure 3. Markov Decision Process Associated with Hopping Channels and Going Idle 10
Figure 4. Rewards Associated with Reactive Jammer Model $\boldsymbol{N_{REACT} = 3}$ 12
Figure 5. Optimal Policies in the Presence of Three Different Reactive Jammers 13
Figure 6. Optimal Policies in the Presence of Two Different Repeater Jammers 14
Figure 7. Uplink System Diagram 17
Figure 8. Example Radiation Pattern of a Directional Antenna 19
Figure 9. SNR at Jammer When Receiving the Ground User's Signal 21
Figure 10. Downlink System Diagram 22
Figure 11. SNR at Jammer When Receiving the Satellite's Signal 23
Figure 12. BER Curve For BPSK in an AWGN Channel With LDPC Coding at Various Rates 25
Figure 13. Time and Frequency Behavior of a Repeater Jammer 28
Figure 14. Maximum Distance the Jammer Can Be From the Ground User in Order to Successfully Jam a Hop, as a Function of $\boldsymbol{\eta}$ 30
Figure 15. Simulation Results Showing the Learning Process Over Time in the Presence of Different Jammers 31
Figure 16. Screenshots of the Uplink Jamming Scenario in STK 33
Figure 17. STK Simulation Showing SNR Measured During the Uplink Attack (at the Jammer) 34
Figure 18. Feasible Region of Countermeasure 35

LIST OF TABLES

Table 1. Summary of How to Cast This Mitigation Approach into an RL Framework 12
Table 2. Distance and Delay Between the Jammer (or Ground User) and Satellite 23
Table 3. Uplink Jamming Attack Link Budget to Calculate JSR 26
Table 4. Downlink Jamming Attack Link Budget to Calculate JSR 27
Table 5. STK Simulation Results Showing JSR and SNR for the Uplink/Downlink Scenarios .. 33
Table 6. Summary of Feasibility Analysis 37

ACKNOWLEDGMENTS

This material is based on research sponsored by Air Force Research Laboratory under agreement number FA9453-13-1-0237. 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

This report describes a novel strategy to avoid or mitigate reactive forms of jamming using a reinforcement learning approach. The mitigation strategy focuses on finding an effective channel hopping and idling pattern to maximize link throughput. Thus, the strategy is well-suited for frequency hopping spread-spectrum systems, and best performs in tandem with a channel selection algorithm. By using a learning approach, there is no need to pre-program a radio with specific anti-jam strategies, and the problem of having to classify jammers is avoided. Instead, the specific anti-jam strategy is learned in real-time and in the presence of the jammer.

Also, this report analyzes the feasibility of reactive jamming (including repeater jamming) in a satellite communications scenario, and describes a countermeasure that takes advantage of the constraints associated with reactive jamming. A reactive jammer is a type of jammer that has the ability to sense a portion of spectrum and immediately transmit a jamming signal when it senses a signal it wants to jam. Thus, a reactive jammer can counter the processing gain associated with frequency-hopping (but not direct-sequence) spread-spectrum.

This report presents a three-step approach to analyzing the primary constraints associated with reactive jamming, as well as detailed example scenarios (both theoretical and simulated) that demonstrate the feasibility of our analysis. Additional clarity is provided by splitting the analysis into uplink and downlink jamming. A strategy to mitigate the effects of reactive jamming is presented, which takes advantage of the geometric constraints of reactive jamming by using a coding and interleaving scheme that results in the transmitted bits appearing at the very beginning of each hop.

This work demonstrates that reactive jamming is a valid threat to satellite communication systems. However, the analysis and simulation results indicate that as long as the geographical area around user terminals is free of reactive jammers, substantial mitigation can be achieved using the proposed mitigation strategy.

2 INTRODUCTION

Wireless communication systems are becoming more prevalent because of their affordability and ease of deployment. Unfortunately, all wireless communications are susceptible to jamming. Jamming attacks can degrade communications and even cause total denial of service to multiple users of a system. As communication technology becomes more sophisticated, so does the sophistication of jammers. Even though jamming techniques such as repeater jamming have been known for decades, recently there has been research into other forms of complex jamming with receiving and processing capabilities (reactive jamming) [1, 2].

Wireless security threats are typically broken up into two categories: (1) cyber-security, and (2) electronic warfare (i.e., jamming). Electronic warfare attacks target the physical (PHY) and/or media access control (MAC) layer of a communication system. Cyber-security attacks are designed to exploit the higher layers of the communications stack. In this project, we were only concerned with jamming, and in particular jamming of an intelligent nature. A series of intelligent jamming attack models are introduced in the paper by Xu *et al.*, including the reactive jammer model [2]. The authors proposed a basic detection algorithm using statistics related to signal strength and packet delivery ratio. For an overview on electronic warfare and jamming, we refer the reader to the text by Adamy [3].

2.1 Reinforcement Learning

A reinforcement learning (RL) or a Markov Decision Process (MDP) approach has been previously used in the wireless domain for channel assignment, general anti-jamming in wireless sensor networks, and jammer avoidance in cognitive radio networks [4-7]. Yang et al. applied reinforcement learning to the problem of channel assignment in heterogeneous multicast terrestrial communication systems [4]. While their discussion did not deal with jamming, it had similar concepts to the proposed techniques.

Zhu *et al.*, proposed an anti-jamming scheme for wireless sensor networks [5]. To address time-varying jamming conditions, the authors formulated the anti-jamming problem of the sensor network as a MDP. It was assumed that there were three possible anti-jamming techniques: (1) transmission power adjustment, (2) error-correcting code, and (3) channel hopping. These techniques were not explored any further; the set of actions available to the radio is simply which technique is used. While their work is similar, it greatly generalized anti-

jamming strategies. In other words, their work did not offer a jamming strategy; it offered a method of choosing the best jamming strategy from a given set.

Wu *et al.*, used a MDP approach to derive an optimal anti-jam strategy for secondary users in a cognitive radio network [6]. For their jammer model, the authors used reactive jammers seeking to disrupt secondary users and avoid primary users. In terms of actions, in each time-slot the secondary user must decide whether to stay or hop frequencies. The authors proposed an online learning strategy for estimating the number of jammers and the access pattern of primary users (this can be thought of as channel availability). Even though the authors used a reactive jammer model similar to the one described in our work, they assumed that the jammer was always successful and that the entire analysis is within the context of dynamic spectrum access. To the best of our knowledge, there have not been any RL or MDP based approaches designed to mitigate a wide range of reactive jamming behaviors.

2.2 Reactive Jammer

A reactive jammer is a type of jammer that has the ability to sense one or more channels and immediately transmit a jamming signal when it senses a signal it wants to jam [2]. In a frequency hopping spread-spectrum (FHSS) system, this allows the jammer to only transmit when the target is transmitting, as well as only transmit on the target's current frequency, without needing to know the hop sequence. The jammer is then able to conserve power, achieving a gain roughly equal to the FHSS processing gain and remaining harder to detect. Typically, a reactive jammer will not simply retransmit the received signal, because this could potentially help the target communications system. Instead, the jammer may *noise modulate* the received signal, or simply transmit noise on the detected center frequency.

In this project, we analyzed the feasibility of performing and mitigating reactive jamming in a satellite communication (SATCOM) type scenario. SATCOM systems are characterized by the long distances that the electromagnetic signals need to travel, and thus directional antennas are typically required. These communications systems typically include satellites in geosynchronous orbit (35,786 km above the Earth's surface), spaced along an orbit above the equator [8]. They also tend to use spread-spectrum techniques such as direct sequence spread-spectrum (DSSS) and/or frequency hopping spread-spectrum (FHSS) for protection against jamming [3, 9]. The analysis in this report assumed a 40 GHz uplink carrier (Earth-to-space) and

a 20 GHz downlink carrier (space-to-Earth). This corresponds to a Ka-band satellite system, which is increasingly popular for commercial SATCOM services. A 2 GHz hopping bandwidth was used as a case study, because this is roughly the average bandwidth of each geostationary SATCOM band [10]. We remained agnostic of whether the satellite had onboard processing (i.e., a regenerative transceiver) or whether it simply implemented a transponder (i.e., a non-regenerative transceiver), because we assume that the reactive jamming occurs on the uplink or downlink signal between the satellite and ground user/station, as shown in Figure 1.

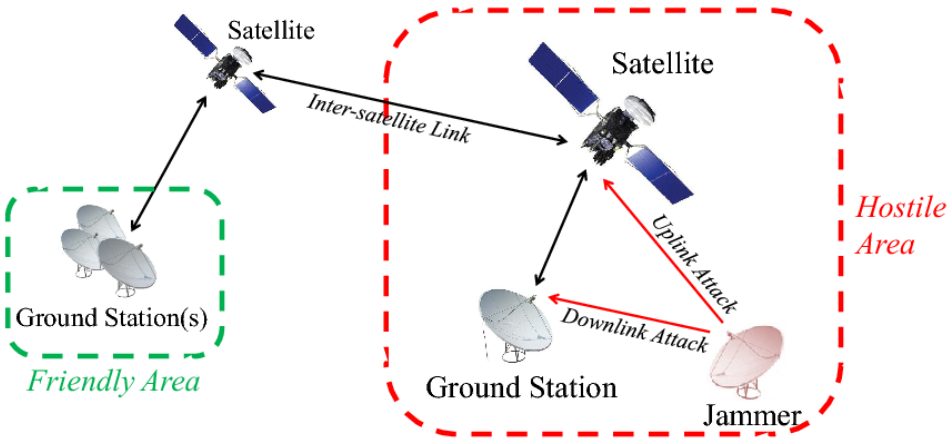

Figure 1. A Satellite Communication Jamming Scenario Involving an Inter-Satellite Link

It is well-known that jamming a FHSS system requires either knowing the hopping sequence beforehand and being time-synchronized, or dynamically detecting the signal at some frequency and immediately transmitting on that frequency [11]. The latter jammers are called reactive, because they react to the signal they are attacking in real-time. In order for reactive jamming to be feasible, there are three main requirements that must be satisfied:

1. The jammer must receive the target signal at a high enough signal-to-noise ratio (SNR) to be able to detect the actual transmission frequency;
2. The jamming signal must reach the target receiver with a high enough jammer-to-signal ratio (JSR) to degrade communications; and
3. The jamming signal must reach the target receiver quick enough to overlap in time and frequency with the target signal.

The goal of any countermeasure is to prevent or mitigate the threat. Therefore, the countermeasure presented in this report will focus on preventing the adversary from meeting one of the above requirements (the one that requires the least amount of effort to foil).

3 METHODS, ASSUMPTIONS, AND PROCEDURES

3.1 Reactive Jamming Mitigation Using Reinforcement Learning

We propose a strategy to mitigate or even avoid reactive forms of jamming using a reinforcement learning (RL) approach. Through a learning approach, the problem of having to detect and classify which type of jammer is present in real-time is avoided. In addition, there is no need to pre-program a radio with specific mitigation strategies. Instead, the strategy is learned in real-time and in the presence of the jammer. The proposed mitigation strategy focuses on finding an effective channel hopping and idling pattern to maximize link throughput. Not only can this approach enable communications, which would otherwise fail in the presence of a sophisticated and reactive jammer, but it can also act as an optimization routine that controls the link-layer behavior of the radio.

The proposed strategy is well suited for frequency hopping spread-spectrum (FHSS) systems, which are widely used in modern wireless communications. The strategy could also be applied to an orthogonal frequency division multiple access (OFDMA) system in which users hop between different subcarriers or groups of subcarriers. Countless users and systems depend on wireless communications and therefore, it is important to secure them against jamming. While there exists many methods to counter *barrage jamming* (the most basic form of jamming), there are few methods that are designed to address the more intelligent behaviors a jammer can exhibit.

3.1.1 System Model and Problem Formulation

Consider the typical wireless communications link with the addition of a jammer that both receives the friendly signal (but not necessarily demodulates it) and transmits a jamming signal, as shown in Figure 2. For the sake of simplicity, we will only consider a unidirectional link, although this analysis also applies to bidirectional links that may be unicast or broadcast, as well as a collection of links. While reactive jamming can take on different forms, we will broadly define the term as any jammer that is capable of sensing the link and reacting to sensed

information. We will assume this sensed information is in the form of the presence or absence of energy, because any additional information such as modulation scheme or actual data would be irrelevant for this mitigation strategy.

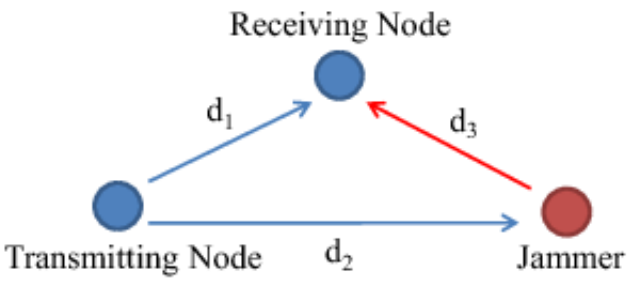

Figure 2. System Model of a Transmitter, Receiver and Reactive Jammer

A simple example of a reactive jammer is one that senses the spectrum for activity and then immediately transmits wideband noise when it senses any activity [2]. This strategy allows the jammer to remain idle while the channel is idle and thus save power and avoid being easily detected. Another form of reactive jamming, commonly known as repeater or follower jamming, works by immediately transmitting what it receives with noise added to it [11]. This can be modeled as a jammer that senses a set of channels, and immediately transmits noise on any channel that appears to be active.

Reactive jamming is only feasible when the geometry of the system is such that the jammer's transmitted signal reaches the target receiver before it hops to a new channel or stops transmitting. As such, reactive jamming is only possible when the jammer is physically located near or between the target transmitter and receiver. If η represents the fraction of each hop duration that must remain un-jammed for communications to succeed, then we have the following inequality limiting the distances d_2 and d_3 [11]:

$$d_2 + d_3 \leq (\eta T_d - T_j)c + d_1, \quad (1)$$

where T_d is the hop duration, T_j is the jammer's processing time, c is the speed of light, and d_1, d_2, and d_3 are the distances indicated in Figure 2. In addition to this limitation, the jammer-to-signal ratio at the receiving node must be high enough to degrade quality of service. In this section, we assume that the jammer is close enough to the transmitter and receiver, and that the jammer-to-signal ratio is significantly high during periods of jamming.

As part of the analysis and simulation, we investigated two specific reactive jamming models. The first, labeled in this section as simply "reactive jamming", is defined as a jammer that successfully jams any transmission that remains active for too long, regardless of the channel or frequency being used. The second jammer model is based on repeater jamming, and it is described as a jammer that successfully jams any transmission that remains on the same channel or frequency for too long. While there are other ways to formulate reactive jamming models, the analysis and simulation in this section will focus on these two. More formal definitions of these two jammer models are as follows:

1. **Reactive Jammer** - Begins jamming any transmission that remains active for more than N_{REACT} time steps, and will only cease jamming once the target is idle for at least N_{IDLE} time steps.
2. **Repeater Jammer** - Begins jamming any transmission that remains on the same channel for more than N_{REP} time steps.

In this analysis, we investigate a transmitter and receiver pair that can hop among a certain number of channels using a FHSS approach, or any other approach that involves radios capable of changing channels. Therefore, at any time step, the transmitter has the option to remain on the channel, change channel, or go idle. Because the actions of the transmitter must be shared with the receiver beforehand, it is expected that decisions are made in advance. It is assumed that channel quality indicators (e.g., whether or not the information was received) are sent back to the transmitter on a per-hop basis. These indicators could be binary (indicating an acknowledged (ACK) or not acknowledged (NACK)), or they could take on a range of values indicating the link quality. Lastly, it is assumed that the receiver is not able to simply detect the presence of a jammer.

3.1.2 Strategy for Mitigation of Reactive Jamming

The mitigation strategy described in this section is based on modeling the system as a MDP, where the transmitter is the decision maker, and then using RL to learn a strategy for dealing with the broad category of reactive jamming. This strategy will be in the form of a channel hopping pattern, where going idle is considered as hopping to the "idle channel" for a certain duration. However, we are not concerned with choosing the best channel to transmit on at any given time, nor identifying corrupt channels that have excessive noise. The mitigation

strategy described in this section is designed to work in tandem with this kind of algorithm (i.e., one that indicates which specific channels are suitable for use and which are not). Likewise, we are not concerned with the PHY layer waveform characteristics that the transmitter or jammer uses (i.e., bandwidth, modulation, or type of noise). Adaptive modulation and coding can be performed alongside the proposed strategy.

3.1.3 Reinforcement Learning Background

RL is the subset of machine learning concerned with how an agent should take actions in an environment to maximize some notion of cumulative reward. The agent is the entity interacting with the environment and making decisions at each time interval, and in this section we will consider the transmitter as the agent (although the actions it chooses must be forwarded to the receiver). An agent must be able to sense some aspect of the environment and make decisions that affect the agent's state. For example, reinforcement learning can be used to teach a robot how to walk without explicitly programming the walking action. The robot could be rewarded for achieving movement in a forward direction, and the robot's action at each time step could be defined as a set of angular servo motions. After trying a series of random motions, the robot will eventually learn that a certain pattern of motion leads to moving forward, and thus, a high cumulative reward. In this section, we apply this concept to a transmitter that learns how to hop or idle in a manner that allows successful communications under a sophisticated reactive jamming attack.

There are four main components of a RL system: (1) a policy, (2) a reward, (3) a value function, and (4) the model of the environment [12]. A policy (denoted as π) defines how the agent will behave at any given time. The goal of a RL algorithm is to optimize the policy in order to maximize the cumulative reward. A policy should contain a stochastic component, so that the agent tries new actions (known as exploration). A reward, or reward function, maps the current state and action taken by the agent to a value and is used to indicate when the agent performs desirably. In a communication system, a possible reward function may combine the throughput of a link, spectral efficiency, and power consumption. While the reward function indicates what is desirable in the immediate sense, the value function determines the long-term reward. A *state* may provide a low immediate reward, but if it leads to other states that provide a high reward, then it would have a high value.

The model of the environment is used to either predict a reward that has not been experienced yet, or simply to determine which actions are possible for a given state. For example, it is possible to create a RL agent that learns how to play chess. The environment would be a model of the chess board, pieces, and the set of legal moves.

In RL, the environment is typically formulated as a MDP, which is a way to model decision making in situations where outcomes are partially random and partially under the control of the decision maker. The probability of each possible next state, s', given the current state s and action a taken, is given by [12]:

$$P_{ss'}^{a} = Pr\{s_{t+1} = s' | s_t = s, a_t = a\}. \quad (2)$$

Equation (2) provides what are known as transition probabilities, and because they are only based on the current state and action taken, it assumes a memoryless system, and therefore, has the Markov property. The expected reward (obtained in the next time step) for a certain state-action pair is given by:

$$R_{ss'}^{a} = E\{r_{t+1} | s_{t+1} = s', s_t = s, a_t = a\}. \quad (3)$$

The goal of a learning agent is to estimate these transition probabilities and rewards, while performing actions in an environment. In order for an agent to take into account the long-term reward associated with each action in each state, it must be able to predict the expected long-term reward. For a certain policy π, we calculate the expected return from starting in state s and taking action a as [12]:

$$Q^{\pi}(s, a) = E_{\pi}\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a\}, \quad (4)$$

where γ is known as the discount rate, and represents how strongly future rewards will be taken into account. Equation (4) is known as the *action-value function*, and in a method known as Q-learning, the action-value function is estimated based on observations. While performing actions in an environment, the learner updates its estimate of $Q(s_t, a_t)$ as follows [12]:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_{a} Q(s_{t+1}, a) - Q(s_t, a_t) \right], \quad (5)$$

where r_{t+1} is the reward received from taking action a, and α is the learning rate, which determines how quickly old information is replaced with new information. Because Q-learning is an iterative algorithm, it must be programmed with initial conditions ($Q(s_0, a_0)$). Optimistically high values are typically used for initialization to promote exploration. However, even once some initial exploration is performed, there needs to be a mechanism that prevents the agent

from simply sticking to the best policy at any given time. An approach known as *Epsilon-greedy* compels the agent to take the "best action" with probability 1-ε and take a random action (using a uniform probability) with probability ε. Epsilon is usually set at a fairy high value (e.g., 0.95) so that a majority of the time the agent is using what it thinks is the best action. For an in-depth tutorial on MDP's and RL, we refer the reader to Sutton and Barto [12].

3.1.4 Markov Decision Process Formulation

We will now formulate a MDP used to model the transmitter's available states and actions. The states exist on a two-dimensional grid, in which one axis corresponds to the time that the transmitter has been on a given channel (including the "idle channel"), and the other axis corresponds to the time that the transmitter has been continuously transmitting. Time is discretized into time steps, and we assume the step size is equal to the smallest period of time in which the transmitter must remain on the same channel. Figure 3 shows the *state-space* and actions available in this MDP.

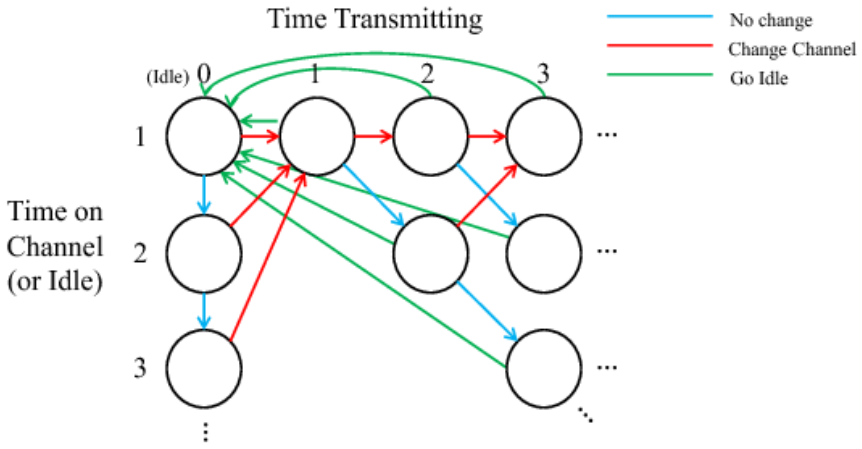

Figure 3. Markov Decision Process Associated with Hopping Channels and Going Idle

The transmitter will always start in the top-left state, which corresponds to being idle for one time step. It then must choose whether to remain idle, or "change channel" (which can be interpreted as "start transmitting" when coming from an idle state). If it decides to change channel, then in the next time step, it must decide whether to remain on that channel or change to a new channel, which we will assume is chosen randomly from a list of candidate channels. It should be noted that the result of each action is deterministic. However, the rewards may contain a stochastic component. Due to what the states represent, the MDP is theoretically infinitely long in directions as indicated by ellipsis in Figure 3. However, in practical systems, the width and height of the MDP must be limited.

The reward associated with each state transition is based on the actual data throughput that occurs during the time step. As such, the rewards are based on the jammer model, and may be stochastic in nature. Figure 4 shows the rewards associated with a transmitter and receiver operating in the presence of a reactive jammer with $N_{REACT} = 3$ and $N_{IDLE} = 1$ (model and parameters defined in the previous section). This example shows that when the radio is transmitting for more than three continuous time steps, the link becomes jammed (red states) and the reward becomes zero until the jammer goes idle and then starts transmitting again (the radio is not rewarded while idle). Although the rewards are shown on top of each state, they are actually associated with the previous state and action taken, and won't always be equal for a given resulting state. The numbers 1, 1.3, and 1.47 are examples to demonstrate the fact that remaining on the same channel is more favorable than switching channels, due to the time it takes to switch frequencies. In a real implementation, the reward would be based on the achieved throughput, or quality of the link, not a model. A summary of how to cast this mitigation approach into a RL framework is given in Table 1.

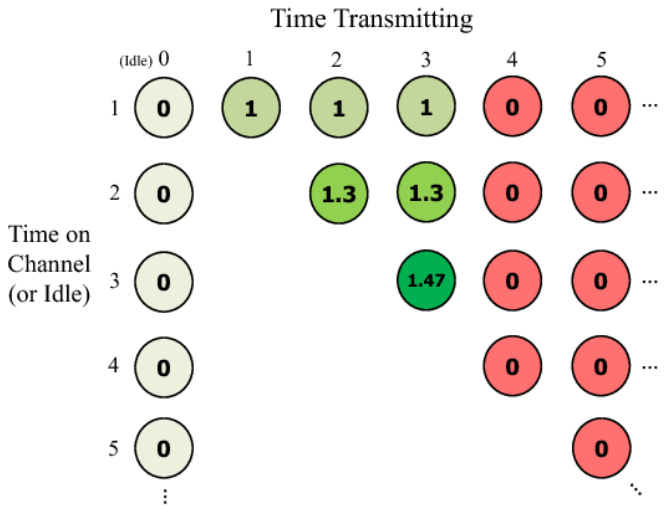

Figure 4. Rewards Associated with Reactive Jammer Model $N_{REACT} = 3$

Table 1. Summary of How to Cast This Mitigation Approach into an RL Framework

Environment States exist on a two dimensional grid, in which one axis corresponds to the time the transmitter has been on a given channel (including the "idle channel"), and the other axis corresponds to the time the transmitter has been continuously transmitting.
Agent's Actions are to either: 1) remain idle, or 2) "change channel", which can be interpreted as "start transmitting" when coming from an idle state.
State Transition Rules are deterministic (although a stochastic component due to external factors could be added) and based on the action taken.
Reward Function is a value proportional to the data throughput that occurred during the time step (not known until feedback is received).
Agent's Observations include the state it is currently in, and the reward achieved from each state-action pair.
Exploration vs. Exploitation is achieved using the Epsilon-greedy approach, in which the agent chooses a uniformly random action a small fraction of the time.
Task type is continuing by nature, but could be treated as episodic where each episode is an attempt to transmit for N time steps.

Now that the states, actions, and rewards are established, we can investigate the learning process of the transmitter in the presence of various types of reactive jammers. In RL, the agent (in this case, the transmitter) learns by trying actions and building statistics associated with each state-action pair. At the beginning of the learning process, the agent has no information about the environment and must try random actions in any state. After a period of learning, the agent eventually chooses what it thinks is the best action for each state in terms of the predicted long-term reward. The Epsilon-greedy approach forces the agent to never consider itself "finished" with learning.

Under a reactive jammer with a certain N_{REACT} and when $N_{IDLE} = 1$, the optimal policy is to remain on the same channel for N_{REACT} time steps, and then go idle for one time step. Three optimal policies are shown in Figure 5, corresponding to $N_{REACT} = 1, 2,$ and 3. Each optimal policy resembles a loop that starts at idle for one time step and proceeds to transmit on the same channel for N_{REACT} time steps. In a real-world scenario, it takes the transmitter many time steps to establish that this is the best policy to take, because it must explore each state-action multiple times to build reliable statistics.

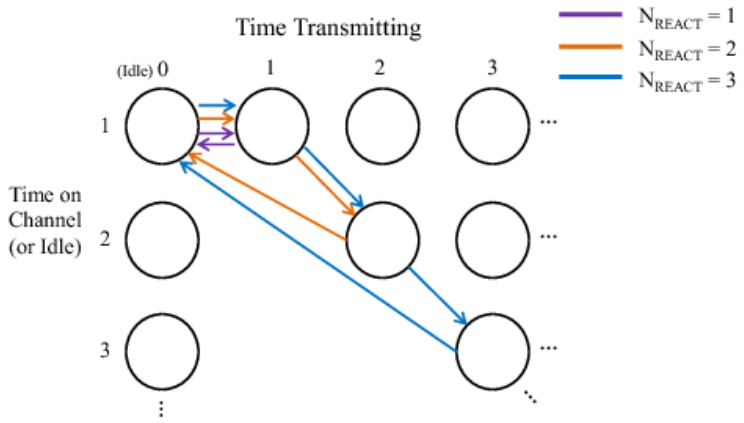

Figure 5. Optimal Policies in the Presence of Three Different Reactive Jammers

The optimal policy for a repeater jammer is shown in Figure 6, using $N_{REP} = 1$ and 2. This zigzag pattern indicates constant-rate frequency hopping, which is well-established as the typical method for avoiding repeater jamming [12]. Unfortunately, the optimal policy will always be infinitely long in the horizontal direction. To take this into account, the learning

process can involve resetting the current state to the top-left state after a certain period of time continuously transmitting. This will have minimal influence on learning the optimal policy as long as the state-space spans enough time steps to take into account the slowest (i.e., the highest value of N_{REP}) perceivable repeater jammer.

Using the approach described in this section, there is no need to perform "jammer classification". As such, the mitigation strategy will remain effective over a wide range of jamming behaviors, and may even be able to deal with new jamming behaviors that were not considered during development.

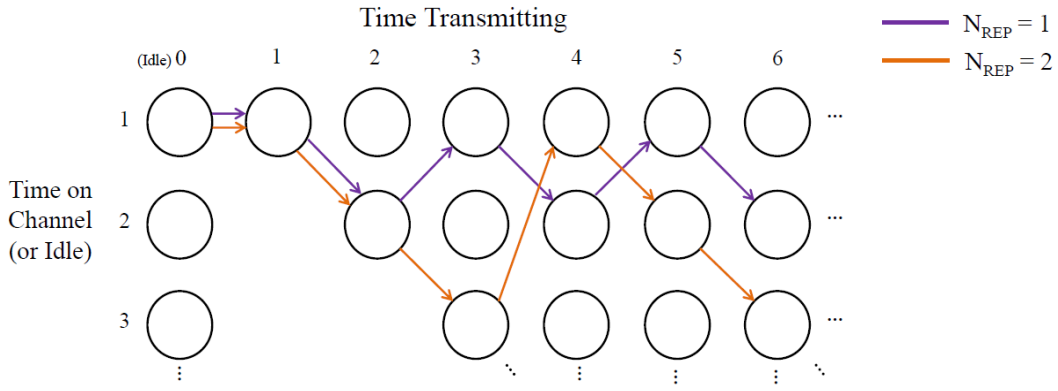

Figure 6. Optimal Policies in the Presence of Two Different Repeater Jammers

3.1.5 Knowledge Decay

The last component of the proposed mitigation strategy is taking into account a changing environment. A given jammer may only be present for a short period of time, and link performance would degrade if the transmitter were sticking to its acquired knowledge. As such, the learning engine must incorporate some form of *knowledge decay*. Due to the nature of Q-learning, the learning rate α can be used as a form of knowledge decay, by setting it low enough so that the learner can react to a changing environment. A proper value for α would be based on how quick the transmitter is able to learn optimal policies for a variety of jammers. A detailed investigation on approaches of knowledge decay or forgetting is beyond the scope of this work, but for more information we refer the reader to Sutton and Barto [12].

3.1.6 Comparison with Traditional Parameter Optimization

Finding an effective channel hopping and idling pattern in the presence of a reactive jammer could also be performed by optimizing the hopping rate and transmission duty cycle. This can also be thought of as adjusting T_{ON} and T_{OFF}; the transmission and idle time of a transmitter, assuming it hops frequencies after each transmission. This type of approach is often used in cognitive radio [13]. If $T_{OFF} = 0$, then T_{ON} becomes the hopping rate. Any number of optimization approaches could be used to tune these two parameters. However, even though this simpler approach can take into account the two specific jammer models described in this section, it does not have the flexibility inherent to the RL approach. For example, consider the scenario involving a reactive jammer with $N_{REACT} = 4$, $N_{IDLE} = 1$ and a repeater jammer with $N_{REP} = 1$, both targeting the friendly node simultaneously. The optimal transmission strategy would be to hop channels every time step, but also go idle for one time step after the fourth consecutive transmission (a strategy that is likely not possible with traditional parameter optimization). In addition, if the actual jammer behavior experienced by the transmitter does not match any models developed during creation of the mitigation strategy, then added flexibility may mean the difference between communicating and being fully jammed.

3.2 Analysis of Reactive Jamming Against Satellite Communications

3.2.1 Received Signal-to-Noise Ratio at the Jammer

For reactive jamming to be feasible, the jammer must be able to receive the target signal at a high enough signal-to-noise ratio (SNR), so that it can either detect the center frequency, or be able to retransmit it without also transmitting excess out-of-band noise. In order to remain agnostic of the specific waveform (e.g., modulation scheme), we will only be concerned with the received signal power at the jammer P_R, the channel noise power P_{NOISE}, and their ratio given by $SNR = P_R/P_{NOISE}$. We will first determine a rough estimate for the minimum SNR needed at the jammer, and then determine how close a jammer would have to be to the transmitter in order to achieve this SNR requirement. The latter analysis is performed for both an uplink and downlink attack separately. It should be noted that the analysis in this section also applies to the general problem of signal interception.

3.2.2 SNR Threshold for Repeating the Signal

The threshold for an acceptable SNR depends on whether the jammer is digital or analog. If it is a digital reactive jammer that detects the center frequency and bandwidth, and then generates a jamming waveform, the SNR must be high enough for this detection process. If we assume the jammer has no *a priori* knowledge of the target waveform, then the only difference between observing a signal and observing noise is the statistical average energy they contain. Therefore, the optimum detector compares the average energy in an observed waveform to a threshold, also known as an energy detector or radiometer [14]. Performance of a Neyman-Pearson type energy detector is parameterized by SNR and number of samples, and given by [14]:

$$P_D(\delta_{NP}) = 1 - \Gamma\left(\frac{n}{2}; \Gamma^{-1}\left(\frac{n}{2}; 1-\alpha\right)(1+SNR)^{-1}\right), \quad (6)$$

where $\Gamma(x, y)$ is the incomplete gamma function, α is the false-alarm probability, and n is the number of samples taken from the observed waveform. Since false-positives and false-negatives are not that critical in this scenario, we can set $\alpha = 0.1$ and $P_D = 0.95$. If the jammer must detect the signal reasonably fast, we can set $n = 10$, which gives us a required SNR of 5 dB.

In the case of an analog jammer, a high SNR would be desired to avoid retransmitting excess noise and thus wasting power. At 5 dB of SNR, the jammer will transmit about 75% of the desired signal and 25% noise ($\frac{0.75}{0.25} = 3 \approx 5$ dB), which is reasonable considering the gain achieved by using reactive jamming. For the remainder of this work, we assumed a 5 dB SNR is required at the jammer's receiver in order to perform reactive jamming. Now that we have an estimated SNR threshold, we can assemble an example link budget for a realistic scenario. We will first analyze an uplink jamming attack, then a downlink jamming attack.

3.2.3 SNR During Uplink Jamming

An uplink attack involves jamming the uplink signal, which is sent from the ground user or station and received by the satellite. Thus, the jammer is receiving the ground user's signal and transmitting noise on the detected signal band towards the satellite's receiver, as shown in Figure 7. In this subsection, we are only concerned with how well the jammer can receive the ground user's signal.

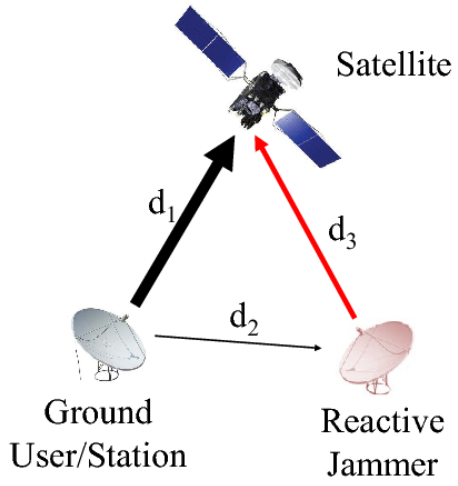

Figure 7. Uplink System Diagram

The link budget for the target signal is given by [15]:

$$P_R = P_T + G_T + G_R - L_{path} - L_{atm} - L_{misc} - L_{process}, \quad (7)$$

and noise power is found using the traditional "*kTB*" method [15]:

$$P_{noise} = k + T + B. \quad (8)$$

The following list describes each parameter:

- P_R - Received signal power
- P_T - Transmitted signal power
- G_T - Transmitting antenna gain
- G_R - Receiving antenna gain
- L_{path} - Path loss
- L_{atm} - Atmospheric attenuation
- L_{misc} - Miscellaneous losses not yet accounted for
- $L_{process}$ - DSSS processing gain
- k - Boltzmann constant (approx. -228.6 dBW/K/Hz)
- T - Noise temperature in dB-Kelvins
- B - Bandwidth of signal in dB-Hz (before spreading).

In order to calculate link budgets for this scenario, we must approximate each parameter listed above. These approximations are made for the sake of obtaining insight into the feasibility of

reactive jamming in a SATCOM scenario. Therefore, when possible (and publicly available), parameters will be taken from existing SATCOM systems. The uplink center frequency f will be centered around 40 GHz as discussed earlier.

P_T is the transmit power of the ground user. To improve traceability of this analysis, we will use information from commercial SATCOM systems in geosynchronous orbit. National Aeronautics and Space Administration (NASA)'s Tracking and Data Relay Satellite (TDRS) has a user equivalent isotropically radiated power (EIRP) of 48.5 dBW in the high data rate mode [16], and the ViaSat VMT-1220 ground terminal states a similar figure [11]. Likewise, the ThinSat Talos Integrated Antenna and Terminal Subsystem is specified to have an equivalent EIRP of 47 dBW [17]. Because $EIRP = P_T + G_T$, we must subtract out the estimated main-lobe antenna gain of 35 dB. This leads us to the rough estimation of $P_T = 13.5$ W or roughly 11 dBW, which will be used for analysis.

G_T is the gain of the ground user's transmitting antenna with respect to the jammer in this case, because we are investigating with what power the jammer receives the ground user's signal. This is a tricky value to estimate, because the ground user will be pointed at the satellite, not the jammer. It is highly unlikely that the jammer will be located in the main lobe of the ground user's antenna, unless it is onboard an airborne platform that is flying directly between the ground user(s) and satellite. This modeling dilemma is depicted in Figure 8. Clearly, there is isolation provided by the directional antenna, but the question is, how much? At any given time, the jammer may be pointed into the peak of a side lobe, or into a null. Regardless, in order to put together a link budget, we need to model the "average gain". For now, a gain of -10 dB will be used for G_T. This can be thought of as isolation provided by the ground user's transmitting antenna. A more elegant solution to this dilemma will be the subject of future research, and may involve a stochastic component.

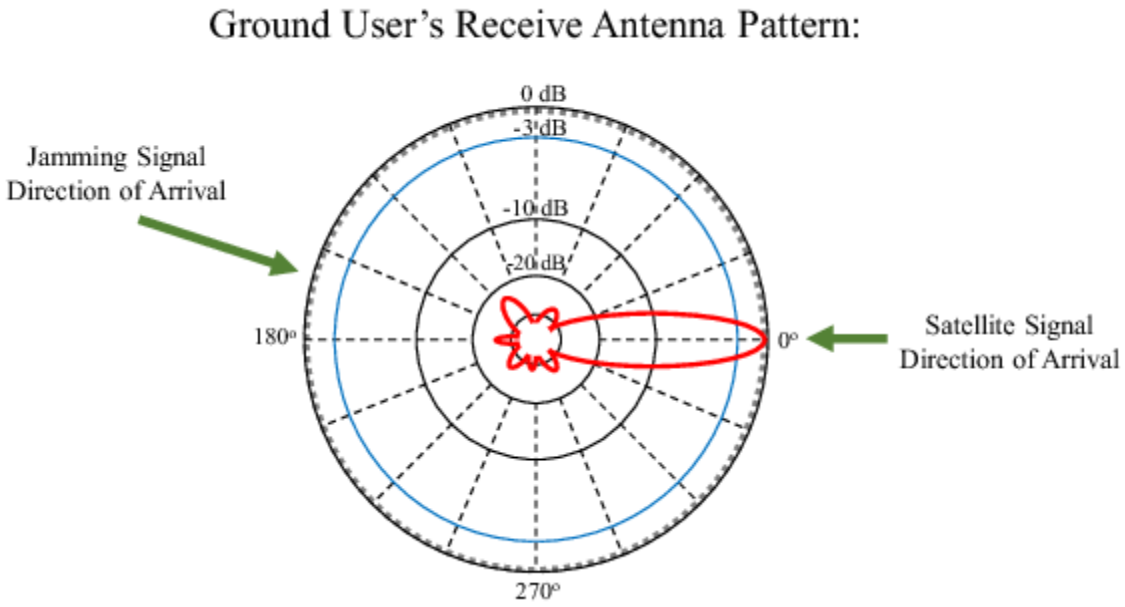

Figure 8. Example Radiation Pattern of a Directional Antenna

G_R is the gain of the jammer's receive antenna, which could be anywhere from omni-directional to highly directional. If the jammer knows roughly where the target ground users are, then it can use a directional antenna pointed at them (although, it would lack any sort of pointing or beamforming feedback). We will therefore assume a 15 dB antenna is used, to take into account a small amount of directionality.

Path loss L_{path} is a function of the channel between the ground user and jammer. In order to perform uplink reactive jamming, the jammer must be within the same spot-beam as the ground user whose signal is being jammed. Therefore, the distance between the ground user and jammer must be no more than the typical beam width, and is likely much less. For the sake of analysis we will vary the distance between 1 - 50 km, so that the jammer is likely to be within the beam. For the channel model, we will use a "line-of-sight (LOS) free space plus reflection and multiple diffraction" channel model to capture a rural type environment at 40 GHz [18]. The attenuation as a function of frequency f in MHz and distance d in kilometers is given in dB using the equation:

$$L_{path} = 32.45 + 20\log_{10} f + 20\log_{10} d + A_r + A_{fr}, \quad (9)$$

where A_r and A_{fr} capture reflection attenuation and Fresnel zone obstruction attenuation, respectively. We will be using 5 dB for both terms for the sake of approximation, as done in [18]. L_{atm} is zero because, in this case, the signal is not passing through the atmosphere. L_{misc} will be set to 2 dB to take into account miscellaneous losses (e.g., cable loss). $L_{process}$ is the processing gain (i.e., spreading gain) associated with the DSSS.

This analysis is focused on detecting the instantaneous center frequency used by the ground user. Therefore, the FHSS processing gain (i.e., the ratio of the hopping bandwidth to the bandwidth after DSSS spreading the narrowband signal) does not come into play in the link budget. As a case study, we will use the processing gain of the 802.11b/g DSSS implementation, which is about 10 dB [19]. This DSSS spreading gain figure can be thought of as the ratio of the narrowband bandwidth to the instantaneous radio frequency (RF) bandwidth.

The system noise temperature T is the sum of the antenna noise temperature and receiver noise temperature, both of which are based on implementation [15]. If the system noise figure is known (denoted as F_{noise} and typically given in dB), then the noise temperature can be calculated by [15]:

$$T = 290(10^{F_{noise}/10} - 1) \quad \text{[Kelvin]}. \tag{10}$$

The value of 290 K is based on typical room temperature. In the case of the jammer pointing at the ground user with a 15 dB gain receive antenna, we will approximate $F_{noise} = 3.15$ or $T = 310$ K, due to the fact that the jammer is pointed towards the surface and not the sky.

The bandwidth before spreading B is largely based on the symbol rate, which we will assume to be 1 MSps. The occupied RF bandwidth will be found as if it were a single-carrier signal utilizing raised cosine filters, using the equation:

$$B = R_S(1 + \alpha), \tag{11}$$

where α is the roll-off factor, which is often around the value of 0.25. Using our assumptions, this comes out to $B = 0.625$ MHz, which is a reasonable amount considering 10 dB of DSSS spreading was assumed, and the DSSS signal is then hopped in frequency across 2 GHz of bandwidth. Thus, the bandwidth after DSSS spreading is $B = 6.25$ MHz, and the FHSS spreading gain is 25 dB using 320 adjacent hopping channels, because:

$$(0.625 \text{ MHz})(10)(320) = 2 \text{ GHz}. \tag{12}$$

Using these many approximations, we can now calculate the SNR while varying the distance between the jammer and ground user. The results are shown in Figure 9, with the horizontal line representing the threshold for frequency detection. It can be seen that our SNR threshold is reached at about 8 km. Therefore, the jammer must be within this distance of the target ground user in order to be able to detect the signal and effectively perform reactive jamming.

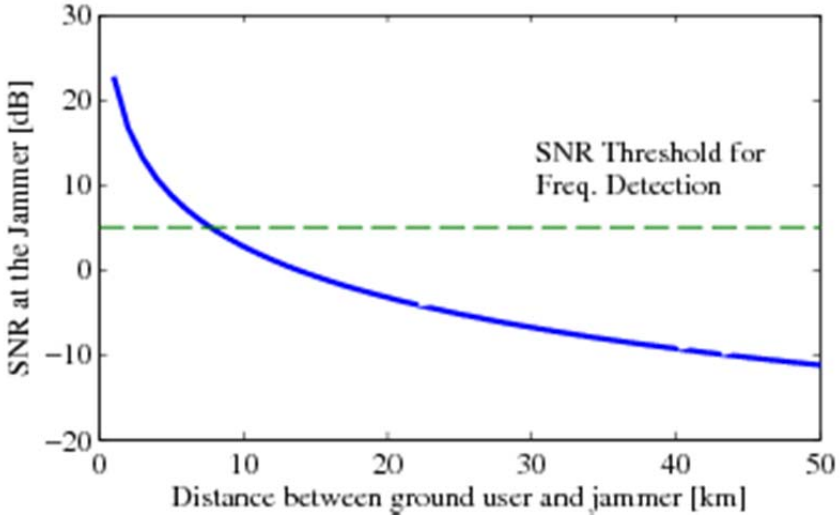

Figure 9. SNR at Jammer When Receiving the Ground User's Signal

3.2.4 SNR During Downlink Jamming

A downlink attack involves injecting noise into the ground user's receiver, and therefore the jammer must receive the signal transmitted by the satellite, as shown in Figure 10. To find how well the jammer can receive this signal, we will reinvestigate the parameters described in the uplink attack subsection, but for the downlink jamming scenario. L_{path} is now the distance between the jammer and satellite, and is considerably larger than that of the uplink attack. Geostationary orbit is located 35,786 km above the Earth's surface, in which the surface is defined by the mean sea level. If we assume that the jammer is in the same meridian as the satellite, then we can use the latitude angle of the jammer to calculate the actual distance d between the two antennas:

$$d = \sqrt{R^2 + r^2 - 2Rr\cos\theta}, \tag{13}$$

where R is the radius of the Earth (roughly 6.37e6 m), r is the orbital radius of the satellite (42e6 m for geostationary orbit), and θ is the latitude of the jammer. Thus, for our analysis this equation can be reduced to:

$$d = \sqrt{1.818e15 - 5.37e14\cos\theta}. \quad (14)$$

To find the distance between the jammer and satellite, we will consider three scenarios, in which the jammer is located at 0°, 45°, and 60° latitude.

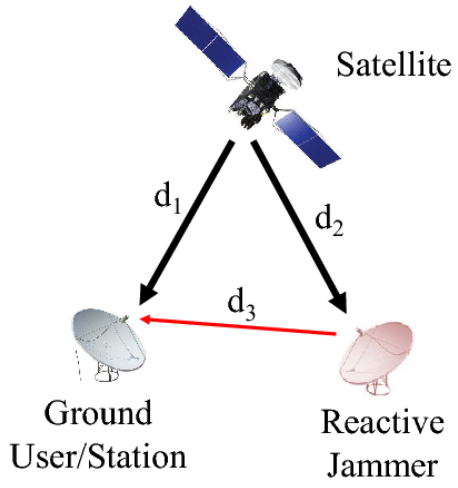

Figure 10. Downlink System Diagram

Table 2 shows the distance d between the jammer (or ground user) and satellite, as well as the propagation delay, which will be used later. The downlink carrier is at 20 GHz as discussed earlier, and in this case, we calculate L_{path} using the free space path loss equation (with an atmospheric attenuation loss added separately):

$$L_{path} = (4\pi df/c)^2. \quad (15)$$

P_T is the transmit power of the satellite, which we will estimate to be 100 W or 20 dBW, based on real-world examples [15]. G_T is the satellite's transmit antenna gain with respect to the ground user's location, but because the jammer must be within a beam width of the ground user (as shown later), we will assume that the jammer experiences the same transmit antenna gain, estimated to be 40 dB. G_R is now much higher because we will assume the jammer has a directional antenna pointed at the satellite. It is reasonable to assume that the jammer has accurate knowledge of the satellite's coordinates, because it is in geosynchronous orbit and the

orbit locations are in the public domain. We will vary this figure between 10 dB and 40 dB. L_{atm} is nonzero now that the signal is passing through the atmosphere. We will use an approximate figure of 1 dB for the uplink and downlink based on the recommendations in [20]. L_{misc} will again be set to 2 dB to take into account miscellaneous losses. The noise temperature T is dependent on the receiver characteristics, but we will approximate it to be 200 K because the jammer's receive antenna is pointed into the sky [15]. We will assume the bandwidth of the downlink signal is the same as the uplink: $B = 0.625$ MHz.

Table 2. Distance and Delay Between the Jammer (or Ground User) and Satellite

Latitude	0°	45°	60°
Distance *d*	35,793 km	37,928 km	39,367 km
Propagation Delay	0.1193 s	0.1264 s	0.1312 s

The SNR while varying G_R for the downlink jamming case is shown in Figure 11, parameterized by the geographical location of the jammer in latitude. Based on the SNR threshold (shown as a red line), the jammer must obtain a receive antenna gain of at least 20 dB for reactive jamming to be feasible. For example, the jammer may choose to have a fairly compact 20 dB gain parabolic reflector antenna pointed towards the satellite.

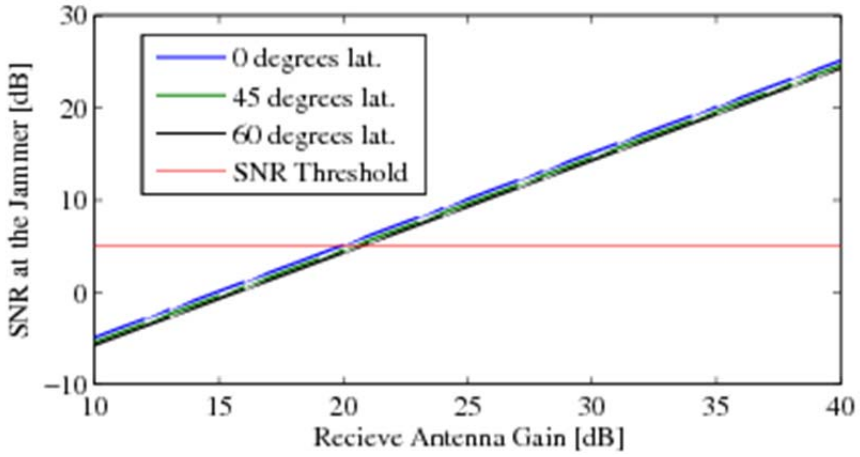

Figure 11. SNR at Jammer When Receiving the Satellite's Signal

3.2.5 Jammer-to-Signal Ratio Component

In order to fully corrupt a communications link, the jammer must cause the bit error rate (BER) after forward error correction (FEC) to be significantly high. Either the data must be significantly corrupted, or the control information must be corrupted to the point at which the data cannot be successfully decoded. While the data error rate can be discussed in terms of BER, it is also common to consider the block error rate (BLER) or packet delivery ratio (PDR) [2]. In FHSS systems, hop error rate can also be used as a metric. Errors on the control information, on the other hand, are more difficult to quantify because the BER metric has much less meaning. For the rest of this analysis, we will only discuss errors on the data and use the BER metric (note, this can be thought of as indirectly taking into account erroneous control information).

The worst-case BER is 0.5, although a BER around 0.1 (after forward error correction) will likely cause denial of service (DOS) due to the number of retransmissions and dropped packets. Even if the information contains real-time voice or video data, which cannot be retransmitted due to their time sensitivity, the coder and decoder (CODEC) would be unable to produce intelligible audio or video. The actual BER or BLER threshold is based off of numerous factors on many different layers and would be best acquired empirically [21]. Therefore, we will continue by assuming a BER threshold to cause DOS of 0.1, or 10%, of the information bits received in error after demodulation and decoding. The minimum JSR to cause a BER of 0.1 is mostly dependent on the modulation scheme and the method of forward error correction (i.e., channel coding). Assuming the communication system supports adaptive modulation and coding, in a harsh environment it will likely ratchet down to binary phase-shift keying (BPSK) or quadrature phase-shift keying (QPSK) using a low rate coding scheme. To gain further insight, we observe a frame error rate (FER) curve for BPSK in an additive white Gaussian noise (AWGN) channel using low density parity check (LDPC) coding at various rates (based on the DVB-S2 protocol), shown in Figure 12. It is clear that around -2 dB of JSR (or 2 dB of SNR), even the most advanced coding reaches a point where the amount of errors would cause link failure.

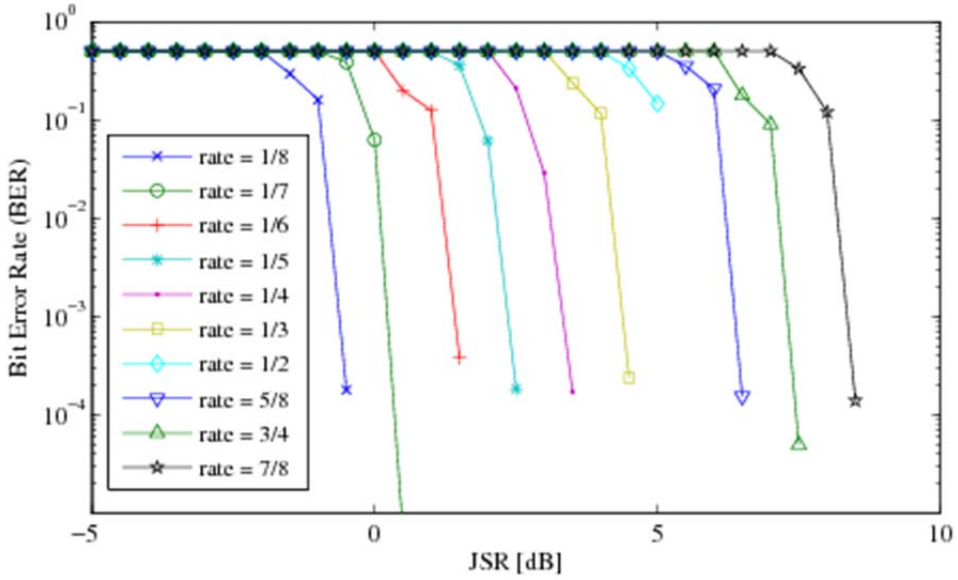

Figure 12. BER Curve For BPSK in an AWGN Channel With LDPC Coding at Various Rates

Given details of a communication link, the JSR can be found by calculating the received friendly signal power P_R and received jamming signal power P_{R-JAM}. Channel noise is ignored because the received jamming power will be significantly higher than the received noise power under successful jamming. P_R and P_{R-JAM} are found using the link budget in Equation (7), except $L_{process} = 0$ for P_R because the communications receiver (either at the ground station or satellite) is able to de-spread the DSSS. In the following two sections, we evaluate these link budgets for the uplink and downlink jamming scenarios respectively.

3.2.6 Uplink Jamming JSR

We will carry over the assumptions made in the previous section in order to map the analysis to a realistic scenario. In addition, we will assume that the jammer has a directional antenna pointed at the satellite. Without this capability, the attack is not feasible. We will also assume that the jammer is located close enough to the ground users such that the jammer is located inside the spot-beam (i.e., region illuminated by the main-lobe of a directional antenna). Previously, we approximated the ground user's transmit antenna gain to have a -10 dB gain in the direction of the jammer, but now we will only care about the main-lobe gain, which we will

approximate to 35 dB after cable losses (a figure which is included in [16]). The transmit power of the jammer is set at 10 dB higher than the ground user's, which accounts for the 10 dB loss associated with the DSSS processing gain. The result is a JSR of 0 dB (due to the other gains and losses being equal), which is above our threshold of -2 dB and will likely cause degradation of the link or full denial of service. Table 3 provides the full link budgets.

In summary, feasibility is achieved by the jammer transmitting at a much higher power than the user, to compensate for the DSSS processing gain. If the SATCOM system does not use DSSS, then the jammer would simply have to transmit at around the same power as the user.

Table 3. Uplink Jamming Attack Link Budget to Calculate JSR

Jammer Power Budget				**Signal Power Budget**		
P_T	21	dBW		P_T	11	dBW
G_T	35	dB		G_T	35	dB
G_R	40	dB		G_R	40	dB
L_{path}	-216	dB		L_{path}	-216	dB
L_{atm}	-1	dB		L_{atm}	-1	dB
L_{misc}	-1	dB		L_{misc}	-1	dB
$L_{process}$	-10	dB				
				P_R	-132	dBW
$P_{R\text{-}Jam}$		dBW				
		JSR =	0.0	dB		

3.2.7 Downlink Jamming JSR

We will now use the same approach for the downlink jamming scenario. The jammer will still use a 15 dB gain antenna pointed toward the target ground user(s), with a P_T = 21 dBW. We will start off by investigating a scenario in which the jammer is 10 km from the ground user. Using the rural path loss model described in Equation (9), this comes out to a path loss of 148.5 dB (for the jamming signal). Table 4 provides full assembly of the link budgets.

Table 4. Downlink Jamming Attack Link Budget to Calculate JSR

Jammer Power Budget				**Signal Power Budget**		
P_T	21	dBW		P_T	20	dBW
G_T	15	dB		G_T	40	dB
G_R	-10	dB		G_R	35	dB
L_{path}	-148.5	dB (10 km)		L_{path}	-210	dB (20 GHz)
L_{atm}	0	dB		L_{atm}	-1	dB
L_{misc}	-1	dB		L_{misc}	-1	dB
$L_{process}$	-10	dB				
				P_R	-117	dBW
$P_{R\text{-}Jam}$	-133.5	dBW				
		JSR =	-16.5	dB		

When the jammer uses a transmit antenna gain of 15 dB, there is no way to achieve a high enough JSR for successful jamming, mainly due to the high antenna gains associated with the signal power budget. For the jammer to achieve an extra 16.5 dB and even out the budgets (i.e., balance the budget), it could either use a 32 dB gain antenna, which is possible if it knows exactly where the target is, or be positioned 1 km from the target instead of 10 km.

3.2.8 Geometric Component

So far, we have investigated the first two reactive jamming constraints. However, even if the SNR and JSR requirements are met, the adversary may still fail to cause denial of service (a desirable outcome for our countermeasure discussed later). In order for a reactive jammer to be effective, the jamming signal must reach the target receiver before it hops to a new frequency. Therefore, the geometry of the system is an important factor when analyzing the feasibility of reactive jamming. In addition, the delay between when the jammer receives the target signal and transmits the jamming signal, known as the jammer's processing delay, should be included in this component. Figure 13 shows the time and frequency behavior of a repeater jammer (red) jamming a target signal (green). Despite the time delay between the friendly signal arriving and jamming signal arriving, the signals are vulnerable to jamming.

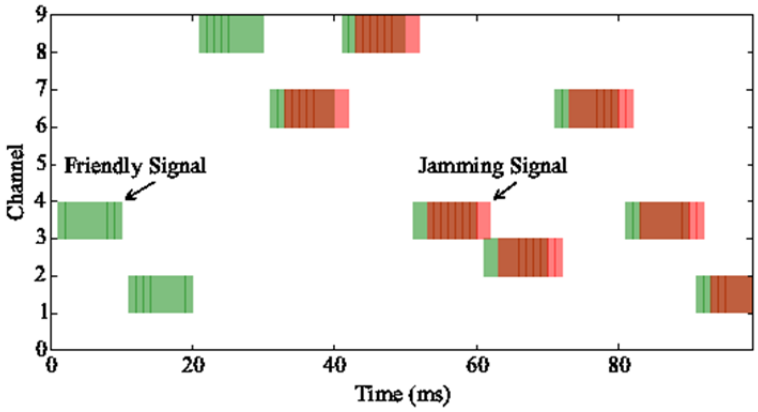

Figure 13. Time and Frequency Behavior of a Repeater Jammer

In order to quantify the impact of jamming, we can abstract the physical layer by assuming a certain fraction of each hop must be jammed at the receiver in order to cause denial of service, a fraction that we will denote as η. The value of η largely depends on the channel coding scheme, interleaving, and JSR. Using this abstraction, for a repeater jammer to be successful we must have:

$$\frac{d_2}{c} + T_j + \frac{d_3}{c} \leq \frac{d_1}{c} + (1-\eta)T_h, \tag{16}$$

where T_j is the processing plus analog delay associated with the jammer, T_h is the duration of each hop, and the distances d are as shown in Figure 7 and Figure 10 for uplink jamming and downlink jamming, respectively. This equation shows that hopping at a faster rate can be used to protect communications from repeater jamming, which is a well-known mitigation strategy [11]. To gain further insight, we must estimate the fraction of each hop that must be jammed, η.

3.2.9 Fraction of Each Hop That Must Be Jammed

Fundamental limitations on the effectiveness of repeater jamming against FHSS are derived in [11]. For example, the authors show that the average symbol error probability can be formulated by introducing P_j, the probability a symbol is jammed given the jamming signal is present during the reception of that symbol. If we assume that the non-jammed symbol error rate is F_{nj} and the symbol error rate under jamming is F_j, then the average symbol error probability is given by [11]:

$$P_s = \frac{T_h - T_{nj}}{T_h} P_j F_j + \left(1 - \frac{T_h - T_{nj}}{T_h} P_j\right) F_{nj}, \quad (17)$$

where T_{nj} is the duration of the hop in which the jamming signal is not present, which is equal to $T_j + (d_2 + d_3 - d_1)/c$. If the hop is not jammed at all, then $P_s = F_{nj}$. It should be noted that F_{nj} is largely determined by the SNR, while F_j is mostly impacted by the JSR.

To estimate T_h, we will assume a symbol rate of 1 MSps. Using 50 symbols per hop, we have a hop duration of about 50 μs, although in the simulation we vary the number of symbols per hop between 10 and 500 to cover all realistic FHSS configurations (we will not take into account a fractional symbol per hop system). For now, we will assume the jammer takes 4 symbols to detect the center frequency and another 1 symbol worth of time to generate and transmit the jamming waveform (this will include the time it takes to tune the transmitter to the right frequency). This results in a jammer processing delay of $T_J = 5\ \mu s$, a figure that is discussed in more detail later.

We will assume that the ground user and jammer are approximately the same distance from the satellite. This is an acceptable assumption because if the jammer is within about 20 km from the ground user, then the most the two distances can deviate is only about 10 - 20 meters, because the distance between the Earth and satellite is three orders of magnitude greater than between the ground user and jammer. In an uplink attack, this distance corresponds to d_1 and d_3, while in a downlink attack, it is d_1 and d_2. Therefore, the feasibility of an uplink and downlink attack can be analyzed simultaneously because of the system's symmetry. In this case, the minimum distance between the ground user and jammer for successful jamming becomes equal to $(T_h - \eta T_h - T_j)c$. Figure 14 shows the maximum distance the jammer can be from the ground user in order to successfully jam a hop, as a function of η, parameterized by varying the number of symbols per hop. The saturation on the right-hand side of each curve is due to the jammer's processing delay, T_J, of 5 μs.

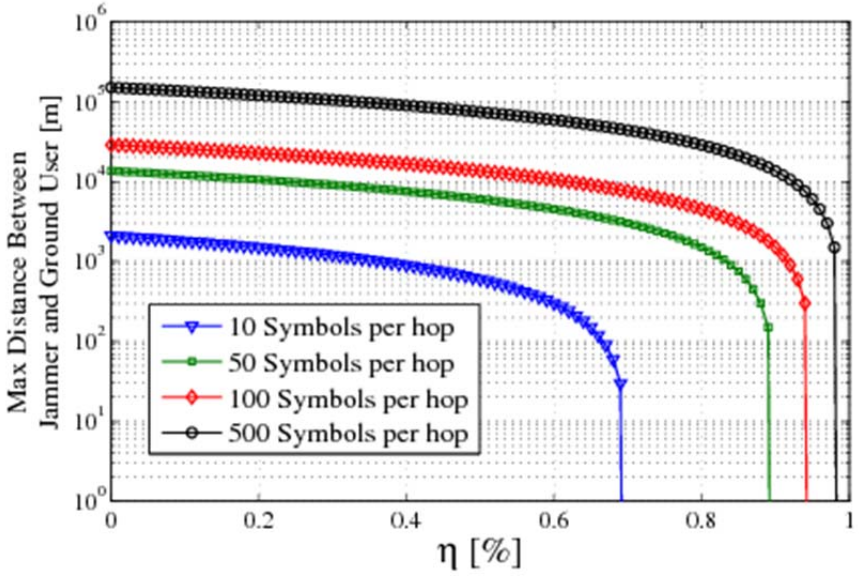

Figure 14. Maximum Distance the Jammer Can Be From the Ground User in Order to Successfully Jam a Hop, as a Function of η

As seen in these simulations, for the 50 symbols per hop case, the jammer will have to be within about 10 km from the ground user in order to perform successful reactive jamming, due to the geometric constraints. This assumes the JSR is received at a high enough level and the SNR at the jammer is high enough to repeat the signal. While these results are largely dependent on the assumptions made, they provide insight to the limitations of reactive jamming SATCOM (e.g., it is highly unlikely to perform successful reactive jamming when d_2 is larger than around 10 km).

4 RESULTS AND DISCUSSION

4.1 Recursive Learning Mitigation Strategy

In this section, we present some simulation results to show proof-of-concept of our proposed technique. To simulate this RL based mitigation strategy, a link layer simulation framework was created, which included the jammer models described in this section. Q-learning was chosen as the RL technique [12]. In terms of Q-learning parameters, a learning rate, α, of 0.95 (the transmitter will quickly use learned knowledge) and discount factor, γ, of 0.8 was used for the simulations. This relatively low discount factor was used because of the cyclic nature of

the optimal policies. Figure 15 shows the reward over time for various jamming models, depicting the learning process until saturating to an effective policy with constant reward. Because the reward from each time step is proportional to link throughput, the results can be interpreted as throughput over time. The barrage jammer was modeled by causing jamming with 20% probability at each time step, regardless of how long the transmitter has been transmitting or on a given channel. This can be thought of as a nonreactive jammer that is always transmitting, but at a jammer-to-signal ratio that is not quite high enough to cause complete denial of service.

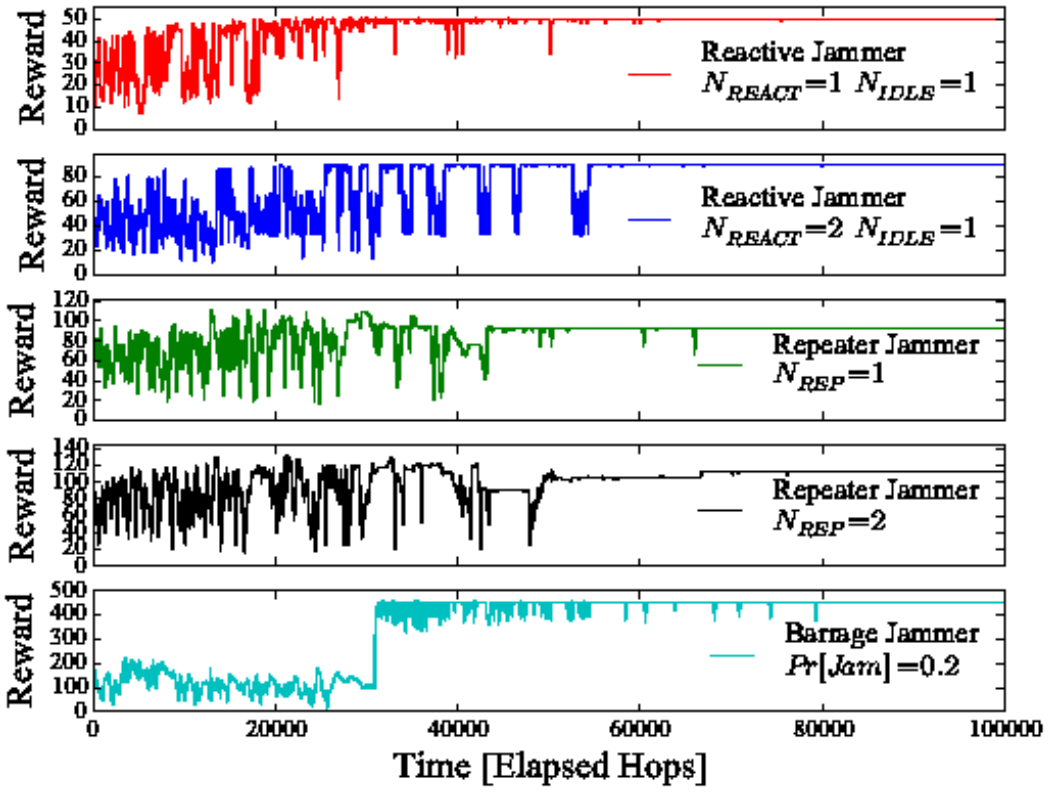

Figure 15. Simulation Results Showing the Learning Process Over Time in the Presence of Different Jammers

The maximum achievable reward under each jamming behavior varies, which is expected (e.g., $N_{REACT} = 2$ will allow using a higher duty cycle than $N_{REACT} = 1$). Although not depicted in Figure 15, it should be noted that the transmitter learned the optimal policy only during the reactive jamming and barrage jamming scenarios. In both repeater jamming scenarios, the learned policy did not traverse the entire zigzag pattern on the MDP, which is the optimal policy

for the repeater jamming model as discussed earlier. Rather, the transmitter would go idle on occasion, which would essentially reset the zigzag pattern. Hence, the reward achieved under repeater jamming was not the maximum possible reward. Under barrage jamming, the optimal policy for the transmitter would be to remain transmitting on the same channel indefinitely, which occurred after around 50,000 time steps, except for the occasional channel hop (as indicated by the small dips in the plot). This demonstrates how the proposed strategy can work under non-reactive jamming, despite not being designed to do so, and even provide better throughput than a constant-rate FHSS approach by avoiding the overhead associated with changing channels.

It should be noted that the time taken to learn an effective strategy for a given jammer is a function of the learning rate parameter and learning technique (Q-learning in this case). Results in Figure 15 show a learning time between 30,000 and 50,000 time steps, which is one or two seconds in a system where the minimum hop duration is on the order of tens of microseconds. While this may seem long compared to a radio that is preprogrammed with specific anti-jam strategies, it is unlikely that the presence of different jammers will change within seconds. In addition, the preprogrammed radio must spend time classifying the type of jammer present in order to know which mitigation scheme to use, a process which is not needed for the proposed strategy. We remind the reader that although wireless channel conditions are known for changing within milliseconds, the proposed strategy is meant to counter the adversary, not traditional channel imperfections such as fading or Doppler shift.

4.2 Simulation Using Systems Tool Kit

As a form of validation, the scenarios discussed in this report were simulated using a software package called Systems Tool Kit (STK). The orbit of a generic geostationary satellite over the Americas was used. Figure 16 shows the 3D view of the uplink jamming scenario. The top portion is zoomed into the ground user and jammer (both transmitting towards the satellite), while the bottom shows the communications satellite in geosynchronous orbit, receiving in the direction of the ground user. The rainbow cones show a 3D projection of the antenna radiation pattern, while the red cone represents the jammer's receiver pointing at the ground user. The jammer was placed on a ground vehicle, so that the distance between the jammer and ground user could be varied between 0 - 50 km (the red line shows the ground vehicle path).

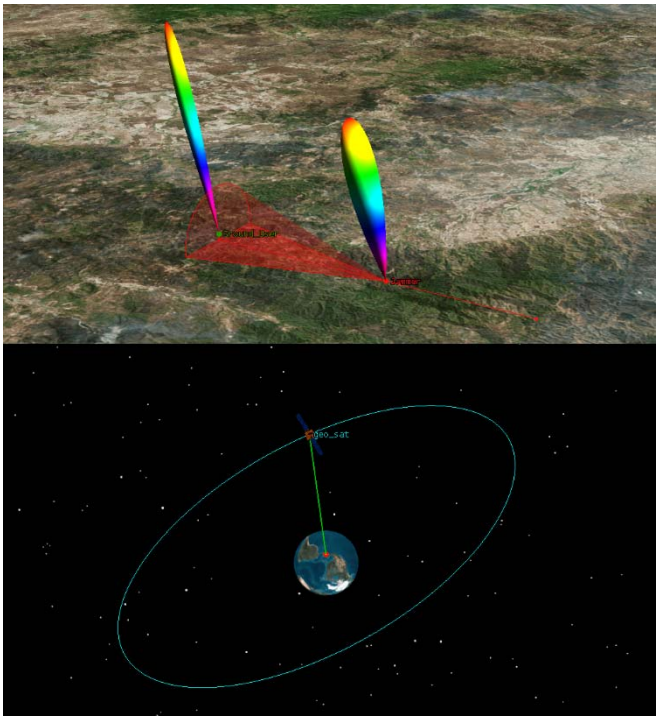

Figure 16. Screenshots of the Uplink Jamming Scenario in STK

These STK scenarios were used to verify the results found previously. Table 5 highlights some of the numerical results found, with notes describing the specific conditions under which the simulations were performed. In all four cases, these simulated results closely match the analysis in prior sections, which is expected considering the more complex models do not deviate too much from the simple models used earlier.

Table 5. STK Simulation Results Showing JSR and SNR for the Uplink/Downlink Scenarios

Scenario	Notes	Result
Uplink SNR	10 km between user and jammer	3 dB
Downlink SNR	at 0° latitude with $G_R = 30$ dB	15 dB
Uplink JSR	same parameters as in link budget	1 dB
Downlink JSR	same parameters as in link budget	-15 dB

Figure 17 shows the results of a simulation that attempts to reproduce the previously presented Figure 9. Both methods show that the SNR threshold of 5 dB is reached at about 8 km. It should be noted that the antenna isolation figure of 10 dB, used throughout this analysis, was included in the simulations.

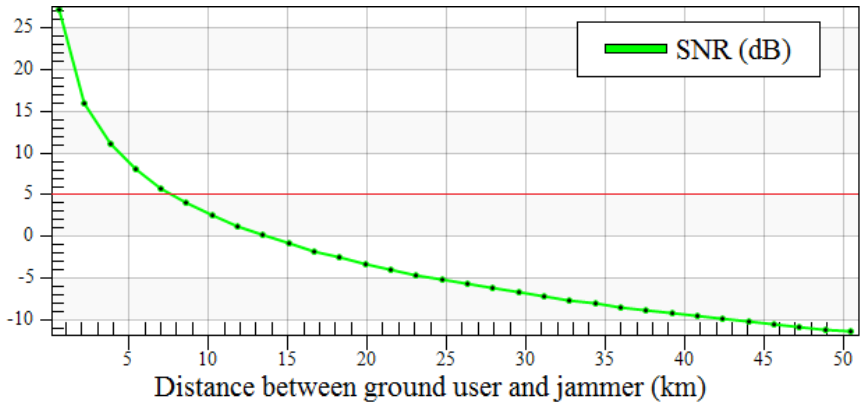

Figure 17. STK Simulation Showing SNR Measured During the Uplink Attack (at the Jammer)

4.3 Coding and Interleaving

While there exists many mitigation strategies for various forms of jamming, we would like to propose one that is based on the insights developed from this project. ***The proposed reactive jamming mitigation strategy is to use a coding and interleaving scheme that results in the transmitted bits (at a 1:1 coding ratio) appearing at the very beginning of each hop.*** The proposed mitigation strategy takes advantage of the geometric constraints of reactive jamming. There would need to be some sort of error detection (e.g., a cyclic redundancy check (CRC) at the higher layer) that checks validity of multiple hops worth of information at once. Most coding literature assumes AWGN or fading channel in which the erasures do not correspond to the last portion of every hop, but as we found out in the geometric constraints portion, jamming the very beginning of each hop is significantly challenging. This strategy would require that synchronization symbols and reference symbols (i.e., *pilots*) not be used in the very beginning of each hop, or at least used sparingly.

As demonstrated using Figure 15 and the related analysis, if the jammer is forced to overlap with over about 90% of the target hop to cause denial of service (i.e., $\eta > 0.9$), then reactive jamming really is not very feasible. While this specific percentage threshold varies based on number of symbols per hop, the jammer's processing delay, and the geometry of the system, the proposed anti-jam strategy can at least attempt to maximize the value of η in order to mitigate most reactive jamming scenarios. As long as the geographical area within close proximity to the ground user is clear of reactive jamming, using this method of coding and interleaving to raise the value of η will provide strong protection against reactive forms of jamming.

Figure 18 shows simulation results in which the *link state* is measured while varying η and the symbols per hop. In every single jammed case, it is the uplink that is denied. This is the type of figure that can be used as a guide in determining how high η must be, based on the number of symbols per hop (or how fast to hop if η is constant).

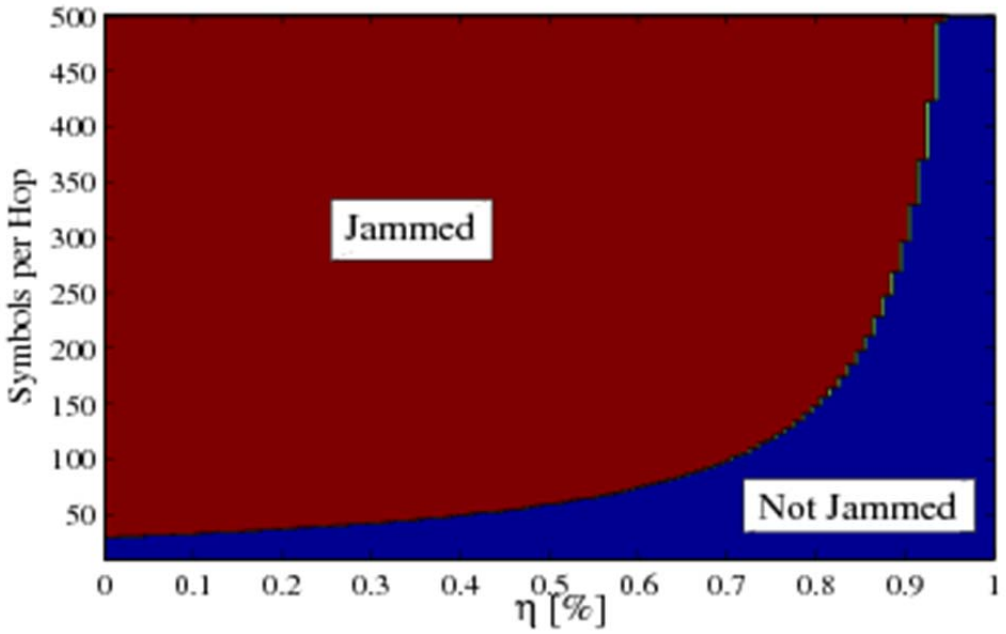

Figure 18. Feasible Region of Countermeasure

5 CONCLUSIONS

Under this research grant project, the authors have developed and described herein a reinforcement learning based strategy that a communication system can use to deal with reactive jammers of varying behavior by learning an effective channel hopping and idling pattern. This is directly applicable to military satellite communications in a contested environment. Simulation results provide a proof-of-concept and show that a high-reward strategy can be established within a reasonable period of time, with the exact time being dependent on the duration of a time step.

An advantage of the approach described in this report is that it can counter a wide range of jamming behaviors, not known *a priori*. Without needing to be preprogrammed with anti-jam strategies for a list of jammers, our approach is able to better adapt to the environment. *The proposed technique is best used in tandem with an algorithm that finds a favorable subset of channels to use, as well as modern optimization techniques such as adaptive modulation and forward error correction.* In future work, we would investigate expanding the Markov Decision Process state-space to take into account additional factors, as well as explore more stochastic jammer models. In addition, it is likely that the reinforcement learning procedure can be tuned to provide even greater performance.

Further, we analyzed the feasibility of performing reactive jamming (including repeater jamming) in a satellite communications scenario, using a three step approach based on the signal to noise ratio, jammer to signal ratio, and geometric jamming constraints. Additional clarity was provided by splitting the analysis into uplink and downlink jamming. Even though reactive jamming is a complex form of jamming that requires receiving capabilities and a low processing delay, it allows a jammer to counter the processing gain associated with frequency hopping spread-spectrum. It is for this reason that we should consider reactive jamming a future threat in the satellite communications domain, especially as software-defined radios become more capable and less costly. Results of each of the three steps are summarized in Table 6.

The overall feasibility of uplink and downlink jamming is highly dependent on the scenario at hand, although we showed an example scenario that would allow for a jammer to deny communications via uplink jamming but not downlink jamming. Because both the uplink and downlink are vital to the operation of a satellite communications link, only one of the two needs to be jammed. Thus, the overall conclusion of this paper is that reactive jamming is in fact

feasible in a satellite communications scenario, and reactive-specific countermeasures should be considered by system designers. However, this work showed that as long as the geographical area around user terminals is free of reactive jammers, substantial mitigation can be achieved using the proposed mitigation strategy.

Table 6. Summary of Feasibility Analysis

Portion	Feasible	Notes
SNR-uplink	Yes	Jammer must be within about 8 km of ground user
SNR-downlink	Yes	20 dB or higher receive antenna gain
JSR-uplink	Yes	Jammer compensates for processing gain with P_T
JSR-downlink	No	Not without described augmentations
Geometry-both	Yes	As long as the jammer is $<$ 10 km from user

6 RECOMMENDATIONS

As a follow-on to analyses and results accomplished during this effort, we recommend conducting research into anti-jam strategies that leverage machine learning and cognition in order to decrease the likelihood of in-band interference, which degrades link quality. Also, we recommend expanding the Markov Decision Process state-space to take into account additional factors, as well as explore more stochastic jammer models. It is likely that the reinforcement learning procedure can be tuned to provide even greater performance. We also recommend developing performance metrics for determining effectiveness in terms of ability to protect communication, computational complexity, and technical feasibility. Satellite communication system design is strongly an optimization, and a metric would be useful in determining the best approach for a given scenario with given constraints.

REFERENCES

1. Boyd, J. A., Harris, D. B., King, D. D., and Welch, H. W. Jr., *Electronic Countermeasures*, Peninsula Publishing, 1978.

2. Xu, W., Trappe, W., Zhang, Y., and Wood, T., "The Feasibility of Launching and Detecting Jamming Attacks in Wireless Networks," in *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 46–57, 2005.

3. Adamy, D. L., *EW 101*, Artech House, 2001.

4. Yang, M., and Grace, D., "Cognitive Radio with Reinforcement Learning Applied to Heterogeneous Multicast Terrestrial Communication Systems," in *Cognitive Radio Oriented Wireless Networks and Communications,* pp. 1–6, 2009.

5. Zhu, Y., Li, X., and Li, B., "Optimal Adaptive Anti-jamming in Wireless Sensor Networks," *International Journal of Distributed Sensor Networks*, Article ID 485345, 2012.

6. Wu, Y., Wang, B., and Lui, K., "Optimal Defense Against Jamming Attacks in Cognitive Radio Networks Using the Markov Decision Process Approach," in *IEEE Global Telecommunications Conference,* pp. 1–5, Miami, FL, 2010.

7. Sodagari, S., and Clancy, T. C., "An Anti-Jamming Strategy for Channel Access in Cognitive Radio Networks," in *Decision and Game Theory for Security*, pp. 34–43, 2011.

8. Wertz, J. R., and Larson, W. J., *Space Mission Analysis and Design*, Microcosm, 1999.

9. Edrich, M., and Schmalenberger, R., "Combined DSSS/FHSS Approach to Interference Rejection and Navigation Support in UAV Communications and Control," in *2002 IEEE Seventh International Symposium on Spread Spectrum Techniques and Applications*, Vol. 3, pp. 687–691, 2002.

10. Ince, A. N., *Digital Speech Processing: Speech Coding, Synthesis and Recognition*, Springer, Vol. 155, 1992.

11. Torrieri, D. J., "Fundamental Limitations on Repeater Jamming of Frequency-Hopping Communications," *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 4, pp. 569–575, 1989.

12. Sutton, R. S., and Barto, A. G., *Reinforcement Learning: An Introduction*, Vol. 1., Cambridge University Press, 1998.

13. Dudley, S. M., Headley, W. C., Lichtman, M., Imana, E. Y., Ma, X., Abdelbar, M., Padaki, A., Ullah, A., Sohul, M. M., Yang, T., and Reed, J. H., "Practical Issues for Spectrum Management with Cognitive Radios," *Proceedings of the IEEE*, Vol. 102, No. 3, pp. 242 – 264, March 2014.

14. Poor, H. V., *An Introduction to Signal Detection and Estimation*," Springer, 1994.

15. Pratt T., Bostian, C., and Allnutt, J., *Satellite Communications*, Electronic Industry Press, Beijing, 2005.

16. Lewis, T., "TDRSS 2nd Workshop," http://msp.gsfc.nasa.gov/TUBE/pdf/infopack.pdf, Omitron, Inc., 1996.

17. ThinKom Solution, "ThinKom's ThinSat Talos Integrated Antenna / Terminal Subsystem," http://thinkom.net/wp-content/uploads/2013/11/TK-TSatTalos-Data-Sheet.pdf, 2013.

18. Anderson, H. R., *Fixed Broadband Wireless System Design*, John Wiley & Sons, 2003.

19. "IEEE 802.11b-1999," http://standards.ieee.org/getieee802/download/802.11b-1999.pdf, 2014.

20. R. ITU-R P.676-2, "Attenuation by Atmospheric Gases," http://www.itu.ch/itudoc/itu-r/rec/p/676-2_29169.html, 1995.

21. Lichtman, M., Reed, J., Clancy, T., and Norton, M., "Vulnerability of LTE to Hostile Interference," in *2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP),* pp. 285–288, Austin, TX, Dec 2013.

LIST OF SYMBOLS, ACRONYMS, AND ABBREVIATIONS

3D	Three Dimension
ACK	Acknowledge
ACM	Association for Computing Machinery
AWGN	Average White Gaussian Noise
BER	Bit Error Rate
BLER	Block Error Rate
BPSK	Binary Phase Shift Keying
CODEC	Coder and Decoder
CRC	Cyclic Redundancy Check
dB	Decibel
dBW	Decibel relative to 1 Watt
DOS	Denial of Service
DSSS	Direct Sequence Spread Spectrum
DVB-S2	Digital Video Broadcasting - Satellite - Second Generation
EIRP	Equivalent Isotropically Radiated Power
FEC	Forward Error Correction
FER	Frame Error Rate
FHSS	Frequency-Hopping Spread Spectrum
GHz	Giga-Hertz
Hz	Hertz
JSR	Jammer to Signal Ratio
K	Kelvin
Km	Kilometer
LDPC	Low Density Parity Check
LOS	Line of Sight
m	Meters
MAC	Media Access Control

MDP	Markov Decision Process
MHz	Mega Hertz
MSps	Million Samples per Second
NACK	Not Acknowledged
NASA	National Aeronautics and Space Adminstration
N_{IDLE}	Number of idle time steps
N_{REP}	Threshold number of time steps to trigger repeater jammer
N_{REACT}	Threshold number of time step to trigger reactive jammer
OFDMA	Orthogonal Frequency-Division Multiple Access
PDR	Packet Delivery Ratio
PHY	Physical Layer
QPSK	Quadrature Phase-Shift Keying
RF	Radio Frequency
RL	Reinforcement Learning
s	Second
SATCOM	Satellite Communication
SNR	Signal to Noise Ratio
STK	Systems Took Kit
TDRS	Tracking and Data Relay Satellite
T_{off}	Idle Time of a Transmitter
T_{on}	Active Transmission Time of a Transmitter
VTM-1220	A moblie two-way ground terminal for remote Internet connection
W	Watt
α	Learning Rate
η	The fraction of each hop duration that must remain un-jammed for communications to succeed
ε	Epsilon, probability of occurance
μ	micro
Γ	Discount Factor

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/RVSV/Steven A. Lane 1 cy