3-14-2014

A Methodology For Measuring Resilience in a Satellite-Based Communication Network

Jonathan S. Turner

A METHODOLOGY FOR MEASURING RESILIENCE IN A SATELLITE-BASED COMMUNICATION NETWORK

THESIS

Jonathan S. Turner, Second Lieutenant, USAF
AFIT-ENS-14-M-31

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-14-M-31

A METHODOLOGY FOR MEASURING RESILIENCE IN A SATELLITE-BASED COMMUNICATION NETWORK

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Jonathan S. Turner, MS

Second Lieutenant, USAF

March 2014

AFIT-ENS-14-M-31

A METHODOLOGY FOR MEASURING RESILIENCE IN A SATELLITE-BASED COMMUNICATION NETWORK

Jonathan S. Turner, MS

Second Lieutenant, USAF

Approved:

//SIGNED//	17 March 2014
Richard F. Deckro, DBA (Chairman)	Date
//SIGNED//	17 March 2014
James W. Chrissis, PhD (Member)	Date

Abstract

According to Argonne National Laboratories, "Resilience is the ability of an entity to anticipate, resist, absorb, respond to, adapt to, and recover from a disturbance." Resilience in a system is important because it allows a system to adapt its operations to unknown and altered operational environments. Presidential Policy Directive 21 states that increasing resilience of critical infrastructures is not only desired, but United States policy. Communications infrastructures are one such critical infrastructure.

The purpose of this research is to develop a methodology for measuring resilience in satellite communication systems for use as a key criterion in the selection and acquisition of new satellite architectures, in accordance with the National Security Space Strategy. The base methodology utilized in this thesis is Extreme Event Modeling implemented through the use of Bi-Level Programming with monotonically nonlinear continuous and mixed integer variables. This model differs from previous efforts applied to other critical infrastructures in that it captures the temporal component associated with multiple events, as well as the repairs, or reconstitution, of infrastructure components. Furthermore, a heuristic based upon a ratio of impact to cost and local searches is developed to solve the resulting continuous bi-level problem.

To my wife, who suffered through my thesis with me and still kept her sanity… mostly.

Acknowledgements

I would like to thank my very patient research advisor, Dr. Deckro, for suffering through my scattered method of writing and continuously pulling me back to the ever-disliked Literature Review, and my reader, Dr. Chrissis, for enduring through the extra work I created for him. I would also like to show appreciation to my sponsor who gave me the utmost freedom to pursue the problem however I chose, supporting any decision that I could explain with a picture.

Jonathan S. Turner

Table of Contents

List of Figures

Figure Page

List of Tables

Table Page

A METHODOLOGY FOR MEASURING RESILIENCE IN A SATELLITE-BASED COMMUNICATION NETWORK

I. The Problem

Why Resilience Matters

In 2010, Washington D.C. was struck by a blizzard which forecasters watched build with the aid of satellite data. It was considered to be one of the largest winter storms that the Mid-Atlantic region had seen in nearly 90 years. As a result of this storm, roads, hospitals, and railways were closed, and three hundred-thousand citizens were without electricity the following day. That was with the aid of the satellites to make preparations. Experiments show that the size of that storm, without the aid of satellites, would have been predicted at less than half of its size and intensity, a prediction that would have left many Americans out in the cold (Cushman).

On July 15, 1996, then-President William J. Clinton signed Executive Order 13010, *Critical Infrastructure Protection*. The very first section was to establish the "President's Commission on Critical Infrastructure Protection."

The purpose of this commission was to identify vulnerabilities, both physical and cyber, as well as to provide "expert guidance to critical infrastructures to detect, prevent, halt, or confine an attack and to recover and restore service" (EO 13010 sect. 7.5.1). What the President was demanding nearly two decades ago is still being worked as a critical issue.

In PPD-21, 16 critical infrastructure sectors are designated, three more than denoted in 2002, but two fewer than in 2009. Those sectors are: Chemical, Commercial Facilities, Communications, Critical Manufacturing, Dams, Defense Industrial Base, Emergency Services, Energy, Banking and Finance, Food and Agriculture, Government Facilities, Public Health, Information Technology, Nuclear, Transportation, and Water. Commercial Facilities, Critical Manufacturing, Dams, and Nuclear are four new sectors, with the former Postal sector being absorbed into Transportation. These new sectors were added in 2009 along with National Monuments and Icons, which, like Postal and Shipping, was absorbed into other sectors in 2013.

On January 11, 2013, the president's National Security Telecommunications Advisory Committee (NSTAC) released a report on its most recent endeavors, which are focused on increasing "the integrity, confidentiality, and availability of Governmental unclassified communications" (NSTAC 2013 p2). In that report, one area of their examinations includes "the interdependencies of networked systems, resulting in higher potential consequences from successful events" (NSTAC 2013 p1). One such network

is the satellite communications network (NSTAC 2013). To eliminate confusion of terms, the mathematical definition of *network*, which is "a set of objects (called *nodes* or *vertices*) that are interconnected", is used throughout. The connections between the nodes are called *edges* or *links*" (Nykamp Network Definition).

More recently, in February 2013 a new Presidential Policy Directive (PPD) was released on the subject of "Critical Infrastructure Security and Resilience", PPD-21. It states that "it is the policy of the United States to strengthen the security and resilience of its critical infrastructure against both physical and cyber threats" (PPD-21 p. 2). Much like in 1996, the goal is to reduce vulnerabilities and minimize the consequences of events (PPD-21). Accompanying that PPD was EO 13636, *Improving Critical Infrastructure Cybersecurity*. In EO 13636, a major threat to critical infrastructure was outlined, the *cyber threat*. As infrastructures become more and more complex, dependence upon computers to aid in daily operations grows as well. The less vulnerable infrastructures are to this threat, like any other, the more resilient they become.

> *"It is the policy of the United States to enhance the security and resilience of the Nation's critical infrastructure and to maintain a cyber environment ... while promoting safety, security, business confidentiality, privacy, and civil liberties." (EO 13636 Section 1)*

On April 9th, 2013, at the 29th National Space Symposium, General William Shelton, Commander, U.S. Air Force Space Command, stated (Moskowitz p1) :

> *"Shrinking government budgets, combined with a growing reliance on space assets by the United States - especially by its military - are putting the country in an undefended position." During his speech, Gen. Shelton "advocated reaching a sweet spot between 'capability, affordability, and resilience'."*

Despite many misconceptions, space systems are not safe. Accidental collisions with other satellites, debris, or meteors are just a small portion of the threats these space assets face. Every satellite must be able to transmit information to and from a terrestrial station, whether directly or indirectly, if they are to be useful to their operators. This means that anything which might interrupt or distort that transmission is also a threat to the capabilities the space assets provide. This can include radio frequency interference (RFI), cyber attacks, and even weather. Recently, a rise in the potential threat of an attack on space assets from kinetic weapons has occurred, as seen by the rise in the technology related to anti-satellite (ASAT) weaponry (NSTAC p. 12).

Resilience is important because it allows a system to adjust and adapt to the unknown, to survive, and to either regain its former status, or to adapt in a useful way of operating in the altered conditions. As the world changes, the preference for infrastructures plans have shifted from resistant plans which focus on withstanding the effects of known threats, to designs that focus more on building systems that can handle an unknown disruptive event.

> *As we invest in next generation space capabilities and fill gaps in current capabilities, we will include resilience as a key criterion in evaluating alternative architectures.*
>
> *National Security Space Strategy*
> (Fact Sheet: Resilience of Space Capabilities)

Defining Resilience

Over the years, an array of definitions of resilience have been proposed, each geared towards the user's purpose for a specific system and type of disruption. In 2000,

Luthar and Cicchetti defined resiliency as a "positive adaption" and that it was "considered a demonstration of manifested behavior on social competence or success at meeting any particular tasks at a specific life stage" (Luthar p. 110). Clearly, their focus was not on critical infrastructure; much of the earliest work in resilience dealt with psychology.

Collins *English Dictionary* defines resiliency for ecology as "the ability of an ecosystem to return to its original state after being disturbed". In physics, the same term is defined as "the amount of potential energy stored in an elastic material when deformed" (Dictionary.com). In 2010, Hill, *et al*, utilized data from United States history to derive a method for measuring regional economic stability based upon not only employment/unemployment rates, but also a variety of components which include, but are not limited to, laws, reaction to previous shocks, or abrupt reduction of system performance.

In this section, existing definitions are reviewed in order to determine the key components that many current definitions, if not all, deem to be pieces to the resilience puzzle. At the conclusion of this section, the definition is to be used in this thesis is specified.

Regardless of how the definition has varied over fields of study and time, they all center on the same root concept. Merriam-Webster defines resilience as "an ability to recover from or adjust easily to misfortune or change". Resilience originated from a Latin word, *resiliens*, which means "to spring back, or recoil". However, the focus of this thesis is not to universally defining resilience. Rather, what is needed is to consider resilience in the context of a critical infrastructure.

In *Resilience and Stability of Ecological Systems*, C.S. Holling, considered to be the first to provide a system level definition for resilience, defines resilience as follows: (Holling p. 14)

> *Resilience is a measure of the persistence of systems and of their ability to absorb change and disturbance and still maintain the same relationships between populations or state variables.*

In *A Framework for Assessing the Resilience of Infrastructure and Economic Systems*, definitions from multiple other works on resilience are provided; however there is one group that best demonstrates the evolution of the definition of resilience (Vugrin p. 84). In 2006, the Department of Homeland Security (DHS) defined resiliency as "the capability of an asset, system, or network to maintain its function during or to recover from a terrorist attack or other incident" (DHS NIPP p. 104). Two years later, their definition had evolved to "the ability to resist, absorb, recover from or successfully adapt to adversity or a change in conditions" (DHS Risk Lexicon p. 23).

A change in the most recent version of the definition may already be seen as a new strategic goal of resilience is "enhanced preparedness". (Keil) Argonne National Laboratory, a leading research laboratory for Department of Energy, has altered its definition to "the ability of an entity -*e.g.*, asset, organization, community, region- to anticipate, resist, absorb, respond to, adapt to, and recover from a disturbance."(Carlson p. 7) They go on to break resilience down into three main components: "Reduced failure probabilities, reduced consequences from failures…, and reduced time to recovery." (Carlson p. 16) Their belief is that, while many confuse resistance with resilience,

resistance should be considered as a subcomponent of resilience, and thus included in the definition. Doing so allows the definition, as well as the resulting measure, to not only measure how efficiently the system "bounces back" from a disturbance, but also takes into account proactive measures taken to protect the system. Hence the addition of *anticipate* and *resist* to the definition.

Mathematically, a measure, μ, on a set, A, is a function from A to the set of real numbers, $\mathbb{R}$, which possesses, at a minimum, the properties of monotonicity and subadditivity. Monotonicity is a property stating that if E_1 and E_2 are two subsets of A where, $E_1 \subseteq E_2 \subseteq A$, then $\mu(E_1) \leq \mu(E_2)$. The subadditive property states that for a countable sequence of sets, E_i, $\mu\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mu\left(E_i\right)$ (Royden p. 31).

For the purposes of this research, the parent set A is the complete set of nodes in the network, all operating at optimal capacity. Each subset E_i, which is referred to henceforth as *states* of the network, is then a state in which one or more of the nodes in the network have been degraded below optimal operational capability, either partially or fully. Given two sets E_i and E_j, $E_i \subseteq E_j$ if and only if every node in the network at state *i* is at or below the operational capability level of the exact same node in state *j*.

The Department of Defense (DoD) also has their own definition of resilience (Fact Sheet: Resilience of Space Capabilities) .

> *Resilience is the ability of an architecture to support the functions necessary for mission success in spite of hostile action or adverse conditions. An architecture is "more resilient" if it can provide these functions with higher probability, shorter periods of reduced capability, and across a wider range of scenarios, conditions, and threats. Resilience may leverage cross-domain or alternative government, commercial, or international capabilities.*

In this definition, four main components are intended (Fact Sheet: Resilience of Space Capabilities).

> ***Avoidance:*** countermeasures against potential adversaries, proactive and reactive defensive measures taken to diminish the likelihood and consequence of hostile acts or adverse conditions
>
> ***Robustness:*** architectural properties and system of systems design features to enhance survivability and resist functional degradation
>
> ***Reconstitution:*** plans and operations to replenish lost or diminished functions to an acceptable level for a particular mission, operation, or contingency
>
> ***Recovery:*** program execution and space support operations to re-establish full operational capability and capacity for the full range of missions, operations, or contingencies

It should be noted this definition has much in common with the stated Argonne National Laboratory definition. While this indicates a move towards a common definition across infrastructure, a commonly accepted one has yet to be attained. *Avoidance* and *robustness* are directly related to the concept of resistance and their breakdown, "to anticipate, resist, absorb". The definition of *avoidance* given by the DoD Fact Sheet introduce a concept that, while included in the realm of resist, was not formally stated,

and that is the use of countermeasures to diminish the impact of an event. *Reconstitution* is the act of replenishing functions in the near future, or to "respond to, adapt to". Finally, *Recovery* is the easiest to compare as the final piece of the Argonne definition is "and recover from a disturbance."

It is evident that, though no one definition has yet to become the standard for resilience, the more work done with the concept, the closer definitions converge to a similar idea of what resilience must be. While the DoD definition did say what they wanted a resilient architecture or system, it does not directly address the subject of what resilience is. However, the intended components illuminated that piece very well. Those components were imbedded in the reviewed definition from Argonne Laboratory, an expansion of the most recent DHS definition available and one that captures all of their desired attributes of the word. Therefore, this thesis will progress with the following resilience definition (Carlson p. 7):

> *Resilience is the ability of an entity -e.g., asset, organization, community, region- to anticipate, resist, absorb, respond to, adapt to, and recover from a disturbance.*
>
> *Argonne National Laboratory*

Key Components of Resilience

In this section, the key components of the DoD definition and the Argonne National Laboratory definition is decomposed in order to describe to the reader the aspects of a network that must be included while measuring resilience. Recall that from the DoD definition, four main components of resilience exist: Avoidance, Robustness, Reconstitution, and Recovery. In this section, a review the necessary nature of resilience measures in order to capture these components is conducted.

Avoidance consists of the steps taken to lessen the likelihood and consequences of an event. Each step taken to reduce the impact of an event generally increases the cost in some other aspect of financial or operational burden. These expenditures need not be on material objects. For example, funds spent to increase friendly relations with an adversary will lessen the likelihood of a negative event occurring from that adversary.

Unfortunately, a network operator cannot and will not always know who the network's potential foes are or may be over the life of a system. Another option is material countermeasures. These could come in the form of missile interceptors, eliminating the threat at its location before the interdiction can be completed, employing additional cyber security assets to an imminent threat, or any other actions which may be taken to deter a diminishing of network capabilities.

Interdiction is defined by the DoD as: (Joint Publication 3-3 p. vii)

1. *An action to divert, disrupt, delay, or destroy the enemy's military surface capability before it can be used effectively against friendly forces, or to otherwise achieve objectives.*

2. *In support of law enforcement, activities conducted to divert, disrupt, delay, intercept, board, detain, or destroy, as appropriate, vessels, vehicles, aircraft, people, and cargo.*

The purpose of interdiction operations is to prevent the adversary from using assets at the time and place of his choosing.

Robustness is the ability of the design to survive and resist degradation. For systems of systems, this can be effectively broken into two distinct pieces, the larger system and then sub-systems. To avoid confusion, in this thesis these levels are referred to as the network and the node respectively, however this process of viewing hierarchical levels separately can be applied to any level of the system with varying levels of difficulty.

In a node, the robustness may be increased by providing more effective inherent components. These could be things such as the antivirus software installed, on-board redundancies for critical components, shielding from radiation, or increased fuel for maneuvering out of danger and readjusting configuration after an event.

In a satellite network, the robustness may be increased by the sheer quantity of ground stations and satellites incorporated as well as on-orbit spares. Moreover, the robustness of the network relies upon the robustness of each satellite in said network. For example, if each satellite in a network is shielded against radiation and electromagnetic pulses, then the network is robust against such anomalies. Regardless of the level at which these protections are implemented, as well as the amount of protection offered by each feature, each comes with its own requirement of resources, both initial and enduring.

Reconstitution is determined by the set of plans and operations needed to restore the network to an acceptable level of operation. What is considered acceptable may be

dependent upon the time and nature of the event. From a DoD perspective, it is first essential to regain sufficient capacity in the network to proceed with a mission or critical function. However, after an initial phase, the network must also be able to be reconstituted to acceptably support its original function. As before, reconstitution of assets requires time, money, and potential operational tradeoffs, but the level of reconstitution needed is time dependent.

Purpose of the Study

The purpose of this research is to provide a measure for network level resilience of Satellite Communications for use as a key criterion in evaluating the implementation of alternative satellite architectures. This measure is divided into multiple distinct time periods, and includes the probability of maintaining mission essential functions, as well as the costs and times associated with maintaining and reconstituting the system, for an assumed level of adversity and a set of individually evaluated intelligent adversary objectives.

In the following chapter, previous work done on measuring resilience in critical infrastructures is reviewed. Particular attention is granted to their measures, how those measures were derived, and how the authors calculated those measures. Those measures and infrastructures will then be related to a satellite communications network to ensure the validity of any translation of methodologies.

In Chapter III a bi-level model is formulated as the primary method for measuring resilience. A heuristic for solving the resulting problem is developed, as well as the method for analyzing the output to achieve the resilience measure.

In Chapter IV a case study is included, which utilizes variations on a satellite network to demonstrate how changes to a network can affect resilience. The results of the case study are analyzed, as well as how the model, measure, and heuristic behaved in regards to the variations. Furthermore, the shortcomings of the method and potential restrictions on its use are investigated.

Chapter V provides a summary of the findings of this research, both respect to developing a resilience measure, as well as the methodology utilized to calculate that measure.

II. Literature Review

This chapter is a review of relevant literature to include existing methods for measuring resilience, as well as the comparison of a satellite network to other critical infrastructures with existing resilience measures. It compares the methods previously used for strengths, weaknesses, and adaptability. The chapter concludes with establishing a baseline minimum measurement of resilience for a notional network.

Threats and Vulnerabilities

A key paper dealing with threats and vulnerabilities is the National Security Telecommunications Advisory Committee's (NSTAC) *Report to the President on Commercial Satellite Communications Mission Assurance* (November 2009) (NSTAC 2009).

In 2005, a Multiple Path Beyond Line of Sight Communications (MUBLCOM), digital communications satellite from the Defense Advanced Research Projects Agency (DARPA), collided with Demonstration for Autonomous Rendezvous Technology (DART), a NASA sponsored project intended to demonstrate automated navigation abilities. In 2009, Iridium 33, a former satellite in the Iridium constellation, collided with

Cosmos 2251, a retired Russian satellite, obliterating both. In January of 2013, a piece of debris from a Chinese missile test collided with a Russian satellite, rendering it unusable.

Accidental collisions are not the only threat. "On January 11, 2007, China conducted its first successful test of an anti-satellite (ASAT) missile to purposely destroy the aging Fengyun-1C meteorological satellite that had been in Low Earth Orbit (LEO) since May 10, 1999." (NSTAC 2009 p. 12) Shortly after in 2008, the US Navy utilized a modified Standard Missile-3 to shoot down a malfunctioning National Reconnaissance Office satellite. The purpose of this US mission was to protect civilians from potentially toxic fuel (Galdorisi p. 1).

In NSTAC's *Report to the President on Commercial Satellite Communications Mission Assurance*, three main categories of threats are outlined (NSTAC 2009 p. 11).

> ***Physical Threats****: Destruction of physical network infrastructure, or physical threats to operational personnel. Examples include explosions, cable cuts, hostage-taking at control centers, natural disasters, power failures, satellite collisions, and space-based attacks.*
>
> ***Access and Control Threats****: Unauthorized access, control, or prevention of the operator's control of its network, underlying devices, control links, and physical plants. Examples include unauthorized commanding of or preventing control of routers, switches, servers, databases, or satellite buses used to control the network; distributed denial of service attacks against network control infrastructure; compromise of network security protocols; and actions by malicious insiders.*
>
> ***User Segment Threats****: Events, such as denial of service attacks, that occur on user traffic paths of the network that degrade or deny service to users by exhausting or preventing customer access to network resources. Examples include botnets, denial of service attacks, route hijacking, viruses, worms, and RFI.*

NSTAC determined that SATCOM networks "often contain the same subsystems as their terrestrial counterparts that are vulnerable to malicious and inadvertent disruption." (NSTAC 2009 p. 11) As a result, *any* vulnerability that is determined as existing in a terrestrial station also likely exists in the corresponding SATCOM. This permits the relation of common ground threats to both the terrestrial stations and the SATCOM network.

Since it is assumed to be easier to disrupt a ground station's functions, NSTAC determined that the ground stations are much more likely to experience an attack. However, these satellite stations generally are not necessarily at any greater risk than those of other communication providers (NSTAC 2009 p. 11). This is due in part to the large number of redundancies that the commercial operators employ.

While terrestrial stations remain unprotected against a variety of intelligent strikes, they often have many geographically diverse redundant stations from which to operate their satellites (NSTAC 2009 p. "ES-2"). Furthermore, each ground station "generally maintains 24-hour guarded access, security fencing, external lighting, registration and clearance of visitors, and security cameras" (NSTAC 2009 p. 17) In addition, 100% of those who participated in NSTAC's questionnaire reported that all of their ground sites were connected by "multiple communication links that provide redundancy and physical path diversity" (NSTAC 2009 p. 17).

In the cyber realm, most satellite operators were in compliance with the National Security Agency's (NSA) approved encryptions for transmissions, and more continue to meet compliance as new satellites are placed in orbit. Along with the encryptions, many

satellite operators utilize “deaf satellites”, which require very large transmission antennas in order for the satellite to receive commands. They also utilize out of band commanding to reduce the risk of shared frequencies. Carrier lockup is a protocol employed to prevent insertion of commands from the redundant ground stations while the primary is still operating, uniqueness in command decoders, autonomy in case of interference, and diversity to allow for redundant telemetry streams (NSTAC 2009 p. 18).

Up to this point, threats that are common for both satellites and ground stations have been reviewed; however there are a significant number of “vulnerabilities and threats of special concern to satellite systems” (NSTAC 2009 p. 18). The greatest of these lies in the physical realm, where there exists the probability of collisions with debris and meteorites. Though the probability of an accidental collision is extremely low, such collisions do occur.

Satellites that are physically damaged can usually not be repaired. While there is no publicly documented precedent to date of a malicious physical attack on another nation's satellites, a more intelligent threat may come in the form of recent advancements in anti-satellite technologies, such as missiles, cyber warfare, and jamming.

China's ASAT launch in 2007 demonstrated ASAT capability. It also created a cloud of debris that expanded to twenty times its original size in under a month. Two years later, the U.S. Space Surveillance Network catalogued 2,378 pieces of debris exceeding 5cm that are a result of the ASAT test, as well as an estimated 150,000 pieces of smaller debris. This one strike resulted in what is now “over 25 percent of all debris in the LEO regime” (NSTAC 2009 p. 13). Figure 1 shows the sharp rise in tracked satellites, much of which is debris.

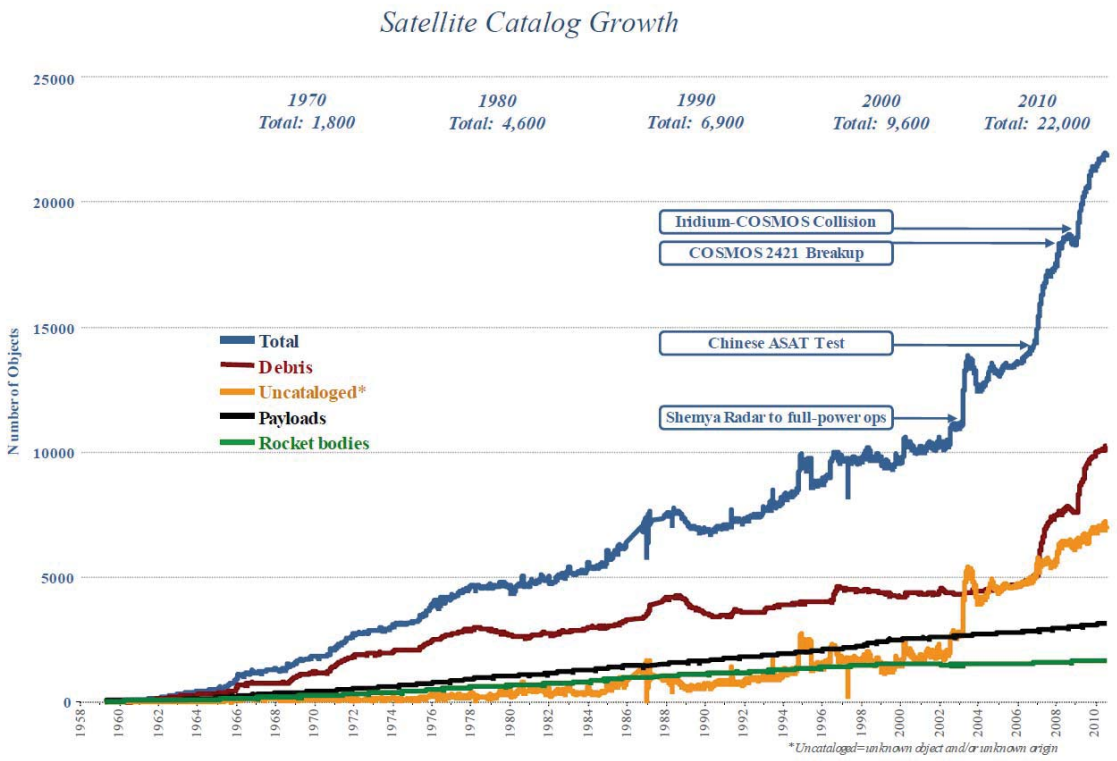


Figure 1: Graph of Tracked Satellites (DoD p. 1)

Another unique threat is that of interference. RFI is an increasing issue for SATCOM. The majority of interference is unintentional, and can be caused by human error, equipment failure, interference from an adjacent satellite or the ground, or even from solar radiation. However, each of these causes can be intentional as well. One of the more common methods of causing interference is to employ jammers.

Jamming and spoofing, the transmitting of a false signal while the original is hijacked, are relatively inexpensive methods for denying the capabilities of a satellite. However, the more protected a satellite becomes, such as by concentrating "its power in a small frequency band" (Wright p. 120), the more power a jammer requires to be effective, thus restricting its mobility. Moreover, a jammer is seen as a legitimate target

during military actions, and the wider an area the jammer attempts to cover, the easier it is to identify (Wright p. 120). As such, the amount of degradation provided by a jammer is inversely proportional to the time that the jammer would be considered effective.

GEO satellites are, however, a slight exception to the last observation. Because the distance from one footprint point on Earth is nearly identical to another, and since GEO satellites maintain the same footprint, it is possible to jam communications from an entirely different nation for an extended period of time. This occurred in 2003 when the United States Telstar 12 satellite was jammed by an installation in Cuba, and again in the same year when China's Shenzhou V was jammed via transmitters in Taiwan. As such, jammers are considered to be more effective against GEO satellites than LEO (Wright p. 122).

Another method for jamming is space-based, via the use of satellites with onboard jamming technology. Because the satellites can remain relatively close to the satellite being jammed, they require significantly less power to be effective (Wright p. 123). The jamming satellite would need to be at an altitude lower than the satellite being jammed. Satellites must move very quickly to maintain orbit; the closer a satellite is to Earth, the faster it must move. While there is a proportional difference in effectiveness based upon the terrestrial area covered, the time in which a jamming satellite is within range of its target is roughly equivalent so long as the distance between them is the same.

> *"A jammer in orbit 1 km below a satellite whose antenna was designed to view the entire section of the Earth below it would cross the broadcast receive area in 2 to 3 hours, whether the satellite was in low earth orbit or geosynchronous orbit."*
>
> (Wright p. 123)

While many commercial operators have shielding against unintentional or extremely low level jamming, few have protection against high powered intentional RFI. Even in the military, not all of the satellites are protected against this type of attack (NSTAC p. 20).

A more recent method employed for denying the use of a satellite is through the use of a high powered microwave (HPM) (Wright p. 130). HPM utilizes a high-powered microwave burst to overwhelm a target satellite's components. This method utilizes the same theory as a Nuclear Burst in that the high intensity radiation is what overwhelms the components. However, an HPM is specifically aimed to the target's receiving frequency, reducing the collateral damage and potentially requiring less energy.

An HPM's energy requirements and effectiveness are proportional to the distance from the weapon to the target. As such, the ideal methods for deployment are either "those based in space or popped up using a suborbital missile" (Wright p. 131). Depending upon the strength of the pulse, as well as the protection of the satellite, an HPM could be classified as either a temporary effect or permanent destruction.

HPM's actual usefulness remains uncertain though as the technology, both for explosive power generation and transmission, is still progressing. "Electronics can be hardened against microwave attacks of moderate levels without a great cost…a hardened satellite can withstand orders of magnitude higher HPM flux than an unhardened satellite" (Wright p. 133).

High powered lasers also pose a growing threat for LEO satellites. While the components and energy required to construct a laser of sufficient power, as well as treaty

agreements, have precluded their research in space-space tests, ground-based megawatt class lasers could damage unshielded robust satellite components in a matter of seconds. Kilowatt class lasers are also able to inflict damage, however, only on unshielded fragile parts over a longer period of time (Wright p. 134). While these terrestrial based lasers pose a threat for LEO satellites, GEO satellites' great distance have protected them thus far as lasers suffer exponential heat loss. Using lasers for temporary denial of service has thus far been restricted to dazzling hyper spectral imagery components (Wright p. 126).

At this time, reasonable methods of partial degradation to communication satellites over a period of greater than a few hours appear to be restricted to either jamming from a neighboring nation or otherwise restricted to imagery components. Therefore, it is reasonable to assume that any goal of medium-long term effects is of the destructive type, which, with the exception of terrestrial nodes with multiple components spaced sufficiently far apart to not be simultaneously affected by a single destructive effort, will have a discrete kill-no kill effect. This assumption only remains valid in the current state of satellite degradation technologies, and only for communication satellites. For example, imaging satellites may already experience a permanent, partial degradation when portions of their imaging devices are damaged.

In this section, a number of threats that currently exist to terrestrial and satellite nodes in a space based communication network were reviewed. The threats included both intelligent and unintelligent occurrences, which fell into a category of either temporary denial or permanent destruction. Based upon the brief literature review, one may conclude that, while many methods exist for temporarily denying capabilities for a

particular node, a medium-long term goal would currently be ill-served with most methods short of destruction.

Key Motivations for Building Resiliently

The need for a more robust, resistant, and adaptive infrastructure has quickly become clear to decision makers, both within the military and in the civilian sector. Moreover, as outlined in the preceding Threats and Vulnerabilities section, the potential for a degrading event, both intentional and accidental, are already present. What remains is to determine which factors are most important to stakeholders in regards to the development of a more resilient infrastructure.

In this section the motivations that a stakeholder may have for constructing a more resilient communications infrastructure are discussed, as well as what rewards would most likely motivate a push for resilience. A key paper in this section by Jennings, Vugrin, and Belasich focuses on the construction of resilient buildings, but many of the key points can be related to a generic network (Jennings p. 1).

In 2012, Jennings *et al.* set out to answer three main questions (Jennings p. 4):

1) Are stakeholders aware of what resilience is and what it means?
2) What would motivate stakeholders to construct more resilient structures?
3) "Do the stakeholders have any opinions about key program features that need to be included or developed?"

To answer these questions, Jennings *et al.* developed an interview questionnaire, and conducted interviews with subject matter experts in a variety of building stability related categories, to include construction, owners, insurance, and certification programs (Jennings p. 6). Because of the qualitative nature of the questionnaire, highly trained individuals conducted the interviews, a single team member rated all of the interviews and answers to ensure consistency, and additional team members rated interviews to ensure unbiased (Jennings p. 9). After 15 interviews, their study concluded with 7 major findings, which are laid out in Table 1.

These findings were followed up with what the interviewed experts considered the greatest motivations for building with resilience included in the design process. Overwhelmingly, the greatest motivators were those directly related to cost, whether through decreased recovery and down time or through decreased insurance rates. Those indirectly related to cost, such as being able to bring in more tenants or users and charge more because of the "attractiveness" of the building, were split almost evenly. Finally, the lowest ranked motivator was "It is 'the right thing to do'", with more experts admitting that it provided no incentive (Jennings p. 15).

Table 1: Stakeholder Interview Findings (Jennings p. 11)

Interview Findings
1. There is either a lack of consensus on the definition of the term 'resilience' or unfamiliarity with the term.
2. A successful program will require an education and training component.
3. Participation in resilience depends directly on costs versus benefits and a demonstrated return on investment.
4. Optimally, resilience should be initiated in the design process and considered throughout the entire building lifecycle.
5. Definitions of resilience tend to vary by industry. For example, the insurance sector strongly ties resilience to 'risk' and 'risk management,' while planners expressed the need for 'recovery' and 'continuity of operations' after an event or disaster.
6. Building resilience extends beyond maintaining the building envelope and includes dependence upon the infrastructure required to operate the business conducted in the building.
7. Interviewees mentioned that a public-private sector partnership model is important to the success of the program.

Jennings *et al.* proceed with recommendations for what should be included in a resilience certification program, including financial incentives, education, historical data, and partnerships between public and private sectors. The most notable recommendation was to "develop a cohesive resilience story across the Federal government's multiple resilience efforts" (Jennings p. 16). By creating a single definition and measure, or at the very least a measurement process, across the entire government, separate components that factor into the resilience of a building, such as the resilience of the infrastructures they use, may be considered in calculations (Jennings p. 11).

Defining Costs Associated With Satellite Networks

In this section, some of the common costs associated with building, maintaining, and operating a satellite network are defined. When determining costs of a satellite, it is possible to categorize costs as either deployment or on-orbit costs (Eremenko p. 2). Common deployment costs are Launch and Non-Recurring Engineering (NRE) costs. Common Operations and Maintenance (OM) costs are Operations and Recurring Engineering (RE) costs (Eremenko p. 15).

To break these down further, an initial inspection of the costs associated with deployment must be completed. Some main costs in deployment are simply the cost to launch the satellite, or group of satellites, into orbit, and the cost of the satellite (Meckling p. 2). However, some background costs which must be completed initially are the development costs and the investment costs (Meckling p. 4). Investment costs are considered because the different manners by which a component is paid for, such as upfront or regular payments, affect the present value cost of the component.

When determining the cost of a ground terminal, a similar division can be used, namely the cost of constructing the station and the cost of operating and maintaining the station. Common construction costs are, of course, the facility itself, along with any associated land, the equipment required to operate, such as antennas, satellites, and computers, the instillation of all equipment, and any initial fees associated with the required utilities and infrastructure of the facility (Meckling p. 11).

The OM costs of a terrestrial station are the same as those seen in most used facilities: "Facility Maintenance, Equipment Maintenance and Parts, Pay and Allowances, Services (Utilities) and Miscellaneous" (Meckling p. 11).

The OM costs associated with a satellite, with the preceding cost allocations, is then diminished. Because the personnel and terrestrial equipment required to operate a satellite is considered in the costs associated with the terrestrial station, what remains is the cost of refueling, and repairing an on-orbit component. Previously, doing either was considered impossible, however recent advancements have made both a reality.

Since 2011, efforts have been placed on refueling and maintaining aging satellites with the aid of high precision robots. As of 2013, NASA and the Canadian Space Agency (CSA) have a prototype robot which has successfully refueled satellites, performed basic fitting, cap, and screw removals, as well as the manipulation of the thermal blanket (SSCO RRM). With the next set of instruments set for delivery to the International Space Station (ISS) by early 2014, the next phase of the Robotic Refueling Mission is to attempt replacement and internal repairs of on-orbit satellite components (SSCO RRM Phase II).

These innovations provide the answers to what the OM costs of a satellite will be: Component delivery, Maintenance, and Refueling. This does imply that the Robotic Refueling Mission (RRM) will spawn a quantity of high precision space maintainers or, at the very least, a consideration for their use when the time arises.

Based upon the NASA and CSA work, the RRM is designed to remain at the ISS until needed, and then return when its mission is complete. This means that whenever maintenance is required for a satellite, the associated OM costs are the cost of the fuel,

components, fuel required by the refueler, and the cost of delivering those item to the ISS. However, this technology is extremely new, and as such many of its potential limitations, such as which orbits it may operate within, remain unknown.

Comparing Satellite Networks to Supply Networks

In any supply network, there exist three main components:

1. Source-where the supply is coming from,
2. Sink-where the demand exists
3. Relays/Transshipment- the points bridging the sources and the sinks.

A source is the point of origination for a supply network. This is a node where the product is stored and/or produced. In a satellite network, products are transmissions which originate from the users of the network.

A sink is a point where the supply chain terminates. The set comprising the sinks may be larger or smaller than the set comprising the sources, but it still follows the general idea of "Where is this product needed?". For a satellite network, a sink is the transmission's destination.

The relay points are all those points that must be traversed in order to transport goods from sources to sinks. Common relay points in supply networks are depots, sea ports, airports, and train stations. Each of these relay points may also offer their own supply as well, or inversely may have a demand of their own which must be filled. For a satellite network, a relay can be anything such as a transmission tower, satellite, gateway,

or even a switchboard. While it is unlikely that a transmission tower is making any phone calls, it may send occasional data bursts to an offsite location detailing its status and use. On the other hand, it is very likely that a terrestrial based gateway station is making and receiving many transmissions as they are generally operated by humans.

A supply chain also has the possibility of degradation of product in route, more commonly known as losses. For many products, this degradation is simply products that were lost in transit, however for items with a short life such as produce, the degradation must also include the products that arrived, but were unusable. This is necessary for a satellite network because every transmitted signal is another potential place for the signal to become further degraded. If too many relays are used, or if the relay used must traverse areas of high interference, the signal may be unusable when it reaches its destination.

As closely related as these two networks are though, there is still a main difference that must be addressed. Satellite networks and supply networks differ in that every source node for a transmission in a satellite network is potentially a sink node as well. Furthermore, while most supply networks have relatively static node locations, satellites, excluding GEO, are constantly changing their relative position with respect to Earth. This means that a transmission from point A to point B may traverse a set of nodes S at time t, but then traverse an entirely different set of nodes R at time $t+1$.

In many communications networks, there are also locations known as "Gateways" (Werner p. 371). In order to determine a user's ability to utilize a network, each transmission device sends out a code when attempting to initiate a connection. Before a transmission may be completed, it must be relayed through the network to a gateway,

where the code may be verified. The gateway then relays the transmission through the remainder of the network to the destination. When attempting to complete a transmission to a user of the network in question, the transmission still must be routed through a gateway to ensure that the receiver has authorization for network use (Werner p. 373). Hence the name “Gateway”; in order to utilize the network, whether as a sender, receiver, or both, every transmission must be routed through a gateway first.

Comparing Satellite Networks to Power Grids

In this section, relations between satellite networks and power grids are established by comparing each of the components that make up the networks. This comparison is necessary for the validity of comparing work done on resilience in power grids, which is seen in the section, Previous Resilience Measurement Methods Used, to the satellite network.

Merriam Webster dictionary defines a power grid as "a network of electrical transmission lines connecting a multiplicity of generating stations to loads over a wide area." Meanwhile, ATIS Telecoms defines a satellite network as "a satellite system, or part of a satellite system, and the cooperating Earth stations" (ATIS).

In *Defending Critical Infrastructure,* the power grid consisted of buses, power lines, transformers, and generating units (Brown 2006). The buses are able to distribute the power down a variety of paths which were defined by the power lines. The power lines were able to transmit electricity in either direction. The transformers adapted the

electricity for travelling long distances, and then reverted the electricity to its common form for every day use. The generating units, or power plants, are where the power originates.

Currently, a direct translation exists for most of these components into a satellite communications network. The majority of these translations come from work done by The President's National Security Telecommunications Advisory Committee (NSTAC). See Figure 2 and Figure 3 for a visual comparison. The power buses are loosely translate to the satellites, capable of sending and receiving a data stream to another connected bus. The power lines in a satellite communication network are not physical wires, but rather the transmissions themselves, radio frequencies being the most common, which are sent and received via uplinks and downlinks. As such, the satellite "power lines" are determined by line-of-sight and distance. The transformers of a satellite network are the assets required to encode/decode the data for transmitting, much like a power transformer converts the energy for long distance travel or how cargo is palletized when being shipped via air.

Finally, the generating units are those nodes from which the transmissions originate. Note that in this case, these generators are the people or system creating the transmission and the transformers are the devices used to transmit the message. In the case of a phone call, the generator would be the human speaking and the transformer the phone. However, for aggregation purposes, these two components are often considered as one in the same for the purpose of modeling, making exceptions only when a gateway is necessary for switching between networks operational encodings.

There are many close translations between a satellite network, a supply network, and a power grid. However, one main difference remains, and that becomes apparent when operating a network handling multiple transmissions instead of only one.

A power grid has one main product for the purposes of these studies, and that is electricity. Like many networks, if the power plant requires any power itself, it supplies itself, thus eliminating its demand and reducing its supply. Likewise, sink nodes such as homes, factories, or office buildings are supplied by these sources. If they produce any power of their own through the use of technology, such as solar panels, windmills, or watermills, that amount is deducted from their demand before additional power is purchased or excess power is sold.

The source nodes and sink nodes in the model make no distinction as to where the power is shipped so long as all demand is met for the lowest possible cost. Another way of thinking about this is that the model does not care where the power came from or where it is going, only the path utilized to meet its objective.

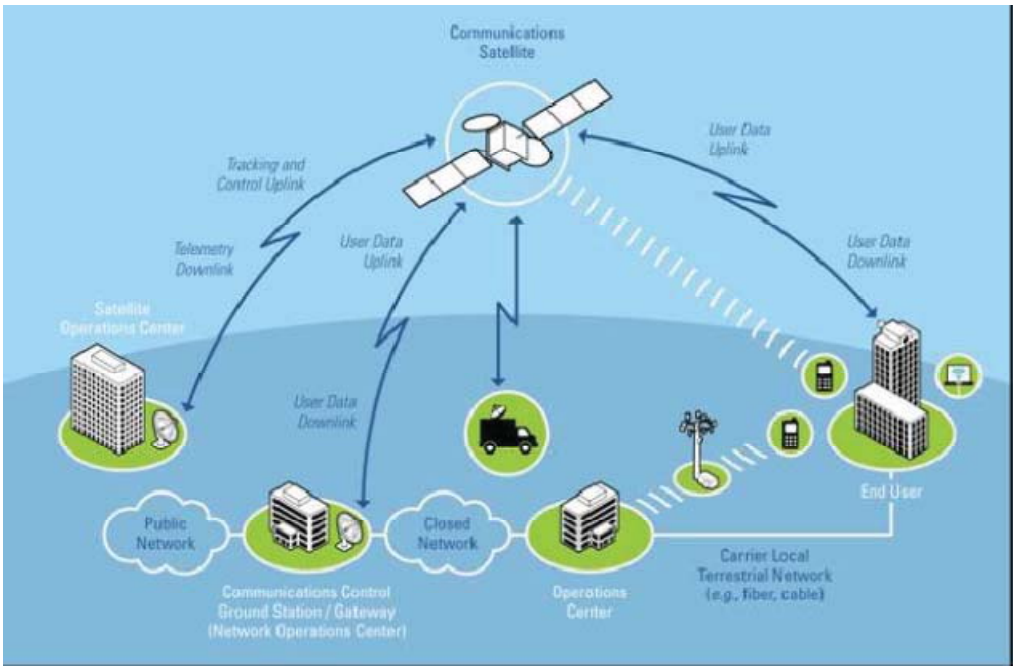


Figure 2: Hybrid Satellite/ Terrestrial Communications Network (NSTAC p. 5)

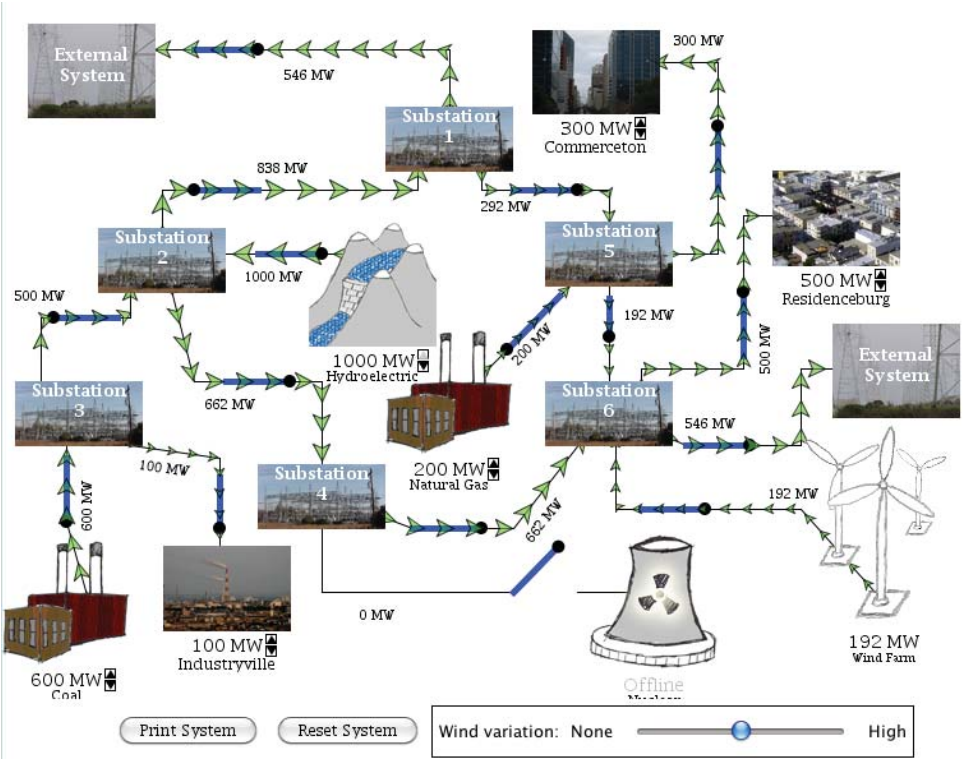


Figure 3: Example Power Grid (CIP Vigilance p. 1)

A satellite network follows this same principle, but only when viewed one transmission at a time, or when a single transmission is being broadcast to a large population of sinks. Whereas a power plant may not care if the power it sends goes ten feet or ten miles, a businessman making a phone call to Texas may be expectedly unhappy if he is connected to a take-out restaurant in Maine. It is because of the need for high-fidelity in transmission source and destination that any adaptation from a clear source-sink model needs to be modified.

Another difference between power grids and satellite networks is the way in which they make transmissions. When viewing a single instance in time, the differences in the methods of transmission are insignificant. However, when viewing the two in a time-dependent fashion, it is clear that while power lines remain fixed, the connections in satellite networks are dynamic.

For Motorola's Iridium satellite network, the average in-view time for a satellite is 10 minutes (Pratt p. 1). This means that every 10 minutes, a ground location essentially breaks one connection and forms another. Furthermore, while each satellite in the Iridium network maintains a connection with the adjacent satellites in its orbital plane, connections are also established to satellites in adjacent co-rotating planes. However, these inter-plane connections do not occur when the geodesic location is in excess of 60^{o} latitude (Pratt p. 8). While this may appear as if the satellite links are dependent upon location as well, the location of each satellite can be estimated using its orbital parameters, starting location, both of which are relatively unchanging, and time. Thus, the connections are time dependent.

Therefore, when viewing a snapshot of single transmission in a satellite network, there is little difference between this network and a power grid. However, every transmission is essentially an additional commodity to be sent through the network, and every new point in time a new network to be solved. Any adaptations of work made must be able to accommodate these differences.

Previous Resilience Measurement Methods Used

In this section, previous works on measuring resilience are reviewed, with an emphasis in the area of critical infrastructure defense and network defense. Major case studies used in previous methods show a focus towards the fields of disaster planning, supply networks, and power grid distributions.

One measure presented as adaptable to measuring any type of infrastructure comes from Argonne National Lab's Decision and Information Systems division, whose definition of resilience is initially presented in Chapter I. Carlson *et al.* utilize an index, aptly named the *Resilience Index*, which may be used to determine the "most important lower level systems" (Carlson p. 20). In this index, four components are combined: Preparedness, Mitigation measures, Response capabilities, and Recovery Mechanisms (Carlson p. 21).

These components were later changed to Robustness, Resourcefulness, and Recovery. The output from this analysis results in numerical values, scaled 0-100 (see Figure 4) showing the index value of each component and the overall Resilience Index.

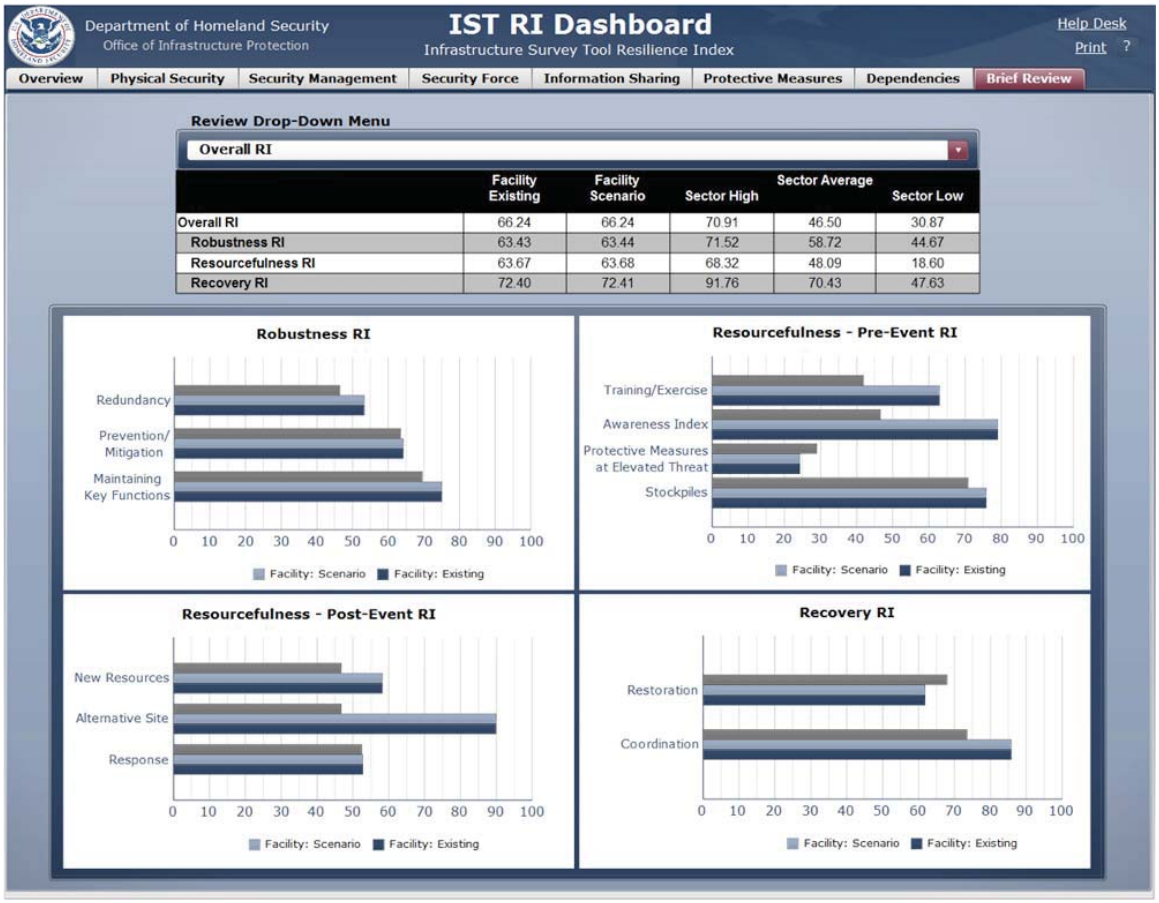


Figure 4: Resilience Index Dashboard (Argonne)

Like the index itself, each of these three components are measured via indices. These three indices are the weighted sum of multiple subcategories, which are to be calculated using the weighted sum of a series of survey questions aimed at analyzing aspects of the individual sites when taken together form a system, as well as the system as a whole (Carlson p. 39). Each question itself is attributed a rank, 1 through 5, and a corresponding weight as determined by the average weight assigned by subject matter experts.

While the resilience index constructed by Carlson *et al.* does have the advantage of being relatively easy to use, it does suffer from the potential uncertainty that comes

with soliciting weights and with survey-based metrics. This effect, minor in itself, is compounded multiple times over, both within the four main components of the index, and then across the indices to construct the overarching resilience index.

Furthermore, this resilience index, while highlighting the area of concern with follow-up analysis, provides relatively little information except for use as a comparison between systems performing identical duties. Even with similar system duties, such as power distribution, weights can vary between providers in two different locations, resulting in identical systems having very different scores. As such, this index may only be compared between networks of the same type operating in an extremely similar, or the same, environment.

In addition, Argonne National Laboratory's resilience index fails to consider the extreme events, which can stress a system in unseen ways, instead focusing on capturing the common or frequent disruptions which a network may face. Where as an analytical measure may be capable of determining combinations of disruption which, together, have a much greater impact, and index of this type would be unable to capture the impact of more than a singular disruption.

Cimellaro *et al.* define resilience as "a function indicating the capability to sustain a level of functionality or performance … over a period defined as the control time" (Cimellaro p. 3640). From this definition, the reader can already see the coming implications of resilience as a time-dependent measure of performance. This is confirmed when they later mathematically define resilience as the area under the time-dependent functionality curve.

To do this, they consider the functionality curve as a piecewise function incorporating a time of operation before an event called the control time, the time after the event until function levels have stabilized, and from that point until the end of the graphed span of time. An example of this function is shown in Figure 5.

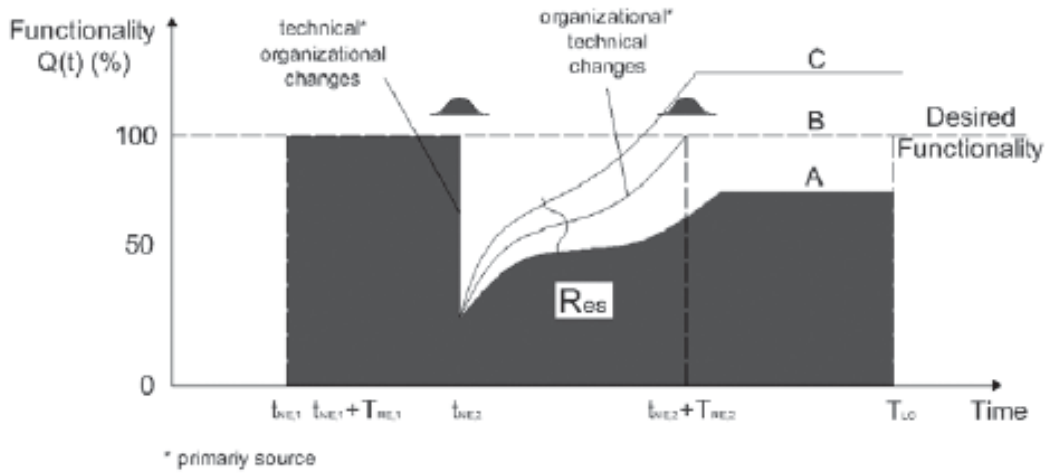


Figure 5: Functionality Curve (Cimellaro p. 3642)

Cimellaro *et al.* go on to describe how the recovery section of this curve can describe the *preparedness* of a community. With a fixed time for recovery, a curve with more area is more prepared, and as a result of their resilience function more resilient (Cimellaro p. 3644). Their resilience measure is presented as a percentage of desired functionality provided over the span of time in regards to a known or predicted event. In their case study, they utilize an earthquake affecting a hospital (Cimellaro p. 3646), as well as an equal time span to compare the resilience provided by multiple construction options. By doing so, the question being answered is "Which option makes me most resilient to event X?".

Zobel and Khansa, employ a similar method in which multiple scenarios are tested against a common event and time span (Zobel p. 83) However, Zobel and Khansa

take into account that multiple degrading events may occur while the system is still recovering. For a visual representation, see Figure 6.

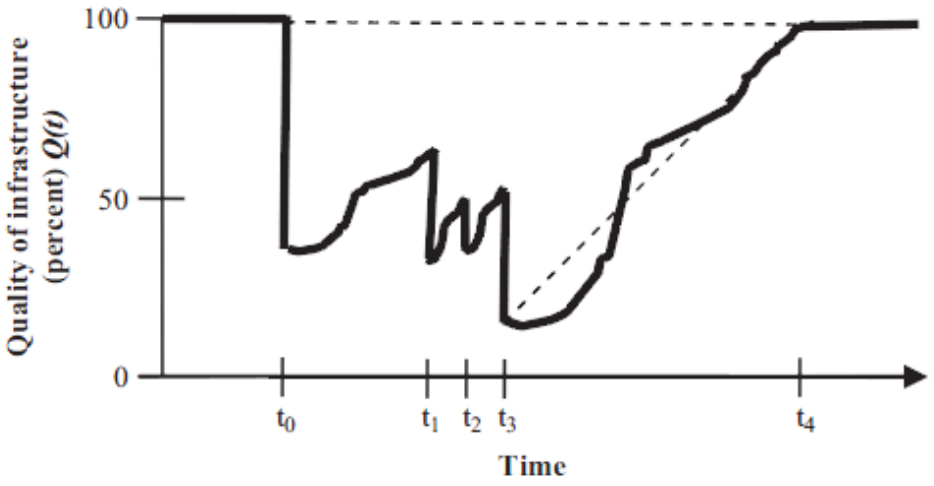


Figure 6: Multi-Event Resilience Graph (Zobel p. 84)

Opposite of Cimellaro *et al*, the measure for resilience is the impact of an event, or how much functionality/quality was lost. In this case, the goal is not to increase the resiliency measure, but to reduce it. Zobel and Khansa also note that very different structures can result in an identical resulting resilience value, specifically pointing out the trade-off between reducing the lost *quality,* as they refer to it, and the time to recover (Zobel p. 87). This two-dimensional tradeoff results in a series of equal-resilience curves, shown in Figure 6.

Zobel and Khansa refer to these variables as effects, dependent upon the robustness and rapidity of a system. Unlike the Cimellaro *et al.* paper, which aimed to determine which safe-guard or defense should be utilized to increase resilience against a specific threat, Zobel and Khansa strive to describe what is making the current system as resilient as it is (Zobel p. 92). As a final note, Zobel and Khansa specify that it is

important to determine what the decision maker considers important, both as criterion for measuring as well as threats for being resilient against (Zobel p. 92).

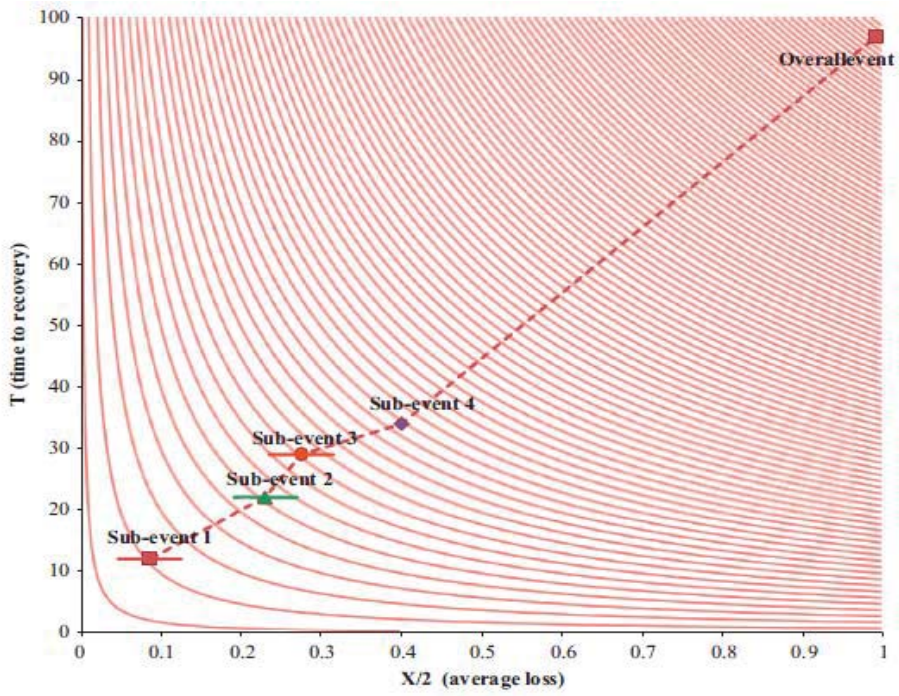


Figure 7: Resilience Curves (Zobel p. 88)

One area that is quickly advancing in resilience models is the area of transportation networks. More specifically, supply networks are a prime candidate for viewing the different methods available for use. Klibi and Martel utilized stochastic methods in order to evaluate the resilience of their network (Klibi p. 1). Their model developed a measure based upon the intensity of a disruptive event, and the time to recover from that event.

Klibi and Martel extended the paradigm by modeling the time to recover as a continuous function of the impact. In addition, they included a term for random error. Inter-arrival times, order sizes, the aforementioned error term, as well as the location and

size of the disruptions are all then associated with probability distributions. Moreover, they included the concept of a temporary demand surge, Figure 7, which follows the recovery of the system (Klibi p. 6). As the reader will note, this is the third time in as many graphs in which a structure utilizing a time-dependent performance measure has been seen as a basis for resilience.

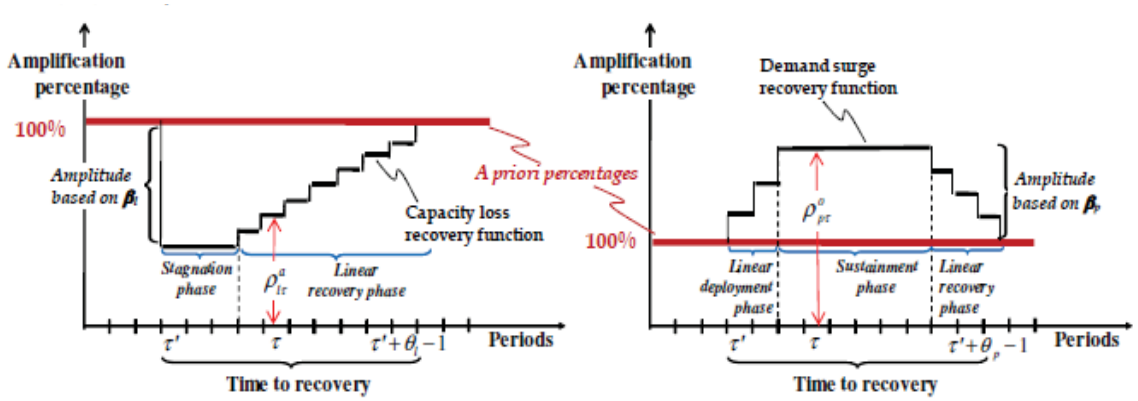


Figure 8: Recovery Function Examples (Klibi p. 6)

Utilizing Monte Carlo simulation, Klibi and Martel use random number draws "first to generate multihazard [disruption] arrivals, second to generate recovery functions, and third to generate daily [demands and capacities]" (Klibi p. 7). The use of simulation was due in part to the scale of the network they were modeling. Were the network very small, a series of distribution convolutions could have provided an exact result. However, with even moderately sized networks, that process is cumbersome and computationally daunting. Through the use of simulation, an approximate final distribution can be attained. In the Klibi and Martel case, an intolerance level was set with the measure being product-days lost. Figure 8 shows a histogram of their results for a large network.

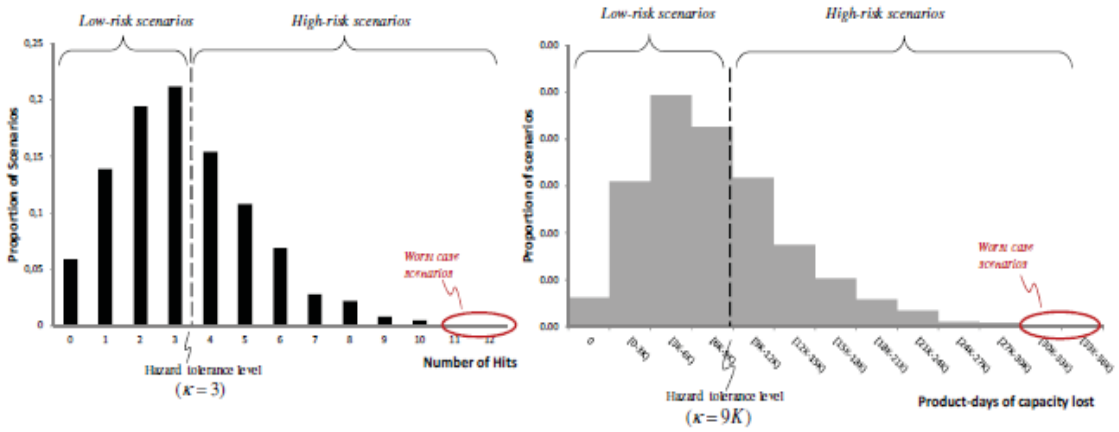


Figure 9: Distribution of Product-days Lost (Klibi p8)

After the event, a separate model following the same constraining guidelines as the model before the event occurred, minus the lost resources of course, determined the optimal manner in which to operate. This is because their networks are built through intelligent design, and are able to be adapted to whatever the situation might be. Though this adaptation may not be sufficient to instantly meet all demand, it does have the effect of lessening the lost network capability during reconstruction.

Klibi and Martel present multiple modeling methods. Another method is their development of a Risk-Neutral Design Model. In this model, the focus is on the expected values of the distributions, and assumes that the modeler has no preference between avoiding the more frequent low-risk scenario nor the less frequent, but more damaging, high-risk scenarios. In this risk-neutral model, much of the probability and intensity portions are lost and the model becomes deterministic at the expected value.

A separate method for measuring resilience conducted at the Naval Post Graduate School by Brown *et al.* and Salmeron *et al.* is discussed next (Salmeron 2004), (Brown 2005), (Brown 2006), (Salmeron 2009). This research line focused primarily on deterministic modeling methods in order to locate a probable worst-case interdiction by

an intelligent adversary on a network, and the actions an operator might take in order to diminish the impact of the event while rebuilding the network. The similarities between the underlying processes of measuring resilience are clear.

The main difference between analyzing a “worst case” model and the "expected values" for an event is that, for a plausible adversary, it puts more emphasis on "how bad *could* the network performance be impacted" instead of "how bad is the network performance *expected* to be impacted" (Salmeron 2004).

Henceforth this type of model is referred to as an *extreme event scenario*. This method of thinking is common with high consequence, low probability events, referred to as the Risk of Extreme Event (Haimes p. 515). One method that is utilized for accounting for these events and evaluating outcomes is to restrict the probability being analyzed to what is commonly referred to as the tail of a distribution (Haimes p. 483).

In line with this action-reaction, extreme event methodology, Brown *et al.* utilized a bi-level optimization model known as Attacker-Defender Models. Attacker-Defender models key in two underlying assumptions. The first is that the defender has an objective they would like to be attained at an optimal level. The second is that the attacker has an objective which conflicts with the defender's objective. In Brown *et al.*, the defender is minimizing the cost of the network and the attacker is attempting to maximize this minimal cost. Note that cost is not necessarily monetary, but could be anything such as resources or time (Brown 2006).

One example Brown *et al.* use is the Strategic Petroleum Reserve Louisiana pipelines, shown in Figure 9. In this complex network, the sources supply the petroleum

to demand nodes (sinks). More often than not, this was done via transfer stations, which had neither demand nor supply.

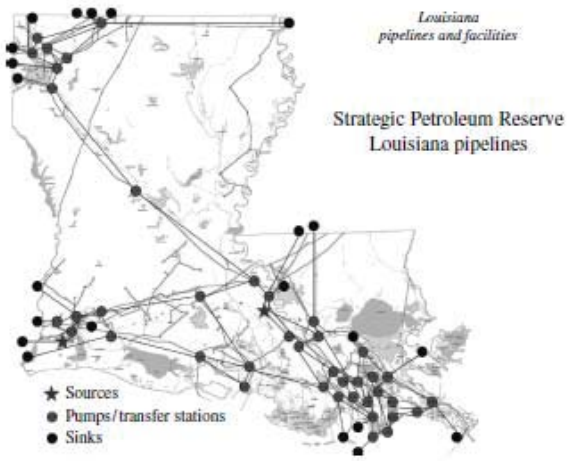


Figure 10: Strategic Petroleum Reserve Louisiana Pipelines (Brown 2006 p. 537)

In this example, they are not looking for the vulnerabilities of the current system specifically, but instead broke their analysis into three distinct pieces, each geared around a separate defense plan. It is those defense plans which were analyzed. Thus, the analyst was able to say "if the system had these protections, how safe are we?" Once again, the measurements of "best" were determined by which plan was least costly, or rather which plan, if optimally interdicted, still cost the least.

Salmeron *et al.* begin by first constructing a simplified version of their network and its operations, and then slowly including more and more details, allowing their model to incrementally expand and increase in accuracy (Salmeron 2009 p. 98). Their first simplification was to assume constant demands and a single repair time in order to eliminate variability. The objective function they utilized was a single minimization problem in which the operator of a network sought the lowest possible generation and load shedding costs for their power grid.

The flow of power and placement of resources resulting in this lowest cost was referred to as the Optimal Power Flow model. This model provides the baseline cost and establishes the values that each relay point experiences based upon the supply and demand of the model. The next step determined the optimal attack. The determined interdiction is the one that maximizes the operating costs of the system.

As part of the constraints of the Salmeron *et al.* model, the nodes where attacks may or may not occur are based upon binary decision variables. Unlike linear programming (LP) models, which are able to arrive relatively quickly at an optimal solution, or alternate optimal solutions, mixed integer programs (MIP) can take significantly longer to solve. Because the latter model is often more complicated than the former, and is intended for an extremely large network, Salmeron *et al.* formulate an improved algorithm for Benders Partitioning, commonly used to solve large integer programs. The improved method is known as Global Benders Decomposition Algorithm (GLBDA).

The main improvement comes with solving bi-level models. Salmeron *et al.* suggest whereas Benders Partitioning is only able to maximize concave functions, GLBD is able to maximize convex functions as well, so long as two requirements are met. The first is that the function in question must be "easy" to evaluate at fixed values. The second is that valid and useful cut coefficients can be defined for the MIP. Salmeron *et al.* go on to validate these requirements for a large-scale power grid (Salmeron 2009 p. 100).

Salmeron *et al.* measure the “difference in operating cost, including penalties for un-served demand” between the initial state of the system and the system after the shock

(Salmeron 2009 p. 98). This methodology permits the allocation or rental of new resources, which combined with the remaining pre-shock grid components, provides the baseline cost to operate the post-shock grid. What remains are the miscellaneous costs associated with clean up, repair, increased security, as well as the penalty costs for demand not met.

As this measure is dissected, it becomes possible to categorize these costs into the components of the definition. Recall that the definition of resilience being utilized is "the ability of an entity -e.g., asset, organization, community, region- to anticipate, resist, absorb, respond to, adapt to, and recover from a disturbance"(Carlson p. vii). In Salmeron *et al.*'s measurement method, there is no explicit accounting for anticipation and resistance. If these components are present, they are absorbed in the operation costs before the disruptive event. There is also no method in their paper for including the acquisition cost for pre-event defensive measures. Response is included in the post-event costs as clean up. Adaptation comes into effect as short term resources or new resources are procured to minimize the event. Finally, recovery can be seen as the repair costs as well as the time to repair.

Salmeron *et al.* introduces the concept of a time persistent resilience model. This permits an adjustment to the model for three possible objectives: Short, medium, and long term outages. Salmeron *et al.* focused on the medium and long term outages, arguing that the effect is “likely to be much greater.” In these methods, the time to repair is considered much more important because the most likely target may no longer be the easy-to-repair bottleneck, but rather a substation that requires a significantly longer period of time before being functional again (Salmeron 2009 p. 99) .

Based upon this approach of including varying time lengths, it can no longer be assumed that an intelligent adversary will utilize all resources in a single coordinated attack. Instead, it is possible that the most optimal solution could be to continuously re-strike a node such as an easy-to-repair bottleneck. For an opponent with multiple strike capabilities, keeping this single node disrupted for the duration of the model time horizon could potentially be more beneficial than disrupting a multitude of nodes elsewhere.

Compare and Contrast Prior Methods

Many of the preceding methods exhibit existed three main similarities:

1) A defense, if any, was established, the attacker strikes, and then the defender recovers.
2) Once set the nodes and paths could not move, but could only be operational or eliminated.
3) There were clear source and sink nodes.

In networks exhibiting specific structures, such as networks consisting of only immobile ground nodes, GEO satellites, and operational requirements precluding the repositioning of GEO satellites, these similarities are expected to hold true. However, in general, a satellite resilience model only follows the first of these three. Unlike terrestrial infrastructure, satellites are able to be repositioned. Each change in location potentially forms an entirely new network, with new arcs being added and former arcs being destroyed. In many situations, such as when the ground components are highly mobile,

the repositioning of a satellite allows it to serve an otherwise under capacitated region, which leads into the violation of the third similarity of clear source/sink nodes.

As was mentioned previously (refer to section Comparing Satellite Networks to Supply Networks), there are clear source and sink nodes in a satellite network only when viewing a single transmission at a time. When viewing the network as a whole, however, each node is a potential source and sink. A simple example of this would be a satellite phone. So long as a satellite is within service range, the phone may transmit or receive a call. Once the call is established, a two-way communication stream is continuously passed back and forth until one side or the other ends the transmission, either intentionally or by leaving a covered area. Any model chosen must be adaptable to these unique operations in which every transmission has a specific source and destination.

One possible way of doing this would be to assume that every end-point of the network be split into two, a transmitter and a receiver. A simulation model could be utilized at this point. Each node may be split into sending and receiving nodes, and each new transmission assigned a random number that would determine its destination. From this simulation, the satellites servicing the most capacity, or acting as bottlenecks, could be identified. Such an approach would require a large increase in the number of nodes.

Unfortunately, while simulation is an excellent tool for providing insight into how the system operates, it can be difficult to use to accomplish large scale optimization, and is even more unwieldy when attempting to make a single model adaptable for multiple architectures or configurations. The desired model should be easily adaptable from the example network to the actual operational system, because its intended purpose is to be used on multiple architecture alternatives to determine which is most resilient.

Another matter are the consideration of *Black Swan* events. A Black Swan is an event which is highly unlikely and causes extreme impact, but an event which humans will investigate retrospectively, "making it explainable and predictable" (Taleb p. xxii). In reviewing the purpose for studying and measuring resilience, the focus falls on maintaining system functionality in the face of unknown and unexpected adversity.

"Black Swan logic makes what you don't know far more relevant than what you do know" (Taleb p. xxiii). As such, seeking predictions from experts while operating in "environments subjected to the Black Swan" provides little information, as if the general population were polled (Taleb p. xxv). Because of the inherent unpredictability of these types of events, a more resilient system will be better able to continue functioning in such an environment.

When considering the strengths that a resilience measurement method should possess, five main aspects can be derived from the methods reviewed:

1. Adaptable: Can it be applied to multiple systems and scenarios with few modifications?

2. Highlights System Vulnerabilities: Will the measure or a bi-product of the method's output alert the analyst to potential network vulnerabilities?

3. Intuitive Method: Similar to adaptability, can the method be easily decomposed into the capabilities/components and recreated?

4. Inclusion of Time: Time is seen as a key criterion in a resilience measure. Can every/most events occur in multiple time periods?

5. Consistent: Is the measurable comparable within networks performing a similar function?

A characterization of the most prominent methodologies, an indexing method, Monte Carlo Simulation, and Worst-Case Interdiction, are displayed in Table 2. These strengths are based upon the methodologies and show which are preferable in regards to the criteria displayed. Cells in Table 2 are marked with an "X" if the method possesses the respective strength.

Table 2: Method Strengths

Strength	Index	Monte Carlo Sim.	Worst-Case
Adaptable	X		X
Highlights System Vulnerabilities	X		X
Intuitive Method		X	X
Inclusion of Time		X	
Consistent		X	X

Based upon the strengths, the Worst-Case Interdiction method used by Salmeron *et al.* appears to be the strongest, though the failure to include the time component is a drawback in measuring resilience. Conversely, the Index method, the most prominent of which comes from Argonne National Labs, shows the least strength. Still, it is acknowledged that this method is the simplest to calculate, with the exception of first requiring input from sufficiently many subject matter experts (SMEs) to arrive at an average weight to be used.

Similarly, there exist a number of shortcomings of each model which may be compared. Each of these shortcomings could hinder the method's ability to achieve a

good resilience measure, whether by under-stating the high impact Black Swan events, discarding the common place disruptions, or in a cacophony of other ways.

Shortcomings 2 and 3, Explicitly Constructed Scenario and Subjective Measure, are contrary to Taleb's assertion that expert predictions provide little usable information. Thus any scenario or weight constructed by an expert may not serve to increase the system's ability to continue operating when faced with an extreme-event.

1. Scenario Dependent: The most common shortcoming, a resilience measure dependent upon scenarios answers the question "Resilient against what?"

2. Explicitly Constructed Scenario: Clearly Scenario Dependent, this shortcoming requires the analyst to construct the adverse events, preventing the model from creating the high impact *perfect storm*.

3. Subjective Measure: Though mitigated with analytical techniques, does the model *require* a subjective estimation of relative value, component durability, or other system parameter?

4. Focuses on Likely Disruptions: The model fails to stress the impact of possible Black Swan events.

5. Discrete Degradation (Kill/No Kill): Are the degradation of system components discrete to the point of Kill or No Kill, or can it exploit synergistic partial degradations?

6. Threat of Compounding Error: Are lower level uncertainties compounded while striving for a network level measure?

7. Deterministic: Is the method incapable (in its current form) of including system reliability or risk?

8. Provides Little Usable Information: Does the method output only allow for resilience comparisons?

Table 3: Method Shortcomings

Shortcoming	Index	Monte Carlo Sim.	Worst-Case
Scenario Dependent	X		X
Explicitly Constructed Scenario	X		
Subjective Measure	X		
Focuses on Likely Disruptions	X	X	
Discrete Degradation (Kill/No Kill)	X	X	X
Threat of Compounding Error	X		
Deterministic (No Reliability/Risk)	X		X
Provides Little Usable Information	X		

Table 3 displays whether each method possesses a shortcoming by marking an "X" in the respective table cell. It can be seen that the Monte Carlo Simulation method utilized by Klibi and Martel has the fewest shortcomings. Of those two shortcomings, the focus on likely disruptions is a direct result of utilizing probability distributions and random number draws. However the cause of that shortcoming is also the same reason for the methods strengths elsewhere. In addition, this is the only method which does not show a dependence upon scenarios. While Klibi and Martel do utilize scenarios to achieve their measure, they utilize a great many number and variations, making the final resilience measure almost independent of the situation.

Once again, the index method falls to the rear, possessing every short-coming listed. This method is unique in that it was the only one of the three to utilize explicitly constructed adverse scenarios, be calculated using subjective measures, whose uncertainty results in compounding errors, and is the only measure which does not

provide greater insight into the behavior of the model. The Resilience Index developed by Argonne National Laboratories is intended for use as a comparison of resilience for facilities performing the same function in highly similar environments. The Resilience Index was constructed to be extremely simple to calculate to promote wide use for this reason. As such, it does not explain system behavior, does not exemplify vulnerabilities, and does not explain which component of resilience is strongest or weakest in the system.

Falling in the middle of the shortcomings was the Worst-Case Interdiction method. Because of its optimization basis, the method is inherently scenario dependent as well as being deterministic. However, the simplifying assumptions made by Salmeron *et al.* also limited the original form of this method to discrete degradations.

Clearly the two main competitors of methods to be utilized are the Monte Carlo Simulation method developed by Klibi and Martel, and and the Worst-Case Interdiction method used by Salmeron *et al*. (Klibi), (Salmeron 2009). While both provide valuable information both about the resilience of a system as well as the behavior under adverse conditions, Worst-Case Interdiction is more adaptable. By modeling the Attacker and Defender capabilities as mathematical constraints, an entirely new system may be modeled with relative ease.

This same adaptability, combined with the intuitive nature of the modeling process, could be exploited in order to include time-dependent Attacker capabilities, as well as removing the requirement of Discrete Degradation. By making these two changes, Worst-Case Interdiction becomes the front-runner method, though admittedly still not ideal. In the next section, solution methods, both to optimality and through the use of heuristics, are explored.

Solving Bi-Level Programs

Bi-level and multi-level programs have been used since the early 1970s in efforts to model hierarchical, interrelated events, and began as an evolution of leader-follower game theory (Moore p. 6). One early method was developed by J.E. Falk in 1973. A key breakthrough in Falk's work was the acknowledgement that while the leader and follower in a bi-level program (BLP) may have competing objectives, they each have their own separate resources and methods that may be employed to achieve their objective (Moore p. 10).

The solution methods have continued to evolve since Falk's initial work in 1973, involving both methods for converting the bi-level program to a standard mathematical program, as well as the development of multiple heuristics and decomposition methods. In 2004, Salmeron *et al.* begin by converting their BLP by utilizing a decomposition-based heuristic to indirectly arrive at a result (Salmeron 2004 p. 907).

Through iteratively solving the defender and the attacker portions of the BLP, Salmeron *et al.* permitted the defender to generate the network, and then allowed the attacker to strike the optimal locations to achieve their objective, followed by the defender's optimal response to the event. While their result is not guaranteed to be optimal, it provides a method for calculating and improving interdiction plans. This method required that the original BLP be broken into two separate models, with each iteration including additional constraints to both models (Salmeron 2004 p. 909).

Indices and Index Sets:

$i \in I$ Buses

$g \in G$ Generating units

$l \in L$ Transmission lines

$c \in C$ Consumer sectors

$s \in S$ Substations

$i \in I_s$ Buses at substation s

$g \in G_i$ Generating units connected to bus i

$l \in L_i^{Bus}$ Lines connected to bus i

$l \in L_s^{Sub}$ Lines connected to substation s

$l' \in L_l^{Par}$ Lines $l' \neq l$ running in parallel to line l

Parameters

o(l), d(l) origin and destination buses of line l

i(g) bus for generator g

d_{ic} load of consumer sector c at bus i

$\overline{P_l^{Line}}$ transmission capacity for line l

$\overline{P_g^{Gen}}$ maximum output from generator g

r_l, x_l resistance, reactance of line l

h_g generation cost for unit g

f_{ic} load shedding cost for customer sector c at bus i

Decision Variables

P_g^{Gen} generation from unit g

P_l^{Line} power flow on line l

S_{ic} load shed by customer sector c at bus i

θ_i phase angle at bus i

$$\min_{P^{Gen},P^{Line},S,\theta} \sum_g h_g P_g^{Gen} + \sum_i \sum_c f_{ic} S_{ic}$$

subject to

$$P_l^{Line} = B_l(\theta_{o(l)} - \theta_{d(l)}) \quad \forall l$$

$$\sum_g P_g^{Gen} - \sum_{l|o(l)=i} P_l^{Line} + \sum_{l|d(l)=i} P_l^{Line} = \sum_c (d_{ic} - S_{ic}) \quad \forall i$$

$$-\overline{P_l^{Line}} \le P_l^{Line} \le \overline{P_l^{Line}} \quad \forall l$$

$$0 \le P_g^{Gen} \le \overline{P_g^{Gen}} \quad \forall g$$

$$0 \le S_{ic} \le d_{ic} \quad \forall i,c$$

Figure 11: Power Flow Model (Salmeron 2004 p. 906)

$M_g^{Gen}, M_l^{Line}, M_i^{Bus}, M_s^{Sub}$	Resource required to interdict generator g, line l, bus i, and substation s, respectively
M	Total interdiction resources available to terrorist

Interdiction Variables

$\delta_g^{Gen}, \delta_l^{Line}, \delta_i^{Bus}, \delta_s^{Sub}$	Binary variables that take the value 1 if [node] is interdicted and are 0 otherwise

Formulation of I-DC-OPF

$$\max_{\delta_g^{Gen},\delta_l^{Line},\delta_i^{Bus},\delta_s^{Sub}} \gamma(\delta_g^{Gen}, \delta_l^{Line}, \delta_i^{Bus}, \delta_s^{Sub})$$

Subject to:

$$\sum_{g\in G} M_g^{Gen} \delta_g^{Gen} + \sum_{l\in L} M_l^{Line} \delta_l^{Line} + \sum_{i\in I} M_i^{Bus} \delta_i^{Bus} + \sum_{s\in S} M_s^{Sub} \delta_s^{Sub} \le M$$

$$\gamma(\delta_g^{Gen}, \delta_l^{Line}, \delta_i^{Bus}, \delta_s^{Sub}) =$$

$$= \min_{P^{Gen},P^{Line},S,\theta} \sum_g h_g P_g^{Gen} + \sum_i \sum_c f_{ic} S_{ic}$$

Subject to

$$\mathrm{P}_l^{Line} = B_l(\theta_{o(l)} - \theta_{d(l)})(1-\delta_l^{Line})(1-\delta_{o(l)}^{Bus})(1-\delta_{d(l)}^{Bus}) \prod_{l\in L_s^{Sub}}(1-\delta_s^{Sub}) \prod_{l'\in L_l^{Par}}(1-\delta_{l'}^{Line}) \quad \forall \mathrm{l}$$

$$\sum_g P_g^{Gen} - \sum_{l|o(l)=i} P_l^{Line} + \sum_{l|d(l)=i} P_l^{Line} = \sum_c (d_{ic} - S_{ic}) \quad \forall i$$

$$\text{-}\overline{\mathrm{P}_l^{Line}}(1-\delta_l^{Line})(1-\delta_{o(l)}^{Bus})(1-\delta_{d(l)}^{Bus}) \prod_{l\in L_s^{Sub}}(1-\delta_s^{Sub}) \prod_{l'\in L_l^{Par}}(1-\delta_{l'}^{Line}) \le \mathrm{P}_l^{Line} \le \overline{\mathrm{P}_l^{Line}}(1-\delta_l^{Line})(1-\delta_{o(l)}^{Bus}) (1-\delta_{d(l)}^{Bus}) \prod_{l\in L_s^{Sub}}(1-\delta_s^{Sub}) \prod_{l'\in L_l^{Par}}(1-\delta_{l'}^{Line}) \quad \forall l$$

$$0 \le P_g^{Gen} \le (1-\delta_{i(g)}^{Bus})(1-\delta_g^{Gen})\overline{P_g^{Gen}} \quad \forall g$$

$$0 \le S_{ic} \le d_{ic} \quad \forall \mathrm{i,c}$$

Figure 12: Interdiction of Power Flow (Salmeron 2004 p. 907)

The Salmeron *et al.* BLP model, shown in Figure 12 works off of the assumed objective of the attacker maximizing the minimal cost to operate the system, the interior optimization model from Figure 11, but is clearly incorporated in Figure 12. By building the Defender's model first, which is the main point of validation for this particular type of modeling method, a large portion of the work for the BLP model was completed. In Figure 12, every constraint followed by (IDC.#) constructs the modified version of the Defender model. Meanwhile, the remaining constraints, which directly precede the internal Defender constraints, are the Attacker's restrictions.

The heuristic utilized by Salmeron *et al.* relied heavily upon the Defender model, which modeled the behavior of the Power Flow network. By optimizing the Defender objective, interdicting the most heavily used nodes, and then resolving iteratively until all

resources had been consumed, their heuristic arrived at a solution, and was able to plot the effect that increasing Attacker resources would have (Salmeron p. 909).

In 2005, Arroyo and Galiana continued the Salmeron *et al.* case study by converting the BLP into an LP, using a two step process to achieve this conversion.

> *"Step 1) the explicit characterization of the inner optimization problem by its Karush-Kuhn-Tucker (KKT) optimality conditions; Step 2) the use of integer algebra results due to Floudas' Nonlinear and Mixed Integer Optimization: Fundamentals and Applications and to Fortuny-Amat and McCarl's A representation and economic interpretation of a two-level programming problem to convert the nonlinear KKT relations into equivalent linear forms."*
>
> (Arroyo p. 790).

What follows is the nomenclature used by Arroyo and Galiana:

A. Indices

j Generator Index

l Transmission line index

n Bus index

B. Sets

J Set of indices of generators

J_n Set of indices of generators connected to bus n

L Set of indices of transmission lines

N Set of indices of buses

C. Functions

L(v) Lagrangian Function of the inner optimization problem, which depends on vector v

D. Constants

A_{nl} Element of the network incidence matrix that is equal to 1 if the bus n is the sending bus of line l, -1 if bus n is the recieving bus of line l, and 0 otherwise

FR(l) Sending bus of line l

M Sufficiently large positive constant

P_n^d Demand at bus n

$\overline{P_l^f}$ Power flow capacity of line l

$\overline{P_j^g}$ Capacity of generator j

$\underline{P_j^g}$ Minimum power output of generator j

TO(l) Recieving bus of line l

x_l Reactance of line l

$\overline{\delta}$ Upper bound for the nodal phase angles

$\underline{\delta}$ Lower bound for the nodal phase angles

$\overline{\mu_l}$ Upper bound for μ_l

$\underline{\mu_l}$ Lower bound for μ_l

ϕ^{Spec} Lower level of total load shed specified by the terrorist

E. Variables

h_l Slack variable used in the linear expression equivalent to the produce of μ_l and v_l

P^f Vector of line power flows

P^g Vector of generator power outputs

s_l^{FR} Slack variable used in the linear expression equivalent to the product of v_l and $\delta_{FR(l)}$

s_l^{TO} Slack variable used in the linear expression equivalent to the product of v_l and $\delta_{TO(l)}$

t_l Variable equal to the product of μ_l and v_l

v_l 0/1 variable that is equal to 0 if line l is destroyed and equal to 1 otherwise

w 0/1 variable used in the linear expressions of the complementary slackness conditions

z_t^{FR} Variable equal to the product of v_l and $\delta_{FR(l)}$

z_t^{TO} Variable equal to the product of v_l and $\delta_{TO(l)}$

$\overline{\alpha_n}$ Lagrange multiplier associated with the upper bound for the load shed at bus n

$\underline{\alpha_n}$ Lagrange multiplier associated with the lower bound for the load shed at bus n

δ Vector of nodal phase angles

ΔP^d Vector of nodal loads shed

One major drawback in Arroyo and Galliana's method for solving the BLP from Salmeron *et al.* is the increase in size and complexity. Adapting a simplified version of the model used by Salmeron *et al* , shown in Figure 13, they converted the linear BLP to a single level mixed integer non-linear program (Salmeron 2009). Using Karush-Kuhn-Tucker conditions, Arroyo and Galliana then converted the nonlinear constraints into

equivalent linear constraints (Arroyo). The resulting model, Figure 14, grew in constraints by a factor of five.

$$\max_{v} \sum_{l \in L} v_l$$

subject to

$$\sum_{n \in N} \Delta P_n^{d*} \geq \phi^{spec}$$

$$\Delta P^{d*} = \arg \left\{ \min_{\delta, P^g, P^f, \Delta P^d} \sum_{n \in N} \Delta P_n^d \right\}$$

subject to

$$\mathrm{P}_l^f = v_l \frac{1}{x_l} \sum_{n \in N} A_{nl} \delta_n \quad \forall \mathrm{l} \in \mathrm{L}$$

$$\sum_{j \in J_n} P_j^g - \sum_{l \in L} A_{nl} \mathrm{P}_l^f + \Delta P_n^d = P_n^d \quad \forall n \in N$$

$$-\overline{P_l^f} \leq P_l^f \leq \overline{P_l^f} \quad \forall \mathrm{l} \in \mathrm{L}$$

$$\underline{P_j^g} \leq P_j^g \leq \overline{P_j^g} \quad \forall j \in J$$

$$0 \leq \Delta \mathrm{P}_n^d \leq P_n^d \quad \forall \mathrm{n} \in \mathrm{N}$$

Figure 13: Simplified BLP (Arroyo p. 791)

$$\max \sum_{l \in L} v_l$$

subject to

$$\sum_{n \in N} \Delta P_n^d \geq \phi^{spec} \qquad \forall n \in N$$

$$P_l^f = \frac{1}{x_l}(z_l^{FR} - z_l^{TO}) \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$z_l^{FR} = \delta_{FR(l)} - s_l^{FR} \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$z_l^{TO} = \delta_{TO(l)} - s_l^{TO} \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$\underline{\delta} v_l \leq z_l^{FR} \leq \overline{\delta} v_l \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$\underline{\delta} v_l \leq z_l^{TO} \leq \overline{\delta} v_l \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$\underline{\delta}(1 - v_l) \leq s_l^{FR} \leq \overline{\delta}(1 - v_l) \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$\underline{\delta}(1 - v_l) \leq s_l^{TO} \leq \overline{\delta}(1 - v_l) \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$\sum_{j \in J_n} P_j^g - \sum_{l \in L} A_{nl} P_l^f + \Delta P_n^d = P_n^d \qquad \forall n \in N$$

$$-\underline{P_l^f} \le P_l^f \le \overline{P_l^f} \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$-\underline{P_j^g} \le P_j^g \le \overline{P_j^g} \qquad \forall j \in J$$

$$0 \le \Delta P_n^d \le P_n^d \qquad \forall n \in N$$

$$\sum_{l \in L} \frac{1}{x_l} A_{nl} t_l = 0 \qquad \forall n \in N$$

$$t_l = \mu_l - h_l \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$\underline{\mu_l} v_l \le t_l \le \overline{\mu_l} v_l \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$\underline{\mu_l}(1 - v_l) \le h_l \le \overline{\mu_l}(1 - v_l) \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$-\lambda_n \mid_{j \in J_n} -\underline{\theta_j} + \overline{\theta_j} = 0 \qquad \forall n \in N$$

$$\sum_{n \in N} A_{nl} \lambda_n - \mu_l - \underline{w_l} + \overline{w_l} = 0 \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$1 - \lambda_n - \underline{\alpha_n} + \overline{\alpha_n} = 0 \qquad \forall n \in N$$

$$\underline{w_l}, \overline{w_l}, \underline{\theta_j}, \overline{\theta_j}, \underline{\alpha_n}, \overline{\alpha_n} \ge 0 \qquad \forall \mathrm{l} \in \mathrm{L}, n \in N, j \in J$$

$$\underline{w_l} \le M w_l^{\underline{w}} \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$P_l^f + \overline{P_l^f} \le M(1 - w_l^{\underline{w}}) \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$\overline{w_l} \le M w_l^{\overline{w}} \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$\overline{P_l^f} - P_l^f \le M(1 - w_l^{\overline{w}}) \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$\underline{\theta_j} \le M w_j^{\underline{\theta}} \qquad \forall j \in J$$

$$\overline{P_j^g} - P_j^g \le M(1 - w_j^{\underline{\theta}}) \qquad \forall j \in J$$

$$\overline{\theta_j} \le M w_j^{\overline{\theta}} \qquad \forall j \in J$$

$$\overline{P_j^g} - P_j^g \le M(1 - w_j^{\overline{\theta}}) \qquad \forall j \in J$$

$$\underline{\alpha_n} \le M w_n^{\underline{\alpha}} \qquad \forall n \in N$$

$$\Delta P_n^d \le M(1 - w_n^{\underline{\alpha}}) \qquad \forall n \in N$$

$$\overline{\alpha_n} \le M w_n^{\overline{\alpha}} \qquad \forall n \in N$$

$$P_n^d - \Delta P_n^d \le M(1 - w_n^{\overline{\alpha}}) \qquad \forall n \in N$$

$$w_l^{\underline{w}} + w_l^{\overline{w}} \le 1 \qquad \forall \mathrm{l} \in \mathrm{L}$$

$$w_j^{\underline{\theta}} + w_j^{\overline{\theta}} \le 1 \qquad \forall j \in J$$

$$w_n^{\underline{\alpha}} + w_n^{\overline{\alpha}} \le 1 \qquad \forall n \in N$$

Figure 14: Single Mixed Integer Linear Program (Arroyo p. 792)

Using the Lagrangian function of the inner optimization problem, the "Defender" or "Follower" problem, Arroyo *et al.* convert the inner problem to an equivalent mixed integer nonlinear dual. Taking advantage of logical binary variables, all nonlinear constraints, of both binary variables multiplied by continuous variables, and of the complementary slackness conditions, are then converted into an equivalent set of linear constraints.

However, it remains unclear whether or not this solution bought anything in the way of efficiency or results. Whereas Salmeron *et al.* ran their algorithm until the improvements were negligible, Arroyo *et al*. allowed their model to stop so long as pre-specified levels were met. This approach, when coupled with their objective of the fewest nodes attacked to achieve these levels, does not guarantee selecting levels the model may need to select the optimal combination. Furthermore, the conversion of the BLP to an MILP required a large increase in the number of constraints, which was accompanied by an even greater increase in the number of integer variables (Arroyo p. 792). As such, this single model, when coupled with the much greater pre-processing time than the heuristic used by Salmeron *et al*., may achieve very little in terms of computational efficiency (Salmeron 2009). However, the model does demonstrate a method for converting a bi-level model to a single-level optimization model.

A similar method to Arroyo *et al.*'s for solving a BLP indirectly is to utilize Benders' partitioning (Brown 2006 p. 533). In the case when the inner problem is an LP, taking the dual of the attacker-defender model will result in a Mixed Integer Linear Program (MILP). Unfortunately, this is the most simplistic BLP case, and as Brown *et*

al. point out, when the BLP is not Attacker-Defender with Defender being modeled as an LP, no general transformation of this form exists (Brown 2006 p. 535).

As one may notice, both Arroyo's and Brown's solution methodologies, which utilized the dual of the problem and incorporated KKT conditions, showed very little effectiveness and were restricted in terms of applicability. This is primarily due to the nonconvexity that all BLP, even those who are entirely linear and continuous, exhibit. However, the heuristic Salmeron *et al.* used in 2004 showed very few restrictions and quickly achieved a result above what Arroyo's proposed optimal method, which had to use the original objective from Salmeron *et al* work as a fixed constraint, could produce (Salmeron 2009). "We are still only able to solve moderate size problems. Heuristics and what have become known as global optimization techniques, offer additional possibilities" (Bard p. 361).

A recent publication by Alderson *et al.* at the Naval Postgraduate School continued the efforts for an exact solution, seeking to now solve a tri-level optimization model, the Defender-Attacker-Defender model. With their setting as a traffic scenario using "The Seven Bridges of Konigsberg" (Alderson p. 38), the general outline for the formulation of constraints remains nearly identical to that used in previous work by Brown and Salmeron. The main difference in this case study from the one utilized in Salmeron *et al.*'s work on Power Flow is size, possessing only a fraction in number of nodes and arcs. Continuing with the efforts applied to the Global Benders Decomposition, a similar method is used in the case of a three tiered optimization model.

This method uses a meta-heuristic, shown in Figure 14, which takes advantages of bounds established by the Global Benders Decomposition developed by Salmeron *et al.*

Algorithm DA$\underline{\text{D}}$-Decomp

Input: Full DA$\underline{\text{D}}$ problem data and optimality tolerances ε, ε_{MP}, $\varepsilon_{\mathbf{A}\underline{\mathbf{D}}} \geq 0$ for the overall decomposition, the DA$\underline{\text{D}}$ master problem, and the A$\underline{\text{D}}$ subproblem, respectively.

/*$\varepsilon \geq \varepsilon_{\text{MP}}$ is assumed.*/

Output: ε-optimal defense plan $\mathbf{w}^*$ and corresponding attack plan $\mathbf{x}^*$;

1. $\text{LB} \leftarrow -\infty$; $\text{UB} \leftarrow \infty$; $K \leftarrow 1$;
2. for (all $(i,j) \in E$)$\{\widehat{w}_{ij}^{d_0 K} \leftarrow 1$; $\widehat{w}_{ij}^{d_0 K} \leftarrow 0, d \neq d_0;\}$; $\mathbf{w}^* \leftarrow \widehat{\mathbf{w}}^K$;

/*That is, choose "no defense" as the initial defense plan and as the incumbent solution.*/

3. ***Subproblem***: Solve DA$\underline{\text{D}}(\widehat{\mathbf{w}}^K, \cdot, \cdot)$ to determine attack plan $\widehat{\mathbf{x}}^K$ given defense plan $\widehat{\mathbf{w}}^K$ such that $z^{\text{UP}}_{\mathbf{A}\underline{\mathbf{D}}} - z^{\text{LO}}_{\mathbf{A}\underline{\mathbf{D}}} \leq \varepsilon_{\mathbf{A}\underline{\mathbf{D}}} z^{\text{LO}}_{\mathbf{A}\underline{\mathbf{D}}}$;

/*We assume $z^{\text{UP}}_{\mathbf{A}\underline{\mathbf{D}}}, z^{\text{LO}}_{\mathbf{A}\underline{\mathbf{D}}} \geq 0$*/

4. if $(z^{\text{UP}}_{\mathbf{A}\underline{\mathbf{D}}} < \text{UB})\{\text{UB} \leftarrow z^{\text{UP}}_{\mathbf{A}\underline{\mathbf{D}}}$; $\mathbf{w}^* \leftarrow \widehat{\mathbf{w}}^K$; $\mathbf{x}^* \leftarrow \widehat{\mathbf{x}}^K;\}$
5. if $(\text{UB} - \text{LB} \leq \varepsilon \text{LB})$ go to **End**;
6. if $\widehat{\mathbf{x}}^K$ repeats any prior attack, i.e., $\widehat{\mathbf{x}}^K \in \widehat{X}^K$, temporarily add one "solution-elimination constraint" to DA$\underline{\text{D}}(\widehat{\mathbf{w}}^K, \cdot, \cdot)$ for each $\widehat{\mathbf{x}}^k \in X^K$, and re-solve for a new $\widehat{\mathbf{x}}^K$;

/*Solution-elimination constraints are described below. For simplicity, the algorithm ignores the possibility that problem in Step 7 could be infeasible.*/

7. $\widehat{X}^K \leftarrow \widehat{X}^{K-1} \cup \{\widehat{\mathbf{x}}^K\}$;
8. ***Master Problem:*** Solve DA$\underline{\text{D}}$-MP $(\widehat{X}^K)$ to determine defense plan $\widehat{\mathbf{w}}^{K+1}$ such that $z^{\text{UP}}_{\text{MP}} - z^{\text{LO}}_{\text{MP}} \leq \varepsilon_{\text{MP}} z^{\text{LO}}_{\text{MP}}$;

/*We assume $z^{\text{UP}}_{\text{MP}}, z^{\text{LO}}_{\text{MP}} \geq 0$*/

9. if $(z^{\text{LO}}_{\text{MP}} > \text{LB})$ $\text{LB} \leftarrow z^{\text{LO}}_{\text{MP}}$;
10. if $(\text{UB} - \text{LB} \leq \varepsilon \text{LB})$ go to **End**;
11. $K \leftarrow K + 1$; go to **Subproblem;**
12. ***End***: print ("ε-optimal defense plan and corresponding attack plan are," $\mathbf{w}^*$, $\mathbf{x}^*$).

Figure 15: Algorithm DAD Decomposition (Alderson *et al.* p. 37)

While Alderson *et al.*'s paper is a clear continuation of previous BLP efforts done at Naval Postgraduate School, the efforts have been focused on improving the method for solving discrete, linear, time independent models.

Little work has been done thus far on solving continuous network interdiction models. A continuous network interdiction is a case in which the nodes or arcs of a network may be partially diminished in a continuous fashion. One of the main difficulties with making these variables continuous is that "we can no longer resort to standard linearization procedures to solve the interdiction problem by a single integer program" (Lim p. 20). Lim and Smith demonstrate two distinct methods for solving continuous network interdiction bi-level programs, an exact approach and a heuristic.

Their exact method is based upon a partitioning algorithm which solves |A| sub-problems where A is the constraint matrix of the defender's mode, and each of these resulting sub-problems are mixed integer linear programs (Lim p. 21). Their partitioning algorithm, laid out in page 21 of their paper, first determines if a solution exists, or rather if the attacker has sufficient resources to interdict any node even partially. Next, Lim and Smith's algorithm solves a model to select and interdict the most beneficial point to eliminate and what impact would be most beneficial for the resources required. If the resulting objective function is better than the previous result, then the current settings are kept and another iteration is performed to find the next best node to degrade. If no resources are left to be expended, or no nodes remain to be eliminated, then the algorithm terminates. Note that this method is close to enumeration and may require a large amount of processing power.

In their heuristic section, Lim and Smith point out that the exact solution method "may not be suitable for solving large-scale problems." (Lim p. 21). To start their heuristic method, Lim and Smith systematically fully eliminate a single node in the network and view the resulting effects. If resources remain, then another node is interdicted, much like in the exact case. However, should resources remain that are insufficient to fully interdict another node, then those resources are utilized to partially interdict the next point which "exhibits the best ratio of objective decrease to budget consumed when interdicted" (Lim p. 22). This heuristic, which follows immediately, demonstrated an improvement in solve time by approximately 35%, with a penalty to optimal solution of roughly 9% (Lim p. 24).

Step 0. Initialize $\bar{x}_h = 0\ \forall h \in A$, define the remaining budget as $\bar{B} = B$, and define the candidate set of arcs that can be interdicted as $\bar{A} = A$.

Step 1. Set $\sigma = 0$, $R = \bar{A}$, and solve the follower's problem given $\bar{x}$. Let v be the optimal objective value of the follower's problem.

Step 2. If $R = \emptyset$, go to Step 4. Otherwise, choose any $h \in R$ and proceed to Step 3.

Step 3. Set $\bar{x}_h = \min\{b_h, \bar{B}\}/b_h$. Solve the follower's problem to obtain the optimal objective value v_h. If $(v - v_h)/\min\{b_h, \bar{B}\} > \sigma$, put $\hat{h} = h$ and $\sigma = (v - v_h)/\min\{b_h, \bar{B}\}$. Reset $\bar{x}_h = 0$ and return to Step 2.

Step 4. Fix $\bar{x}_{\hat{h}} = \min\{b_{\hat{h}}, \bar{B}\}/b_{\hat{h}}$, update $\bar{B} = \bar{B} - \min\{b_{\hat{h}}, \bar{B}\}$, and remove $\hat{h}$ from $\bar{A}$. If $\bar{B} = 0$, terminate with the heuristic solution $\bar{x}$. Otherwise, return to Step 1.

Heuristic Algorithm from Lim and Smith p. 22

Unfortunately, Lim and Smith found that their heuristic, while operating quickly with an acceptable margin of optimality error, were applicable only to "grid structured problems" and were found to be unreliable for "general topologies generated by Mnetgen" (Lim p. 25).

One heuristic method that is quickly gaining recognition is ant colony optimization. Ant colony optimization is a metaheuristic that was first proposed in 1992 by Dorigo as a method for solving the Travelling Salesman Problem (TSP) (Solnon p. 1). Based upon swarm intelligence, in a travelling salesman problem ants are sent out from the point of origin and probabilistically select a path to travel. When a path is travelled, a "pheromone" correlating to the objective value derived from travelling that path is assigned to it (Solnon p. 109). This "pheromone" then alters the probability of that particular path being chosen again, and every time the path is selected, the "pheromone" is increased (Solnon p. 110). To ensure that premature convergence to an answer is

avoided, a degradation function is placed on the pheromones so that their effect does not explode in a short period of time and the method can converge to an overall best solution (Solnon p. 111).

Ant Colony Optimization was recently used to solve a BLP by Calvete *et al*. Their method took advantage of the nature of BLP which solves first the attacker portion before solving the responsive defender model. Using an adapted form of ACO, Calvete *et al.* send their *ants* down the many possible paths in their production-distribution case, solve the resulting defender model, calculate the attacker's objectives, and finally update the pheromone trails based upon the attacker's objective values (Calvete p. 324). This method, which is similar in concept to that used by Salmeron *et al.* in 2004, explores the many possible combinations the attacker may exploit, and slowly converges to the area of a best solution. Calvete *et al.* close by saying that, with modifications, more complex problems may be applied to the lower level defender problem (Calvete p. 327).

While Calvete *et al.* have used more advanced techniques than Salmeron *et al.* used nine years prior, their method at the core is an improvement in the rate of convergence. Where Salmeron *et al.* eliminated whole combinations when it was showed that a better solution existed, Calvete *et al.* used the nature of ACO to force their heuristic to trend towards the best individual options, which together created the preferred sets. However, the Calvete *et al*. paper still shows the same restriction of discrete binary options of where to place a supply node based upon the location of the demand nodes (Calvete p. 325). As was discussed in Lim and Smith's work, the restricted discrete case is significantly easier to solve due to the restricted number of options and the nature of the current solution methods being pursued to solve BLP.

Heuristics for Large Scale Set Covering Problems

In this section, prominent works for heuristics designed to solve the Set Covering Problem for large scale models are reviewed. Key works for this section are *A Heuristic Method for the Set Covering Problem (SCP)* by Caprara *et al.*, *Ant Colony Optimization and Constraint Programming* by Christine Solnon, and *A New Model for Planning Emergency Facilities in Shanghai* by Luo *et al*.

The most common heuristic method across all optimization problem types is the greedy algorithm. In a SCP the greedy algorithm selects the placement which gains the most value, regardless of resource cost. Being an approach commonly tested against, the greedy approach rarely shows as the best method, and, as in the following papers to be reviewed, is frequently the worst. However, due to its simplicity, the greedy algorithm does serve as a fall-back when all else fails.

In *A Heuristic Method for the Set Covering Problem* by Caprara *et al.*, a "Langrangian-based heuristic" (Caprara p. 730) was developed to solve very large scale SCP. The key points of this heuristic are that it utilizes dynamic values and column fixing to find "improved solutions" (Caprara p. 730). Caprara *et al.* focus solely on a continuous linear covering problem which utilizes the dual constraints and a Benders type convergence method to obtain a near-optimal solution.

While the method showed little improvement in final solution versus the "classical" strategies, which refer to a greedy approach, the method presented by Caprara *et al*. does converge to that solution in roughly one quarter of the iterations (Caprara p. 733). Furthermore, by utilizing the Lagrange of the dual, the method will always

converge to a good solution if allowed to run long enough. For case studies, Caprara *et al*. utilized massive sets based upon the Italian railway company. The largest of the sets run on a personal computer was 507x63,009 nodes, which required 634.8 computer seconds. The remaining sizes required a high-performance computer to complete (Caprara p. 735).

Some restrictions to the method utilized by Caprara *et al*. is the need to find the dual, and for the dual problem and the original problem to converge to a solution. While this is always true for linear problems, non-linear problems often struggle with both restrictions.

Another, more recent method to solving set covering problems is through the use of Ant Colony Heuristics. "The whole is more than the sum of its parts." (Solnon p. 106). The concept Solnon is conveying is known as *holism* and is an underlying principal of Swarm Intelligence, and the later evolution to Ant Colony Optimization (ACO). Swarm Intelligence is the "collective ability to achieve global tasks" (Solnon p. 107). Through basic interactions, the entities of a swarm influence each other as well as the objective value of the system, either directly or indirectly. These interactions occur at an elementary level, such as in the example of a shortest path problem. In such a problem, the elementary decision to be made is which path to proceed down next, which in many cases is countable. This decision then leads to a new set of paths which may be tread.

Many swarms operate by using pheromones to transmit paths travelled; however if a path has never been travelled then no pheromone exists to be followed. The individual entities in a swarm begin with initially no preference of a pathway or option,

making the initial decision random. As time progresses, the better paths become frequented more often, and an increased quantity of pheromones are deposited at a faster rate than the less used pathways (Solnon p. 108).

Having been originally developed for the Travelling Salesman Problem (TSP) in 1992 (Solnon p. 1), this metaheuristic shows its roots in the original formulations. Given *v,* the currently occupied vertex, *v'* the destination vertex, and finite number of feasible pathways J, an entity will travel between these two points via edge e_i , $i \in$ J, with probability:

$$p_{ve_i}(t) = \frac{\left[\phi_{ve_i}(t)\right]^{\alpha}}{\sum_{j \in J}\left[\phi_{ve_j}(t)\right]^{\alpha}},$$

where $\phi_{ve_i}(t)$ is a function determining the strength or quantity of pheromone along edge e_i based upon the number of entities who have selected edge e_i previously, and α is the sensitivity of the entities to the pheromone (Solnon p. 110).

Since the time of its original use in TSP, ACO has become more generalized as a greedy heuristic with biased probabilities. The pheromones used biasing the probability trend the paths towards the more desirable orders to such an extent that the heuristic may "learn" to more often place some object j after object i (Solnon p. 117). This ordering method is frequently used in "vehicle routing problems, car sequencing problems, and job scheduling problems" (Solnon p. 117).

At every point, the objective of the ACO is to find the best combination of objects to optimize the objective, beginning unbiased. Using local searches, combinations may be improved much more rapidly, which may allow a greater rate of improvement and a stronger pheromone trail (Solnon p. 118).

Intensification encompasses the need to increase the strength of the pheromone applied to a trail based upon its resulting value (Solnon p. 128). Diversification is the need to allow less travelled routes to be re-inspected at regular intervals, either through branching or through steady degradation of the pheromone to prevent an excess bias (Solnon p. 136). It is important that intensification and diversification be balanced so that no trail becomes overpowering, as well as reducing the likelihood that a potentially better trail is ignored.

In 2013, Luo *et al*. utilized Ant Colony Optimization techniques to determine the placement of first aid emergency facilities in Shanghai, China, which they modeled as an integer program (Luo p. 224). With an objective function of minimizing the expected cost of delayed responses, as well as the cost of operating each station and vehicles, Luo *et al.* utilized a set of permissible locations and respective radii based upon equipment to be used at the location. Any location that fell within that radius was considered to have no delay, while any location outside of the radius and not covered by any other station suffered a delay dependent upon the distance to the nearest station (Luo p. 225).

To adapt the ACO algorithm to set covering problem, Luo *et al.* determined that every visited vertex was an activated supply station while any vertex not visited remained closed/unused (Luo p. 226). The probability that a node would be visited was defined as:

$$p_{js}^{k}=\begin{cases}1, & \text{if } q<q_0 \text{ and } \mathrm{y}_j = \arg\max\{\tau_{js}*\eta_{js}^{\beta} \mid s=o,1..y_{\max}\} \\ 0, & \text{if } q<q_0 \text{ and } \mathrm{y}_j \neq \arg\max\{\tau_{js}*\eta_{js}^{\beta} \mid s=o,1..y_{\max}\} \\ \dfrac{\tau_{jy_{random}}\eta_{jy_{random}}^{\beta}}{\sum_{h=1}^{y_{\max}}\tau_{jh}*\eta_{jh}^{\beta}}, & \text{if } q\geq q_0 \text{ and } y_{\text{random}}=0,1,...y_{\max}\end{cases}$$

In this equation, *p* represents the probability of being selected, *y* the number of ambulances to be located at the vertex in question, τ the gained likelihood of a vertex being activated based upon previous iterations, η a static measure showing the deviation from y, and β the significance of the measures τ and η (Luo p. 226).

The following two equations are used by Luo *et al.* to ensure that a minimal number of ambulances within an area are sufficient to permit the double coverage of their requirement. Here, y'_j signifies the minimal number of ambulances required by an area W_j.

$$\tau_{ir}=\begin{cases}1, & \text{if } \mathrm{y}_j = \mathrm{y}'_j \\ \max(0.001, 1-0.1*|\mathrm{y}_j - \mathrm{y}'_j|), & else\end{cases}$$

$$\eta_{ir}=\begin{cases}1, & \text{if } \mathrm{y}_j = \mathrm{y}'_j \\ \max(0.001, 1-0.1*|\mathrm{y}_j - \mathrm{y}'_j|), & else\end{cases}$$

The results of this heuristic show that the cost of providing care could actually be reduced by half increasing the number of emergency facilities from 29 to 47 (Luo p. 227).

In this section, heuristics designed to solve the large scale Set Covering Problems were reviewed. These heuristics were based on Lagrange, Duality, and Ant Colony Optimization techniques. It has been determined that, while the use of dual problems

guaranteed convergence in the models reviewed, it required that all constraints be linear. The method derived from Ant Colony Optimization, on the other hand, made no restriction to the type of constraints, but also is unable to guarantee convergence to a single best point or set of best points.

Is an Empty Network Resilient

In this section, a step that is often used in Fourier Transforms of functions is used, and that is to define the bounds. Because the number of nodes in a satellite network can become large, the only bound requiring a hard definition is the lower bound of a non-existent or empty network. Defining this bound is important from a mathematical standpoint because it will create a baseline for the analysis.

As the framework for the model developed in this thesis is constructed, there is an issue that remains to be discussed, and that is if a network with no nodes is considered to be resilient. Recall that Salmeron *et al.* measure resilience as "the difference in operating costs, including penalties for un-served demand, after and before interdiction." (Salmeron) Consider a network where there are no nodes that can be attacked.

This network is certainly resistant to interdictions as there is no change from before an event to after an event, regardless of how difficult it would be to discern an event on a system that does not exist. Because there is no change in the system from before and after the shock, there is no change in un-served demand, and thus no change in penalty for un-served demand. There is also no increased cost for clean up and repair.

Based upon Salmeron *et al.*'s measure, a network that does not exist is certainly the best network, an error that must be accounted for.

Utilizing the measure for resilience presented by Cimellaro *et al*., in which the measure is the integral of the network performance over time given an event, the opposite occurs. In this situation, one can see intuitively that an empty network, which is permanently unable to provide any performance, is the least resilient. An exception to this is the situation in which there is simultaneously zero function and zero need for the function. In such a situation, the network is both the least and the most resilient, as the desired level of performance is always achieved.

In line with the Argonne definition followed in this work, as well as the components of resilience outlined by the DoD, the effects that increasing network capacity is viewed, whether for relaying or being a source of commodities, by examining the change that increasing the number of homogenous nodes has on resilience under each measure. Recall from Chapter I section Defining Resilience, the four components of resilience are avoidance, robustness, reconstitution, and recovery.

Increasing the number of homogenous nodes in the network will have a combination of two effects, acting as a spare should a primary node collapse and satisfying previously unsatisfied demand. Note that it may wholly fall under a single of these effects. In the situation when a node is a spare, the change in costs would be lessened as there is less change in un-served demand. However, if the new node is satisfying a set of only previously unsatisfied demand, then after an event in which a node is lost, the change in un-served demand is the same as if the new node had not been included.

The effect this change would have on the measure utilized by Cimellaro *et al*. is now inspected (Cimellaro). In case of a new node acting as a spare, the change in network performance would experience a lessened shock as a result of the degrading event. As such the resilience measure would increase. If the new node is satisfying a unique set of previously unsatisfied demand, then the network performance curve is translated up, though the lost performance as a result of an event remains the same. In this situation, the network is still considered to be more resilient than if the new node had not existed.

Clearly, including a new node should increase resilience so long as it is performing some function, whether as a spare or as a new primary relay/source node. As such, the resilience measure should increase, though not necessarily by the same amount, in both situations. The measure utilized by Cimellaro *et al*. measure satisfies this while that used by Brown *et al*. does not (Cimellaro), (Brown 2006).

As no specification has been made as to the number of nodes in the network before the addition, this remains true for transitioning from a nonexistent network to a network with only one node, making exception for the obvious irrational situations in which a new node is useless, such as a supply network with only a relay node and no source. Since a network with one node should be more resilient than a network with no nodes, then a nonexistent network cannot be the most resilient. As such, the nonexistent, or empty, network must be the least resilient.

In this section, the measurement methods of Brown *et al.* and Cimellaro *et al.* were utilized to exemplify setting an empty network as possessing the theoretical lower

bound on resilience. Furthermore, the concept of increasing network size having a positive effect on robustness and recovery, and as a result on resilience, was presented.

Measuring Resilience

In this section, inherent measure characteristics and behaviors prescribed by the DoD definition, as well as those apparent in *The Black Swan* by Nassim Nicholas Taleb, and the Argonne National Laboratory definition, are reviewed.

A portion of the measure guides are laid out in the DoD definition of resilience (Fact Sheet: Resilience of Space Capabilities).

> *Resilience is the ability of an architecture to support the functions necessary for mission success in spite of hostile action or adverse conditions. An architecture is "more resilient" if it can provide these functions with higher probability, shorter periods of reduced capability, and across a wider range of scenarios, conditions, and threats.*

From this definition, the DoD outlines that the measure of resilience of a system should improve is it can maintain its functionality in the face of a wide range of adversities. The measure should also improve if recovery time is decreased or if the probability of failure when faced with adversity is decreased. Note that the improvement may be negligible or nonexistent, but never negative.

From *The Black Swan* by Nassim Nicholas Taleb, which investigates robustness and fragility, another set of guidelines is exposed. First, robustness, a component of resilience as determined by DoD and the Argonne National Laboratory definitions, is improved as redundancies are included (Taleb p. 312). By extension, the resilience

measure should improve as well. Mitigating actions, actions which lower the probability or the impact of an event, improve resilience by improving robustness and avoidance (Taleb p. xxvii).

Consider the Argonne National Laboratory resilience definition (Carlson p. 7):

> *Resilience is the ability of an entity -e.g., asset, organization, community, region- to anticipate, resist, absorb, respond to, adapt to, and recover from a disturbance.*
>
> *Argonne National Laboratory*

It is clear that any action which results in reduced recovery time, increased adaptation, and an increase in a system's ability to resist or absorb degrading events will also increase resilience.

An obvious characteristic of the resilience measure is its ability to capture the performance and functionality of the system when faced with adversity. It must also incorporate the probabilistic nature system survival and reliability, but must do so while acknowledging the possibility of a Black Swan event. Recall the measure utilized by Klibi and Martel (Klibi p. 6), which was the network performance over time. This measure is also utilized by Cimellaro *et al.* (Cimellaro p. 3642) and provided the basis for Zobel and Khansa's measure of resilience (Zobel p. 84).

Another system aspect seen to improve resilience is diversification. Diversifying components of a system, be they resources, locations, or capabilities, reduces the impact that environment changes and degrading events may have on a system (NSTAC p. ES-2).

Utilizing network performance as a basis, established and well known measures are incorporated. Incorporating the time component, both for the duration of the

degrading event as well as the recovery and post-operation portions, the resilience measure may also capture the behavior and functionality of the system over time.

Redundancies and mitigation efforts may be captured through the use of scenarios, though a level of scrutiny should be utilized in order to not over inflate their effect. Including such a component or defense may serve to only shift the point of degradation and not actually improve resilience (Taleb p. xxvii; Salmeron 2004 p. 911).

When constructing scenarios, it is important to capture the Black Swan events, which stress the system beyond its normal operational environment. As such, explicitly constructed scenarios or those scenarios constructed through the use of probability distributions and past knowledge are less equipped than those models which are capable of seeking the "Worst-Case" system degradation combinations.

Finally, a resilience measure should be at its worst for systems which do not exist, as was determined in the section Is An Empty Network Resilient. As such, the measure should not directly rely on a change in the network or aspects of the network which do not alter the functionality of the network.

Based upon this information, the following resilience measure is proposed.

> *Resilience is measured as the time-averaged expected network performance under extreme-event degradation.*

This proposed resilience measure has a number of implications. The most obvious of these is that it is scenario based. By utilizing a scenario, resilience measures of varying systems must be compared with the same adversary scenario. However, this also allows for gauging the impact of redundancies and mitigation efforts, as well as viewing system behavior when faced with adversity.

This measure also assumes the incorporation of time. Time is a necessary component of resilience, as seen in the components respond, adapt, and recover of the resilience definition. However, multiple events and repetitive degradation can also occur across a wide range of time, further impacting an already degraded system. Time can also be a hindrance though, requiring certain modeling restrictions such as a time horizon.

The probabilistic nature of the system and degrading events should also be considered. A common method for reducing probabilistic curves to a single measure is the expected value. It is well known that reducing a curve to a single value results in a loss of information, but there is no requirement on the remaining information being entirely discarded. Though the resilience measure is reduced to the expected value, the remaining output may be retained for potentially invaluable information and insight.

In this chapter, the threats that communication satellites face were reviewed, as well as the similarities and differences that these satellites have when compared to other well known networks. A number of previous efforts and methods used in measuring resilience were presented with particular emphasis on their strengths, weaknesses, and commonalities. Finally, the chapter was concluded by noting desirable characteristics of a resilience measure based on the DoD and Argonne National Laboratory definitions which may be applied to not only satellite communication networks, but any system in which resilience is measured.

In the following chapter, the many concepts and insights presented here are used to develop a methodology for measuring resilience. The chapter will begin by adapting methods presented in this chapter, working to pull the many strengths of the methods

together. At the end of the chapter, a measure for resilience, and a method for attaining the measure is developed.

III. Methodology

In this chapter, two bi-level models are developed, one of which is used to optimize the effects of an attack over a single network cycle, and another which does the same for a much longer time span. A heuristic based heavily upon Lim and Smith's work is developed to solve the BLPs. This methodology is demonstrated with a case study, which utilizing two variations on an approximated network.

Model Notation

The indices, variables, and parameters utilized in this chapter are presented succinctly in this section, as well as in Appendix A. Note that some parameters are exclusively utilized by the medium/long term model to be presented later in this chapter and may not appear in short term model. In addition, a change of indices occurs with respect to time.

Short-Term Model:	Primary Time is Short Term Interval (STI)
	Secondary Time is always 1
Medium/Long-Term Model:	Primary Time is Long Term Interval (LTI)
	Secondary Time is Short Term Interval (STI)

Indices

i, j, k	All three are to be used to denote the node.
t	Source node of the transmission
l	The network system the transmission last experienced
s	The security level of the transmission
d	Time (Primary)
∂	Time (Secondary)
δt	Duration of a single STI
a	Attack type
e	Active Defense node
r	Orbital Radius {0,1,2}

Variables

$y_{i,j,t,l,s,\partial,d}$	Flow of transmission from node t currently flowing from node i to node j with security level s and node i operating in network system l
$\partial y_{i,t,s,\partial d}$	Amount of unsatisfied demand at node i with security level s
$\text{TCactive}_{\partial,d}$	Binary variable noting whether an active TC station exists at time d, ∂ (1=yes, 0=no)
$\text{Excess\_Cap}_{s,\partial,d}$	Amount of capacity of security level s in the network that remains after all demand has been filled at time d.
$\text{Excess\_Value}_{\partial,d}$	Value of remaining capacity in network at time d
$Excess\_Cap\_PCT_{\partial,d}$	Percent capacity remaining in Network at time d
Total_Net_Value	Total value of the attempted calls in the network
$\varphi_{\text{i,a,g,d}}$	Amount of resources of type g to be used against node i in attack type a $\left[0, RqRsc_{\text{i,a,g}}\right]$
$\omega_{i,a,d}$	Binary Variable denoting an attack of type a against i at time d
$\text{ADFired}_{e,a,d}$	Binary Matrix specifying if defense e, a was used at time d
$x_{i,\partial,d}$	Degradation of node i at LTI time d
$Fuel_i\left(DistPos_i\right)$	Fuel burned by node i given the distance repositioned
$\text{empty}_{i,d}$	Binary variable denoting if node i is out of fuel at time d
$RepoD_i(DistPos_i)$	Time to reposition node i given the distance repositioned
$RepoS_i$	LTI time node i begins its repositioning
$DistPos_i$	Distance repositioned by node i

Parameters:

$b_{i,t,s}$	Supply/Demand at node i from source node t with security level s
$c_{i,t,s}$	Value of transmission to node i
$cap_{i,l,s}$	Capacity of node for network type l and security level s
$x_{i,\partial,d}$	Operational Status of node i at time d,∂
TC_i	Vector denoting if node i is a Tracking and Control station
$Network\_Value$	Sum of network capacity multiplied by value of security rating
ec_s	Value of excess capacity with security level s
$Conn_{i,j,\partial,d}$	Binary parameter denoting the connection between node i and j at time interval d,∂
$costO_i$	Cost of operating node i for 1 time step
$costR_{i,a}$	Cost of reparing node i from attack type a
GlobeMax	Global capacity of network
RegMax	Regional capacity of network
CostMax	Maximum possible cost resulting from a probable event
$AdvRsc_g$	Adversary Resources of type g
$RqRsc_{i,a,g}$	Required Resources of type g to eliminate node i via attack type a
$f_{i,a,d}(\varphi_{i,a,g,d}, ADFired_{e,i,a,d})$	Function computing node i effectiveness at time d based upon amount of resources used and active defenses used
$IntDur$	Duration of each time step
$TimeR_{i,d}(x_{i,a,d})$	Time to repair node i after an attack degrading to $x_{i,a,d}$
$CostR_{i,d}(x_{i,a,d})$	Cost to repair node i after an attack degrading to $x_{i,a,d}$
$TimeRpc_i$	Time to replace node i
$CostRpc_i$	Cost to replace node i
$ST_{e,a}$	Minimum threat of type a worthy of considering defense from e
$ADFP_{e,a}$	Maximum distance defense e can protect against attack type a
$ADRC_{e,a}$	Cost to use defense node e against attack type a
$ADRT_{e,a}$	Time between uses of defense e given it was used to protect against type a
$ADProt_{e,i,a}$	Binary matrix denoting if node i is protected from attack type a by defense node e
$Mobile_i$	Binary variable denoting if node i is mobile
$maxpos_i$	Maximum distance node i may be respositioned to
$minpos_i$	Minimum distance node i may be respositioned to
$maxfuel_{i,d}$	Maximum fuel available to node i at time d
$FuelUseR_i$	Amount of fuel used by node i per time step d
$footprint_{i,r}$	Radius of footprint of node i for orbital radius r
OD	Polar Orbit Direction of orbital plane at some starting time d=0 in regards to a contiguous half sphere of Earth (1 if North, -1 if South)
CN	Number of lateral orbits completed during each cycle
OT	Orbit Time
CT	Cycle Time

Adapting to a Satellite Network

Recall that in Chapter II, two major differences were discussed that differentiated satellite networks from other common distribution networks. In this section, the methodology used to adapt previous methods to a satellite network is presented.

The first difference to discuss is the time-dependent nature that a satellite network exhibits. Because satellites are in constant motion, orbiting at different speeds, in different planes, and travelling in different directions as adjacent orbits, the links with other satellites and with terrestrial areas are brief and temporary. This is especially true with LEO satellites which must move at high velocities to maintain a stable orbit due to their relatively closeproximity to the Earth's surface. While many properties determine when a satellite is able to establish a connection; the main three are line of sight, distance, and rotation of orbital plane.

Line-of-sight is much easier to account for with inter-satellite links (ISLs) than with up/downlinks. Geographical features such as mountains and valleys can have a significant impact on if a link is viable. Terrain can obscure large plots of land from the satellite's view.

The connections between terrestrial nodes and satellites are highly dependent upon distance. Using distance alone, the footprint of a satellite on a smooth surface can be determined. On a perfectly smooth surface, if the terrestrial node falls inside of that footprint, then a connection is established.

As discussed in Chapter II, a snapshot of a satellite network possesses the same stability as a common distribution network, and the location of a satellite can be

estimated when given a specific time. Therefore the model views the satellite network at short, discrete time intervals. Doing so, the average usage for those discrete intervals of time must be used in place of a higher fidelity and more accurate continuous method, resulting in a slight loss of accuracy.

The second major difference between a satellite network and common distribution networks is the need to force transmissions from one specific node to another equally specific node. Reaching back to network flow models, a transmission is forced between a chain of nodes if a node on one end of the chain has a supply, and the node on the other end has a demand. This is why, for one transmission, a satellite network's performance can be optimized the same as any other network.

However, when combining many thousands of transmissions into a single model, the simple transmission can no longer be the single encompassing commodity. The resolution is to not view the network as a large quantity of transmissions moving through the nodes, but rather a large set of transmission products defined by their origin, which can be handled by multi-commodity network flow.

By making a transmission from every point of origin in the network a separate commodity, there is control to where it is sent, maintaining the limitations of the network by retaining nodal capacities. Moreover, this high-fidelity method can be simplified, if needed, by translating the demands and supplies of the lowest level nodes up to the next level in the network, or by aggregating multiple connected origination nodes into a single node, adjusting the capacity of the satellites as necessary.

For example, as with a satellite phone network, the points of origin are every single phone and ground station. This network is massive to the point that any general

model built would likely be intractable. If instead the satellite phones are approximated to grids on Earth, a very large problem is reduced to a smaller one. However, the capacity of the covering satellite needs to be reduced by the number of calls whose origin and destination were within the same grid. Note that this adjustment is a simple preprocessing of values.

Making another step towards improving model performance, those grids may be eliminated and their supply and demand translated up to the satellite covering their respective area. Once again, the capacity of the satellite needs to be reduced by any demand now served from within the covered area. These two short steps can reduce a model size considerably.

Approach

Brown *et al.* state that "For many situations, a linear program will provide an adequate model of the defender's system and its operations" (Brown 2005 p. 106). Over the course of this section, the prescribed method for measuring resilience and all of its components is formulated, including assumptions as they become necessary. This method results in not a singular equation, but rather a step-by-step process that presents multiple equations and objective functions to arrive at the final measures.

Bi-level programs allow for optimizing the solution of directly competing objectives. One general bi-level program is the *max-min* problem (Brown 2005 p. 106). In this model, the defender, interior objective function, is minimizing some measure, and

the attacker, outer objective function, seeks to maximize that minimum (Brown 2005 p. 106).

Recall Salmeron *et al.* define short term as the period between the time of the event and the time to repair a cascading failure. In the model short term is defined as a single network cycle. This provides a clear break for when a short term model may be terminated, as well as setting firm bounds on the model, while still being adaptive to the network. With networks experiencing no cycle time, such as pure GEO networks, this would essentially reduce the short-term model to a single time step.

The next issue is to address cascading failures. Patera states that as the amount of debris in a satellite's orbiting area increases, so does the probability of a collision (Patera p. 716). For the purpose of this research, the following assumptions are made. The first is that the capacities of the satellites cannot be exceeded, thus precluding a cascading overload from usage spikes, the most likely cause for a cascading failure in a power grid.

The second assumption is that the only cascading effects possible, in regards to satellites, is the increased probability of collision with an orbital object, which may be achieved through either the raised quantity of debris, or through the elimination of Tracking and Control stations whose purpose is to perform maneuvers to reduce the probability of collisions.

The basis of this assumption comes from the inherent defenses employed by satellite operators, which prevent a dangerous overload of capacity, as well as allow the satellites to, for a duration, autonomously maintain their orbits. While this begins down the path of reliability and probabilistic effects, it is an important piece to consider as the probability of collision may spike dangerously high as a result of an event. As there may

be collateral damage as the result of an attack, this term is restricted to referring to effects that come as a direct result of the primary attack, which may include time lingering effects, area of effect, or a combination of both.

Unlike power grids, satellites also experience the benefit of being mobile. It would seem ideal to define medium term as the time after the short term has concluded, and until the satellites are able to migrate in order to compensate for the losses, should the operators choose to do so. Unfortunately, this method is not viable due to the multitude of ways in which such a period may be affected by choices made by the operator. Mobilizing a satellite requires the burning of fuel for rapid repositioning, or in some cases the allowance of a slow drift.

Many satellites, especially those in an elliptical orbit such as HEO and LEO satellites, experience a slow longitudinal drift (Kumar p. 719). However, repositioning via this method requires a significant amount of time, and may only apply for relatively close orbital planes. For example, Iridium used a combination of a fuel burn and drift method to position new spares Iridium 90, 94, and 96 in February 2002, and again in June 2002 to reposition Iridium 98. In addition, in 2005, Iridium 98 was maneuvered from one stable orbit to another via a drift-burn combination. The trip, which moved only a third of the circumference of the LEO realm required the time from June 2005 till May 2007 (Sladen).

For a movement from spare to operational orbit burn, another precedence is found in Iridium, after the 2009 Iridium 33-Cosmos collision. Immediately following the collision, efforts began to move the spare, Iridium 91, up to operational range. The transition took approximately one month (Sladen). To this point, Iridium has neither

utilized an orbit to orbit burn nor have they repositioned an operational and in-use satellite, instead choosing to drift spares or to position replacement satellite packages nearest the orbital plane most in need. Therefore, a separate medium term model or definition is not included, but is instead presented as a modification of the long term model.

Salmeron *et al.* define the long term, being the period after the satellites have been repositioned and until the network's performance has returned to a level at or above its previous state, which may include procuring replacements or repairing destroyed nodes. (Salmeron p. 99). While this period seems ideal, if sufficient satellites are eliminated, reconstituting the network to a state equivalent to, or exceeding preceding status could potentially take years, at which point the advancement of technology, as well as the usage of newer architecture models, may create complexities unable to be properly captured by the model or analyst. Moreover, the inherent complications that arise from allowing multiple time-dependent degrading events creates far too much chaos, greatly blurring the lines of periods defined in such a manner.

As such, the long term timeline is limited to include sets of short term periods, starting from a predetermined time and extend up to the point where changes in technology and policy is considered to be significantly great to warrant high unpredictability in network components that may be utilized by the conclusion of the model duration. In extension, the medium term may be considered to be any period of significance greater than one period, but shorter than the duration described as the long term. However, mathematically this results in no significant change to a long term model, allowing us to construct a single medium/long-term model to accommodate both.

While defining these partitionings is important, the true separation between medium and long-term is irrelevant when maintaining a worst-case scenario model. This is because a worst case event may involve recurring disruptions, which blur the split between the two epochs. As such, the medium and long term models remain connected while making the assertion that any long term model must contain a medium term model. If only a medium term model is desired, the model developed later in this chapter for the long term scenario may have the reconstitution component removed and the time duration adjusted accordingly.

Due to restrictions in time and resources, the model is constructed with the definition simplified to:

> *Resilience is measured as the time-averaged ~~expected~~ network performance under extreme-event degradation.*

The measure was simplified in this manner because of the choice to utilize a deterministic modeling approach. As such, the resilience measure which arrives as an output of the yet-to-be developed model tend to be a slight overestimation of the true resilience of the system.

Building The Defender Model

To measure network operating levels, two preexisting measures are employed: the percent of transmissions blocked and the percent of excess capacity. These two are selected because they directly translate to measures which DISA uses to measure network performance, which is confirmed by Eremenko *et al.* in their own cost-benefit analysis (Eremenko p. 4).

In DISA's *Telecommunications Service Level Agreement*, the main measure utilized to specify performance levels is Percent Management Threshold, which states the minimum acceptable level of performance, in their case availability, that must be ensured at all times (DISA 2012 p. 1). In this thesis the means for measuring percent availability when the calls have dropped below 100% connected is to use the *value weighted percent of transmissions blocked*. Since many methods exist for eliciting these weights from decision makers and the determination of weights is outside the scope of this research effort, no further detail is provided in this work as to the many methods for soliciting weights, nor is there any restriction on the weights to be used at this point. Because this measure cannot account for any capacity in excess of all-calls-connected, this measure is coupled with the percent of capacity remaining in the network. For example, if using the level of 90% of the capacity in the network to complete all transmissions, operational or otherwise, then the percent capacity remaining is 10%.

The first step is to calculate the network operating levels for an empty set of failed nodes by constructing the inner Defender IP. The purpose of this step is to determine the network performance of the fully operational state of the network.

$$\underset{y}{Min}\ z=c\bullet\partial y$$

Subject to:

$$a_{j\bullet}y+\partial y=b_j$$

$$Fy \le U(1-x)$$

$$y \ge 0,\ x=\{0,1\}$$

Model 3.1

In this linear programming model, which is formulated as general a fashion as possible, $a_{j\bullet}$ is the j^{th} row of constraints, y is the main decision variable adjusting commodity flow, and ∂y is the total amount of demand not met. Another way of thinking of this demand is the number of calls blocked. Because calls can have varying levels of importance, the cost vector, *c*, is included to denote the varying levels of importance of those calls. Note that if the model is pulled away from the lowest level and aggregated to some higher level, the cost is no longer associated with a transmission from a single node, but now with an area or region depending on how low or high the fidelity is.

In the following model, multi-commodity flows, varying levels of security requirements, a supply/demand Tracking and Control commodity, and the need for a gateway to pass calls between nodes operating incompatible frequencies, procedures, authority, and so forth are included to increase model validity. It shall be assumed that, in an unaltered network configuration, those operating the networks are aware of or are able to estimate, what connections exist at each time step, or are capable of determining these connections on their own, both within the network nodes as well as terrestrial zones. Thus, $Conn_{i,j,d}$ is a binary parameter denoting the connection between node i and j

at time interval d. While in most cases $Conn_{i,j,d}=Conn_{j,i,d}$, this need not always be the case.

Though the equations are much more explicit, a close read will show that the only major change between the following model and the general linear model presented before is the addition of extra subscripts. As such, they shall be defined very carefully.

$$\underset{y}{Min}\ z = \sum_s \sum_t \sum_i c_{i,t,s} \bullet \partial y_{i,t,s} \Big/ Total\_Net\_Value$$

Subject To:

(1) $\sum_j y_{(j,i),t,l,s} \le cap_{i,l,s} \quad \forall i,l,s$

(2) $\sum_l \sum_s \sum_j y_{(j,i),t,l,s} \le cap_i \bullet x_i \quad \forall i$

(3) $-\sum_l \sum_k y_{(t,k),t,l,s} \ge b_{t,t,s} \bullet x_t$ (Supply node t) $\quad \forall t,s$

(4) $\left(\sum_l \left(\sum_j y_{(j,i),t,l,s} - \sum_k y_{(i,k),t,l,s} \right) \right) + \partial y_{i,t,s} = b_{i,t,s} \quad \forall i,t,s$

(5) $cap_i \bullet Conn_{j,i,d} \ge y_{(j,i),t,l,s} \ge 0$ (6) $\partial y_{i,t,s} \ge 0 \quad \forall i,t,s$

(7) $\partial y_{i,t,s} \le ((\sum_i (TC_i \bullet x_i)) - 1) \bullet (cap_i \bullet (1 - TC_t)) \bullet TCactive + (cap_i \bullet (x_i + (1 - TCactive))) \quad \forall i,t,s$

(8) $TCactive \le \sum_i (TC_i \bullet x_i)$

(9) $Excess\_Cap = \sum_i \left(cap_i \bullet x_i - \sum_l \sum_s \sum_j y_{(j,i),t,l,s} \right)$

(10) $Excess\_Cap\_PCT = Excess\_Cap \Big/ Network\_Cap$

(11) $Total\_Net\_Value = \sum_s \sum_t \sum_i c_{i,t,s} \bullet b_{i,t,s}$

(12) $c_{t,t,s} = 0 \quad \forall t,s$

Model 3.2

In the preceding IP, recall that x_i is known, currently a zero-vector, and is implemented to the model in order to determine the network operating level, or the importance of the blocked calls. This is different from a true Attacker-Defender bi-level

model which would select those degradation levels, x_i. Meanwhile, *Total_Net_Value* is a constant and actually has no effect on what the optimal configuration will be. Excluding or including it in the objective function will, computational error notwithstanding, only serve to adjust the final answer by a predetermined scalar. If a node is unable to transmit to or receive from a particular system, or if its security level is insufficient, then $cap_{i,l,s}$ will prevent the flow through that node via constraint 3.2.(1).

Constraint 3.2.(2) ensures that the total capacity of the node is not exceeded. Constraint 3.2.(3) allows the supply the vary, which is necessary when ∂y begins to grow. Note that supply is negative, and the equations are flow-in minus flow-out. Constraint 3.2.(4) is the standard network flow constraint. Constraint 3.2.(7) forces at least 1 Tracking and Control (TC) transmission to be transmitted to every demanding node so long as the node is still active and an active TC station exists. Constraints 3.2.(9) and 3.2.(10) build the second measure, the percent excess capacity in the network. Constraint 3.2.(11) measures the total value of the calls being placed in the network. Note that, as specified in constraint 3.2.(12), the value of supply is 0. This is done so that transmissions are not double counted. This constraint may easily be preprocessed out to conserve memory as the c array is a parameter and not a variable.

$$\underset{y}{Min}\ \ z = \frac{1}{Total\_Net\_Value}\sum_{s}\sum_{t}\sum_{i} c_{i,t,s} \bullet \partial \mathrm{y}_{i,t,s}$$

Subject To:

(1) $\sum_{j} y_{(j,i),t,l,s} \le cap_{i,l,s} \quad \forall i,l,s$

(2) $\sum_{l}\sum_{s}\sum_{j} y_{(j,i),t,l,s} \le cap_i \bullet x_i \quad \forall i$

(3) $-\sum_{l}\sum_{k} y_{(t,k),t,l,s} \ge b_{t,t,s} \bullet \mathrm{x}_t$ (Supply node t) $\forall t,s$

(4) $\left(\sum_{l}\left(\sum_{j} y_{(j,i),t,l,s} - \sum_{k} y_{(i,k),t,l,s}\right)\right) + \partial y_{i,t,s} = b_{i,t,s} \quad \forall i,t,s$

(5) $\mathrm{cap}_i \bullet Conn_{j,i,d} \ge y_{(j,i),t,l,s} \ge 0$ (6) $\partial y_{i,t,s} \ge 0 \quad \forall i,t,s$

(7) $\partial y_{i,t,s} \le ((\sum_{i}(TC_i \bullet x_i)) - 1) \bullet (cap_i * (1 - TC_t)) \bullet TCactive + (cap_i \bullet (x_i + (1 - TCactive))) \quad \forall i,t,s$

(8) $TCactive \le \sum_{i}(TC_i \bullet x_i)$

(9) $Excess\_Cap_s = \sum_{i}\left(cap_i * x_i - \sum_{l}\sum_{j} y_{(j,i),t,l,s}\right)$

(10) $Excess\_Value = \sum_{s} ec_s \bullet Excess\_Cap_s$

(11) $Weighted\_Excess\_Cap\_PCT = Excess\_Value \Big/ Network\_Value$

(12) $Total\_Net\_Value = \sum_{s}\sum_{t}\sum_{i} c_{i,t,s} \bullet b_{i,t,s}$

(13) $c_{t,t,s} = 0 \ \forall t,s$

Model 3.3

The value weighted percent of calls blocked and the value of the percent excess capacity associated with this network state are recorded in an array for later use. These values are used in order to calculate the Attacker objective value that results from the current state of the system.

At the moment, the network is currently modeled as operating under perfect conditions. However, to proceed a baseline value for the impact that each node in the

satellite network possesses is required. To do this, the preceding model is resolved, but with a single "deactivated" node each run.

This is done because, though the method utilized here is to construct a continuous mixed integer non-linear bi-level program (CMINLBLP), currently the only well established method known for solving such problems, large or small, is to utilize heuristics. As discussed in Chapter II, this is due to the non-convexity that all BLP in general are subject to, even when both components are linear. Determining a baseline importance for each node will allow the CMINLBLP heuristic to more rapidly arrive at a solution by establishing initial node preferences for selection.

In this section, the inner Defender model to the CMINLBLP was developed and the values necessary to determine initial preferences to be used in the heuristic were calculated. In the following section, this Defender model is utilized to construct the Attacker-Defender Model.

Extreme Event Attacker-Defender Model

In the previous section, the interior Defender model was developed. In this section, a bi-level Attacker-Defender model is built around that IP in order to determine where the vulnerabilities in the network are and the effects of those weaknesses being exploited.

At the moment, the most probable disturbance to a satellite communications network is a natural or accidental one, as no known malicious attack on a satellite have

been publically recorded to date. However, the Attacker-Defender model employs the aspects of an intelligent adversary in order to determine the worst cases in which a disruption may occur given a fixed level of resources. Another way of viewing this problem is the occurrence of a "perfect storm" against the network with a feasible level of severity.

Recall from the previous section that the final integer programming model was Model 3.3. To apply a bi-level program to this IP, first an objective, or set of objectives, for an adversary must be determined. These objectives should be as simple as possible to reduce complications and specificity, which could result in unintentional restraint on the attacker's actions, but broad enough to be valid. An example of three such objectives which can easily encompass many more specific objectives include:

1. Reduce global capacity
2. Reduce coverage of a particular region of the planet
3. Increase the cost of operating the network.

Each of these objectives is simple, pertain to the network, and are reasonable objectives for an adversary to pursue. To compare these objectives in the model, each of them is divided by their unconstrained optimal level. This means that one must first know what each of the optimal solutions are. Note that it is not necessary to utilize these specific objectives; a different set may be chosen.

Cost has yet to appear in the model in any way; this includes the costs of a satellite, a Tracking and Control (T&C) station, or the cost of a node failing. However, these costs are fundamentally the same as the values placed on transmissions and capacity. If the values of transmissions are converted to units of cost, or vice versa, then

the cost of losing/rebuilding a node is in the same units as the value of transmissions unable to be accomplished.

To proceed, the following assumptions are made. The first is that all costs of operating, procuring, and repairing all nodes are known by both the defender and the attacker. The second assumption is that the time for all necessary procurements and repairs is known by both the defender and the attacker.

The first two assumptions, while not always true in reality, are necessary in order to apply a BLP. A BLP is a deterministic optimization model; thus it is assumed that all information is known with certainty. If these costs and times were permitted to be unknown or highly variable, then a method of modeling other than BLP would need to be employed. With these assumptions in mind, *GlobeMax, RegMax,* and *CostMax,* along with a few cost and resource parameters, are included:

While *GlobeMax and RegMax* may be quickly calculated from pre-existing data, *CostMax* is not as simple. In order to determine this value, first a model with this as an attacker's sole objective must be formulated.

$\underset{x}{Max}\ \mathrm{v} = Cost$

Subject To:

[1]Cost=$\sum_i \left(\left(\sum_a (x_{i,a} \bullet (\mathrm{costR}_{i,a}) \right) + (1 - x_i) \bullet \mathrm{costO}_i \right)$

[2] $x_i = \sum_a x_{i,a} \quad \forall i$

[3] $\sum_i \sum_a RqRsc_{i,a,g} \bullet x_{i,a} \le AdvRsc_g \quad \forall g$

[4]$\underset{y}{Min} \quad z = \frac{1}{Total\_Net\_Value} \sum_s \sum_t \sum_i c_{i,t,s} \bullet \partial \mathrm{y}_{i,t,s}$

Subject To:

(1) $\sum_j y_{(j,i),t,l,s} \le cap_{i,l,s} \quad \forall i,l,s$

(2) $\sum_l \sum_s \sum_j y_{(j,i),t,l,s} \le cap_i \bullet x_i \quad \forall i$

(3) $-\sum_l \sum_k y_{(t,k),t,l,s} \ge b_{t,t,s} \bullet \mathrm{x}_t$ (Supply node t) $\forall t,s$

(4) $\left(\sum_l \left(\sum_j y_{(j,i),t,l,s} - \sum_k y_{(i,k),t,l,s} \right) \right) + \partial y_{i,t,s} = b_{i,t,\mathrm{s}} \quad \forall i,t,s$

(5) $\mathrm{cap}_i \bullet Conn_{j,i,d} \ge y_{(j,i),t,l,s} \ge 0$ (6) $\partial y_{i,t,s} \ge 0 \quad \forall i,t,s$

(7) $\partial y_{i,t,s} \le ((\sum_i (TC_i \bullet x_i)) - 1) \bullet (cap_i \bullet (1 - TC_t)) \bullet TCactive + (cap_i \bullet (x_i + (1 - TCactive))) \quad \forall i,t,s$

(8) $TCactive \le \sum_i (TC_i \bullet x_i)$

(9) $Excess\_Cap_s = \sum_i \left(cap_i \bullet x_i - \sum_l \sum_j y_{(j,i),t,l,s} \right)$

(10) $Excess\_Value = \sum_s ec_s \bullet Excess\_Cap_s$

(11) $Weighted\_Excess\_Cap\_PCT = Excess\_Value \Big/ Network\_Value$

(12) $Total\_Net\_Value = \sum_s \sum_t \sum_i c_{i,t,s} \bullet b_{i,t,s}$

(13) $c_{t,t,s} = 0 \ \forall t,s$

Model 3.4

Model 3.4 is a bi-level optimization program, specifically an Attacker-Defender model. This multileveled optimization can be seen in the overarching maximization objective, as well as the embedded objective at constraint 3.4.[4]. Just as the whole

model has constraints it must satisfy, so too does the inner Defender model. In this scenario, each level of the model controls a different set of decision variables, which are denoted below the Max/Min. The Attacker then has perfect knowledge of the Defender's responses, whereas the Defender may only act in response to the attacker.

Note that, excluding x_i changing from a parameter to an Attacker controlled variable, Model 3.4 required no change to the pre-existing portions. This is because the Attacker's objectives and restrictions do not affect the defender's. In this model, the Attacker's sole objective is to maximize the cost of operating the network, which is calculated in 3.4.[2]. Another simple way of thinking about the degradation, x_i, is the percent of capacity remaining for usable transmissions.

If the Attacker's resources are extremely limited, or if the defender's network is very large, then one may pursue this same method when determining the remaining two objectives, *GlobeMax* and *RegMax*. Doing so reduces the amount of rounding error that may occur when dealing with extremely small decimals.

$$\underset{x}{Max}\ \mathrm{v} = \frac{Cost}{CostMax} + \frac{\sum_i \sum_t \sum_s c_{i,t,s} \bullet \partial \mathrm{y}_{i,t,s}}{GlobeMax} + \sum_{i \in R} \frac{\sum_t \sum_s c_{i,t,s} \bullet \partial \mathrm{y}_{i,t,s}}{\mathrm{RegMax}}$$

Subject To:

$$[1]\mathrm{Cost} = \sum_i \left(\left(\sum_a (x_{i,a} \bullet (\mathrm{costR}_{i,a}) \right) + (1 - x_i) \bullet \mathrm{costO}_i \right)$$

$$[2]\ \mathrm{x}_i = \sum_a x_{i,a} \quad \forall i$$

$$[3]\ \sum_i \sum_a RqRsc_{i,a,g} \bullet x_{i,a} \le AdvRsc_g \quad \forall g$$

$$[4]\underset{y}{Min}\ \ z = \frac{1}{Total\_Net\_Value} \bullet \sum_s \sum_t \sum_i c_{i,t,s} \bullet \partial \mathrm{y}_{i,t,s}$$

Subject To:

(1) $\sum_j y_{(j,i),t,l,s} \le cap_{i,l,s} \quad \forall i,l,s$

(2) $\sum_l \sum_s \sum_j y_{(j,i),t,l,s} \le cap_i \bullet x_i \quad \forall i$

(3) $-\sum_l \sum_k y_{(t,k),t,l,s} \ge b_{t,t,s} \bullet \mathrm{x}_t$ (Supply node t) $\forall t,s$

(4) $\left(\sum_l \left(\sum_j y_{(j,i),t,l,s} - \sum_k y_{(i,k),t,l,s} \right) \right) + \partial y_{i,t,s} = b_{i,t,\mathrm{s}} \quad \forall i,t,s$

(5) $\mathrm{cap}_i \bullet Conn_{j,i,d} \ge y_{(j,i),t,l,s} \ge 0$ (6) $\partial y_{i,t,s} \ge 0 \quad \forall i,t,s$

(7) $\partial y_{i,t,s} \le ((\sum_i (TC_i \bullet x_i)) - 1) \bullet (cap_i \bullet (1 - TC_t)) \bullet TCactive + (cap_i \bullet (x_i + (1 - TCactive))) \quad \forall i,t,s$

(8) $TCactive \le \sum_i (TC_i \bullet x_i)$

(9) $Excess\_Cap_s = \sum_i \left(cap_i \bullet x_i - \sum_l \sum_j y_{(j,i),t,l,s} \right)$

(10) $Excess\_Value = \sum_s ec_s \bullet Excess\_Cap_s$

(11) $Weighted\_Excess\_Cap\_PCT = Excess\_Value \Big/ Network\_Value$

(12) $Total\_Net\_Value = \sum_s \sum_t \sum_i c_{i,t,s} \bullet b_{i,t,s}$

(13) $c_{t,t,s} = 0 \ \forall t,s$

Model 3.5

Currently, the model is still restricted to the same base assumption from Brown *et al*., where every attack is fully successful and there is no option for partial degradation.

However, this assumption is becoming less and less valid as advances are made in the cyber realm, which holds the potential for limiting usage without entirely eliminating a node. Therefore, the binary requirement of $x_{i,a}$ is now relaxed so as to permit the variable to take on a continuous bounded range.

Let $x_{i,a}=f_{i,a}(\varphi_{i,a,g}),\ \varphi_{i,a,g} \epsilon\ [0,RqRsc_{i,a,g}]$, be a function such that $f_{i,a}[0,RqRsc_{i,a,g}]$ is a monotonic decreasing continuous function covering [0,1]. This function must be decreasing simply because $x=1$ has been selected in this work as the node being operational and $x=0$ as the node being eliminated. If the opposite is selected, then along with a few constraint modifications, the function must be made increasing.

$RqRsc_{i,a,g}$ specifies the resources needed to fully eliminate a node in the network, though less may be used to achieve a partially degraded result.

$$\underset{x}{Max}\ \mathrm{v} = \frac{Cost}{CostMax} + \frac{\sum_i \sum_t \sum_s c_{i,t,s} \bullet \partial \mathrm{y}_{i,t,s}}{GlobeMax} + \sum_{i \in R} \frac{\sum_t \sum_s c_{i,t,s} \bullet \partial \mathrm{y}_{i,t,s}}{\mathrm{RegMax}}$$

Subject To:

$$[1]\mathrm{Cost} = \sum_i \left(\left(\sum_a (x_{i,a} \bullet (\mathrm{costR}_{i,a}) \right) + (1 - x_i) \bullet \mathrm{costO}_i \right)$$

$$[2]\ \mathrm{x}_i = \sum_a x_{i,a} \quad \forall i$$

$$[3]\ \sum_i \sum_a RqRsc_{i,a,g} \bullet x_{i,a} \le AdvRsc_g \quad \forall g$$

$$[4]\ \mathrm{x}_{i,a} = f_{i,a}(\varphi_{i,a,g}) \quad \forall i,a$$

$$[5]\underset{y}{Min}\ \ z = \frac{1}{Total\_Net\_Value} \bullet \sum_s \sum_t \sum_i c_{i,t,s} \bullet \partial \mathrm{y}_{i,t,s}$$

Subject To:

(1) $\sum_j y_{(j,i),t,l,s} \le cap_{i,l,s} \quad \forall i,l,s$

(2) $\sum_l \sum_s \sum_j y_{(j,i),t,l,s} \le cap_i \bullet x_i \quad \forall i$

(3) $-\sum_l \sum_k y_{(t,k),t,l,s} \ge b_{t,t,s} \bullet \mathrm{x}_t$ (Supply node t) $\forall t,s$

(4) $\left(\sum_l \left(\sum_j y_{(j,i),t,l,s} - \sum_k y_{(i,k),t,l,s} \right) \right) + \partial y_{i,t,s} = b_{i,t,\mathrm{s}} \quad \forall i,t,s$

(5) $\mathrm{cap}_i \bullet Conn_{j,i,d} \ge y_{(j,i),t,l,s} \ge 0$ (6) $\partial y_{i,t,s} \ge 0 \quad \forall i,t,s$

(7) $\partial y_{i,t,s} \le ((\sum_i (TC_i \bullet x_i)) - 1) \bullet (cap_i \bullet (1 - TC_t)) \bullet TCactive + (cap_i \bullet (x_i + (1 - TCactive))) \quad \forall i,t,s$

(8) $TCactive \le \sum_i (TC_i \bullet x_i)$

(9) $Excess\_Cap_s = \sum_i \left(cap_i \bullet x_i - \sum_l \sum_j y_{(j,i),t,l,s} \right)$

(10) $Excess\_Value = \sum_s ec_s \bullet Excess\_Cap_s$

(11) $Weighted\_Excess\_Cap\_PCT = Excess\_Value \Big/ Network\_Value$

(12) $Total\_Net\_Value = \sum_s \sum_t \sum_i c_{i,t,s} \bullet b_{i,t,s}$

(13) $c_{t,t,s} = 0 \ \forall t,s$

Model 3.6

The basic bi-level model associated with an extreme event under the Brown *et al.* assumptions of a single coordinated, fully effective strike with one attacker objective was presented. Multiple possible attacker objectives were also introduced, and the model was formatted for a generalized function to scale the impact of a partial attack.

Including the Time Component

In this section, integer programming based scheduling methods is incorporated into Model 3.6 so that it may select not only where and how to attack, but also when. By including this time component, the model is now capable of degrading a node multiple times, if desired.

With respect to introducing a time component, one approach would be to first view the problem as a scheduling problem. In essence, the model attempts to schedule when and where the notional adversary strikes. Along with allowing the attacker to choose when to strike, the model must allow the defender to respond.

The most basic response is repair. To begin, the following assumptions are made:

1) The duration and cost of repair is independent of both time and the state of other nodes in the network.
2) If repair is possible, then along with the associated costs, there will also be a necessary time component over which the repairs occur.

These assumptions are for simplification of parameters, though cost and time-to-repair as a function of the current time period, which is known, would still be valid as

long as appropriate minor alterations to the deterministic model are made. To accomodate these time-independent repair functions, two functions must be formulated.

Let $CostR_{i,d}(x_{i,a,d})$ be a one-to-one continuous decreasing function with $CostR_{i,d}(1)=0$ and $CostR_{i,d}(0)=CostRpc_i$ where $CostRpc_i$ is the cost of replacing node i. Let $TimeR_{i,d}(x_{i,a,d})$ be a one-to-one continuous decreasing function with $TimeR_{i,d}(1)=0$ and $TimeR_{i,d}(0)=TimeRpc_i$ where $TimeRpc_{i,d}$ is the time required to replace node i. Once again, these functions must be decreasing because of the specified $x_{i,a}$ bounded [0,1] with 1 being operational. The function $TimeR_{i,d}(x_{i,a,d})$ is guaranteed to be positive, because $TimeR_{i,d}(x_{i,a,d})$ is decreasing, continuous, and $TimeR_i(max(x_{i,a,d})) = min(TimeR_{i,d}(x_{i,a,d})) = TimeR_{i,d}(1)=0$. $CostR_{i,d}(x_{i,a,d})$ is also guaranteed positive for similar reasons.

Next, the model must be adjusted to allow for attack time selection. To do so, the model must regrettably include more variables and parameters. Let *IntDur* be the constant stating the duration of each time interval in the same units used in *TimeR*. Let *D* be a sufficiently large number restricting the number of intervals the adversaries may strike within. *D* may be set to any integer such that:

$$D \geq \frac{1}{IntDur} \max\{\sum_i \sum_a TimeR_{i,a}(f_{i,a,d}(\varphi_{i,a,g})) \mid \sum_i \sum_a \varphi_{i,a,g} \leq AdvRsc_g\} \quad \text{(EQ 3.1)}$$

to ensure sufficient size, however it is recommended that the model be permitted to run for a longer duration.

Restricting *D* as set forth in EQ 3.1 permits the full time for longest recovery period possible in the model. While this does permit the calculation of network performance over time, in many cases, namely where the repair times of nodes are relatively similar, such a small D could preclude the re-degrading of nodes. In such a

case, there is no significant difference between this model and the one developed by Brown *et al*, which permits only a single event.

Next, create a binary variable matrix which will specify when a node is available for attack and when it is not. Assume that a node will only be interdicted via one method at any given time. This assumption, while potentially invalid in reality, is used to simplify the model and reduce the dimensionality, as well as allowing the occurrence of an event to act as a trigger in later constraints. Let *Tgt* be a *n*x*D* matrix where n is the number of nodes in the network. For every $\omega_{i,a,d} \in Tgt$, if $\omega_{i,a,d}$, node *i* at time *d* by method *a*, is targetable then $\omega_{i,a,d}=1$, otherwise $\omega_{i,a,d}=0$.

Unless otherwise specified, each node is targetable at the initialization of the model. For any node i that is deemed non-targetable at any given time, include a constraint in the model to the effect of $\sum_{a}\sum_{d}\omega_{i,a,d}=0$. If a node is only non-targetable for a specific time interval or set of time intervals, *NT*, by attack type, *a*, then include constraints $\sum_{a,d\in NT}\omega_{i,a,d}=0$. For example, if an anti-satellite missile has a finite footprint in which it may strike, and its firing location if fixed, then any satellite i may only be attacked by weapon *a* at time d if satellite i is within weapon a's footprint.

The time subscript, d, is now included on variable $\varphi_{i,a,g}$, making this variable $\varphi_{i,a,g,d}$ so the adversary may select when to use resources. In extension, all associated variables will require the same addition. Using this, the following pseudo-constraints are included:

$$(1)\sum_{g}\varphi_{\mathrm{i,a,g,d}} \le M\omega_{i,a,d}$$

$$(2)\ \text{If } \omega_{i,a,d'} = 1, \text{ then } \forall\ d \in (d', d' + \frac{1}{IntDur} TimeR_{i,d}(x_{i,a,d'})], \omega_{i,a,d} = 0$$

Model 3.7

Constraint 3.7.(2) restricts the targeting matrix Tgt by forcing some of its elements to be non-targetable. It is inside this constraint where the computer is prevented from continuously targeting the same node for an infinitesimally small value infinitely many times. Unfortunately, 3.7.(2) is not in a form that may be confidently solved by most deterministic solver programs. This is due to its if-then formatting, which is currently only a pseudo constraint format. These pseudo-constraints may be reformatted equivalently as:

$$(1)\sum_{g}\varphi_{\mathrm{i,a,g,d}} \le M\omega_{i,a,d}\ \ \forall i,a,d$$

$$(2)\ \sum_{d'=d}^{d+\bar{d}}\sum_{a}\omega_{i,a,d'} = 1 + M(1-\phi_i)$$

$$(3)\phi_{i,d} \le TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,g,d}))$$

$$(4)\bar{d} = \left\lceil \frac{TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,g,d}))}{IntDur} \right\rceil$$

Model 3.8

Note that constraint 3.8.(2) allows for only one attack on a single node for the duration of its degradation by taking advantage of the way in which $TimeR_{i,d}$ was constructed. When a node experiences a repair time, or is destroyed as such an event was built into the repair time, then the node becomes non-targetable for $TimeR_{i,d}$ time units. However, if $TimeR_{i,d}$=0, then no further constraints are placed on the network. Constraint

3.8.(3) is a switching constraint that triggers the binary variable ϕ, which is necessary in constraint 3.8.(2).

Finally, the model must allow immediate reactionary methods the defender may utilize to mitigate the event, such as interception of conventional weaponry. These methods is referred to as Active Defenses, not to be confused with Passive Defenses such as structural hardening or on-board antivirus software. These defenses will follow much the same principals as the other nodes in the network in that they will have associated costs, repair or reload times, and a footprint in which they may operate. If the defense is global, then the footprint is considered to be of sufficient size to encompass the planet.

Let index *e* denote an individual Active Defense node. Let $ADFP_{e,a}$ denote the maximum distance in which the defense e is effective against attack type a. Let $ADRC_{e,a}$ denote the cost to operate and reload the defense. In the case of a physical defense, this cost may refer to the ammunition while in a cyber event this cost may be the cost of personnel and resources dedicated to combating the threat. Let $ADRT_{e,a}$ denote the time required to reload the defense e, or the time required for the defense e to have been effective. In a physical scenario, this time would be associated with reloading and retargeting. In a cyber event, this would be the time required for the threat to be eliminated.

Let $ST_{e,a}$ be a value denoting a minimum quantity of utilized resources of attack type *a* below which a threat is considered insufficient to allow the model to utilize Active Defense *e*. If there is no such minimum and every threat is sufficient in magnitude, then let $ST=1$. Lastly, let $ADFired_{e,a,d}$ be a binary matrix specifying if Active Defense node e was used against a threat type a during time step *d*.

In words, the following constraints are constructed. If node *i* is within distance $ADFP_{e,a}$ of node *e* during an attack of type *a*, then allow protection of node *i* at the cost of $ADRC_{e,a}$. If node e is used against attack of type *a* in time step *d*, then force $ADFired_{e,a,d}=1$. If the sum of $ADFired_{e,i,a,d}$ over all nodes *i* is 1, then force $ADFired_{e,i,a,d'}=0$ for all *i* and all *d'* in the set $[d+1,d+ADRT_{e,a}]$.

Notice that much of these constraints are similar to when the model's attack constraints were formulated. Using some initial distance calculations, it would be prudent to now calculate the distances from nodes to active defenses. For each node, active defense combination, let $ADProt_{e,i,a}$ be a binary matrix denoting if node I is protected by node e against attack type a. These operational requirements and behaviors lead to the following constraints:

$$f(\varphi_{i,a,g,d}) \rightarrow f(\varphi_{i,a,g,d}, ADFired_{e,i,a,d})$$

$$ST_{e,a} * ADFired_{e,i,a,d} \leq f(\varphi_{i,a,g,d}, 0) * \mathrm{ADProt}_{e,i,a}$$

$$\sum_{d'=d}^{d+\hat{d}} \sum_{i} ADFired_{e,i,a,d} \leq 1$$

$$\hat{d} = \left\lceil \frac{ADRT_{e,a}}{IntDur} \right\rceil$$

With this last addition and modification of constraints in place, along with the addition of usage costs to the before defined Cost variable, the following bi-level model is formulated.

$$\underset{\varphi}{Max}\ \mathrm{v}=\frac{Cost}{CostMax}+\frac{\sum_i\sum_t\sum_s c_{i,t,s}\bullet\partial \mathrm{y}_{i,t,s}}{GlobeMax}+\sum_{i\in R}\frac{\sum_t\sum_s c_{i,t,s}\bullet\partial \mathrm{y}_{i,t,s}}{\mathrm{RegMax}}$$

Subject To:

$$[1]\mathrm{Cost}=\sum_i\left(\left(\sum_a(x_{i,a}\bullet(\mathrm{costR}_{i,a})\right)+(1-x_i)\bullet\mathrm{costO}_i\right)+\sum_i\sum_d CostR_{i,d}(x_{i,a,d})$$
$$+\sum_e\sum_a ADRC_{e,a}\left(\sum_d ADFired_{e,a}\right)$$

$$[2]\ \mathrm{x}_{i,d}=\sum_a x_{i,a,d}\quad\forall i$$

$$[3]\ \sum_i\sum_a\varphi_{i,a,\mathrm{g}}\le AdvRsc_g\quad\forall g$$

$$[4]\ \mathrm{x}_{i,a,d}=f_{i,a,d}(\varphi_{i,a,g,d},ADFired_{e,i,a,d})\quad\forall i,a,d$$

$$[5]\ \sum_g\varphi_{\mathrm{i,a,g,d}}\le M\omega_{i,a,d}\quad\forall i,a,d$$

$$[6]\ \sum_{d'=d}^{d+\bar{d}}\sum_a\omega_{i,a,d'}=1+M(1-\phi_i)\quad\forall i,d,\ \text{where}\ \bar{d}=\left\lceil\frac{TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,g,d}))}{IntDur}\right\rceil$$

$$[7]\ \phi_{i,d}\le TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,g,d}))\quad\forall i,d$$

$$[8]\ \underset{y}{Min}\ \ z=\frac{1}{Total\_Net\_Value}\sum_d\sum_s\sum_t\sum_i c_{i,t,s}\bullet\partial \mathrm{y}_{i,t,s,d}$$

Subject To:

(1) $\sum_j y_{(j,i),t,l,s,d}\le cap_{i,l,s}\quad\forall i,l,s,d$

(2) $\sum_l\sum_s\sum_j y_{(j,i),t,l,s,d}\le cap_i\bullet x_{i,d}\quad\forall i,d$

(3) $-\sum_l\sum_k y_{(t,k),t,l,s,d}\ge b_{t,t,s,d}\bullet\mathrm{x}_{t,d}$ (Supply node t) $\forall t,s,d$

(4) $\left(\sum_l\left(\sum_j y_{(j,i),t,l,s,d}-\sum_k y_{(i,k),t,l,s,d}\right)\right)+\partial y_{i,t,s,d}=b_{i,t,\mathrm{s},d}\quad\forall i,t,s,d$

(5) $\mathrm{cap}_i\bullet Conn_{j,i,d}\ge y_{(j,i),t,l,s,d}\ge 0$ (6) $\partial y_{i,t,s,d}\ge 0\ \forall i,t,s,d$

(7) $\partial y_{i,t,s,d}\le((\sum_i(TC_i\bullet x_{i,d}))-1)\bullet(cap_i\bullet(1-TC_t))\bullet TCactive_d+(cap_i\bullet(x_i+(1-TCactive_d)))\ \forall i,t,s,d$

(8) $TCactive_d\le\sum_i(TC_i\bullet x_{i,d})$

(9) $Excess\_Cap_{s,d}=\sum_{i,d}\left(cap_i\bullet x_{i,d}-\sum_l\sum_j y_{(j,i),t,l,s,d}\right)\ \forall s,\mathrm{d}$

(10) $\mathrm{x}_{i,d}\le 1\ \forall\mathrm{i},\mathrm{d}$

(10) $Excess\_Value_d = \sum_s ec_s \bullet Excess\_Cap_{s,d}$

(11) $Weighted\_Excess\_Cap\_PCT = \sum_d Excess\_Value_d \Big/ Network\_Value$

(12) $Total\_Net\_Value = \sum_d \sum_s \sum_t \sum_i c_{i,t,s} \bullet b_{i,t,s,d}$

(13) $c_{t,t,s} = 0 \ \forall t, s$

(14) $ST_{e,a} \bullet ADFired_{e,i,a,d} \leq f(\varphi_{i,a,g,d}, 0) \bullet \text{ADProt}_{e,i,a,d} \quad \forall e, i, a, d$

(15) $\sum_{d'=d}^{d+\hat{d}} \sum_i ADFired_{e,i,a,d} \leq 1 \quad \forall e, a, d$ where $\hat{d} = \frac{ADRT_{e,a}}{IntDur}$

Model 3.9

In this section, an Attacker-Defender model, Model 3.9, was developed with which to model an extreme event intelligent adversary strike on a satellite infrastructure. This model is constructed to be operated for a single orbital cycle, and as such currently only exhibits the "short-term" scenario. In the following section, adaptations is developed to account for "medium-term" and "long-term" recovery options.

Increasing the Time-Span

In this section the preceding Attacker-Defender model, Model 3.10, is adapted for "medium-term" and "long-term" recovery options. Major adaptations are the inclusion of spares and reconstitution.

In the preceding Model 3.10, no indication was made as to whether a node was actually operating, only if it could operate. However, if on-orbit satellite spares or terrestrial station spares exist, then they will not be operating until after an event occurs, though they are able to operate at any given point there-after. When investigating

adaptations to be made, the model must be able to differentiate between a terrestrial spare or an orbital spare, the main difference of which is the mobility of a satellite and its ability to be repositioned in order to eliminate major gaps in coverage.

First, a distinction must be made between the time-intervals involved in the "short-term" and the "medium/long-term". Recall that the "short-term" time intervals (STI) constructed were partitioning on the time to complete a single cycle. However, as shown in the 2009 Iridium-Cosmos collision recovery (Sladen), a satellite may require many cycles in order to be repositioned. Depending upon personnel issues, a terrestrial node may also require multiple cycles before being activated, though it is desired that the number of cycles for such is relatively low.

As such, the "medium/long-term" time interval (LTI) is defined as a multiple of the network cycle, with the assumption that some initial event must occur within the first cycle, but that later cycles may also experience degrading events. This assumption makes no actual change to the freedom of the model, but instead reduces the dimensionality of it by restricting the amount of memory utilized at the start of the model. The model also requires an upper bound on the number of cycles, one which is sufficient enough to capture the reconstitution of the network, but not so long that complicating network changes become overwhelming. A common phrase associated with this limit is a Planning Horizon.

In a perfect world with unlimited memory and computational power, one would now increase the dimensionality of the model to incorporate these longer time-steps, breaking each LTI into STI. Unfortunately, the model is already cumbersome and increasing the dimensionality potentially thousands of times is impractical as well a

potentially intractable. It is for this reason that the method must differentiate between the two very distinct types of objectives, short-term and medium/long-term. The short-term model is already constructed, and unless a replacement may be included in a single cycle, no alterations need be made.

For the long-term model, the nodes of a network must be considered as falling within two distinct categories, stationary footprint or cyclical footprint. These categories require the assumption that long duration erratic satellite movements do not occur in the network. Erratic movements, which are corrected via on-board computers or Tracking and Control stations, are probabilistic in nature. In essence, this assumption approximates those probabilistic positional variations to the expected path of a satellite, which, due to correctional movements, remain on the projected assumed paths.

One must account for the flow of transmissions in the network, no longer viewing the nodes at discrete time steps as points, but instead as the cyclical paths they traverse over a single LTI. As always, when aggregating constraints in order to conserve memory, an amount of fidelity and precision is lost.

To tackle this problem, a common LEO orbit type is examined, the polar orbit. Polar LEO satellites travel from the south pole to the north pole, their latitude independent of the rotation of Earth. Because of this, an orbit of satellites may visit the entire planet one or more times over the period of a single cycle without ever significantly breaking their cyclical path. What this means is that it is no longer possible to examine specific transmissions and their flow through the network nodes over the model's new time periods.

As Pratt *et al.* point out in their review of the Iridium constellation, the discrete loss of a single satellite operating in a cyclical nature exhibits a time loss of coverage for an area, for each area it covers in its cycle (Pratt p. 5). For a continuous case, the time loss does not accurately capture the effect of partial coverage, and it also fails to align with the network performance measures. The model will utilize the time-dependent user rates and will adjust the number of successful transmissions by the degraded operating capacity of the satellite. Note that through pre-processing, if only the number of transmissions for an area in a cycle is available, and is not time-depended, then the transmissions may be approximated based upon a probability distribution of operating times.

An additional sequence of preprocessing is pursued in which the number of transmissions attempting to be serviced by the cyclical satellite are summed. Continuing with *d* denoting the time steps, though now for LTI, $0 \leq x_{i,d} \leq 1$ denoting the degradation level of node *i* at time step *d*, $c_{i,t,s,d}$ the cost or relative importance of transmission from node t to node *i* with security rating s at LTI time step *d*, and $b_{i,t,s,d}$ denoting the number of attempted calls from node *i* to node *t* of security rating *s* during LTI *d*.

$$\max \sum_{i,t,s,d} \left(c_{i,t,s,d} * b_{i,t,s,d} - c_{i,t,s,d} \bullet \partial y_{i,t,s,d} \right)$$

Such That

$$\sum_{s} \sum_{t} b_{i,t,s,d} - \partial y_{i,t,s,d} \leq \text{cap}_i \bullet x_{i,d}$$

$$\sum_{t} \left(c_{i,t,s,d} \bullet b_{i,t,s,d} - c_{i,t,s,d} \bullet \partial y_{i,t,s,d} \right) \leq \text{cap}_{i,s} \bullet x_{i,d}$$

Model 3.10

This model may be restructured for equivalent results, while requiring fewer calculations and slightly faster running time as follows:

$$\min \sum_{i,t,s,d} \left(c_{i,t,s,d} \bullet \partial y_{i,t,s,d} \right)$$

Such That

$$\sum_{s} \sum_{t} b_{i,t,s,d} - \partial y_{i,t,s,d} \leq \text{cap}_i \bullet x_{i,d}$$

$$\sum_{t} \left(c_{i,t,s,d} \bullet b_{i,t,s,d} - c_{i,t,s,d} \bullet \partial y_{i,t,s,d} \right) \leq \text{cap}_{i,s} \bullet x_{i,d}$$

Model 3.11

This linear program simply allows the model to choose the most beneficial or most important transmissions and process those based upon priority. Using this information, which is completed before the actual model long-term model has been executed, a great deal of information may be discerned. Note that though $x_{i,d}$, the degradation level of node *i* at time step *d*, is included in this model, there has been no allowance for any attacks. This means that only the baseline degradation of the node, that which is already in place or expected to be in place during the time phases that these calculations are used for, is used in the preceding linear program.

In order to account for a spare being mobile or immobile, with respect to its place in the network, the model requires yet another variable with which to discriminate between operating nodes and stand-by nodes, or spares. Let $Mobile_i$ be a variable specifying the maximum distance a node *i* may traverse for repositioning. The time required for a short term time interval is denoted as STI, and a long term time interval as LTI.

In order to limit the inherent complexities in calculating the distance actually traversed by an orbiting spare to a new location in the orbit, or a new orbital plane entirely, the distance traversed is reduced to simply the great circle distance between its starting location and its desired final location at the same STI, though not the same LTI. While there are very few theoretical limitations on how far a satellite may move, especially as the satellite may utilize drift to alter position, operationally it would not be unreasonable to restrict the repositioning of a satellite to within *n* orbital planes to regain service faster, especially following an event on the satellite network. If the operator has no such preference or limitation, then the fuel available and predicted fuel consumption rates may be utilized to set such a bound.

Let $State_i$ be an integer variable of the set {0,1,2,3} such that 0 specifies permanently destroyed, 1 specifies temporarily disabled, 2 specifies operating and in use, either degraded or fully operational, and 3 specifies an operational spare, also either degraded or fully operational. Furthermore, let $MobileRFB_i$ be the fuel consumption per distance unit of moving node i via a burn. Let $MobileIFB_i$ be an initial fuel consumption of moving node i via a burn. Let $MobileTB_i$ be the time required per unit distance of moving node i via a burn.

Similarly, let $MobileTD_i$ be the time required per unit distance of moving node i via a drift. Let $MobileIFD_i$ be an initial fuel consumption of moving node i via a drift, or in other words to raise or lower the node to an engineering altitude to allow for the drift, if necessary. The model must account for a node switching from state 1 to state 0, state 2 to state 1, state 1 to state 2, and state 3 to state 1. Any node, *i,* currently in states 2, or 3 must be permitted to travel up to $Mobile_i$. Furthermore, any mobile node must be

allowed to utilize a combination of burn and drift, thus resulting in a trade off for fuel use and time operationally active.

Based on precedence from historical satellite movements, Iridium satellites currently being repositioned, either from a spare state to operational state, or from one operational location to another, is unable to be utilized operationally for the duration of the transition (Sladen Website). In such a scenario, it is unnecessary to account for a node moving directly from state 3 to state 2, or vice versa. All satellites in such a network moving from 2 to 3 or 3 to 2 must pass through 1 first. However, if this is not the case, or rather, if the satellite may continue its operational duties during the duration of the repositioning, then no such requirement exists. As a result, no state switching would need to occur.

The imbedding of a time-dependent covering problem into the defender model greatly increases the complexity of the problem. The imbedded covering problem is determining the optimal location to place satellites which will utilize the least fuel and permit the greatest coverage.

One method of considering fuel consumption is the quantity of units of fuel used. In this manner, the closest satellite is always the best option. Another way is to instead view the consumption as percent of remaining fuel used, in which a close satellite with the largest fuel remaining is more likely to be selected. Both are viable options, especially as the efforts in robotic refueling progress, and as both can be directly translated to different perspectives of life-time remaining.

By viewing the amount of fuel utilized, and knowing how much fuel the satellite requires over a period of time, it is relatively simple to convert from fuel used to instead

life-time used. Similarly, the percent of remaining fuel used may be converted to the percent of remaining life used, or the percent reduction in life. As both are equally valid methods, the Chapter IV case study, as well as the following model formulations, will proceed with using the raw quantity of life-time consumed which is converted into an integer number of LTI. This decision may be changed with minor alterations to the model and associated equations.

It is a requirement for all orbiting satellites to have their operating frequency and orbit, or in the case of GEO satellites their geospatial coordinates, internationally registered to reduce the probability of a collision. However, no requirement could be located as to exactly where these positions or orbits must be so long as the position is not currently occupied and the satellite's residence is approved. As such this problem is continuous.

One common method utilized for solving continuous covering problems is to establish a multi-parameter grid in which the parameters are continuous and the resulting area of influence is a function of those parameters. In a spherical grid, the most taxing problem is the need for the boundaries of the grid to be equivalent, even if the values when approaching from opposite directions is extreme. To overcome the consideration the two parameter polar coordinate system is utilized. The polar coordinate system is most commonly thought of as three parameters, however with the radius of the satellites being relatively constant based upon orbit type, (terrestrial, LEO, MEO, GEO), the radius is fixed as a property of the node in question (Wright p. 30).

Standard notation for polar equations are $f(\varphi,\theta)$, however those symbols have already been used in the formulation. Therefore, $f(\Phi,\Theta)$, $0< \Phi <\pi$ for latitudinal and $0< \Theta <2\pi$ for longitudinal placement is utilized.

As the transmissions being attempted are based upon the terrestrial and nodal components and not specifically related to the positioning of the nodes, the main tribulation is to determine what nodes fall within the satellite's footprint and at what time, and how many satellites or terrestrial components can serve the demand of that node at each respective time period. By knowing those, one can accurately determine how many transmissions must enter the engineered portion of the network through that node at that time period.

Determining what areas a satellite may serve at any given time is dependent upon the time-dependent location of the satellite and the footprint of that satellite in that location. While it is well known that the terrain affects the size and shape of the footprint, the simplification that the footprint suffers no obstruction is made in this thesis. While this assumption may not be valid for high fidelity models, it does provide a suitable approximation. Future efforts may be employed to increase the fidelity of the footprint in this analysis, if required.

Next it is necessary to determine how many network components capable of transmitting the signal are in range of a given terrestrial area. If the preceding calculations are already completed, then this step is trivial. However, what is not trivial is determining a function which will complete all of these calculations.

As the reader will recall, a linear model, Model 3.6, was utilized to determine how transmissions would flow throughout the network by taking advantage of the operator's

assumed knowledge of their own network. In this thesis' case study, an approximation of the time-dependent connections is assumed known or determinable via knowledge of footprint size and node location. The calculations is completed by knowing first where a satellite is at a given time, the orbit time, *OT*, orbit direction, *OD*, cycle time, *CT*, and the number of lateral orbits completed during each cycle, *CN*.

To do this, it is necessary to imbed the short-term time steps in the long-term model, denoted as subscript δ, and the duration of each short-term time step as δt. This blends the covering model with the long term model. The following constraints will use the starting locations of the nodes as the decision variables, with the remaining orbital parameters acting as discrete functions.

$$
\begin{aligned}
&1)\ \mathrm{POS}_{i,\partial,0} = \Phi_i + OD * \frac{\partial * \delta t}{OT} \quad \forall \mathrm{i},\partial \\
&2)\ \mathrm{POS}_{i,\delta,1} = \Theta_i + \frac{\partial * \delta t * CN}{CT} \quad \forall \mathrm{i},\partial \\
&3)\ \ \mathrm{dist}_{i,j,\partial} \geq (1 - Conn_{i,j,\partial}) * \mathrm{footprint}_i \quad \forall \mathrm{i,j},\partial \\
&4)\ \ \mathrm{dist}_{i,j,\partial} \geq (1 - Conn_{i,j,\partial}) * \mathrm{footprint}_j \quad \forall \mathrm{i,j},\partial
\end{aligned}
\qquad \text{(EQ 3.2)}
$$

Here $\mathrm{POS}_{i,\partial,0}$ is the North-South position of node i at time ∂, $\mathrm{POS}_{i,\partial,1}$ is the East-West position of node i at time ∂, $dist_{i,\mathrm{j}}$ is the distance from node *i* to node *j* at time ∂, $footprint_i$ is the radius of the footprint of node *i*, and $Conn_{i,\mathrm{j},\delta}$ is a binary variable equaling 1 if a connection can exist between nodes *i* and *j* at time ∂ and 0 otherwise. In the preceding constraints, the distance from node *i* to *j* must be less than both footprints in order for a connection to exist. However, it is not required that, if in range, a connection be established. If the connection is beneficial at the time step, then the model, as a result

of optimizing the objective within the constraints, will activate the connection and utilize it as it determines to be optimal.

The constraints in EQ 3.2 are not yet ready to be integrated into the rest of the model though. At the moment, they are the generic approximation of positioning and connection possibility. To determine an actual position it is necessary to take into account the initial positions of the nodes in the network, specifically the mobile nodes such as satellites. To do this, $\Phi_{i,0}$ and $\Theta_{i,0}$ are indexed as thus to denote the initial positions of the nodes, making their counterparts in the preceding constraints 1 and 2, $\Phi_{i,1}$ and $\Theta_{i,1}$ respectively. If limited or unreliable information is available as to the exact positioning of each node as they currently exist or are expected to exist, then this same method may be employed to estimate the current positioning of the nodes. The constraints in EQ 3.2 were constructed to approximate location based upon the new position of the nodes.

While it is left to the user to determine the best method for computing distances, two limits are certain. The first limit is a maximum distance that the satellite may be repositioned from its current location, $maxpos_i$, and the second is a minimum distance for repositioning, $minpos_i$. The distance travelled by a node is denoted as $DistPos_i$, the fuel consumed in such a move as a function of the distance, but possibly distinct for each node, $Fuel_i(DistPos_i)$, and the accompanying max fuel as $maxfuel_i$.

Another important aspect to repositioning is the LTI when the repositioning occurs. Define $RepoS_i$ to be the decision variable denoting the time step in which the repositioning is initiated. $RepoD_i(DistPos_i)$ is the number of time steps required for repositioning and is a function of the distance being traversed.

There is a clear need for consideration of the user's values. In the $Fuel_i(DistPos_i)$ and $RepoD_i(DistPos_i)$ functions, a single output is determined from a single input, however the satellites may be repositioned through a combination of burns and drifts. This results in a fuel-time tradeoff, as the increased fuel usage results in a faster repositioning and reactivation of service. While one could allow the model to determine the tradeoffs, preprocessing this function would conserve memory and allow the model to run faster. In the case study, $Fuel_i(DistPos_i)$ and $RepoD_i(DistPos_i)$ are set as preprocessed functions of the illusory user's preferences and tradeoff values.

Parameters :

MobileRF(B/D)_i Initial fuel consumption of moving node i via a burn/drift

MobileIF(B/D)_i Initial fuel consumption of moving node i via a burn/drift

MobileT(B/D)_i Time required per unit distance of moving node i via a burn/drift

Functions :

$$RepoD_i\left(DistPos_i\right) = (TO_i)*\left(\text{MobileTD}_i + \text{MobileRTD}_i * DistPos_i\right) + \\ +(1-TO_i)\left(\text{MobileTB}_i + \text{MobileRFB}_i * DistPos_i\right) \quad \text{(EQ 3.3)}$$

$$Fuel_i\left(DistPos_i\right) = (TO_i)*\left(\text{MobileIFD}_i + \text{MobileRFD}_i * DistPos_i\right) + \\ +(1-TO_i)\left(\text{MobileIFB}_i + \text{MobileRFB}_i * DistPos_i\right)$$

In these equations for $Fuel_i(DistPos_i)$ and $RepoD_i(DistPos_i)$, the approximation of linear requirements of fuel used and tradeoffs, TO_i, was used. However, this method will work for any function of distance travelled. The tradeoff is that, while increasingly nonlinear approximations may be more accurate, they are also computationally intractable. One inherent restriction to using this preprocessed function tradeoff method is that the functions themselves must be injective and continuous.

The calculation of distances between nodes, with the specific method chosen left to the user, is denoted as $dist_{i,j,\delta}$, the distance from node i to node j. The subscript, δ, denotes the STI time step in which this distance is being calculated for. Note that all of these connections are symmetric, meaning that $dist_{i,j,\delta}=dist_{j,i,\delta}$.

As a note, though a distance component is provided for use between two terrestrial components, the maximum connection distance between two terrestrial nodes may be irrelevant due to land line connections. Because of the massive land based communications infrastructure of many nations, as well as the submarine communication cables spanning the oceans, and the model only repositions mobile nodes, the vast majority of those connections is considered as constant. However, capacities may be adjusted for the amount of data sent from one terrestrial node to another, both overall and for each security setting.

When compared to the short term model, thus far there are three major adaptations to the medium/long term model. The first is conversion of the STI to LTI and the necessity to force some event to occur within the first LTI. The second adaptation is the method for approximating the value of transmissions, and as a result the expected lost value due to degradation of a node. The third and most important change thus far is allowing the nodes of the model to be repositioned, which resulted in need for a systematic method in which the model may construct an entirely new network after each move.

To begin formulating the medium/long term model, a potentially computationally cumbersome method is utilized, and that is to include every STI inside of each LTI. Along with showing how the number of constraints quickly grows, building the model

this way first aids in the adaptation to a model form capable of being performed much more quickly.

As the reader will notice, the majority of the following model, Model 3.12, is very similar to the short term model. This is because to accurately determine the value of failed transmissions, it is essential to know the optimal routing of transmissions in each new network created by mobilizing a node. The majority of the new constraints begin with constraint 3.12.(17) which restricts the use of a node based upon the current state it is in. Constraints 3.12.(18)-3.12.(21) are utilized solely for the purpose of estimating the new location of nodes after a move occurs.

Constraints 3.12.(22) and 3.12.(23) calculate the distance from node i to node j at STI ∂ during LTI d. The model implements the Haversine formula for great circle distance around a sphere. However, as was specified before, the analyst may utilize whichever distance calculation method the analyst determines to be best suited to the situation; for example line of sight if appropriate. Constraints 3.12.(24)-3.12.(26) utilize the distance calculated in 3.12.(23) to determine if a connection can exist between any two nodes.

Combined, constraints 3.12.(18)-3.12.(26) permit the model to independently construct new networks based upon the $\partial = 0$ position a node exists at for each *d*.

Constraints 3.12.(27)-3.12.(30) model the fuel usage of a node, ensuring that only a mobile node may move, and that once a satellite, or other mobile node, is no longer fueled it may no longer be used. For terrestrial components not requiring fuel to continue operations, $FuelUseR_i$ may be set to 0 and $maxfuel_{i,0} > 0$. Since the node has fuel, does not lose fuel over time, and is immobile preventing spikes in fuel usage, terrestrial nodes

will continue to operate until destroyed. Constraint 3.12.(28) ensures simultaneously that only mobile nodes move and that they only move once.

Constraints 3.12.(31)-3.12.(34) model the requirement of the node to be inoperable for the duration of the repositioning. Constraints 3.12.(32) and 3.12.(33) work in tandem to ensure that the node is disabled for at least the duration of its move and that the disabling occurs at the same time that the move begins.

Constraints 3.12.(35) and 3.12.(36) link the change of location to the LTI in which the change occurs, and ensures only one bounded move occurs per mobile node. While the model expresses this in a compact form, $dist_{(i,1),(i,|d|)}$ is the distance from node i at the start of the model and the end of the model. The decision variables used in this distance is $\Theta_{i,d}$ and $\Phi_{i,d}$ for the respective node and times. This method is only valid when a single move is allowed, but may be expanded to permit multiple moves.

$$\max_{\varphi} \text{v} = \frac{Cost}{CostMax} + \frac{\sum_i \sum_t \sum_s c_{i,t,s} \bullet \partial \text{y}_{i,t,s}}{GlobeMax} + \frac{\sum_{i \in R} \sum_t \sum_s c_{i,t,s} \bullet \partial \text{y}_{i,t,s}}{\text{RegMax}}$$

Subject To:

$$[1]\ \text{Cost} = \sum_i \left(\left(\sum_a (x_{i,a} \bullet (\text{costR}_{i,a}) \right) + (1 - x_i) \bullet \text{costO}_i \right) + \sum_i \sum_d CostR_{i,d}(x_{i,a,d}) + \sum_e \sum_a ADRC_{e,a} \left(\sum_d ADFired_{e,a} \right)$$

$$[2]\ \text{x}_{i,d} = \sum_a x_{i,a,d} \quad \forall i$$

$$[3]\ \sum_i \sum_a \varphi_{i,a} \le AdvRsc$$

$$[4]\ \text{x}_{i,a,d} = f_{i,a,d}(\varphi_{i,a,d}) \quad \forall i,a,d$$

$$[5]\ \sum_g \varphi_{\text{i,a,d}} \le M \omega_{i,a,d} \quad \forall i,a,d$$

$$[6]\ \sum_{d'=d}^{d+\bar{d}} \sum_a \omega_{i,a,d'} = 1 + M(1 - \phi_i) \quad \forall i,d, \text{ where } \bar{d} = \left\lceil \frac{TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,g,d}))}{IntDur} \right\rceil$$

$$[7]\ \phi_{i,d} \le TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,d})) \quad \forall i,d$$

$$[8]\ \min_{y}\ z = \frac{1}{Total\_Net\_Value} \bullet \sum_d \sum_\partial \sum_s \sum_t \sum_i c_{i,t,s} \bullet \partial \text{y}_{i,t,s,\partial,d}$$

Subject To:

(1) $\sum_j y_{(j,i),t,l,s,\partial,d} \le cap_{i,l,s,\partial,d} \quad \forall i,l,s,\partial,d$

(2) $\sum_l \sum_s \sum_j y_{(j,i),t,l,s,\partial,d} \le cap_i \bullet x_{i,\partial,d} \quad \forall i,\partial,d$

(3) $-\sum_l \sum_k y_{(t,k),t,l,s,\partial,d} \ge b_{t,t,s,\partial} \bullet \text{x}_{t,\partial,d}$ (Supply node t) $\forall t,s,\partial,d$

(4) $\left(\sum_l \left(\sum_j y_{(j,i),t,l,s,\partial,d} - \sum_k y_{(i,k),t,l,s,\partial,d} \right) \right) + \partial y_{i,t,s,\partial,d} = b_{i,t,s,\partial,d} \quad \forall i,t,s,\partial,d$

(5) $\text{cap}_i \bullet Conn_{j,i,\partial,d} \ge y_{(j,i),t,l,s,\partial,d} \ge 0$ (6) $\partial y_{i,t,s,\partial,d} \ge 0 \quad \forall i,t,s,\partial,d$

(7) $\partial y_{i,t,s,\partial,d} \le ((\sum_i (TC_i \bullet x_{i,\partial,d})) - 1) \bullet (cap_i * (1 - TC_t)) * TCactive_{\partial,d} + (cap_i \bullet (x_{i,\partial,d} + (1 - TCactive_o))) \quad \forall i,t,s,\partial,d$

(8) $TCactive_{\partial,d} \le \sum_i (TC_i \bullet x_{i,\partial,d})$

(9) $Excess\_Cap_{s,\partial,d} = \sum_{i,d} \left(cap_i \bullet x_{i,\partial,d} - \sum_l \sum_j y_{(j,i),t,l,s,\partial,d} \right) \quad \forall s,\partial,d$

(10) $\text{x}_{i,\partial,d} \le 1 \quad \forall \text{i},\partial,d$

(11) $Excess\_Value_{\partial,d} = \sum_{s} ec_s \bullet Excess\_Cap_{s,\partial,d}$

(12) $Weighted\_Excess\_Cap\_PCT = \sum_{d}\sum_{\partial} Excess\_Value_{\partial,d} \Big/ Total\_Net\_Value$

(13) $Total\_Net\_Value = d_{\max} \bullet \sum_{\partial}\sum_{s}\sum_{t}\sum_{i} c_{i,t,s} \bullet b_{i,t,s,\partial}$

(14) $c_{t,t,s} = 0 \ \forall t,s$

(15) $ST_{e,a} \bullet ADFired_{e,i,a,\partial} \le f(\varphi_{i,a,g,d}, 0) \bullet \mathrm{ADProt}_{e,i,a} \quad \forall e,i,a,d$

(16) $\sum_{d'=d}^{d+\hat{d}} \sum_{i} ADFired_{e,i,a,\partial} \le 1 \quad \forall e,a,\partial, \text{ where } \hat{d} = \frac{ADRT_{e,a}}{IntDur}$

(17) $\mathrm{x}_{i,\partial,d} \le \mathrm{State}_{\mathrm{i,d}} \bullet (\mathrm{State}_{\mathrm{i,d}} - 1) \quad \forall \mathrm{i},\partial,d$

(18) $\mathrm{POS}\Phi_{i,\partial,d} = \mathrm{Sin}[(\delta t \bullet \partial \bullet 360 / OT) - 90 \bullet (OD - 1) + OD \bullet \mathrm{Arcsin}(\Phi_{i,d} / 90)] \bullet 90 \quad \forall i,\partial,d$

(19) $\mathrm{POS}\Theta_{i,\partial,d} = \Theta_{i,d} + 180 - 360 \bullet (\delta t \bullet \partial / (CT / CN)) + 360 \bullet Wrap_{i,\partial,d} \quad \forall i,\partial,d$

(20) $0 \le \mathrm{POS}\Theta_{i,\partial,d} \le 360 \quad \forall i,\partial,d$ (21) $\mathrm{Wrap}_{i,\partial,d} = \{0,1\} \quad \forall i,\partial,d$

(22) $\mathrm{HA}_{i,j,\partial,d} = Sin^2[(\mathrm{POS}\Phi_{i,\delta,d} - \mathrm{POS}\Phi_{j,\delta,d}) \bullet 0.5] +$

$+Cos[\mathrm{POS}\Phi_{i,\delta,d}] \bullet Cos[\mathrm{POS}\Phi_{j,\delta,d}] \bullet Sin^2[(\mathrm{POS}\Theta_{i,\partial,d} - \mathrm{POS}\Theta_{j,\partial,d}) \bullet 0.5] \quad \forall \mathrm{i,j},\partial,d$

(23) $dist_{i,j,\partial,d} = \mathrm{Atan2}\left(\sqrt{1 - \mathrm{HA}_{i,j,\partial,d}}, \sqrt{\mathrm{HA}_{i,j,\partial,d}}\right) \quad \forall i,j,\partial,d$

(24) $dist_{i,j,\partial,d} \le \mathrm{footprint}_{i,r_j} \bullet ChckI_{i,j,\partial,d} + M \bullet (1 - ChckI_{i,j,\partial,d}) \quad \forall \mathrm{i,j},\partial,d$

(25) $dist_{i,j,\partial,d} \le \mathrm{footprint}_{j,r_i} \bullet ChckJ_{i,j,\partial,d} + M \bullet (1 - ChckJ_{i,j,\partial,d}) \quad \forall \mathrm{i,j},\partial,d$

(26) $\mathrm{Conn}_{i,j,\partial,d} \le 0.5 * (ChckI_{i,j,\partial,d} + ChckJ_{i,j,\partial,d}) \quad \forall \mathrm{i,j},\partial,d$

(27) $\mathrm{maxfuel}_{i,d} = \mathrm{maxfuel}_{i,d-1} - FuelUseR_i - D_{i,d} \bullet Fuel_i(DistPos_i) \quad \forall \mathrm{i},d$

(28) $\sum_{d} \mathrm{D}_{i,d} \le Mobile_i \quad \forall \mathrm{i}$

(29) $\mathrm{maxfuel}_{i,d} \le M \bullet empty_{i,d} \quad \forall \mathrm{i},d$

(30) $\sum_{\partial} \mathrm{x}_{i,\partial,d} \le empty_{i,d} \quad \forall \mathrm{i},d$

(31) $\mathrm{State}_{i,d} \le M \bullet x_{i,d} \quad \forall i,d$

(32) $\mathrm{R}_{i,d} \le \mathrm{R}_{i,d-1} + D_{i,d} \quad \forall i,d$

(33) $\sum_{d} \mathrm{R}_{i,d} \ge \mathrm{RepoD}_i(DistPos_i) \quad \forall i$

(34) $\mathrm{State}_{i,d} \le 3 - 2 \bullet R_{i,d} \quad \forall i,d$

(35) $\mathrm{DistPos}_i = dist_{(i,1),(i,|d|)} \quad \forall i$

(36) $\mathrm{minpos}_i \bullet \mathrm{D}_{i,d} \le dist_{(i,d-1),(i,d)} \le \mathrm{maxpos}_i \bullet \mathrm{D}_{i,d} \quad \forall i,d$

Model 3.12

In this section, a medium/long term bi-level program was developed which models the operational parameters of the network, as well as capturing the possible methods of degradation. The key differences of this model, Model 3.12, versus the short term model, Model 3.9, are the inclusion of nodal repositioning/reconstitution, and network creation/destruction based upon those modification of nodes' parameters.

Reducing Model Size

In the preceding sections, a short term model and a medium/long term model were developed. These bi-level programs model the ways in which the network may react and be degraded over different spans of time, however in their current form they may be practically intractable, or require high performance computers to solve if every aspect is included. In this section, the medium/long term model, the larger of the two, is analyzed and modifications are made with which to reduce its size. Many of the modifications presented here may be applied to the short term model as well. Final versions of both models with the improvements made in this section may be found in Appendix A.

Though constraints in bi-level programs do not directly correlate to any constant multiple of constraints in single level programming, a bi-level program that can be converted to a single level more often requires a sharp increase in the number of constraints. As such, the number of constraints serve to show a lower bound on the amount of memory required by using this method of imbedding the short term model in the long term model. The preceding bi-level medium/long term model has:

$$4+i\left(4+d\left[9+a\left(2+e\right)+\partial\left(6+5j+s\bullet t\left[l+3\right]\right)\right]\right)+t\bullet s+d\left(e\bullet a+\partial\left[2+s\right]\right) \text{(EQ 3.3)}$$

constraints.

As an example, a constellation with:

Network Parameters	Index Max
90 nodes and a ground fidelity of 1x1 degree ground grid	i=j=t=64890
Five security levels	s=5
Two network encodings	l=2
Time fidelity of every ten minutes on a 24-hour cycle	∂=144
Three years with time steps at each cycle	d=1095
Three methods of attack	a=3
Six active defense nodes	e=6

has at least $1.9918\bullet 10^{16}$ constraints. Reducing the fidelity of the terrestrial grid to 2x2 degree components, i=j=t=16290 and the number of constraints is reduced to $1.255\bullet 10^{15}$, which is a reduction in size of approximately 93.7%, but is still prohibitively large repeated for application.

If the short term time components, ∂, could be removed, even without removing all associated terms that would be eliminated by separating the short term network operations from the long term mobility constraints, the number of constraints would be reduced to $2.345\bullet 10^{09}$. This is a reduction of over 99.9%. Making both changes simultaneously, the number of constraints is reduced to $5.889\bullet 10^{08}$.

Our final change to this example is single security level, single operational encoding single attack method, and no active defenses. Even with these simplifications, which reduce the generality of the model and fail to capture the many options available to both the Attacker and Defender, the minimal number of constraints is still $1.962 \bullet 10^{08}$. As the reader can see, even in an idealized situation in regards to the model, a strict optimization of this bi-level problem, even if an exact method is determined, would be unreasonable in many, though not necessarily all, situations.

One exact constraint number, assuming no exploitation of special structures, is the constraints of the Defender model. The Defender model, in which the operational options of the network exist, have $1.9918 \bullet 10^{16}$ constraints in the original example, but only $1.249 \bullet 10^{08}$ constraints in the idealized modification. This indicates that the majority of the constraints exist in the Defender model, which can be seen intuitively when viewing the Attacker's seven constraints versus the Defender's thirty-six.

Therefore, the key to reducing the dimensionality of the problem is not to reduce the number of nodes in the network, which is presently insignificant when compared to the ground grid constructed, but is instead to reduce the fidelity of ground coverage, and to reduce the number of time indices, ∂ and d, whether by reducing fidelity or reducing the time-span. However, reducing the STI intervals, as well as reducing the time modeled below a single cycle, may cause a gap in coverage modeling, making ∂ essentially exempt from alteration. Reducing the ground fidelity to 4x4 grids and the LTI to 10 cycles, and leaving all other parameters in the example as they were originally specified, reduces the number of constraints to $8.108*10^{08}$. While this value is still

almost eight times the idealized number of constraints, it was reached by reducing accuracy, but maintaining validity.

It is important to view another example with a very different structure. Up till now, no assumptions have been made upon which types of satellites are included in the network, meaning that the model had to be adaptive enough to handle both LEO, MEO, and GEO. However, many networks involve only GEO satellites, which are much less mobile relative to a terrestrial location. The question may be posed if as many constraints would remain.

To answer that question, a new example is formed based only upon GEO and terrestrial nodes. This example begins by setting sizes for the dimensions of the new network.

Network Parameters	Index Max
6 GEO nodes, 8 terrestrial nodes, and a ground fidelity of 1x1 degree ground grid (64800 grids)	i=t=64814 j=14
Five security levels	s=5
Two network encodings	l=2
Time fidelity of every ten minutes on a 24-hour cycle	∂=144
Three years with time steps at each cycle	d=1095
Three methods of attack	a=3
Six active defense nodes	e=6

As the reader can see, with no further adjustment to the model other than the alteration in network components, very little has changed in terms of size of each

dimension, the only change being in the number of nodes, being reduced by 76, which by itself is notably insignificant. The model may be further improved for these conditions by removing constraints which are unnecessary due to the static nature of the nodes remaining.

With these alterations in mind, the majority of STI parameters, save the time-dependent transmissions, are irrelevant and thus may be omitted. This is because GEO satellites and terrestrial nodes maintain a relatively constant footprint by design. This property allows us to reasonably reduce the number of distance calculations needed. With satellites changing footprints as quickly as LEOs do, it was necessary to continue allowing the model to calculate new orbiting and connection parameters. However with GEOs, calculations are only necessary when a node has been moved.

With this reduced model, the number of constraints is now defined as:

$$3 + i \bullet (4 + d \bullet (13 + 2 \bullet a + \partial \bullet (1 + 3 \bullet t \bullet s + l \bullet s) + t \bullet s) + 5 \bullet j^2) + \\ + d \bullet (e \bullet a \bullet (i + 1) + t \bullet s \bullet \partial + 1) \qquad \text{(EQ 3.4)}$$

where j marks only the mobile nodes. In the example, this correlates to $9.959*10^{15}$ constraints, compared to $1.9918*10^{16}$ from the earlier example. A common thought would be that many of the remaining requirements, such as the high fidelity transmission constraints, are unnecessary. Since a node may still be degraded at any LTI, *d*, and the transmissions in the network, which may be time dependent, flow in according to the relay nodes available, the model must still be permitted to optimize the flow of transmissions through remaining nodes at each time step. Moreover, the model allows the transmissions to be STI time dependent, which require the optimization of flows.

As for the remainder of the constraints, the satellites are still mobile, meaning that the model must still be able to reposition them, reduce their fuel, determine new connections, and adjust their states accordingly. While the model no longer needs to map the movement of the satellites around the planet, the elimination of these constraints constituted $2.044 \bullet 10^{10}$ of the original constraints. While that number is large in and of itself, it is still only 0.0001% of all constraints.

The largest reduction of the alteration occurred with the elimination of the need for the STI components in the bulk of the constraints and were able to be reduced to only the LTI component in which a move occurred. This alteration resulted in a reduction of $i \bullet j \bullet (\partial \bullet d - j)$, which correlates to $1.43 \bullet 10^{11}$, 0.0007 % of original constraints. To show this visually, Figure 16 provides a visual representation of the original example.

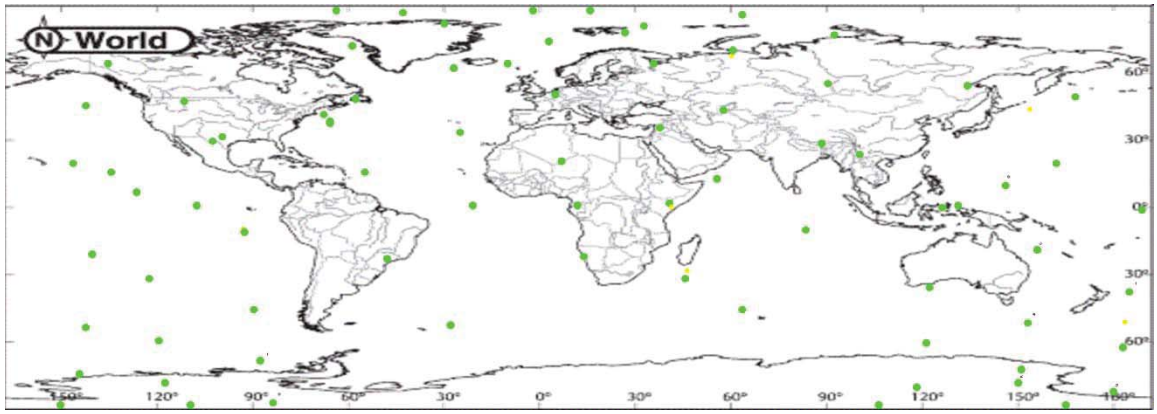


Figure 16: Snapshot Projected Location of Example 1 Nodes

In Figure 16, each circle represents an active node and its location with respect to Earth. Understandably, there appear to be a number of nodes to contend with, and they appear to be randomly spread across the surface. They are, however, following a common LEO distribution of assets, with three GEO satellites spaced evenly around the

Equator. Figure 17 views this same map and network, but with the associated terrestrial regions.

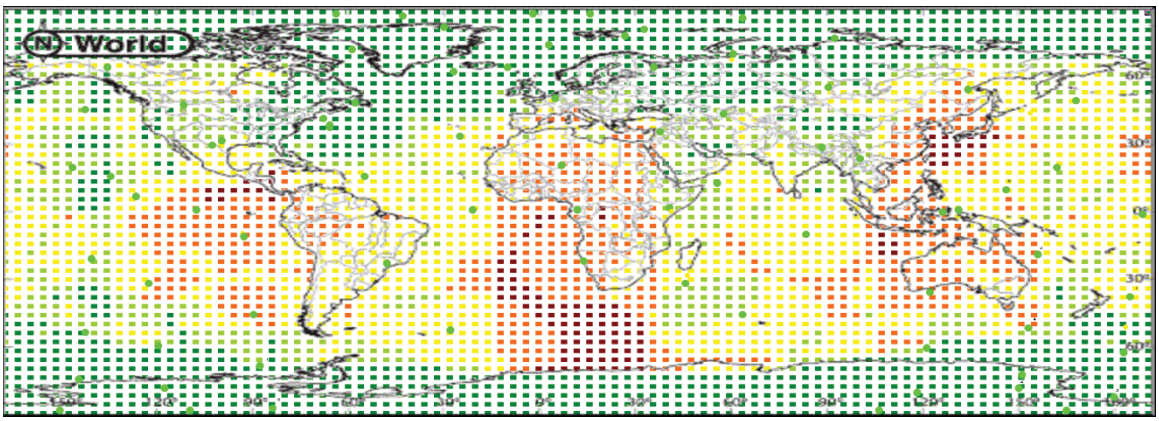


Figure 17: Example 1 Nodes with Associated Terrestrial Connections

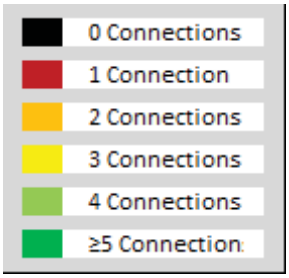


Figure 18: Color Legend of 17

The squares on this map each represent a 2x2 Lat-Long grid, and only one of every four grids is colored to maintain visibility. Even so, the sheer number of colored squares is noticeably greater than the number of network nodes, which previously seemed large. Here the reason behind the minor reduction in constraints in relation to the number of network nodes can be seen. The changing of that relatively small number is near insignificant in relation to the quantity of terrestrial regions.

The question remains as to where in the model the majority of the constraints are attributed to. This is addressed in regards to the original model, Model 3.12.

Table 4: Long Term Model Constraints by Function

Function	Constraint Numbers	Number of Constraints In Variable Terms	Primary Example's (%) of Total
Attacker Model	[1]-[7]	$2+i \bullet (d \bullet (2+2 \bullet a))$	$5.69 \bullet 10^{-6}$
Transmission Flow	(1)-(8)&(10)	$d \bullet (i \bullet (\partial \bullet (2+3 \bullet t \bullet s+1 \bullet s)+t \bullet s)+ \partial \bullet (t \bullet s+1))$	99.99
Value Calculations	(9) &(11-14)	$3+s \bullet (t+\partial \bullet d)$	$1 \bullet 10^{-8}$
Active Defense	(15),(16)	$e \bullet a \bullet d \bullet (i+1)$	$1.28 \bullet 10^{-5}$
Position	(18)-(21)	$4 \bullet i \bullet \partial \bullet d$	$4.10 \bullet 10^{-4}$
New Network	(22)-(26)	$5 \bullet i \bullet j \bullet \partial \bullet d$	$7.17 \bullet 10^{-3}$
Mobility	(27)-(30)&(35),(36)	$i \bullet (2+4 \bullet d)$	$2.84 \bullet 10^{-6}$
States	(17)&(31)-(34)	$i \bullet (d \bullet (\partial+3)+1)$	$1.04 \bullet 10^{-4}$

Clearly, with over 99.99% of all constraints originating from the optimizing of transmissions through the network, this is where the focus should fall. The model that follows is the set of constraints, as well as objective function, that come directly from the unaltered medium/long term model.

$$[8]\underset{y}{Min}\quad z=\frac{1}{Total\_Net\_Value}\bullet\sum_{d}\sum_{\partial}\sum_{s}\sum_{t}\sum_{i}c_{i,t,s}\bullet\partial\, y_{i,t,s,\partial,d}$$

Subject To:

(1) $\sum_{j} y_{(j,i),t,l,s,\partial,d} \le cap_{i,l,s,\partial,d} \quad \forall i,l,s,\partial,d$

(2) $\sum_{l}\sum_{s}\sum_{j} y_{(j,i),t,l,s,\partial,d} \le cap_{i}\bullet x_{i,\partial,d} \quad \forall i,\partial,d$

(3) $-\sum_{l}\sum_{k} y_{(t,k),t,l,s,\partial,d} \ge b_{t,t,s,\partial}\bullet x_{t,\partial,d}$ (Supply node t) $\forall t,s,\partial,d$

(4) $\left(\sum_{l}\left(\sum_{j} y_{(j,i),t,l,s,\partial,d}-\sum_{k} y_{(i,k),t,l,s,\partial,d}\right)\right)+\partial y_{i,t,s,\partial,d}=b_{i,t,s,\partial,d} \quad \forall i,t,s,\partial,d$

(5) $cap_{i}\bullet Conn_{j,i,\partial,d} \ge y_{(j,i),t,l,s,\partial,d} \ge 0$ (6) $\partial y_{i,t,s,\partial,d} \ge 0 \quad \forall i,t,s,\partial,d$

(7) $\partial y_{i,t,s,\partial,d} \le ((\sum_{i}(TC_{i}\bullet x_{i,\partial,d}))-1)\bullet(cap_{i}\bullet(1-TC_{t}))\bullet TCactive_{\partial,d}+$

$+(cap_{i}\bullet(x_{i,\partial,d}+(1-TCactive_{\partial}))) \quad \forall i,t,s,\partial,d$

(8) $TCactive_{\partial,d} \le \sum_{i}(TC_{i}\bullet x_{i,\partial,d})$

(10) $x_{i,\partial,d} \le 1 \quad \forall i,\partial,d$

Model 3.12.[8]

This model was initially formulated by utilizing common network connection techniques, in which each node of the network may be connected, but with parameter restrictions. Those parameters could potentially limit the capacity of a flow to zero, such as the connection binary variable, *Conn*. However, according to the operating procedures of the network, one can never transmit directly from one ground grid to another without passing through another node first, unless otherwise permitted. This is where the first modification is made. Let *j, k, t* define the combined set of all nodes and terrestrial grids, and *i* define the nodes in the network capable of relaying transmissions.

New constraint 3.13.(1.1) now transmits from any relay node to any node in the network, so long as transmission requirements are satisfied. Constraint 3.13.(1.2) does the reverse. The size of constraint 3.13.(5) has been reduced by summing over the

transmissions. Moreover, the size of all constraints have been reduced by limiting the effects of the network to the nodes, excluding 3.12.[8].(3) which requires that all transmissions leaving a source must be less than the number originating at that source.

$$[8]\underset{y}{Min}\ \ z=\frac{1}{Total\_Net\_Value}\bullet\sum_{d}\sum_{\partial}\sum_{s}\sum_{t}\sum_{i}c_{i,t,s}*\partial\, y_{i,t,s,\partial,d}$$

Subject To:

$$(1.1)\ \sum_{t,j}y_{(i,j),t,l,s,\partial,d}\le cap_{i,l,s,\partial,d}\bullet x_{i,\partial,d}\quad \forall i,l,s,\partial,d$$

$$(1.2)\ \sum_{t,j}y_{(j,i),t,l,s,\partial,d}\le cap_{i,l,s,\partial,d}\bullet x_{i,\partial,d}\quad \forall i,l,s,\partial,d$$

$$(2)\ \sum_{l}\sum_{s}\sum_{j}y_{(j,i),t,l,s,\partial,d}\le cap_{i}\bullet x_{i,\partial,d}\quad \forall i,\partial,d$$

$$(3)\ -\sum_{l}\sum_{k}y_{(t,k),t,l,s,\partial,d}\ge b_{t,t,s,\partial,d}\bullet x_{t,\partial,d}\quad \text{(Supply node t)}\ \ \forall t,s,\partial,d$$

$$(4)\ \left(\sum_{l}\left(\sum_{j}y_{(j,i),t,l,s,\partial,d}-\sum_{k}y_{(i,k),t,l,s,\partial,d}\right)\right)+\partial y_{i,t,s,\partial,d}=b_{i,t,s,\partial,d}\quad \forall i,t,s,\partial,d$$

$$(5)\ \ cap_{i}\bullet Conn_{j,i,\partial,d}\ge\sum_{j,t,l,s}y_{(j,i),t,l,s,\partial,d}\ge 0\ \ \forall i,\partial,d\qquad (6)\ \partial y_{i,t,s,\partial,d}\ge 0\ \ \forall i,t,s,\partial,d$$

$$(7)\ \sum_{t,s}\partial y_{i,t,s,\partial,d}\le((\sum_{i}(TC_{i}\bullet x_{i,\partial,d}))-1)\bullet(cap_{i}\bullet(1-TC_{t}))\bullet TCactive_{\partial,d}+$$
$$+(cap_{i}\bullet(x_{i,\partial,d}+(1-TCactive_{\partial})))\ \ \forall i,\partial,d$$

$$(8)\ TCactive_{\partial,d}\le\sum_{i}(TC_{i}\bullet x_{i,\partial,d})\quad \forall\partial,d$$

$$(10)\ x_{i,\partial,d}\le 1\ \ \forall i,\partial,d$$

Model 3.13.[8]

The transmission flow function, which was just remodeled , now has $d\bullet(\partial\bullet(1+j\bullet s)+i\bullet(2\bullet s\bullet\partial\bullet(j+l)+4\bullet\partial))$ constraints, the indices redefined as listed before. This results in $9.22\bullet10^{12}$ constraints, a significant reduction from before with no loss of fidelity. This change is applied to the remainder of the model, and, in terms of writing the model, nothing else is affected due to the selection of indices. However, as a whole

the model now contains$1.38 \bullet 10^{13}$ constraints, again with no loss of fidelity. Recall that the model began with $1.9 \bullet 10^{16}$ constraints under this example, making this a reduction of 99.9%. These updates reallocate constraint percentages as presented in Table 5.

Table 5: Constraint Groupings of Model 3.13

Function	Constraint Numbers	Number of Constraints In Variable Terms	Primary Example's (%) of Total
Attacker Model	[1]-[7]	$2+i \bullet (d \bullet (2+2 \bullet a))$	1.08E-03
Transmission Flow	(1)-(8)&(10)	$d \bullet (i \bullet (\partial \bullet (2+3 \bullet t \bullet s+l \bullet s)+t \bullet s)+ \partial \bullet (t \bullet s+1))$	66.63
Value Calculations	(9) &(11-14)	$3+s \bullet (t+\partial \bullet d)$	1.37E-03
Active Defense	(15),(16)	$e \bullet a \bullet d \bullet (i+1)$	2.49E-03
Position	(18)-(21)	$4 \bullet i \bullet \partial \bullet d$	7.81E-02
New Network	(22)-(26)	$5 \bullet i \bullet j \bullet \partial \bullet d$	33.27
Mobility	(27)-(30)&(35),(36)	$i \bullet (2+4 \bullet d)$	5.41E-04
States	(17)&(31)-(34)	$i \bullet (d \bullet (\partial+3)+1)$	1.98E-02

To proceed, the medium/long-term model is reformatted in such a manner that, at the very least, the Defender model may be solved. As such, the focus shall rest on the Defender model for the remainder of this section.

Since it is clear that the majority of the short term constraints cannot remain within the model due to the large number of resulting constraints, and since the size of the model as well as current solving techniques suggest the development a heuristic to solve the bi-level program, instead the Defender model is reconstructed utilizing Goal Programming, which is imbedded in the bi-level program heuristic. This is developed in the following section, Heuristic H-1.

It is assumed priorities of certain regions of the planet may vary to the extent that priority *n* areas are of much greater significance than priority *n+1*. This method is known as Lexicographic Goal Programming, or Pre-emptive Goal Programming (Ignizio p. 5). While this explanation is sufficient for this research, the reader may read more on this method in James Ignizio's *Goal Programming and Extensions* and others. One method for prioritizing the regions is based upon the value of the transmissions originating or concluding there. With this method, varying thresholds must be set for value cutoffs such that if a region's transmission values fall within the range, then the region is part of the priority, otherwise it is assigned to a different priority group.

Another method for prioritizing regions is to rank the regions themselves based upon some weighting structure. For example, in the military a nation at war may be of higher concern than the nation providing backdoor support which is higher than a neutral nation, and so on until a region not requiring coverage would be prioritized lowest. In this method, the prioritization is based upon the importance of the possible transmissions and not a region's cumulative transmission value.

Clearly, either of the methods, as well as many more, may be appropriate in a variety of situations. The Chapter IV case study will utilize the second method suggested here with a rating for each region, which is discuss further in the Chapter IV section, Case Study. Because each region has a rating, sets I_n are created such that if a region, j, is within set I_n, then region j has a priority rating n, with 1 being the highest priority rating. However, the lowest priority rating will continue being assigned to any region sending/receiving no transmissions. While the optimization model will automatically place these as the lowest level internally, creating a separate priority will reduce the

amount of memory required with no significant change to system configuration and performance.

Model 3.14 may also be referred to hereafter as the lexicographic model. In the Model 3.14, the constraint numbers appear as they did in Model 3.12 solely for the purpose of showing how many constraints were removed. A new binary variable, $C_{j,\partial,d}$, was introduced along with new parameters, w_j and $\text{Req}_{j,\partial,d}$, which denote the time independent importance of a region and the time-dependent number of required satellite connections respectively. Priority parameter, w_j must reflect the specified importance of a region by varying greatly between priority levels, but only slightly within levels. If the number of connections is independent of time, then the dimensionality of Req may be reduced to one.

Note that the objective function was changed from maximizing the number of completed transmissions to maximizing the number of priority n regions without meaningful connections. In addition, recall that, as focus is only on the Defender model in Model 3.14, the level of degradation, x, is considered to be known.

$$\underset{\Phi,\Theta}{Max}\ \ \mathrm{v}_n = \sum_{j\in I_n,\partial,d}\left(w_j C_{j,\partial,d}\right)$$

Subject To:

(17) $\mathrm{x}_{i,\partial,d} \le \mathrm{State}_{\mathrm{i,d}} \bullet (\mathrm{State}_{\mathrm{i,d}} - 1) \ \ \forall \mathrm{i},\partial,d$

(18) $\mathrm{POS}\Phi_{i,\delta,d} = \mathrm{Sin}[(\delta t \bullet \partial \bullet 360/OT) - 90 \bullet (OD-1) + OD \bullet \mathrm{Arcsin}(\Phi_{i,d}/90)] * 90 \ \ \forall i,\partial,d$

(19) $\mathrm{POS}\Theta_{i,\delta,d} = \Theta_{i,d} + 180 - 360 \bullet (\delta t \bullet \partial / (CT/CN)) + 360 \bullet Wrap_{i,\delta,d} \quad \forall i,\partial,d$

(20) $0 \le \mathrm{POS}\Theta_{i,\delta,d} \le 360 \quad \forall i,\partial,d$ (21) $\mathrm{Wrap}_{i,\delta,d} = \{0,1\} \ \ \forall i,\partial,d$

(22) $\mathrm{HA}_{i,j,\partial,d} = Sin^2[(\mathrm{POS}\Phi_{i,\delta,d} - \mathrm{POS}\Phi_{j,\delta,d}) \bullet 0.5] +$

$$+Cos[\mathrm{POS}\Phi_{i,\delta,d}] \bullet Cos[\mathrm{POS}\Phi_{j,\delta,d}] \bullet Sin^2[(\mathrm{POS}\Theta_{i,\delta,d} - \mathrm{POS}\Theta_{j,\delta,d}) * 0.5] \ \ \forall \mathrm{i,j},\partial,d$$

(23) $dist_{i,j,\partial,d} = \mathrm{Atan2}\left(\sqrt{1-\mathrm{HA}_{i,j,\partial,d}}, \sqrt{\mathrm{HA}_{i,j,\partial,d}}\right) \quad \forall i,j,\partial,d$

(24) $dist_{i,j,\partial,d} \le \mathrm{footprint}_{i,r_j} \bullet ChckI_{i,j,\partial,d} + M \bullet (1 - ChckI_{i,j,\partial,d}) \ \ \forall \mathrm{i,j},\partial,d$

(25) $dist_{i,j,\partial,d} \le \mathrm{footprint}_{j,r_i} \bullet ChckJ_{i,j,\partial,d} + M \bullet (1 - ChckJ_{i,j,\partial,d}) \ \ \forall \mathrm{i,j},\partial,d$

(26) $\mathrm{Conn}_{i,j,\partial,d} \le 0.5 \bullet (ChckI_{i,j,\partial,d} + ChckJ_{i,j,\partial,d}) \ \ \forall \mathrm{i,j},\partial,d$

(27) $\mathrm{maxfuel}_{i,d} = \mathrm{maxfuel}_{i,d-1} - FuelUseR_i - D_{i,d} \bullet Fuel_i(DistPos_i) \ \ \forall \mathrm{i},d$

(28) $\sum_d \mathrm{D}_{i,d} \le Mobile_i \ \ \forall \mathrm{i}$

(29) $\mathrm{maxfuel}_{i,d} \le M \bullet empty_{i,d} \ \ \forall \mathrm{i},d$

(30) $\sum_\partial \mathrm{x}_{i,\partial,d} \le empty_{i,d} \ \ \forall \mathrm{i},d$

(31) $\mathrm{State}_{i,d} \le M * x_{i,d} \quad \forall i,d$

(32) $\mathrm{R}_{i,d} \le \mathrm{R}_{i,d-1} + D_{i,d} \quad \forall i,d$

(33) $\sum_d \mathrm{R}_{i,d} \ge \mathrm{RepoD}_i(DistPos_i) \quad \forall i$

(34) $\mathrm{State}_{i,d} \le 3 - 2 \bullet R_{i,d} \quad \forall i,d$

(35) $\mathrm{DistPos}_i = dist_{(i,1),(i,|d|)} \quad \forall i$

(36) $\mathrm{minpos}_i \bullet \mathrm{D}_{i,d} \le dist_{(i,d-1),(i,d)} \le \mathrm{maxpos}_i \bullet \mathrm{D}_{i,d} \quad \forall i,d$

(37) $\mathrm{Req}_{j,\partial,d} \bullet C_{j,\partial,d} \le \left(\sum_i Conn_{j,i,\partial,d} \bullet \mathrm{x}_{i,\partial,d}\right) \quad \forall i,\partial,d$

(38) $\mathrm{v}_m = \sum_{i\in I_m,\partial,d}\left(w_j \bullet C_{j,\partial,d}\right) \quad \forall m < n$

Model 3.14

The determination if at least one meaningful connection exists is completed by new constraint 3.14.(37), and the requirement of completing all previous optimal priority

settings is constraint 3.14.(38). Note that no requirements for the higher priority node to always be covered have been developed, but rather to be covered optimally as the previous model settings allowed. This is to reduce infeasibilities which may arise from a node's connections being momentarily severed due to a malicious event. While a node may not be continuously connected, the model does force it to reconnect to the network when possible and optimal.

By doing this, if $v_n>0$ for any priority excluding the last, then these alterations have reduced the size of the problem by all regions that the model failed to connect to while optimally, with respect to the priorities, connecting to all others. With widely varying capacities on satellite components, it may be prudent to incorporate the connected capacity versus a strict connection so as to reduce the likelihood of a sufficiently connected region remaining under capacitated. Formatting the problem in this manner, equation 3.14.(37) may be rewritten as:

$$(37)\ \left(\sum_{k,s} b_{j,k,s,\partial,d}\right)\bullet C_{j,\partial,d} \le \sum_{i}\left(Conn_{j,i,\partial,d}\bullet\left(\sum_{l,s} cap_{i,l,s,\partial,d}\right)\bullet x_{i,\partial,d}\right)\ \forall j,\partial,d$$

Equivalently, this change may be created by defining:

$$\mathrm{Req}_{j,\partial,d} = \frac{\left(\sum_{k,s} b_{j,k,s,\partial,d}\right)}{\left(\sum_{l,s} cap_{i,l,s,\partial,d}\right)} \qquad \text{(EQ 3.5)}$$

If the transmission values reflect the priorities established through the region ranking, and if they follow the inherent assumption that transmissions from region n are greatly more important than transmission from region n+1, then this method will provide an optimal solution to the original objective of maximizing the cumulative value of

completed transmissions. However, if the transmission values are not widely separated between priority settings, then this method, while providing a good network configuration, may not necessarily provide the optimal configuration.

Using the resulting network configurations from the Lexicographic Goal Programming, one then utilizes the constraints of the Defender model that were removed. These constraints, which are primarily focused on the flow of the transmissions through a given network, are very similar to that of the Short-Term Defender Model. This model is referred to hereafter as the LTI Defender model.

$$\underset{y}{Min}\ \ z = \frac{1}{Total\_Net\_Value} \bullet \sum_{d}\sum_{\partial}\sum_{s}\sum_{t}\sum_{i} c_{i,t,s} \bullet \partial\, y_{i,t,s,\partial,d}$$

Subject To:

(1.1) $\sum_{t,j} y_{(i,j),t,l,s,\partial,d} \le cap_{i,l,s,\partial,d} \bullet x_{i,\partial,d} \quad \forall i,l,s,\partial,d$

(1.2) $\sum_{t,j} y_{(j,i),t,l,s,\partial,d} \le cap_{i,l,s,\partial,d} \bullet x_{i,\partial,d} \quad \forall i,l,s,\partial,d$

(2) $\sum_{l}\sum_{s}\sum_{j} y_{(j,i),t,l,s,\partial,d} \le cap_{i} \bullet x_{i,\partial,d} \quad \forall i,\partial,d$

(3) $-\sum_{l}\sum_{k} y_{(t,k),t,l,s,\partial,d} \ge b_{t,t,s,\partial} \bullet x_{t,\partial,d}$ (Supply node t) $\forall t,s,\partial,d$

(4) $\left(\sum_{l}\left(\sum_{j} y_{(j,i),t,l,s,\partial,d} - \sum_{k} y_{(i,k),t,l,s,\partial,d}\right)\right) + \partial y_{i,t,s,\partial,d} = b_{i,t,s,\partial,d} \quad \forall i,t,s,\partial,d$

(5) $\text{cap}_i \bullet Conn_{j,i,\partial,d} \ge \sum_{j,t,l,s} y_{(j,i),t,l,s,\partial,d} \ge 0 \ \forall i,\partial,d$ (6) $\partial y_{i,t,s,\partial,d} \ge 0 \ \forall i,t,s,\partial,d$

(7) $\sum_{t,s} \partial y_{i,t,s,\partial,d} \le ((\sum_{i}(TC_i \bullet x_{i,\partial,d})) - 1) \bullet (cap_i \bullet (1 - TC_t)) \bullet TCactive_{\partial,d} +$
$+ (cap_i \bullet (x_{i,\partial,d} + (1 - TCactive_{\partial}))) \ \forall i,\partial,d$

(8) $TCactive_{\partial,d} \le \sum_{i}(TC_i \bullet x_{i,\partial,d}) \quad \forall \partial,d$

(9) $Excess\_Cap_{s,\partial,d} = \sum_{i,d}\left(cap_i \bullet x_{i,\partial,d} - \sum_{l}\sum_{j} y_{(j,i),t,l,s,\partial,d}\right) \ \forall s,\partial,d$

(10) $x_{i,\partial,d} \le 1 \ \forall i,\partial,d$

(11) $Excess\_Value_{\partial,d} = \sum_{s} ec_s \bullet Excess\_Cap_{s,\partial,d}$

$$(12) Weighted\_Excess\_Cap\_PCT = \sum_{d}\sum_{\partial} Excess\_Value_{\partial,d} \Big/ Total\_Net\_Value$$

$$(13) Total\_Net\_Value = d_{\max} \bullet \sum_{\partial}\sum_{s}\sum_{t}\sum_{i} c_{i,t,s} \bullet b_{i,t,s,\partial}$$

$$(15)\ ST_{e,a} \bullet ADFired_{e,i,a,\partial} \le f(\varphi_{i,a,g,d}, 0) \bullet \text{ADProt}_{e,i,a} \quad \forall e,i,a,d$$

$$(16) \sum_{d'=d}^{d+\hat{d}} \sum_{i} ADFired_{e,i,a,\partial} \le 1 \quad \forall e,a,\partial \quad \text{Where } \hat{d} = \frac{ADRT_{e,a}}{IntDur}$$

Model 3.15

The lexicographic model contains $i \bullet (d \bullet (\partial \bullet (6 + 5 \bullet j) + 7) + 2)$ constraints, and the LTI Defender model contains $d \bullet (\partial \bullet (1 + j \bullet s) + i \bullet (2 \bullet s \bullet \partial \bullet (j + l + 1) + 4 \bullet \partial) + e \bullet a \bullet (i + 1)) + 3$ constraints.

Figure 19: Binary Land Matrix

Figure 19 is a binary matrix such that all entries darkened sections depict either land or areas of water with extremely heavy usage, and all blank space is sparsely travelled water. Using the previous example for comparison, and an estimated six priorities of equivalent size, with Figure 19 white space as the least preferred sixth priority, which encompasses 57.4% of the total area, Antarctica the fifth priority, which encompasses another 21.1% of the total area, and lowering the fidelity of regions from

1x1 to 2x2 grids, a reduction to $6.82 \bullet 10^{10}$ constraints can be obtained. Reducing the fidelity of LTI from every day to every week then results in $9.75 \bullet 10^{09}$ constraints, or 975 billion.

As one can tell from the equation for the current example model, the number of constraints is linearly related to each of the parameters. As it is relatively easy to reduce the size of a exceedingly large constraint than a small constraint, the majority of the focus should fall on i_n, ∂, or d. This observation agrees with the earlier observation that the number of nodes in the network is insignificant in terms of number of constraints when compared to the terrestrial fidelity, or the time components.

In this example, the oceans were simplified to being the lowest priority. However, there is no restriction on priority groupings being contiguous, which allows for important regions being pocketed in otherwise insignificant zones. In this situation, if the path travelled is well used, then a strip may be increased in priority so as to accommodate. Similarly for Antarctica, stable locations such as research facilities may be classified as higher priority pockets of interest.

It is also possible to assign objects other than plots of land as terminal nodes. For example, if a vessel is traversing a rarely used path, then it might be prudent to utilize a node that has a unique time-dependent location following the planned path of that asset. However, pursuing this for even moderately sized sets of nodes is ill advised as it will rapidly increase the size of the model.

As the size of the model is still cumbersome, even with Lexicographic Goal Programming, the use of another simplification is recommended, which is the use of only

discrete allowable locations of placement for mobile nodes. This will essentially restrict Θ, Φ to be integers, with corresponding constant scalars attached for higher/lower fidelity.

Restricting the placement decision variables, decided from a move but not the resulting positions in orbit, to discrete locations, it is possible to utilize a multitude of simple and well investigated heuristics. The heuristic chosen will reside in the same place in the BLP heuristic that an optimization solution would, only now optimality cannot be guaranteed for the Defender's model.

To proceed, a greedy algorithm is modified by the distance from the current location. Imbedded inside of this algorithm is the connection and fuel constraints from the Defender model, specifically 3.15.(22)-(26) and 3.15.(37), along with a new constraint which simply requires that the node being repositioned to satisfy a specific requirement actually fills that requirement.

This heuristic, modified to account for the mobility and time dependencies of the network, utilizes many smaller calculations and memory storages to reduce the amount of memory required at any given time or used for any single model. Step SCP.{3}, the only optimization model in the heuristic, utilizes $6 \bullet j \bullet \partial$ constraints. Recall that all j in this model are $j \in I_m$, or the regions only of the current or higher priorities.

Steps SCP.{2} and SCP.{5.1}, while appearing large, are only minor calculations which pulled constraints out of the model to maintain consistency. Step SCP.{5.2} utilizes the results from step SCP.{3} to prevent unnecessary recalculations of previously determined connections.

{1} *If* $\sum_{i \in I_m} \left(C_{i,\partial',d'} \right) < \left| I_m \right|$ then

Select $\left\{ i \mid \text{Req}_{i,\partial',d'} \bullet C_{i,\partial',d'} \le \left(\sum_j Conn_{j,i,\partial',d'} \bullet \text{x}_{j,\partial',d'} \right), i \in I_m \right\}$

Select $\left\{ \text{j} \mid \min_j \left(\text{dist(j,i)-footprint}_j > 0 \right), j \notin J \right\}$ and J=J+{j}

{2} Calculate expected position of all mobile nodes except for j

[Constraints $(18)-(20)$]

{3} $\max_{\Theta\Phi} \left\{ \sum_{k \in I_m, \partial, d'} \left(w_k C_{k,\partial,d'} \right) \mid POS_{j,0,d'} = (\Theta, \Phi) \right\}$

Subject to:

(22) $\text{HA}_{i,j,\partial,d'} = Sin^2[(\text{POS}\Phi_{i,\delta,d'} - \text{POS}\Phi_{j,\delta,d'}) \bullet 0.5] +$

$+Cos[\text{POS}\Phi_{i,\delta,d'}] \bullet Cos[\text{POS}\Phi_{j,\delta,d'}] \bullet Sin^2[(\text{POS}\Theta_{i,\delta,d'} - \text{POS}\Theta_{j,\delta,d'}) * 0.5] \quad \forall \text{i},\partial$

(23) $dist_{i,j,\partial,d'} = \text{Atan2}\left(\sqrt{1-\text{HA}_{i,j,\partial,d'}}, \sqrt{\text{HA}_{i,j,\partial,d'}} \right) \quad \forall i,\partial$

(24) $dist_{i,j,\partial,d'} \le \text{footprint}_{i,r_j} \bullet ChckI_{i,j,\partial,d} + M \bullet (1 - ChckI_{i,j,\partial,d'}) \quad \forall \text{i},\partial$

(25) $dist_{i,j,\partial,d'} \le \text{footprint}_{j,r_i} \bullet ChckJ_{i,j,\partial,d} + M \bullet (1 - ChckJ_{i,j,\partial,d'}) \quad \forall \text{i},\partial$

(26) $\text{Conn}_{i,j,\partial,d'} \le 0.5 \bullet (ChckI_{i,j,\partial,d'} + ChckJ_{i,j,\partial,d'}) \quad \forall \text{i},\partial$

(37) $\text{Req}_{i,\partial,d'} * C_{i,\partial,d'} \le \left(\sum_j Conn_{j,i,\partial,d'} * \text{x}_{j,\partial,d'} \right) \quad \forall i,\partial$

(new) $\text{Conn}_{i,j,\partial',d} = 1$

{4} *If* $Fuel_j \left[\text{dist(j,}(\Theta\Phi)) \right] > \text{maxfuel}_{j,d} - FuelUseR_j \bullet (d_{\max} - d')$

Then J=J-{j}, J'=J'+{j}, and return to {1}

{5} Calculate time for movement subject to constraints and Record Connections from {3}

$\left[\text{constraints } (32)-(34) \right]$ simplify to

set $\text{State}_{j,d} = 1$ for all $d' \le d < d' + \text{RepoD}_j(DistPos_j)$

$\left[\text{constraints } (22)-(26) \right]$

{6}*C*alculate Value function $\sum_d z_d$ subject to linear program:

$$\underset{y}{Min}\ z_d = \frac{1}{Total\_Net\_Value} \bullet \sum_{\partial}\sum_{s}\sum_{t}\sum_{i} c_{i,t,s} \bullet \partial\, y_{i,t,s,\partial}$$

Subject To:

(1.1) $\sum_{t,j} y_{(i,j),t,l,s,\partial,d} \le cap_{i,l,s,\partial,d} \bullet x_{i,\partial,d} \quad \forall i,l,s,\partial,d$

(1.2) $\sum_{t,j} y_{(j,i),t,l,s,\partial,d} \le cap_{i,l,s,\partial,d} \bullet x_{i,\partial,d} \quad \forall i,l,s,\partial,d$

(2) $\sum_{l}\sum_{s}\sum_{j} y_{(j,i),t,l,s,\partial,d} \le cap_i \bullet x_{i,\partial,d} \quad \forall i,\partial,d$

(3) $-\sum_{l}\sum_{k} y_{(t,k),t,l,s,\partial,d} \ge b_{t,t,s,\partial} \bullet x_{t,\partial,d}$ (Supply node t) $\forall t,s,\partial,d$

(4) $\left(\sum_{l}\left(\sum_{j} y_{(j,i),t,l,s,\partial,d} - \sum_{k} y_{(i,k),t,l,s,\partial,d}\right)\right) + \partial y_{i,t,s,\partial,d} = b_{i,t,s,\partial,d} \quad \forall i,t,s,\partial,d$

(5) $cap_i \bullet Conn_{j,i,\partial,d} \ge \sum_{j,t,l,s} y_{(j,i),t,l,s,\partial,d} \ge 0 \ \forall i,\partial,d$ (6) $\partial y_{i,t,s,\partial,d} \ge 0 \ \forall i,t,s,\partial,d$

(7) $\sum_{t,s} \partial y_{i,t,s,\partial,d} \le ((\sum_{i}(TC_i \bullet x_{i,\partial,d})) - 1) \bullet (cap_i * (1 - TC_t)) \bullet TCactive_{\partial,d} +$

$+ (cap_i \bullet (x_{i,\partial,d} + (1 - TCactive_{\partial}))) \quad \forall i,\partial,d$

(8) $TCactive_{\partial,d} \le \sum_{i}(TC_i \bullet x_{i,\partial,d}) \quad \forall \partial,d$

(10) $x_{i,\partial,d} \le 1 \quad \forall i,\partial,d$

{7}*If* $(\partial < \partial_{\max})$ and $\left(J + J',\ J' \neq \varnothing \text{ contains all mobile nodes or } \sum_{i \in I_m,d}\left(C_{i,\partial',d}\right) \ge d_{\max} \bullet |I_m|\right)$,

then set ∂'=∂'+1 and return to step {1}

{8}*If* $J + J'$ contains all mobile nodes, then set J=J'=∅ and proceed to next d' such that there exists a node i where $\left|x_{i \in I_m,d'} - x_{i \in I_m,d'-1}\right| > 0$

Return to step {1}

Repeat until d'=d_{max}

Heuristic SCP

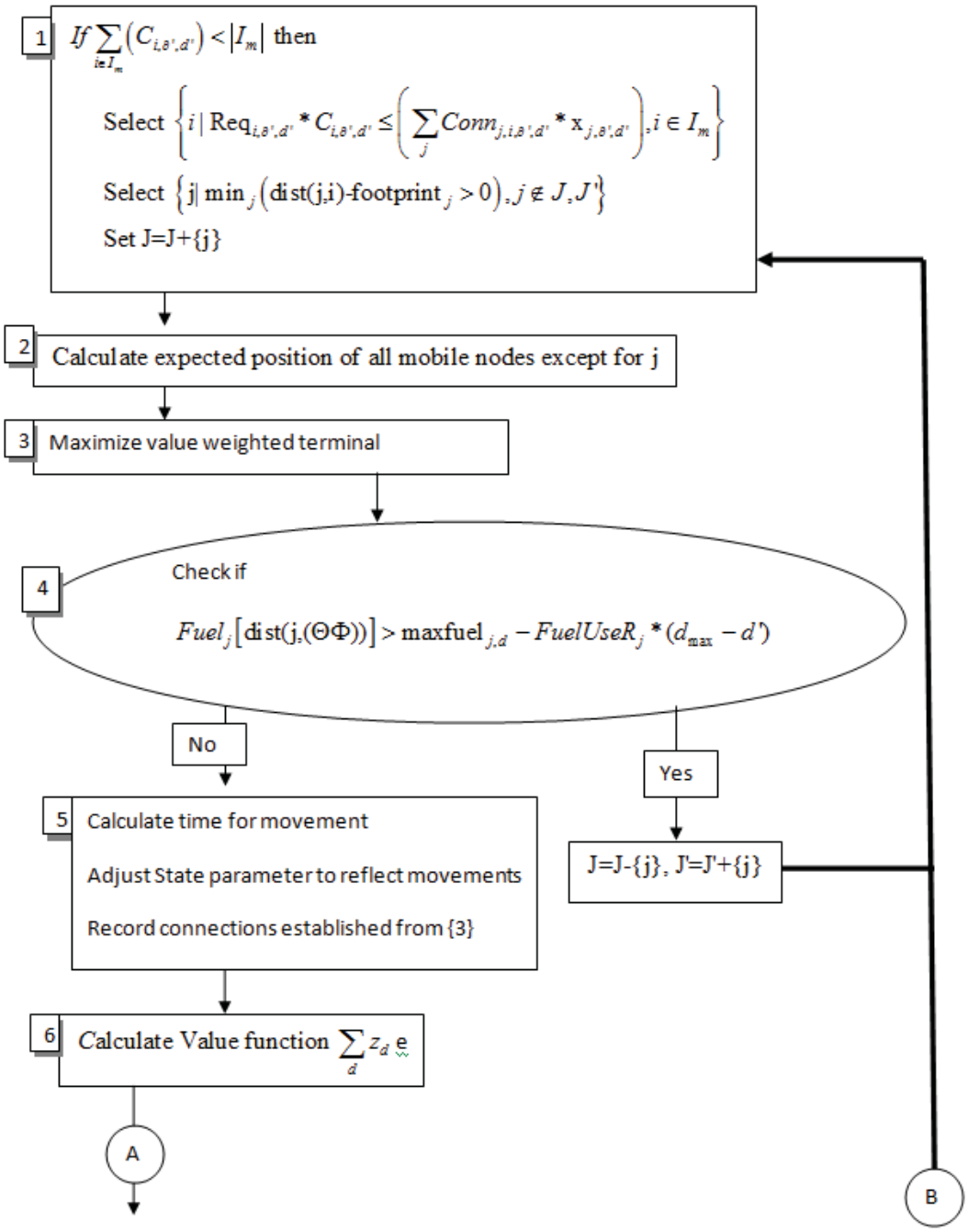
While SCP.{3} was focused solely on node j, meaning that initial connections would need to be calculated before the first run of this heuristic, though those connections may be accomplished trivially outside of the model, cumulating the

connections dependent only upon the changing node reduce the number of calculations and memory required by the computer. Step SCP.{6}, while appearing as an optimization model, consists of only known parameters, reducing the region to be searched from an infinite space to a single value.

Steps SCP.{7} and SCP.{8} allow the model to step through the various nodes available for repositioning, switching the focus first on each STI, and then on each LTI. This simple heuristic, when imbedded in the sequential inclusion of lower priorities that comes with the use of Lexicographic Goal Programming, provides a good solution to the long term model.

While acknowledging that the greedy heuristic is often dominated by other methods (Capara p. 733), this type of heuristic was selected for the sole purpose of its simplicity and adaptability. With the focus of this work on the development of methodology for measuring Resilience in satellite communication networks, the improvement of this time-dependent mobile network set covering heuristic is left to future work.

While the model is now able to successfully reposition, and reconstitute a network, there is still work left to be done in the area of the Defender value model. The Defender value model, which is primarily comprised by the Transmission Flow function, still poses a problem of having a size of $1.9*10^{10}$, even in this reduced fidelity situation. However, there is still a key piece of the networking operations which has yet to be exploited.


1
If $\sum_{i \in I_m} (C_{i,\theta',d'}) < |I_m|$ then
Select $\left\{ i \mid \text{Req}_{i,\theta',d'} * C_{i,\theta',d'} \le \left(\sum_j Conn_{j,i,\theta',d'} * x_{j,\theta',d'} \right), i \in I_m \right\}$
Select $\left\{ j \mid \min_j \left(\text{dist(j,i)-footprint}_j > 0 \right), j \notin J, J' \right\}$
Set J=J+{j}
2
Calculate expected position of all mobile nodes except for j
3
Maximize value weighted terminal
4
Check if
$Fuel_j \left[\text{dist(j,}(\Theta\Phi)) \right] > \text{maxfuel}_{j,d} - FuelUseR_j * (d_{\max} - d')$
No
Yes
5
Calculate time for movement
Adjust State parameter to reflect movements
Record connections established from {3}
J=J-{j}, J'=J'+{j}
6
Calculate Value function $\sum_d z_d \underset{\sim}{e}$
A
B

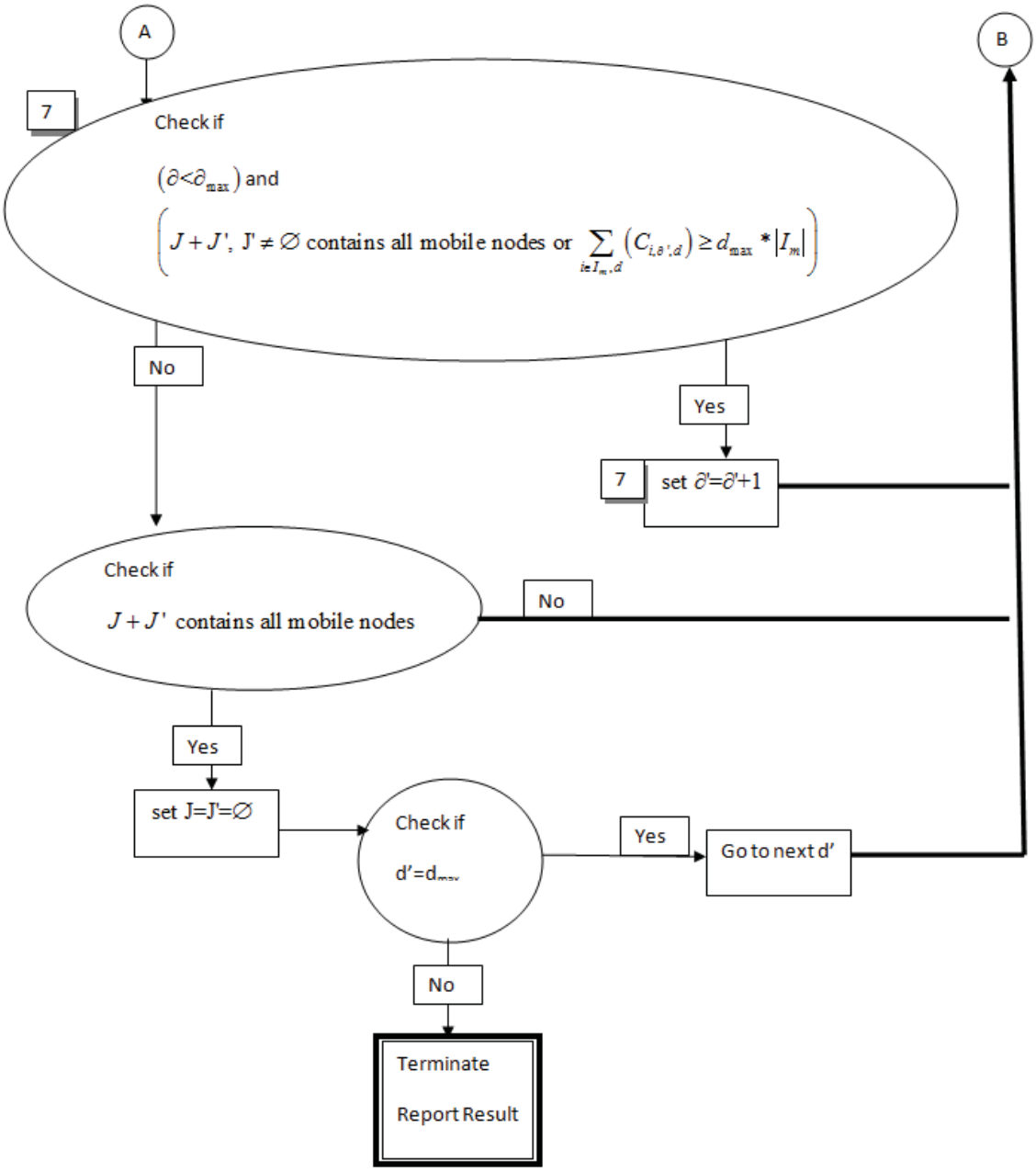


Figure 20: Time-Dependent Covering Problem Heuristic Flow Chart

Gateways, as was discussed in Chapter II section Comparing Satellite Networks to Supply Networks, must be traversed for the majority of known satellite communication networks. As of 1995, the only documented gateways were terrestrial based, however that is no longer true as gateways may exist both on Earth and in orbit aboard satellites.

Regardless of their position though, the basic network operating parameters hold. The main function of a gateway is to permit a user to access the full network, and as a result, every transmission must either originate from, or be passed through a gateway.

Since every transmission must pass through a gateway before it may be passed along to the destination, the gateways will now be defined as the sink nodes in the network, with all previous terminal nodes being source nodes. In conjunction with this, every transmission which previously had a single source and a single destination will now be modeled as originating from the original source and destination, with the new destination being any gateway. Because this change is essentially doubling transmissions, but reducing the transmission distance, no change need be made to the satellite capacities, however all transmission values must be halved. If they were set as rankings, the rankings remain unaffected.

To model this, a single node is made with unlimited capacity and complete demand as the sink of the network. Furthermore, this single node will have constant, non-destructible connections to all gateways and only to gateways. If all gateways are terrestrial based, if there are sufficiently many such that their capacity and location are only minor considerations, or if every orbital gateway in the network has a continuous connection to a terrestrial gateway or relay node, then modifications may stop here. Note that only one of these three conditions need be met, and that, while a network may be constructed specifically to not meet any of these, many if not all of the networks operate within at least one of these categories.

Using the described alterations to the model, which preserve validity in transmission quantity, type and network operational behavior, following medium/long term model Transmission Function constraints are achieved.

$$\underset{y}{Min} \quad z = \frac{1}{Total\_Net\_Value} \bullet \sum_{d} \sum_{\partial} \sum_{s} \sum_{j} c_{j,s} \bullet \partial \, y_{j,s,\partial,d}$$

Subject To:

(1.1) $\sum_{j} y_{(j,i),l,s,\partial,d} \le cap_{i,l,s,\partial,d} \bullet x_{i,\partial,d} \quad \forall i,l,s,\partial,d$

(2) $\sum_{l} \sum_{s} \sum_{j} y_{(j,i),l,s,\partial,d} \le cap_{i} \bullet x_{i,\partial,d} \quad \forall i,\partial,d$

(3) $-\sum_{l,i} y_{(j,i),l,s,\partial,d} - \partial \, y_{j,s,\partial,d} = b_{j,s,\partial,d} \quad \forall j,s,\partial,d$

(4) $\left(\sum_{l} \left(\sum_{j} y_{(j,i),l,s,\partial,d} - \sum_{k} y_{(i,k),l,s,\partial,d} \right) \right) \le b_{i,s,\partial,d} \quad \forall i,s,\partial,d$

(5) $\text{cap}_{i} \bullet Conn_{j,i,\partial,d} \ge \sum_{j,l,s} y_{(j,i),l,s,\partial,d} \ge 0 \quad \forall i,\partial,d$

(8) $TCactive_{\partial,d} \le \sum_{i} (TC_{i} \bullet x_{i,\partial,d}) \quad \forall \partial,d$

Model 3.16 Transmissions

The reader may note that constraints 3.15.(1.2), (6), and (7) were removed. All remaining constraints were also restructured by removing the dependency on the origination of a transmission, which required extra effort on constraints 3.16.(3) and (4) so that the failed transmissions are still properly attributed to the appropriate value. This new version utilizes $\partial \bullet d \bullet (i \bullet (2 + s + s \bullet l) + j \bullet s + l)$.

Additional insight is gained in reviewing the SCP heuristic. In SCP.{6} of this heuristic, there is no direct dependency between d and ∂, and the flow of transmissions. As such:

$$\sum_{d}\sum_{\partial} \underset{y}{Min} \quad z = \frac{1}{Total\_Net\_Value} \bullet \sum_{s}\sum_{j} c_{j,s} \bullet \partial \, \mathrm{y}_{j,s,\partial,d}$$

Subject To:

(1.1) $\sum_{j} y_{(j,i),l,s,\partial,d} \le cap_{i,l,s,\partial,d} \bullet x_{i,\partial,d} \quad \forall i,l,s$

(2) $\sum_{l}\sum_{s}\sum_{j} y_{(j,i),l,s,\partial,d} \le cap_{i} \bullet x_{i,\partial,d} \quad \forall i$

(3) $-\sum_{l,i} y_{(j,i),l,s,\partial,d} - \partial \, \mathrm{y}_{j,s,\partial,d} = b_{j,\mathrm{s},\partial,d} \quad \forall j,s$

(4) $\left(\sum_{l}\left(\sum_{j} y_{(j,i),l,s,\partial,d} - \sum_{k} y_{(i,k),l,s,\partial,d}\right)\right) \le b_{i,\mathrm{s},\partial,d} \quad \forall i,s$

(5) $\mathrm{cap}_{i} \bullet Conn_{j,i,\partial,d} \ge \sum_{j,l,s} y_{(j,i),l,s,\partial,d} \ge 0 \ \forall i$

(8) $TCactive_{\partial,d} \le \sum_{i} (TC_{i} \bullet x_{i,\partial,d})$

Model 3.17 Transmissions

This change reduces the model size, but includes the need for $\partial * d$ models, each of size $(i \bullet (2 + s + s \bullet l) + j \bullet s + l)$. However, there exists a subtle change to the models optimization behavior precluding the substitution of Model 3.17 Transmissions in the full Long-Term Model. By pulling the time components outside of the minimization as a summation, Model 3.17 is now attempting to optimize each time period myopically. If included in the full model, such an alteration would prevent a change in the network which sacrificed a small amount of performance early on for a greater performance later.

Still, Model 3.17 Transmissions is still valid and fully applicable in a situation in which the connections and locations are parameters and not variables, as occurs in the SCP heuristic. Using the most recent example of i=90, now i=91 due to the required inclusion of a new sink node, s=5, l=2, j=16291, $\partial = 144$, $d = 156$, this results in 22,464 linear models, each of size 82,987 constraints. While this may appear to be a large

model, because it is linear it runs fairly quickly, even on a personal computer. If each linear program requires only 4 seconds to solve on a personal computer, which is a reasonable estimation, then the complete model will require approximately 25 hours.

While manageable in itself, when run through the heuristic this must be solved an undetermined and potentially large number of times, which could require an exorbitant amount of time. In the short term model, with d=1, this model only requires 72 minutes, just over one hour. In the medium/long term model, the disruptive events can only occur in regards to the LTI, *d*. This was exploited in the SCP heuristic by only analyzing the network at the time of the next disruptive event. As such, a further exploitation is used to reduce run time.

If this same method is utilized, then the number of LTI, *d*, is reduced to the number of distinct time periods in which degradation, repositioning, or reconstitution occur. However, the same value may be retained by multiplying the cycle network value by the number of LTI between that event and the next network change. For example, if the events occur within 6 distinct time periods and the model completes repositioning/reconstituting in another 12, then the time to run, under the same approximation of a 4 second LP duration, is $144 \cdot 18 \cdot 2 \text{ sec}$, or 2.88 hours on a personal computer.

In a great many cases, considering the high costs associated with building and operating satellite networks, this level of time, which would be considerably less on a high performance computer, is certainly viable. However, as time on high performance computers can be costly and results of this form of analysis could lead to changes in design, the Chapter IV case study will show how aggregation and small sacrifices to

fidelity can preserve enough memory to be accomplished on a personal computer in a reasonable amount of time.

In this section, the models constructed in previous sections were improved by reducing the memory requirements, increasing the viability of using the models as more than tools for viewing a network's operating procedures. This was accomplished by first using an example to demonstrate how large the medium/long term model could become and why it is not always possible to solve as it was previously constructed. After analyzing the constraints it was determined that, in the model's original form, Model 3.12, the number of components in the network were insignificant in regards to memory consumption, and that the true problem laid with the ground fidelity and the time dependencies. This led to the decomposing of the model into its many functions, which showed how the Transmission Function was the true obstacle in the constraints.

After remodeling the Transmission Function constraints to reduce memory consumption and the number of constraints, Lexicographic Goal Programming was then pursued to prioritize regions/nodes and drastically reduce the scope of the problem and provide a more manageable model to be solved. Finally, the simplest known heuristic type, a greedy heuristic, was used to reduce the amount of memory required by an optimization model at the cost of increasing time through multiple out-of-model calculations. This, in effect, provides a suite of modeling options which offer a trade off of fidelity versus computational speed.

Heuristic H-1

In this section, the heuristic developed by Lim and Smith is adapted in order to account for monotonically continuous non-linear functions by including aspects of Ant Colony Optimization. The reader may notice that a large portion of the heuristic comes from the Attacker model while the Defender model acts as a calculation stage to be optimized repeatedly.

Recall from the Chapter II section Solving Bi-Level Programs, the heuristic developed by Lim and Smith for solving continuous linear BLP is:

Step 0. Initialize $\bar{x}_h = 0 \; \forall h \in A$, define the remaining budget as $\bar{B} = B$, and define the candidate set of arcs that can be interdicted as $\bar{A} = A$.

Step 1. Set $\sigma = 0$, $R = \bar{A}$, and solve the follower's problem given $\bar{x}$. Let v be the optimal objective value of the follower's problem.

Step 2. If $R = \emptyset$, go to Step 4. Otherwise, choose any $h \in R$ and proceed to Step 3.

Step 3. Set $\bar{x}_h = \min\{b_h, \bar{B}\}/b_h$. Solve the follower's problem to obtain the optimal objective value v_h. If $(v - v_h)/\min\{b_h, \bar{B}\} > \sigma$, put $\hat{h} = h$ and $\sigma = (v - v_h)/\min\{b_h, \bar{B}\}$. Reset $\bar{x}_h = 0$ and return to Step 2.

Step 4. Fix $\bar{x}_{\hat{h}} = \min\{b_{\hat{h}}, \bar{B}\}/b_{\hat{h}}$, update $\bar{B} = \bar{B} - \min\{b_{\hat{h}}, \bar{B}\}$, and remove $\hat{h}$ from $\bar{A}$. If $\bar{B} = 0$, terminate with the heuristic solution $\bar{x}$. Otherwise, return to Step 1.

In addition, recall that Ant Colony Optimization effectively operates in a shortest-path-problem by adjusting the probability that an edge is chosen based upon the value of the solutions achieved by selecting the edge in preceding iterations. However, there is also a fading function to reduce the likelihood of premature convergence.

The essence of a shortest-path problem is choosing the set of edges which will connect two nodes with the shortest possible distance/time/cost, etc. In Salmeron *et al.*'s model, the Attacker is selecting the set of nodes which will reduce their resource pool to zero with the greatest possible degradation objective. This simple comparison is because of Salmeron *et al*.'s use of discrete binary degradation levels.

The initial step to the heuristic follows directly from Lim and Smith's heuristic, which is to first approximate the impact that each node in the network has on the objective function. To do this, the heuristic initializes with calculating the defender model at various states, and considers each time interval, d, as a separate network. The first state is fully operational, in which all applicable nodes are operating as if there is no negative event. Following that, each node in the network is "turned off", and only one node is eliminated for any given run. This will provide an initial preference set that may be used to increase the efficiency of the heuristic by essentially aiming it at the most valuable nodes.

For the initial preferences, given node i, and attacker objective value v_i, the probability of a node being interdicted is denoted as follows:

$$p_{i,d} = \frac{v_{i,d}}{\sum_{i=0}^{n} v_{i,d}} \quad \text{For Each d} \qquad \text{(EQ 3.6)}$$

where $p_{i,d}$ is the conditional probability of node i being interdicted given the incident will happen in time-step d, and i=0 is the baseline operating state where no nodes are degraded.

Next, Lim and Smith's heuristic for solving continuous linear BLP is adapted. Their heuristic, in essence, is to use the remaining budget at each round and apply it to

the node which requires the minimum budget to eliminate and improves the attacker's solution. To proceed, a random number draw is utilized to select which nodes to degrade based upon the derived probabilities.

Given nodes i and j, random number $x \in [0,1]$,and remaining budget B:

$$\varphi_{i,d} = \begin{Bmatrix} 0 & \text{if } p_{i,d} = 0 \text{ or } p_{i,d} < \max_{d}\{p_{i,d}\} \\ \min\{f_i^{-1}(1), B\} & \text{if } x \in [\sum_{j=1}^{i} p_{j,d}, p_{i,d} + \sum_{j=1}^{i} p_{j,d}) \\ 0 & \text{Otherwise} \end{Bmatrix} \qquad \text{(EQ 3.7)}$$

Note that EQ 3.7 fully interdicts a set of nodes, and then uses the remaining budget to partially interdict a final node. This is the same method used by Lim and Smith, though altered such that less concern is placed on the resources required and more for the impact of the node. This method still permits little freedom in regards to time though.

To allow for this, one must first determine the period of time that a node is degraded, and thus non-targetable, based upon the intensity of the previous incident. Since this structure has already been developed in the Attacker-Defender model, given by constraints 3.12.[5], [6] and [7]:

$$[5] \sum_{g} \varphi_{i,a,g,d} \le M \omega_{i,a,d} \quad \forall i,a,d$$

$$[6] \sum_{d'=d}^{d+\bar{d}} \sum_{a} \omega_{i,a,d'} = 1 + M(1-\phi_i) \quad \forall i,d \qquad \text{Where } \bar{d} = \frac{TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,g,d}))}{IntDur}$$

$$[7] \phi_{i,d} \le TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,g,d})) \quad \forall i,d$$

one may use the same method in restricting the heuristic.

For a given time interval, d, $p_{i,d}$=0 if $\sum_{a} \omega_{i,a,d} = |A|$ where |A| is the number of attack types, a. This equation restricts a node from being targeted if all attack types have been used on the node thus far. However, the model has yet to prevent the same attack type to be used repeatedly during a degraded state. To do this, simply make the use of an attack type an inclusion criteria in selecting which node to degrade. This is done by only selecting those nodes who have an attack method usage variable, $\omega_{i,a,d} = 0$.

Next, one must establish the terminating conditions for the initial search. The termination condition is the occurrence of results within some value epsilon of the current best result found by the heuristic for a set number of iterations, m. The following heuristic, which is referred to henceforth as Heuristic-1 (H1), is:

1) Set B to the amount of resources, z=0, ε be an arbitrarily small number,
m be a sufficiently large integer, and v'=$\max_i(v_i)$

2) For every i,d combination, set $x_{i,d'} = 0$,
(or expected with no incident) for all d' ≥ d,
and record resulting Attacker objective as value $v_{i,d}$ '

3) Reset all $x_{i,d}$ to initial values and set $x_{i,d}' = x_{i,d}$, $x_{i,a,d}' = 0$

4) Set $p_{i,d} = \frac{v_{i,d}}{\sum_{i=0}^{n} v_{i,d}}$ and $p_{i,d}' = p_{i,d}$ for each i,d

5) Set $p_{i,d} = \frac{p_{i,d}}{\sum_{i=0}^{n} p_{i,d}}$

6) If $\sum_a \omega_{i,a,d} = |A|$, then set $p_{i,d} = 0$ and return to step 5.

7) If $x_{i,d}' = 1$, then set $p_{i,d} = 0$ and return to step 5.

8) Select a random number $x \in [0,1]$

9) Set $\varphi = \begin{cases} 0 & \text{if } p_{i,d} = 0 \text{ or } p_{i,d} < \max_d \{p_{i,d}\} \\ Min_d \left\{ \left(\min_{a \mid \omega_{i,a,d}=0} \{f_{i,a,d}^{-1}(1) - f_{i,a,d}^{-1}(x_{i,d}')\} \right), B \right\} & \text{if } x \in [\sum_{j=1}^{i} p_{j,d}, p_{i,d} + \sum_{j=1}^{i} p_{j,d}) \\ 0 & \text{Otherwise} \end{cases}$

10) $x_{i,a,d} = f_{i,a,d}(\varphi_{i,a,d} = \varphi + f_{i,a,d}^{-1}(x_{i,d}'))$ for each a

11) If $x_{i,a,d} = \max_{a \mid \omega_{i,a,d}=0} \{x_{i,a,d}\}$ and no other $x_{i,a',d} = \max_{a \mid \omega_{i,a,d}=0} \{x_{i,a,d}\}, a' \neq a,$ exists,
then set $x_{i,a,d}' = x_{i,a,d}$. If there exists $x_{i,a',d} = x_{i,a,d} = \max_{a \mid \omega_{i,a,d}=0} \{x_{i,a,d}\}, a' \neq a,$
then set $x_{i,a,d}' = x_{i,a,d}$ such that $TimeR_{i,d}(x_{i,a,d}) = \max_a \{TimeR_{i,d}(x_{i,a,d})\}$,
and set $\omega_{i,a,d} = 1$.

12) Set $x_{i,d}' = \sum_a x_{i,a,d}', B = B - \varphi, p_{i,d} = 0$

13) If B>0, then return to step 5, else continue

14) Calculate resulting Attacker objective value and denote value as v.

$$\text{Set failedv}_d = \frac{\sum_s \sum_t c_{t,t,s} \left(\sum_l \partial y_{t,l,s,t,d} \right)}{\sum_{i=1}^{n} \sum_s \sum_t c_{i,t,s} \left(\sum_l \partial y_{i,l,s,t,d} \right)}$$

15) If $1+\varepsilon > \frac{v}{v'} > 1-\varepsilon$ then z=z+1

16) If v>v', then set v'=v and $X_{i,a,d} = x_{i,a,d}' \ \forall \ i,a,d.$

17) If z>m then terminate, else continue

18) Set $p_{i,d} = 0.5 \bullet \left(1 - \frac{v}{v'} \right) \bullet \frac{p_{i,d}'}{\sum_{i=0}^{n} p_{i,d}'} +$

$$+0.5 \bullet \left(\left(\frac{v \bullet (1 - failedv_d)}{v'} \right) \bullet \frac{\left(\sum_s \sum_t c_{i,t,s} \left(\sum_j \sum_l y_{(i,j),l,s,t,d} \right) \right)}{\sum_{i=1}^{n} \left(\sum_s \sum_t c_{i,t,s} \left(\sum_j \sum_l y_{(i,j),l,s,t,d} \right) \right)} + \left(\frac{v \bullet failedv_d}{v'} \right) \bullet \frac{x_{i,d}}{\sum_i^n x_{i,d}} \right)$$

19) Every ζ itterations, utilize previous ζ degradation, and corresponding resilience, cost, and repair times, to determine the current best time dependent resilience per unit cost.

19.1) Denote this degradation combination as $\text{BIN}_{\bullet,1}$. For each *i* from 2 to the number of degraded nodes in $\text{BIN}_{\bullet,1}$, let $\text{BIN}_{\bullet,i} = \text{BIN}_{\bullet,1}$ except for the *ith* degraded node, which will be set equal to 1.

19.2) Set $\text{BIN}'_{\bullet,1}$ equal to the degradation vector of the best time dependent resilience per unit cost from all $\text{BIN}_{\bullet,i}$. If $\text{BIN}'_{\bullet,1} \neq \text{BIN}_{\bullet,1}$, then return to step 19.1 and repeat. Else go to step 19.3.

19.3) Set $p_{i,\bullet} = \{1 - \text{BIN}_{i,1}\}$ and go to step 20

20) Set $p_{i,d}' = p_{i,d}$ for each i,d, reset all $x_{i,d}$ to initial values, and set $x_{i,d}' = x_{i,d}$, $x_{i,a,d}' = 0$ and return to step 6.

Heuristic-1 (H1)

Though this method looks much larger than Lim and Smith's heuristic, the increased size comes primarily from accounting for the multiple time steps, the inclusion of probability ranges based upon objective value, the more fully expanded steps, and the logical requirements which more precisely choose the best option among otherwise

equivalent attack types. Note that the termination criterion is experiencing a change of less than ε more than m times, regardless of if the heuristic is improving or not. At the point of termination, the choice exists to either reset and begin again, or to press on. This decision is left to the analyst measuring resilience.

Step 18 in H1 is the equation which determines the probabilities for the nodes in the next iteration. It is broken first into two distinct pieces, which is the current probability and the probability gained from the most recent iteration. In this second half, the gain in probability is determined through two criteria. The first is how much value a particular node moved at each time step. The second piece, however, is the opposite, instead providing higher probability to degraded nodes based upon how much value failed to even enter the network. Either piece alone would likely lead to a premature convergence, either by targeting only transit nodes, or by targeting only those nodes degraded previously. Together, they permit the model to locate a synergy between supply and transit nodes.

Finally H1 step 19 performs a small local search around previous run iterations. It selects the combination which showed the greatest drop in resilience per unit time and cost, and then searches around that point for improvements by removing extraneous degradations. In this manner, the heuristic is able to pull the Attacker-Defender model from overwhelming a small time span in exchange for continuous degradation, but at a lower level.

Heuristic H1 is displayed in flow chart form in Appendix A. In the flow chart, a rectangle refers to a process, an oval to the checking of a logical variable, and a circle with a single letter inside is used to bridge the flows over page breaks.

Note that, in H1, the majority of the complexities arise from step 9, which is the large equation driving the selection of nodes. It has been broken down as such to better show how it may be implemented in a computer program which is unable to understand the compressed equation form. Steps requiring multiple blocks have been grouped using a dotted box line.

In this section, the base operating concepts of Ant Colony Optimization were utilized to adapt Lim and Smith's heuristic for continuous linear bi-level programs to a monotonically nonlinear situation.

Methodology for Output Analysis

In the preceding sections, the methodology for measuring time-dependent network performance using multiple measures, and over varying time intervals was developed. In this section, the resulting output is explained along with the process for developing it. It is in this section that the suggested single-value measure for resilience is presented, as well as time-dependent network performance chart.

The previous section concluded with a heuristic locating one or more degradation and network reaction plans. As a consequence of these results, two measures can be attained, value weighted percent of blocked calls, the Defender model objective value, and value weighted percent of remaining capacity, model constraint 3.12.(11).

The first step is to determine what value excess capacity provides, as well as what value a blocked transmission loses. Because values were placed on the individual

components in the model, what remains is to do is the combining of the ending measures that the model returned. As such a weight function may be applied. However, the function must be piece-wise continuous and either non-decreasing or non-increasing.

The combined value of these measures, percent blocked transmissions and percent excess capacity, is the resilience measure. Note that there will actually be two values, the resilience of the network in regards to a short term event and a long term event. Both measures are the time-dependent performance of the network under an extreme event, and both provide equally important information, though to slightly different questions.

However, both results provide an answer to resilience in the face of extreme event network degradation. A time-dependent network performance was also utilized because, based upon the definition of resilience, a network may be *more* resilient if it can resist an event which many models including ours examine, but also if it can be repaired quickly, or can quickly adapt to operating within the new state, which relatively few models have captured thus far. Furthermore, the inherent defenses, as well as the active defenses, that a network may incorporate to lessen the impact of an event, if not prevent it altogether, were included so that their effectiveness to an extreme event could be gauged.

This time-dependent network performance under extreme event degradation captures the ability of the network to resist, adapt, absorb, respond to, and recover from an unlikely, yet possible occurrence. The attentive reader may notice that the model failed to properly address the "anticipate" portion of the definition. Unfortunately that is something that is currently outside the capabilities of the presented deterministic model, but which may be included later with future work.

While single value measures are useful, there is also a great deal of data within the model that may aid a decision maker. Two vastly different networks may experience an equivalent measure of resilience, though they may achieve this measure in very different ways. For example, a fragile network which can recover rapidly can be as resilient as a hardened network which takes a long period of time to recover.

To determine how a network is degraded and how it recovers, it is recommended the analyst utilize the variables $x_{i,d}$, and v_d and $Excess\_Value_d$, which combine into the measure of Network Performance. By plotting each of these variables over time, the analyst may see what aspect of their network struggled, as well as what aspect may have save them.

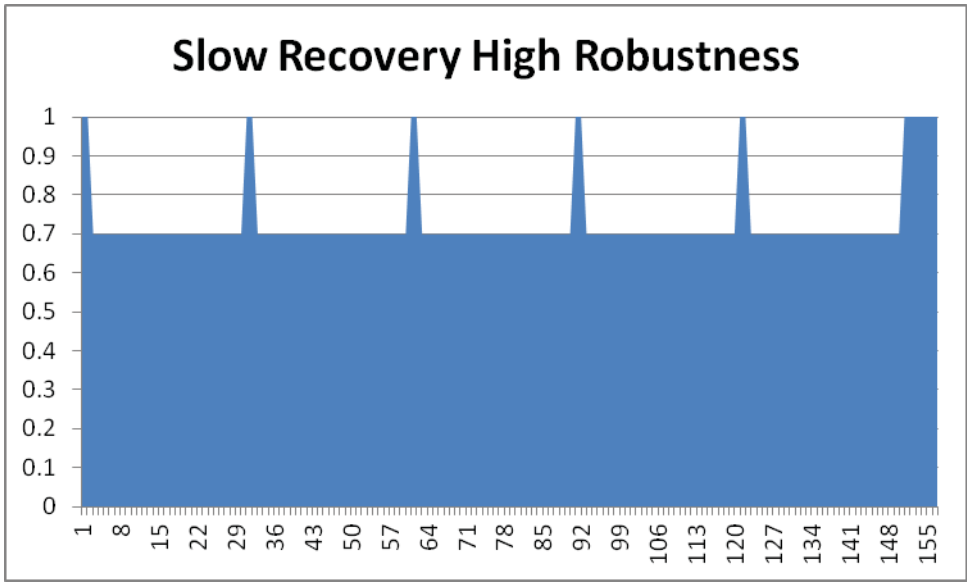


Figure 21: Example Slow Recovery High Robustness

For example, if a network began with a high Network Performance, but that value degraded over a long time period, and $x_{i,d}$ took a long time to recover, then the network is resilient because of an aspect such as its massive size, or capacity, but may become more

resilient if its ability to recover is improved. Figure 16 displays a notional chart of network performance over time in such a situation when repair/replacement time is long.

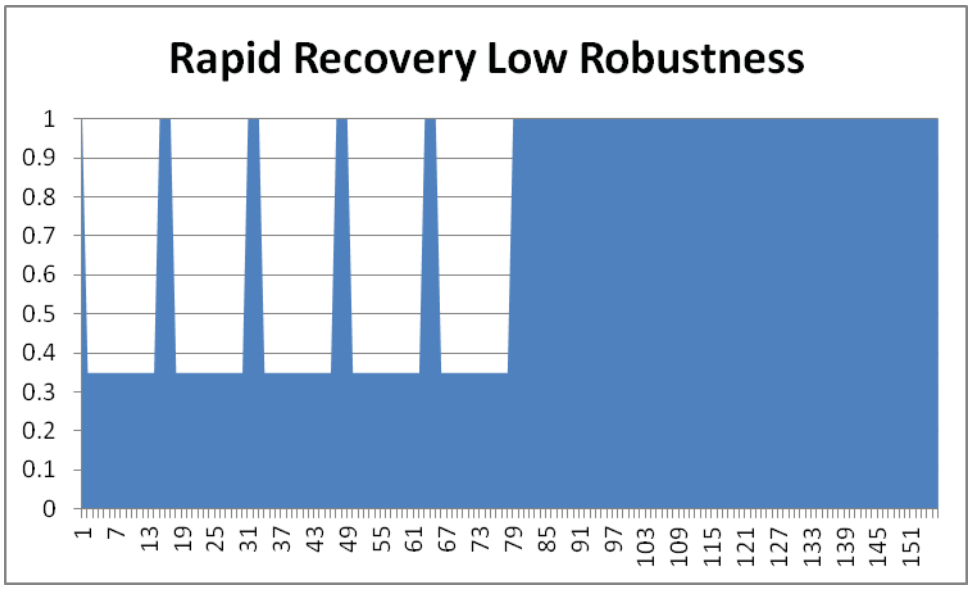


Figure 22: Example Rapid Recovery Low Robustness

If the network experiences frequent changes in Network Performance, and those changes show a high correlation to changes in $x_{i,d}$, then the network is resilient because of its ability to recover, and may become more resilient by increasing its initial capacity or hardening against some form of attack. Figure 22 shows an example of network performance over time when a node experiences frequent degradations, but quickly recovers. Note that while Figure 21 and Figure 22 examples behave very differently, both exhibit the same average resilience measure of approximately 0.709 over the bounded time period.

Both situations may warrant a greater emphasis on defensive measures, though more in depth analysis into the changes of Network Performance as a result of expended Attacker resources would be required to make such a determination. Finally, if a node *i* rarely or never experiences degradation, then it may imply that the node is either

extremely well protected, superfluous, experiences a recovery time that precludes the resources expended to degrade it, or a combination of the three. Again, further efforts and sensitivity analysis would need to be conducted before a conclusion as to which situation is true.

In this section, the resilience measure is defined as the time-dependent network performance under an extreme event, keeping the short term and long term measures separate as they provide very different information as to the behavior of the network under different scenarios. The utilization of the measure of network performance and the degradation variables from the model were shown as a method for drawing basic conclusions about the network's behavior under a degrading event.

In this chapter, a bi-level Attacker-Defender model was formulated to capture the options and constraints available to the network operator and the degrader. Once formulated, this model was adjusted to allow for reduced memory consumption. Heuristics were then employed to reduce the bi-level program to an algorithm containing a linear program. Finally, the methodology for analyzing the outputs from the main heuristic, H-1, were presented with example graphs. This approach offers the user a suite of model options with varying fidelity and precision at varying computational requirements.

IV. Analysis and Results

In this chapter, a notional satellite communication network of GEO satellites and terrestrial nodes is utilized as a case study. Using the results of that case study and its variations, analysis is conducted on the behavior of the network performance over time to determine the approach proposed in this thesis. By doing so, insights may be gained as to the changes one could expected in the network performance based upon a change to the network. After the case study has been completed, the performance of the model and heuristic during those variations is analyzed.

Case Study

A notional satellite communication network constructed of GEO satellites and terrestrial nodes serves as the basis for a case study. The base configuration and various values of the GEO satellite communications network is fully laid out in Appendix B. The notional network consists of six GEO satellites, and nine terrestrial nodes which may communicate within, but not across, continents, and two sink nodes, both of which are terrestrial.

Primary variations affect the number of nodes in the network, as well as values such as recovery time, attacker resources, and initial nodal placement. These variations are run systematically, their network performance charts and degradation levels presented

in the following section and as well as the variation data tables that may be found in Appendix B.1.

Secondary variations are performed on the base settings, and involve including more unique changes to the network, such as the inclusion of a satellite gateway or the use of active defenses. Results are fully displayed in Appendix B.2; a selection of those results are displayed in the following section.

The case study COMSAT network operates six GEO satellites, all positioned evenly around the equator at an altitude of 35,786 km. Each of these satellites may operate only as a relay in the base scenario, however that is altered in one of the secondary variations.

Most equatorially based geosynchronous satellites can communicate with locations as far north or south as 75 degrees (Geo-orbit p. 1). The resulting radius is 10,348 km, which is equivalent to a footprint of 336,395,288 km^2. Secondly, a maximum inter-satellite link range of 90,000km is permitted, which is approximately one third of the total GEO orbital circumference.

Since this resilience measure is intended for use as a key criterion in selecting a new satellite architecture, the use of a high performance computer is not unreasonable at an operational level during a design stage. However, a base level analysis, with a reduction in fidelity, can still be performed on a personal computer while continuing to provide valuable insight.

To attain the memory and time reduction required to implement this model on a personal computer in a reasonable amount of time, preprocessing aggregation of transmissions is conducted. By distributing the transmissions of many regions to their

connected network nodes, the cumbersome effect generated by the granularity of the terrestrial partitioning is effectively eliminated. However, the shortcoming of doing this is that the high fidelity of associating the exact regional transmission value is lost.

To compensate, the costs of the transmissions being preprocessed to a satellite are averaged to arrive at a new cost-per-transmission. For example, assume the analyst is presented with a situation in which six regions, {A, B, C, D, E, F}, and their associated transmission values, v_i, and number of transmissions, n_i, are connected to a single satellite. Then the value of transmissions originating from that satellite, *v'*, in the aggregated model is defined as:

$$v' = \frac{\sum_i v_i n_i}{\sum_i n_i} = \frac{v_A n_A + v_B n_B + v_C n_C + v_D n_D + v_E n_E + v_F n_F}{n_A + n_B + n_C + n_D + n_E + n_F} \quad \text{(EQ 4.1)}$$

When a region is connected to multiple satellites, the transmissions are divided evenly across all connected satellites. This method may be substituted to filling satellite capacity with high value regions first and working down, however this may leave some regions transmitting to satellites with no remaining capacity while other satellites have excess.

To compute the distance traversed, both for transmission linkages as well as nodal movement, the Haversine Formula for great circle distances is used (Miller p. 134):

$$\text{distance} = 2 \bullet r \bullet \arcsin\left(\sqrt{\left(\sin^2\left(\frac{\Phi_{i,1} - \Phi_{i,0}}{2}\right) + \cos(\Phi_{i,0}) \bullet \cos(\Phi_{i,1}) \bullet \sin^2\left(\frac{\Theta_{i,1} - \Theta_{i,0}}{2}\right)\right)}\right) \quad \text{(EQ 4.2)}$$

The Haversine formula has known problems with rounding errors with extremely small distances or as the two points approach opposite ends of the sphere. However, this may be alleviated in the model by preventing the satellites from approaching those two conditions. To do this, set a minimum distance a satellite may be repositioned, $minpos_i$, and a max, $maxpos_i$. In this formulation, r is the radius of the nodes from the center of the sphere, or in this case, the center of the Earth.

The distance to determine connections, $dist_{i,j,\delta}$, is calculated using two methods. First, if the two nodes are satellites, not necessarily of the same altitude, then use:

$$dist_{i,j} = \sqrt{\left(r_i - r_j\right)^2 + \left(2 \bullet r_i \bullet \sin\left(\frac{H_{i,j}}{r_i^2}\right)\right)^2} \qquad \text{(EQ 4.3)}$$

where $H_{i,j}$ is the Haversine distance between two nodes of the same radius, r_1 is the radius of satellite i, and r_2 is the radius of satellite j. As a note towards wording, altitude denotes the distance from the surface of the Earth, and radius denotes the distance from the center of the Earth. For simplification, if the satellites are at approximately the same altitude, then the Haversine Formula alone may suffice.

Along with the six homogenous GEO satellites, nine terrestrial locations are included. Seven of these locations act solely as relay points, while the other two are the gateways through which all transmissions must flow. The terrestrial locations may communicate with satellites in range; however they may only communicate directly with another terrestrial location if both sites are on the same continent. This is simply an assumption of the case study which encourages use of the satellite network by reducing intercontinental transmissions via another method, and may be altered to suit the network's operational abilities.

For simplicity, it is assumed in this case study that the cost to attack and time to repair all terrestrial nodes is the same. In the base scenario, the cost to fully destroy a terrestrial node is $1.41M, and the time to rebuild after destruction is 42 days, or 6 weeks. All terrestrial relays will follow degradation function $D=x^5$ where I is the percent of permanent destruction resources used. Meanwhile all gateways will follow degradation function $D=(1/3)\bullet(5\bullet x^2-2\bullet x^5)$. All terrestrial nodes will follow a linear repair function.

The GEOs in this notional scenario require an estimated $180M to be fully destroyed, and will follow degradation function $D=3\bullet(x)^2-2\bullet(x)^3$ which simulates an early high return on even a small allocation of resources towards degradation. Unlike terrestrial nodes, GEOs always require the full time to recover, which is assumed to be 26 weeks in this notional case study. Note that this six month period is likely an underestimation of the time required. The different degradation and repair functions were utilized to demonstrate that the nodes may be degraded in a variety of ways, and that those degradations may be modeled via a mathematical function.

Each variation of this baseline case study is processed in the long term for a duration of three years, the short term for one week, and is placed against three magnitudes of attacks. The smallest attack has a resource capacity of $1.44B, with the medium set at $2.16B and the largest set at $2.88B. These values were derived from a run with three GEO satellites with the smallest resource capacity being the minimal approximated adversity level required to reduce that scenario measure to 0 Resilience.

To reduce the size and runtime of the model, the time-dependent transmissions are averaged to a single time-independent set, and their value determined by the location from which they originate. Furthermore, the defender's value model will only be run

each time a change occurs in the network, such as a new degradation, repair, or repositioning.

In this notional example, gateways, which are acting as the sink nodes are not targetable. When permitted as targetable, and even bolstering the cost of eliminating them to equivalent with GEOs, the model, as constrained, always strikes them.

This is because with only two gateways, the cost to the Attacker to reduce flow within the network is extremely small, making any other targets, be they terrestrial or orbital, less desirable targets. Therefore, it is clear that the first step to improving any satellite network is to increase the number of gateways, or nodes which may act as gateways, and for their defense. This could, however, present other security and operational difficulties that need to be considered.

If the network is one in which many structures exist which were not required to access a gateway to proceed, such as is the case when the source or destination of a transmission may act as gateways as well or are permitted to bypass that restriction, then making the sinks in the network non-targetable is a valid alteration. This alteration is a valid approximation because transmissions are still flowing around the network, but are required to pass through a set of nodes that must always exist. As always, special and rare situations may be constructed which would call into question the validity of the method after alteration.

In such an uncommon scenario, other adaptations to the heuristic may be made, or the original model, Model 3.12, still remains valid and may be optimized as is. Both methods are valid, though the memory requirements for both are vast, and are

recommended only in extremely high cost situations. As it happens, constructing satellite networks is often an extremely high cost endeavor.

The majority of the focus on the case study is on processing the long-term model. Due to time and resource restrictions, the array of options available to short term degradations is not coded. Instead, both time periods will rely on physical events in the case study. However, the model and heuristic constructed in Chapter III are capable of operating under a conglomerate of options.

The node HQ (Bethesda, MD), which may act only as a terrestrial relay, was also labeled as a non-targetable node. This was done to exemplify that some nodes may be of such importance that the network operator will have taken sufficient steps to assure that the site would be extremely difficult to eliminate. This was the only non-gateway node permitted this "perfect defense".

Table 2 shows the major baseline parameters. Table 3 lists the aspect of the model to be altered, as well as the corresponding name associated with each variation. In Table 3, "Satellites Redistributed" denotes that the remaining satellites in the network were repositioned at equivalent intervals around the equator with at least one satellite positioned at 0^{o} Longitude. This position being filled is simply part of the case study and is in no way a requirement of the model, as seen in variation Shifted, which has no satellites positioned at 0^{o} Longitude. In this notional model, it is assumed that the orbital positions of each satellite has receives the necessary approvals.

			Linking Paramters		Degrees					
Node Name	Satellite (Y/N)	Target(Y/N)	InterSat Link (km)	Up/Downlink Radius(km)	Latitude (NS)	Longitude (EW)	Funds ($US Thousands)	Sink (Y/N)	Capacity	Rebuild
TTAC-1 (Svalbard, Norway)	N	Y		12000	80.238166	12.447236	1410	N	172000	42
TTAC-2 (Fairbanks, AK)	N	Y		12000	66.8350185	-149.65307	1410	N	172000	42
TTAC-3 (est Vancouver, canada)	N	Y		12000	51.25	-126.1	1410	N	172000	42
TTAC-4 (est Toronto, canada)	N	Y		12000	45.652527	-82.381961	1410	N	172000	42
TTAC-5 (est reykjavik, iceland)	N	Y		12000	66.1333	-24.9333	1410	N	172000	42
Sat Network Ops Center (SNOC) (Leesburg, virginia)	N	Y		12000	41.252181	-80.744541	1410	N	172000	42
Commercial Gateway (Tempe, Arizona)	N	N		12000	35.414842	-114.909319	1410	Y	172000	42
DoD Gateway (Wahiawa, HI)	N	N		12000	23.502574	-161.022938	1410	Y	172000	42
HQ (Bethesda, MD)	N	N		12000	40.98472	-80.09472	1410	N	172000	42
GEOSAT1	Y	Y	90000	7768	0	-120	180000	N	20000	182
GEOSAT2	Y	Y	90000	7768	0	-60	180000	N	20000	182
GEOSAT3	Y	Y	90000	7768	0	0	180000	N	20000	182
GEOSAT4	Y	Y	90000	7768	0	60	180000	N	20000	182
GEOSAT5	Y	Y	90000	7768	0	120	180000	N	20000	182
GEOSAT6	Y	Y	90000	7768	0	180	180000	N	20000	182

Table 6: Baseline Major Parameters

	Primary Change	Accompanying Change
Baseline 6	None	
Shifted	GEO Long -30	
Repair	Rebuild/2	
Reduced 4	-2 GEO	Satellites Redistributed
Increased 8	+2 GEO	Satellites Redistributed

Table 7: Variation Changes

Another point of interest is how the values of the transmissions are set. As the model is running on a basis of Lexicographic Goal Programming, it is important to know where the priorities are in order to assign the transmissions from those locations appropriate values.

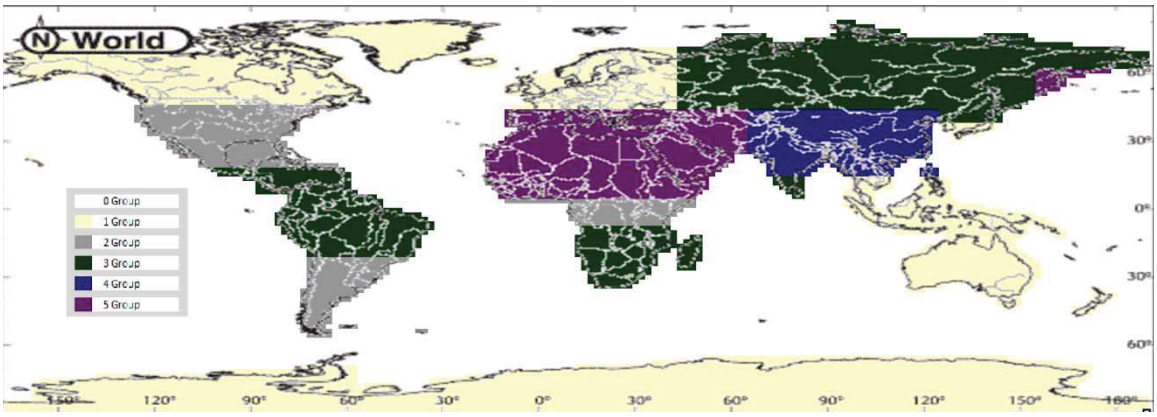


Figure 23: Location Transmission Priority

Figure 23 shows the priority of transmissions originating from the surface. In this notional case study, the greatest value is placed on those areas colored purple. Two or more separate and geographically diverse locations were used in most of the priority levels to demonstrate that the priorities need not be continuous.

Another variation of interest is the inclusion of Active Defense nodes in the network. Active Defense nodes are stations which may be used to protect other nodes and prevent degradation. While they are operational and ready for use, an attack made by the Attacker which the active defense is prepared to defend would be rendered inert, though resources are consumed. As a result of using one of these nodes, a "reload" time is incurred, leaving the nearby nodes vulnerable for the duration. The Attacker has perfect knowledge of these sites, but still may pursue to strike protected nodes.

The inclusion of Active Defense nodes is processed as a secondary variation, and as such, is only processed against the Attacker budget of $2.16B resources, and only with the original baseline case study parameters. This variation utilizes three Active Defense nodes spread relatively evenly around the world. In this notional case, the nodes are located in Australia, Djibouti, and Montana.

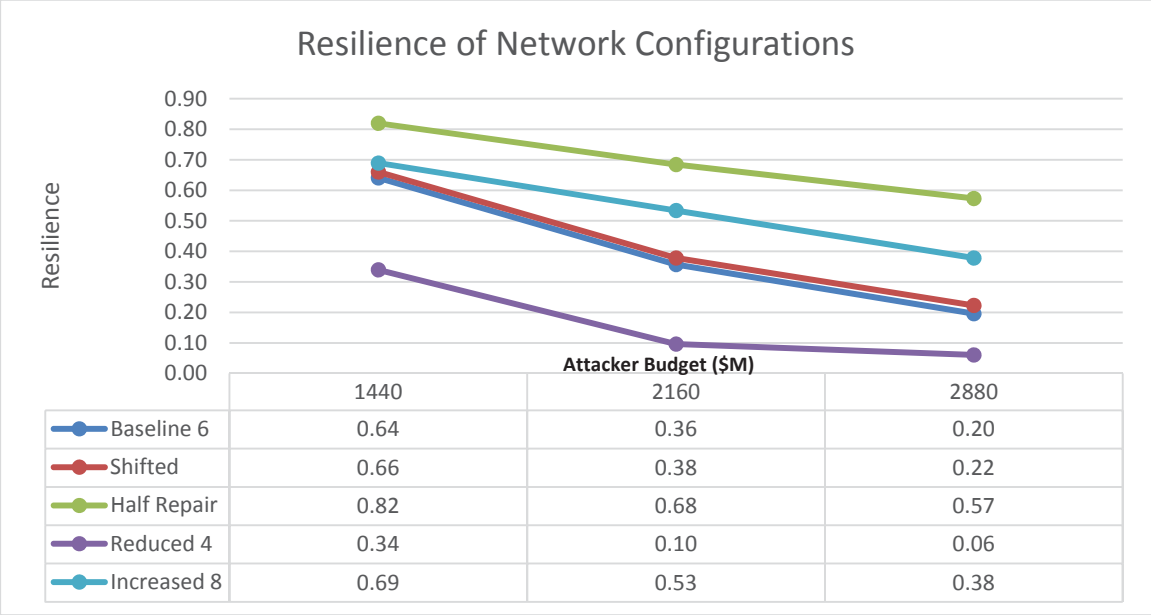
Case Study Output Analysis

The previous section presented the case study and its variations, establishing a base to perform the demonstrative resilience analysis. In this section, the outputs from the baseline case study and variations are presented and analyzed.

The baseline GEO SATCOM network consisted of six satellites, seven terrestrial relays, and two terrestrial gateways, which acted as the sink nodes. Recall from Chapter III that the output from the model would consist of a single value, which was the average network performance for the duration of the model, but that the graph of that network performance over time could be of equal or greater importance than the final number.

Figure 24 shows the resilience with the baseline case study parameters as well as the four primary variations made to the case study. The values graphed in Figure 24 are the single values under a long term model, with the vertical axis displaying the resilience measure, and the horizontal axis showing the change in attacker resource capacity, presented in millions of dollars.

As expected, the level of resilience drops in response to the increased level of adversity. This shows the inescapable nature of a resilience measure based upon extreme-event or worst-case. When utilizing a method such as this, the probability of events is eliminated, allowing exposure of events that may rarely, if ever, occur. As such, this resilience measure, like other measures developed in a similar fashion, is dependent upon the scenario, which in this case is represented by the level of resources the attacker may implement. It is for this reason that a spread of adversity magnitudes is utilized in this case study, and recommended for use in future analysis of this type.



	1440	2160	2880
Baseline 6	0.64	0.36	0.20
Shifted	0.66	0.38	0.22
Half Repair	0.82	0.68	0.57
Reduced 4	0.34	0.10	0.06
Increased 8	0.69	0.53	0.38

Figure 24: Resilience of Case Study and Primary Variations

Along with the drop of resilience in response to increased adversity, another trend can be seen, that of diminishing returns on the part of the attacker. Though the increase of resources is the same, the decrease in resilience from $2.16B-$2.88B is less than $1.44B-$2.16B.

These diminishing returns are expected as eliminating a smaller set of scattered transmissions is more difficult than eliminating the large dense pockets. Knowing this, the variations Reduced 4, Baseline 6, and Increased 8 may be viewed together. Recall the varied parameters and names of the variations are outlined in Table 6.

Having the greatest difference in the change of resilience, Reduced 4 goes from a 0.24 drop to a 0.04 drop in resilience. This is because the scenario exaggerates the situation with high value targets being eliminated, having relatively few high value targets to destroy thereafter. With the increase to $2.16B, the high value assets remain mostly destroyed for the duration of the model. Increasing the resources further to $2.88B allows the model to strike at the significantly lower value assets, further decreasing resilience.

Figure 26 displays the time-dependent network performance of variation Reduced 4 under each of the three resource capacities. The legend for this resilience overlay, as well as all that follow, is shown in Figure 25.

As predicted from the Resilience chart in Figure 24, the majority of the network performance was already eliminated in 2160, leaving very little to be eliminated in 2880. Moreover, it is clear that the degradations were carried out for the entirety of the time period. For reference, all Network Performance charts (such as Figure 26) begin with a two period warm-up of no degradation. This is so that, in a scenario such as Reduced 4, one can determine if the network performance is low because of degradation or if the network experienced imperfect network performance even before degradation. This two-step warm-up is not utilized in the resilience measure calculations though.

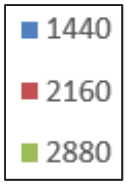


Figure 25: Attacker Budget ($M)

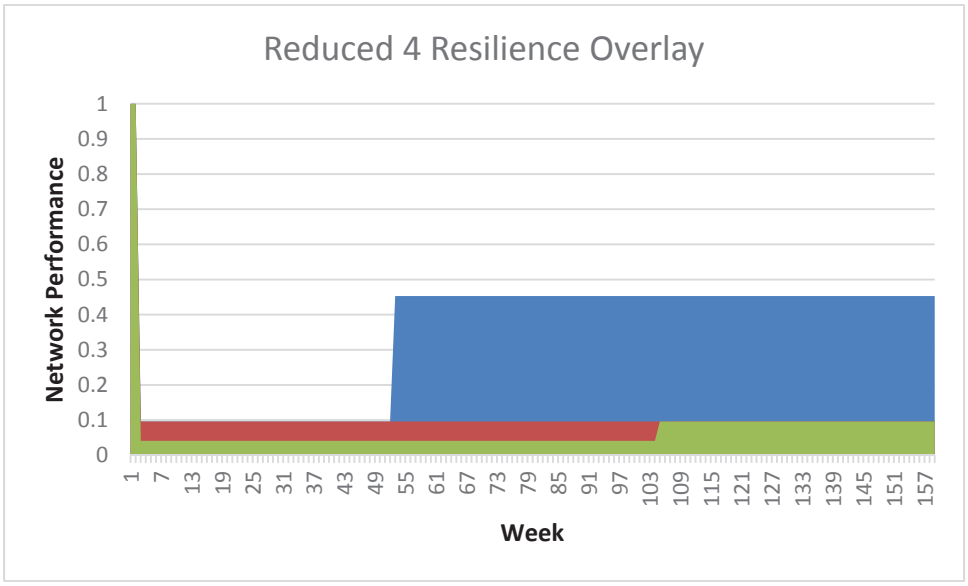


Figure 26: Reduced 4 Resilience Overlay

Increased 8 has the theoretical lowest drop, with no change in the drop of resilience from 1440-2160 to 2160-2880. This suggests that at the first two initial levels, and possibly the third, the model was still using all of its resources to eliminate the high value nodes in the notional communications network.

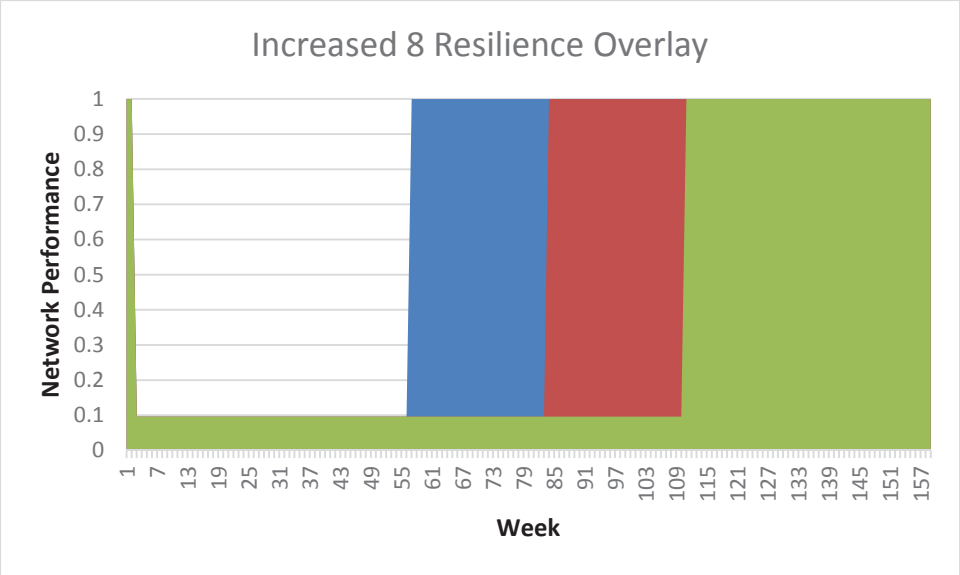


Figure 27: Increased 8 Resilience Overlay

To examine if this is the case, Figure 27 is presented. In Figure 27, it is clear that the model is degrading one or more nodes because of the sharp drop in performance. However, because of the large cost to degrade a GEO satellite, and the inherent construction of the network which, more often than not, places a GEO as a high value target, the attacker has insufficient resources to keep the degradation for the duration seen in the Reduced 4 scenario with the notional repair time.

For comparison, it is valuable to also view the connections occurring before and after a degradation, to inspect the topological effects of an event. The before and after coverage for Increased 8 are shown in Figure 28 and Figure 29 respectively. A green circle represents an active node while a red circle represents a destroyed node. The remaining colors follow the same legend as used in Figure 18, though simply put follow a basic stoplight chart.

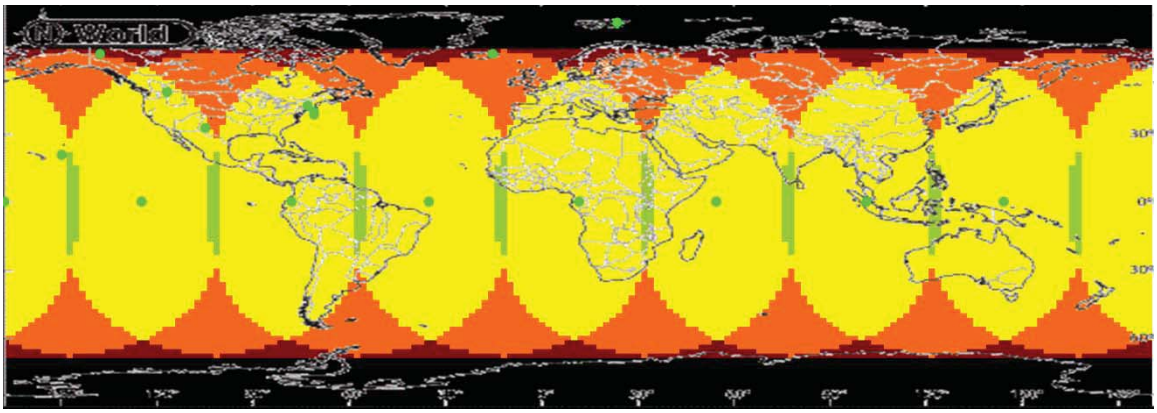


Figure 28: Pre-Event Coverage

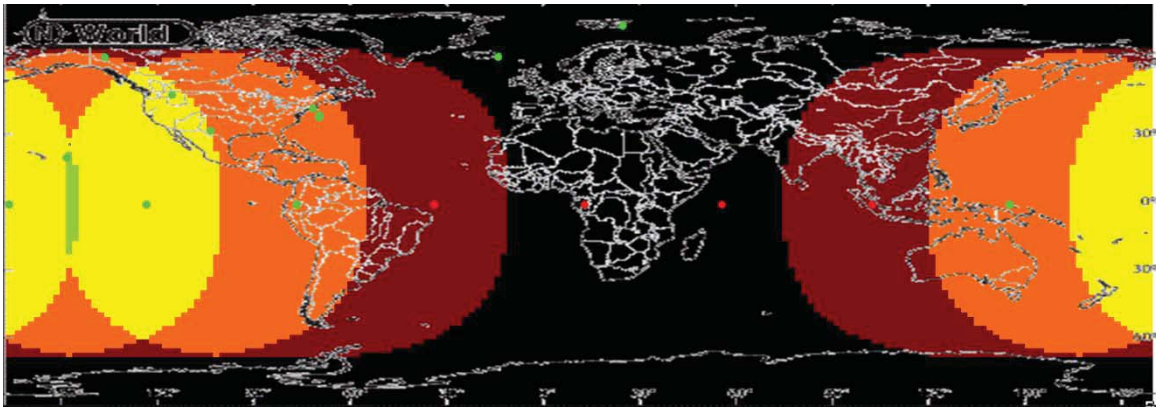


Figure 29: Post Event Coverage

Using these two maps, further insight is gained about the network. The first thing one might notice is the green dot located north of Europe. This point represents a terrestrial location, and is clearly outside of the reach of any other point in the network. A review of parameters may remove that node, or may find that it was incorrectly placed. In this case, the former is true.

Another insight gained from these two maps is the manner in which the "Attacker" behaves. Even with eight GEO satellites, each of which is very costly to remove, the model has selected the four nodes best suited to cover northern Africa and

western Asia, which include the largest plot of land associated with Priority 1 as well as roughly half of Priority 2. While the model is going after high-value areas, it is focusing on the nodes directly serving demand. The question may be asked if it would be better to eliminate the nodes servicing the gateways instead.

In Figure 28, this notion is addressed by the coverage of the gateway in Hawaii, which also is in range of four GEO satellites. The second gateway, located in Arizona, is serviced by three of those four, and so removing those GEO satellites appears to be a promising strategy. However, intra-continental terrestrial connections were permitted.

This information, coupled with the terrestrial nodes in Virginia and Maryland being covered by another satellite, which serves neither of the gateways, striking the relays or downlink nodes would require the striking of minimum five satellites to achieve equivalent or greater degradation within a single time epoch. Even if the Maryland node was not assumed indestructible in the example, the cost would still require destroying two additional terrestrial nodes, and keeping them destroyed for the duration.

Doing so, the model could achieve a greater raw impact, however the cost of doing so is prohibitive. Instead, the model selected four GEOs which provided a greater reduction per dollar expended for this notional data set.

Baseline 6 and its Shifted variation fall in the middle of this change in resilience drops. Losing 0.28 and then 0.16 for the increase of $1.44B to $2.16B and $2.16B to $2.88B attacker resources respectively, both are able to remove a material amount of the high value network assets and move on to lower valued assets; however, the trend suggests that there was still room for improvement event at the $2.88B level.

Figure 30 and Figure 31 show the Baseline 6 and Shifted network performance over time respectively. Analyzing any one of them independently shows something slightly different than what was predicted from the resilience measures alone. Based on those measures and trends seen in Figure 24: Resilience of Case Study and Primary Variations Figure 24, it was thought that the high value assets may have been degraded already and that the model had moved onto some lesser nodes. However, as can be seen in the specified figures, this is not the case.

Instead, there exist time steps in which high value nodes may continue to be degraded. The most enlightening portion of this graph to a first time viewer is the change that occurs in $1.44B at approximately time step 55. Before that jump, more than 90% of the network performance is eliminated, and yet afterwards less than one quarter is. To examine why, the degradations are inspected.

In Baseline 6, three GEOs, those located at -60^{o}, 0^{o}, and 60^{o} latitude, are fully degraded from timestep 1 and until timestep 54. At timestep 55, the -60^{o} GEO is reactivated and remains so, leaving the other two degraded until timestep 82. In Shifted, a similar event occurs, only now with GEOs shifted $+30^{o}$ latitude. This change aligns with the jumps in network performance seen in Figure 30 and Figure 31 respectively.

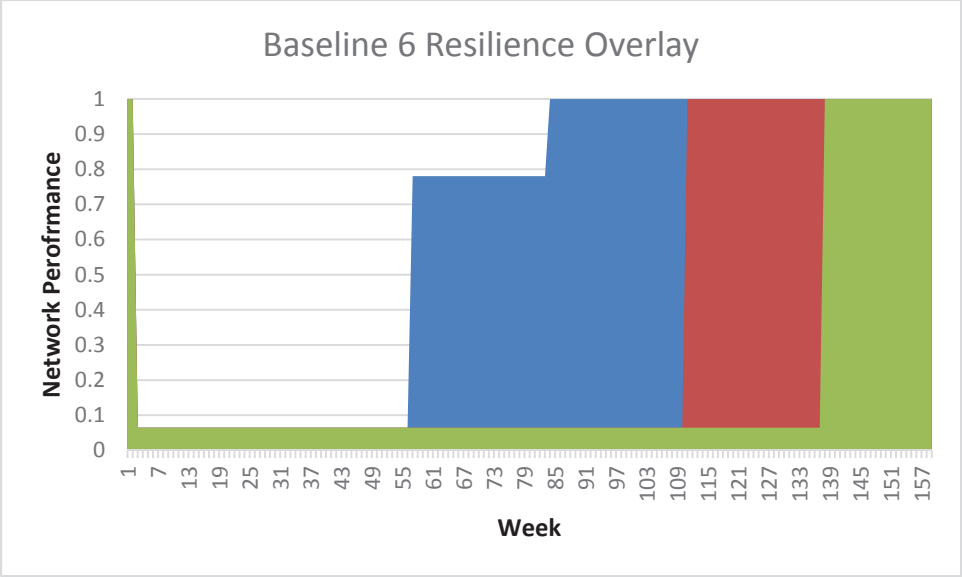


Figure 30: Baseline 6 Resilience Overlay

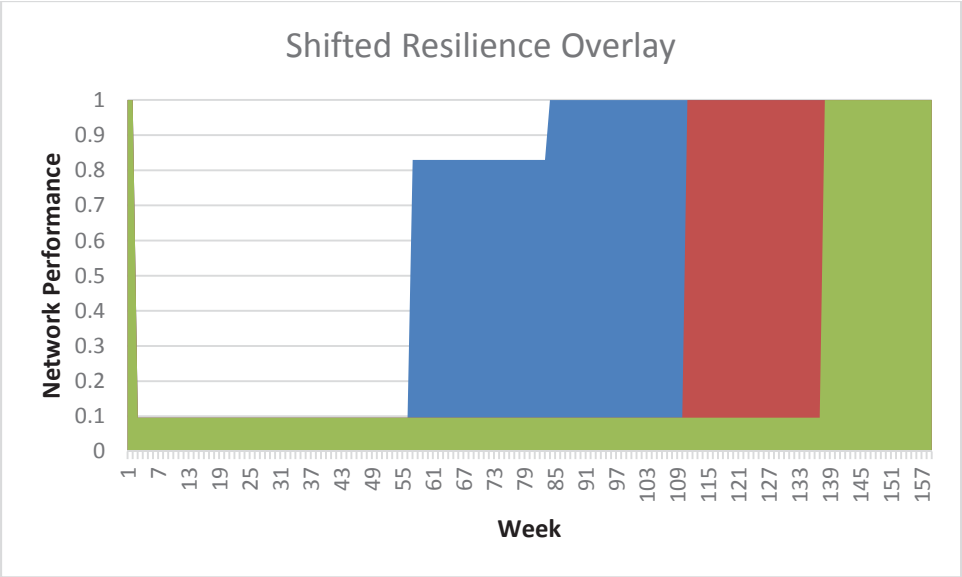


Figure 31: Shifted Resilience Overlay

If the effects of degradation were simply additive, then one might expect the time period spanning from 55-82 to have a more significant loss of performance as two nodes remain destroyed. Especially when compared to the time period preceding this when only one more was destroyed.

However, this is a common phenomenon in networks known as synergy. From the point of view of the defender, keeping any one of the three GEO satellites operational provides a safety net in case of unseen events. The attacker may attack any one, leaving two to fill the gap and maintaining roughly 0.95 of the network performance. Given the resources to eliminate two nodes, the network value plummets to 0.78. After eliminating three, this drops even further to 0.06.

This synergy is exemplified in the variation Repair, whose network performance over time is shown in Figure 32.

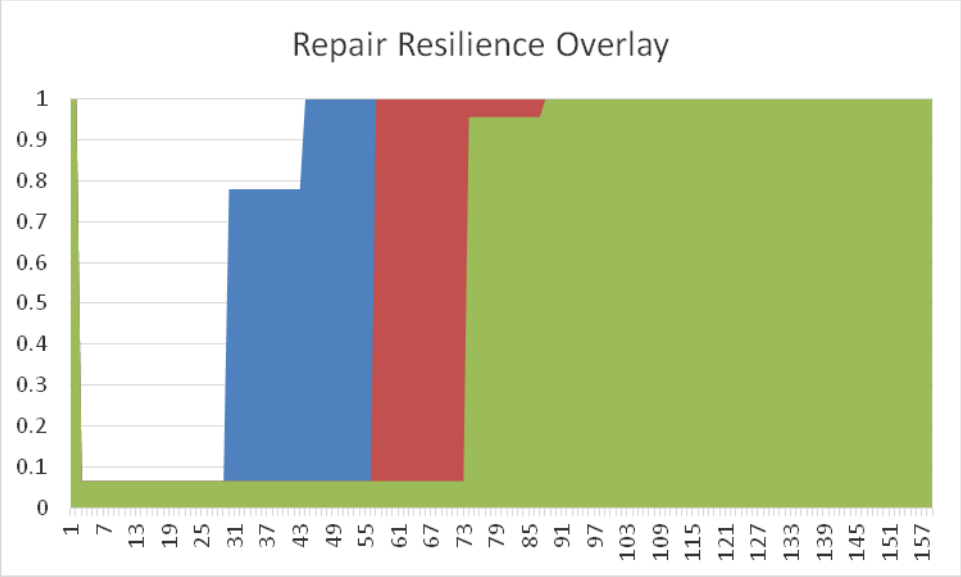


Figure 32: Repair Resilience Overlay

In the variation, Repair, the time to repair the nodes was cut in half, all else remaining equivalent to Baseline 6. Intuitively, one might believe that the degradations would simply be left shifted, the resilience doubled. However, it can be seen that the model and heuristic focused on drawing out the triplet of degraded nodes for as long as possible.

Viewing the Repair variation resource capacities of 1440 and 2880 together, four distinct plateaus are seen. The lowest occurs in the majority of the scenarios involving six GEOs, and occurs when three are simultaneously eliminated. The second lowest was seen in Baseline 6 and Shifted, and occurs when only two are eliminated. The third lowest only occurs in Repair when a single GEO is destroyed, and the fourth is the ever present baseline that the network restores itself to. The levels are shown in Table 8.

GEOs Simultaneously Eliminated	Resulting Network Performance
0	1
1	0.956
2	0.780
3	0.065

Table 8: Repair Variation Network Performance Under Adversity

From 0.045, to 0.176, to 0.715, the change in loss of performance is clearly not linear. The synergistic effects can be best seen in expanding the number of simultaneously degraded nodes. However, notice that the Attacker never pushed beyond degrading three nodes in the baseline case study. Synergy is common in many distribution networks. The fact that the model was able to locate and exploit this effect increases the validity of the method.

As the heuristic tested degradation combinations, it was common for the model to strike six or seven nodes simultaneously. In the best case though, the one which provided the most degradation over the model duration, the Attacker stopped at three nodes. This is because of diminishing returns of the resilience measure as previously discussed and

the time-dependent nature the model is allowed to pursue. Were the model only permitted to strike once, then there would be no reason to consider the efficiency of the strike and it may have continued pursuing a full assault on the network. Instead, it was better to only degrade up to the maximal return vs cost, and then save the remaining resources until the next time the node or nodes were activated again.

The Figure 32 shows the state of the Repair network under these three plateaus, beginning with the greatest loss of performance.

The inclusion of Active Defense nodes is also a point of interest for many network operators. As the reader will recall, these special nodes may be used to protect other network components, but at the cost of a reload time, during which those components are once again vulnerable.

This secondary variation on the case study resulted in a resilience of 0.52. Recall that the same network against the same level of adversity without these Active Defense nodes had a resilience of 0.36 (see Figure 24). Including Active Defense increased the resilience of the notional network by 45%.

To view how Active Defenses influenced the Attacker, consider Figure 34. Originally, the attacker could keep the network at a degraded level of performance for 108 weeks (see Figure 30). However, because the Attacker needed to force the Active Defenses to fire before taking advantage of the short window of vulnerability, this degradation could only be continued for 82 weeks.

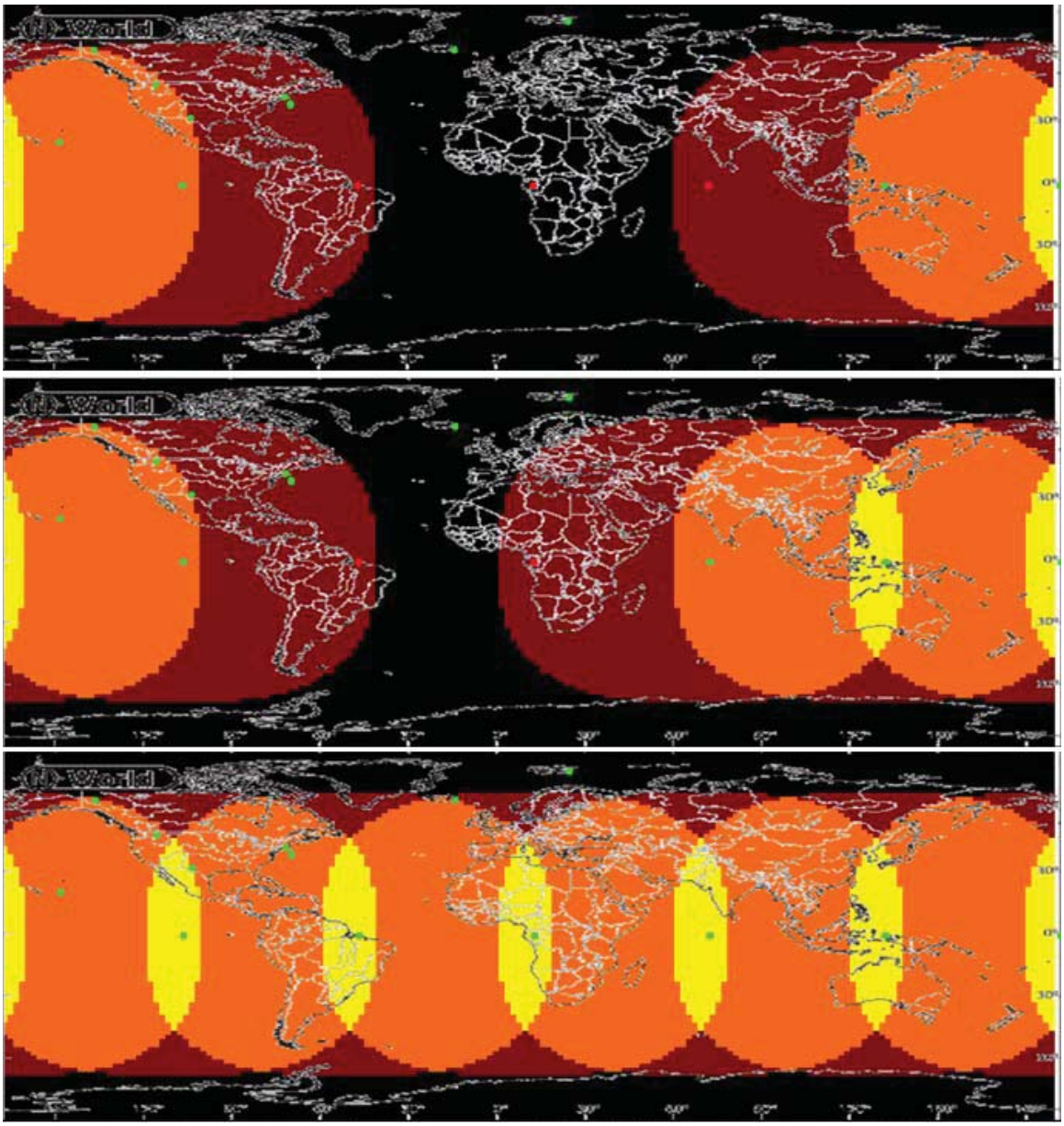


Figure 33: Degradation Progression of Repair 1440

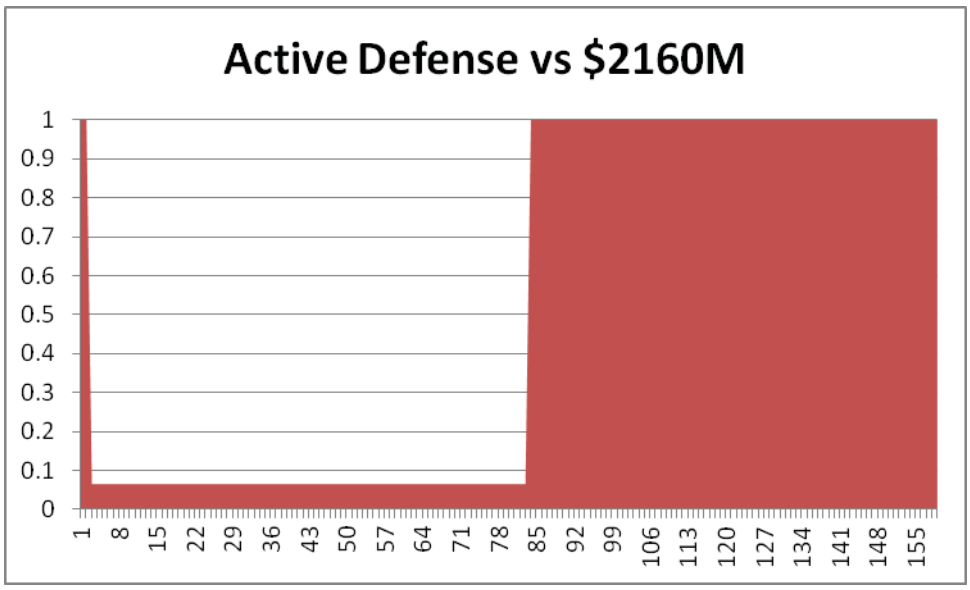


Figure 34: Active Defense vs. $2.16B

Therefore, for the case of the pure geosynchronous satellite communications network used in this example, the inclusion of Active Defenses did increase resilience. In this case study, and with only three such nodes, the increase was material.

For a numeric comparison of resilience measures, the same assumption of one single strike which Salmeron *et al.* made in their "long term models" is utilized against the Baseline case in the study. In this situation, with all else equivalent, the model is capable of eliminating all attackable nodes, reducing the network performance to nearly 0. However, the terrestrial nodes are rebuilt after six time steps and the GEOs in 36 time steps.

In this situation, the attacker required only $1089.87M to eliminate the network, but it is not capable of utilizing the remaining funds. The resulting network performance chart appears in Figure 35. The resulting resilience measure is an impressive 0.822.

While this number may encourage a network operator, the reader is reminded that the Attacker is seeking to minimize the network performance. Comparing models when that with the Single-Event assumption is incapable of effectively utilizing the full resources may not be a fair match though, so a test is run with $1090M. The results of the baseline case study without the single-attack assumption are shown alone in Figure 36, and alongside in Figure 37.

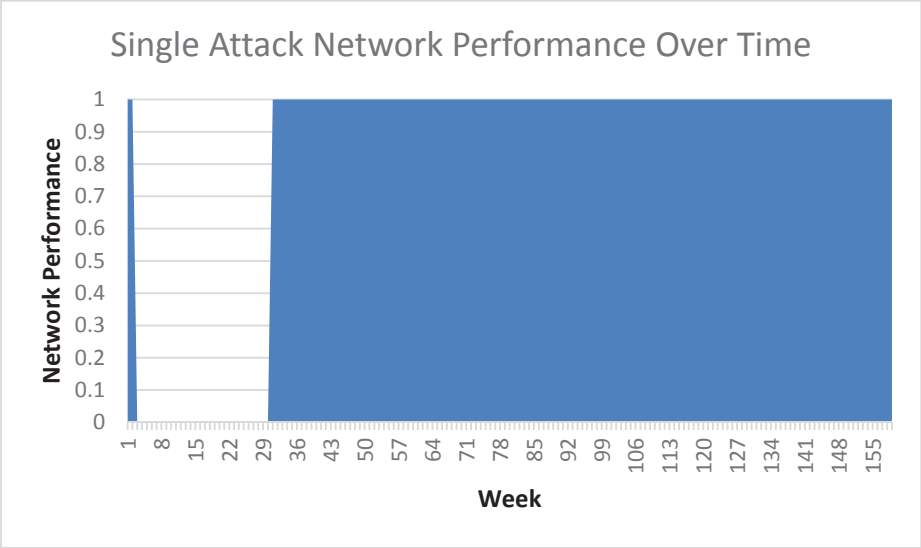


Figure 35: Single Attack Network Performance Over Time

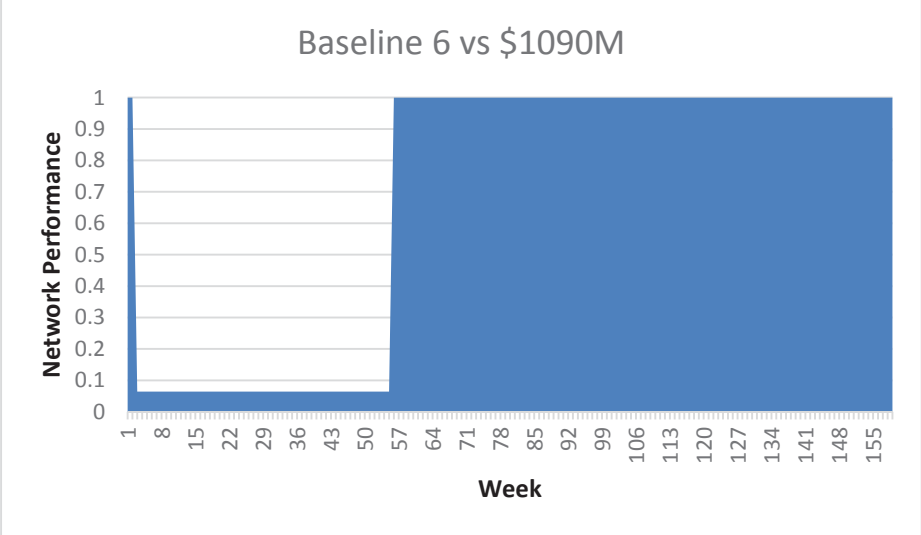


Figure 36: Baseline 6 Against $1090M Attacker

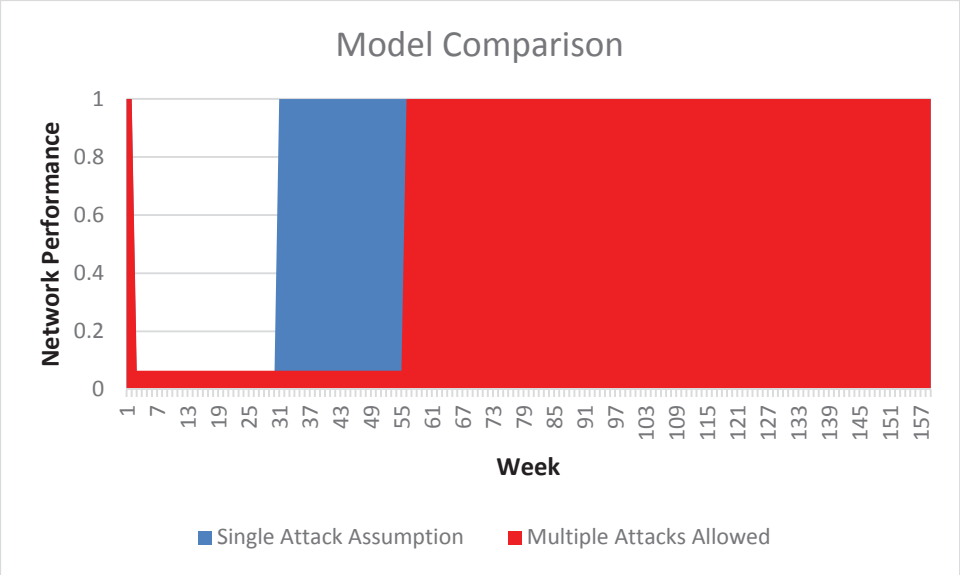


Figure 37: Comparison of Models

From the start of the model and until the GEO nodes are first able to be rebuilt and operational at time step 28, Multiple Attacks performs worse Single Attack at a comparison of 0.06-0.00. For the next 28 time steps, when the Single Attack model shows the network fully operational once more, the comparison changes to 0.06-1. After time step 55, they are equivalent.

While higher network performance is desirable, attaining such a value due to inadequacies in the model is not desirable. The purpose of the model is to degrade the network performance as far as possible. Thus a lower, feasible value, with no change to the system, is preferable and more accurate. With a simplifying Single Event assumption, the model is incapable of locating a much more significant impact, thereby inflating the resilience measure.

Already it can be seen that repetitive strikes, even under equivalent conditions, may prove to be more troublesome than a one-hit attack. To determine how much more troublesome, the resilience under each condition is calculated. Recall that the

assumptions made by the model developed by Salmeron *et al*. model would have reported a resilience of 0.822 under the measure used here (Salmeron). Instead, the true resilience is closer to 0.684, and could be even lower than that. This is an improvement of model performance, and by extension accuracy of the calculated resilience measure to that proposed, of at least 77.5%.

To calculate the improvement, the simplest way is to calculate the percent difference in the calculated measure and the actual measure. Because of the simplifying assumption of only physical attacks, as well as the removal of network reliability from the modeled measure, the true resilience is expected to be lower than that calculated. The lowest the resilience can be is 0. Thus, the lowest possible increase in accuracy is:

$$\text{Accuracy Improvement} = \frac{1-R'}{1-R} \bullet 100\% \qquad \text{(EQ 4.4)}$$

In EQ 4.4, R is the model with the simplifying Single Event assumption and R' is the resilience of the model without the simplifying assumption.

Due to the restrictions placed on the case study short term attacks, the only method available to the Attacker model was the use of physical degradation. With relatively large resources, it would be expected that the model would strike the synergistic nodes, and then continue with the remaining nodes, either locating lesser combinations, or single high value assets.

This effect was seen on the short term model runs, which were conducted on the baseline case study. The highest level of adversity was set to $800M Against this level of adversity, the resulting resilience measure was 0.0589. Note that this is only a drop of

0.0061 compared to when only the three best nodes are degraded, exemplifying the diminishing returns the Attacker must cope with.

The next short term attacker level run was $500M. This value is low enough to preclude a full degradation of the desired triplet, but high enough that it may still pursue a two GEO combination and either partially degrade another, or select terrestrial nodes instead. Another option theoretically available is to partially interdict each node of the triplet, however the heuristic developed by Lim and Smith forces the full degradation of the preferred nodes first and utilizes the remainder to partially interdict the last.

With $500M resources, the best combination the model was able to locate was to fully degrade GEOSATs 3 and 4, which were two of the triplet, refer to Table 2 for locations. For the partial degradation, the node selected was GEOSAT 6, which was not part of the commonly hit triplet. This plan resulted in a short term resilience of 0.448. To determine if GEOSAT 6 was a throwaway node or actually of value, the reader is referred to Table 8, in which the best 2-node resilience of 0.78 is listed. This means that even partially degrading GEOSAT 6 provides an additional 0.34 impact when used in junction with GEOSATs 3 and 4.

Because these models were run with physical degradations and at the short term level on a GEO network, there is no need for a network performance chart as it is a single plateau, making it equivalent to the resilience measure. However, in a network with time-dependent locations or transmissions over the course of a single cycle, this chart may still show multiple plateaus and valleys and provide valuable insight.

In this section, the output from the case study and its variations were analyzed. The results showed a 77.5% in attacker performance due to recurring attacks over the

same method under previous works' assumption of single event. Moreover, the Attacker was capable of exploiting synergistic degradations, both in the long term and short term models, increasing the validity of utilizing a modified heuristic based upon Lim and Smith's.

Analyzing Costs

In this section, the costs associated with the results of the Case Study and its variations are analyzed.

One major parameter which was not altered in this study was that of the resources required to degrade or destroy nodes. Recall from Table 7 that the cost of destroying a GEO satellite is set at $180M. This value is constructed primarily from the cost required to launch a satellite into GEO orbit.

> *... the Atlas 5 and Delta 4's potential, with the launchers selling for more than **$160 million** in the last two years to government customers like the Air Force and NASA.*
>
> *Clark*

However, this cost was increased for the inclusion of a missile borne into orbit on the rocket. The costs were estimated based upon the Standard Missile 3 unit cost of $13.4M (Eshel p. 1), bringing the total cost to roughly $173.4M. This number was rounded up to an even figure, $180M.

For the destruction a terrestrial node, the cost of a Tomahawk cruise missile was utilized. The total cost of each missile is reported to be $1.41M (Weinberger p. 1). However, these missiles are highly accurate, having a hit rate of 85% during the Gulf

War, and carrying a payload strong enough to strike even heavily defended targets (Weinberger p. 1).

Though not used as decision variables in the case study, there were also costs associated with the Defender's responses. The largest of these was the cost to build and launch replacement GEOs. Looking to the most recent DoD GEO satellites developed, the Advanced Extremely High Frequency (AEHF), the cost to simply build a replacement is $975M (Lockheed p. 1). It then must be launched into orbit, the current method being to deliver it with an Atlas V (Lockheed p. 1), which has already shown a cost of over $160M. For this demonstration, each AEHF that must be replaced is estimated to cost $1135M.

Furthermore, the cost of Active Defenses must be inspected. Utilizing the US missile which was used in an ASAT capacity, the Standard Missile 3 is the clear choice to intercept a physical attack aimed at an orbital component. As previously stated, the cost of one such missile is $13.4M (Eshel p. 1).

Recall Figure 34, in which the time-dependent network performance for the Active Defense variation was presented. In this setting, Active Defenses were utilized three times, each one stopping an Attacker strike costing $180M in the demonstration. Each of these uses was modeled at a cost to the Defender of $13.3M.

However, under the same resources without Active Defense, the Attacker was able to strike an additional three GEOs. As the cost of replacing a GEO is estimated at $1135M, this means in the scenario that expending $40M allowed the Defender to waste $540M of the Attacker's resources while simultaneously saving $3405M. In this section, the costs associated with the case study were presented and explained. The cost effect of

the inclusion of Active Defenses for both the Attacker and the Defender were calculated, showing that the rewards of using such defenses far outweigh the costs. It should be noted, however, replacement costs do not represent the potential value of the lost communications. These could be added if they are quantified.

Heuristic Performance

In this section, the performance of H-1, the heuristic developed in Chapter III and based upon Lim and Smith's heuristic, is analyzed. Along with its running efficiency, its strengths and weaknesses are reviewed, as well as interesting aspects discovered while processing the case study.

To begin, the computer and programs which all of the scenarios were modeled on have the following specifications:

Model:	Windows 7 x86 AFIT LAB Image v1.2
Processor:	AMD Athlon™ II x2 215 Processor 2.70 GHz
Installed Memory (RAM):	4.00 GB
System Type:	64-bit Operating System
Solver Software:	Lingo 11© 2008 Lindo Systems Inc.
Coded In:	Microsoft Office Excel 2007 VBA © Microsoft

The average time to run a long term pure GEO model was 35.18 minutes with a standard deviation of 10.28 minutes. The minimum time across all long term runs was 15.8 minutes, which occurred on Reduced 4 against 2880 Attacker resources. The

longest run time was 51.25 minutes, which occurred on Increased 8 against 2880 Attacker resources. With all of the other variations of the case study falling between these two values, it is possible that the run time is correlated to the number of nodes in the network. This observation is in-line with the method in which the heuristic selects its targets, as well as the increased memory requirements for larger models discussed in Chapter III.

The average number of iterations requiring processing for the long term model, which would be modeled over10 runs, was 7.167, with the mode being 8 runs. This means that most of the runs exhibited 8 degradation variations which differed from the preceding iteration in some manner. However, this in no way guarantees that the 8 runs were unique. On average, the best degradation method was located in 4.167 iterations, with the mode being 4.

The average run time per iteration was 5.00 minutes. This number does not take into account the increased time requirements for performing the local search which occurred every twice in every run. It should be noted that, in the long term models, the best value was always returned by one of these searches, though it did not always occur on the first one.

In the two short term model runs, the average time to complete a pure GEO one attack method model was 6.09 minutes. Interestingly, the half resources, $500M versus $1000M, required almost double the time. This is likely due to the reduction of options which $1000M permitted. With a resource capacity so close to the amount required for full interdiction, the focus changed to what not to attack instead of what to attack. In the

$500M case, the resource capacity was fairly restrictive, forcing the model to determine what the best combinations were.

The amount of time required per iteration was fairly close though, with the average time at 0.79 minutes, the max occurring in the $500M case at 0.89 minutes. It was determined unnecessary to perform a local search on the short term scenario in a case of physical interdiction as any event of that type would be preferred front loaded. Surprisingly the $1000M variation's best combination was located on the initializing iteration, while the $500M best option was located on the 7th iteration.

As noted earlier in this chapter, the gateways were denoted as non-targetable. When targeting of gateways was permitted, the model selected them for destruction every time, even without the local search.

With the gateways unable to be targeted, the focus primarily stayed with the satellites in the network. However, when excess resources were available, the terrestrial node most commonly attacked was TTAC-2 (Fairbanks, AK). This node was selected for interdiction so often because it was commonly used as a relay point, being on the same continent as a gateway and thus having a connection to the sink nodes. However, because the satellite servicing this station also directly serviced Commercial Gateway (Tempe, Arizona), destroying the node had little to no impact in the notional case.

Destroying the servicing GEO also had little impact because of the four relay nodes, TTAC-3 (Vancouver, Canada), TTAC-4 (Toronto, Canada), Sat Network Ops Center (Leesburg, Virginia), and HQ (Bethesda, MD), and the large amount of overlap able to cover said nodes in the demonstration. As such, the model commonly chose to

instead focus on the satellites showing the highest uplink value, always finding its best value in such a situation.

An inherent weakness to the heuristic, H-1, comes from the heuristic it was based upon. H-1 was not structured to overcome the basic operational behavior of allocating full resource requirements to degrading nodes in the order they were selected. As such, it may suffer a disadvantage when the optimal solution is to partially degrade a set of nodes.

The heuristic also shows a preference for front-loading the degradations instead of spreading the degradation more uniformly over a period of time. This was the motivation for the inclusion of a small local efficiency search in the heuristic, which in the case study was processed every fourth iteration. With the inclusion of this efficiency search, the performance of the heuristic is greatly improved, however it comes at a sacrifice in run time.

One strength of this heuristic is its ability to locate the combinations of nodes which provide the greatest drop in network performance. This result was observed many times over, being greatest exemplified in the variation Repair and in the short term model run against $500M. Both in the situation of a set containing full degradation, and a set with one partial degradation, the model exploited the redundancy of network nodes to yield a synergistic attack plan.

Other strengths of this heuristic are its ability to be run on a personal computer for relatively small networks as well as its ability to locate the strengths and weaknesses of a network. Though the runs commonly required a half hour or more, the time requirement to run a model is far outweighed by its ability to locate the strengths or weaknesses of an

extremely costly network. In the case study, the network's weakness was its small number of GEO satellites, which could be destroyed and kept destroyed for a long time span. However, it showed the strength of a network which had, in most of the variations, anywhere between two and five satellites covering a high value location. In that situation, losing any single or pair of satellites had a relatively small impact.

The greatest strength of the heuristic, though, is its adaptability. From its original version as developed by Lim and Smith, to the version developed in this thesis based upon the bi-level model, and even for the addition of imbedded local searches to improve its performance, the heuristic is capable of being altered to suit the situation.

In this chapter, a notional GEO SATCOM network was created to utilize as in a case study to demonstrate the model developed in this thesis. In calculating the resilience of this notional network, it was determined that the relatively few gateways within the network created a key point of concern for transmission flows in adverse operational conditions. After assuming these points to be non-targetable, multiple variations and levels of adversity were employed so as to view the behavior of the model and the heuristic developed in Chapter III.

V. Summary and Conclusion

Contributions

In this chapter, the contributions made by this thesis to the area of resilience research are summarized. The contributions include the measure itself, which is time-averaged network performance under an extreme-event situation, the Attacker-Defender model, which accurately models the majority of options available to the operator of a satellite communications network, as well as an intelligent adversary seeking to degrade said network for one or more reasons, and the heuristic used to solve the bi-level program.

At the beginning of this research, it became clear that selecting a definition for resilience, from the many that exist, was required if a measure was to be attained. After a review of many definitions and their underlying components, the definition developed by Argonne National Laboratories was selected:

> *Resilience is the ability of an entity -e.g., asset, organization, community, region- to anticipate, resist, absorb, respond to, adapt to, and recover from a disturbance.*
>
> *Argonne National Laboratory* (Carlson p. 7)

With this definition of resilience in mind, previous measures utilized in distribution and transportation networks were analyzed. As a result, two key methods

came to the forefront. The first was that of "Worst-Case Interdiction", with major works conducted by Salmeron *et al.* and Brown *et al.* (Salmeron 2004), (Brown 2005), (Brown 2006), (Salmeron 2009). The second was the inclusion of probability and risk in a time-dependent situation, which was seen in the work of Klibi and Martel (Klibi).

Both of these methodologies provided a valuable component to the extreme-event time-dependent network performance developed in this work; however, each still had weaknesses. The work performed by Klibi and Martel, which focused on the probability and severity of events, underplayed the potentially disastrous situations, while the work performed by Salmeron *et al.* was lacking in the handling of a time component which underpins resilience (Klibi), (Salmeron 2009). Building on key elements of both works, it was determined that any degradation must capture the worst-case scenario, but must do so in a manner which exemplifies the role of time in a measurable fashion.

As such, the measure developed in this work is the time-averaged expected network performance under extreme-event degradation. To capture the worst case event and the best case response, a bi-level mathematical program known as an Attacker-Defender model was constructed. As the measure is being applied to a satellite communications network, the nuances and vulnerabilities of a satellite network were included in the model, increasing its validity. To reduce the model's size, minor optional sacrifices were made to its fidelity and accuracy, such as aggregation of transmissions.

From review of the previously developed measures, as well as long term and short term vulnerabilities, it was clear that along with the time component, there was a distinct difference between objectives existing in each time span. As such, a second model was developed, extending the first. This resulted in two variant bi-level programs, one which

models the short-term Attacker and Defender options, and a second which does the same for a prolonged period of time.

Finally, a heuristic developed by Lim and Smith to solve bi-level models was modified for use with this model (Lim). The models and the heuristic are presented in Appendix A.

Methods for analyzing the outputs of the model, which include a time-dependent measure of network performance, were presented. Using the time-dependent network performance, the reader was shown how the strengths and weaknesses of the network could be located and presented. Furthermore, the synergies available to the attacker and redundancies existing in the network are highlighted by the outputs, allowing the analyst and network operator to focus their attention to these commonly less-visible combinations.

To show the performance of both the model and the heuristic, a case study consisting of a notional global satellite communications network was analyzed. In this case study, variations and sensitivity analysis were performed on the major parameters of the network. This method showed how processing multiple scenarios under identical threats allows a comparison of the resilience of differing networks, as well as showing which network configuration is the most or least resilient.

Furthermore, a secondary variation on the Baseline case study in which Active Defense nodes were included was performed. The review of the literature did not reveal any previous measure of resilience which includes such a network component; previous efforts restricted themselves to passive or continuous defenses. The inclusion of Active

Defenses showed an improvement to the long term resilience of approximately 45% for the specific example case, suggesting the effect may be generalized to other settings.

Along with the outputs from the case study, the performance of the heuristic was analyzed, increasing the understanding of the operation of the network, the model formulation, and the behavior of the heuristic. It was noted that the key strengths of the heuristic were its adaptability, its ability to locate synergistic degradation combinations, and its exploitation of network weaknesses.

Future Research

Throughout this thesis, strengths and weaknesses were presented and addressed, followed by improvements and compromises made where necessary to provide an operational model. Any compromise suggested, especially in regards to accuracy, may be optional, depending upon the availability of a high performance computer. However, as with any modeling efforts, improvements are possible. As such, major areas of future research are presented in this section.

The heuristic used in this thesis operates on a basis of linear constraints. Since this is an approximation, it may be improved upon to better handle the non-linear satellite communications network. Research may be conducted into modifying this heuristic for a non-linear situation, or to develop a follow-up local search to do the same. One possible method for permitting such an improvement can originate from utilizing low-adversity

scenarios, extending the degradations and adversity levels together. Such a search may provide a different set of nodes than seen with higher levels of adversity.

Another potential improvement to the model may come in the form of Dynamic Programming. Dynamic Programming permits successive decision making, which is essentially what the model is striving to optimize. By converting the model into a Dynamic Programming representation, it may be possible to locate an optimal solution, which currently is not possible with the prescribed Attacker-Defender model and known BLP solution methods. In addition, the use of Dynamic Programming would allow for locally optimal policies at different points in time, which could be used for network planning purposes.

The resilience measure calculated in this thesis is currently a deterministic value, but note the measure itself was proposed as:

> *Resilience is measured as the time-averaged expected network performance under extreme-event degradation.*

The inclusion of probability into this measure exemplifies the need for inclusion within the modeling methodology itself. A simple addendum may include the reliability of the final degradation combination based upon the time, location, and type of attack. However, a more in depth continuation may consider how reliability analysis may be included into the model itself to improve accuracy and validity. Such reliability-based modeling is used in other areas of design and analysis, such as aerospace and structural design.

Another aspect of resilience measuring which is of interest to many operators is deciding how to improve the resilience of their network. In this thesis, a simple analysis

of the outputs was presented which could provide minor insight into the weaknesses of the model. Further research into evaluating options to improve their network would be of value. This may include the adaptation to a tri-level model, or a method for testing a variety of situations for improvement under a cost constraint. The model proposed here could be used to identify potential weaknesses while a second model might locate options available to mitigate the vulnerabilities. This also indicates the potential for Dynamic Programming, since as the levels of modeling increase, so does the complexity of multi-level solution approaches.

The coding which processed the heuristic and the model may also benefit from increased attention. Improving the coding for reduced memory consumption, eliminating redundant commands, or containing the analysis within one program instead of three could reduce the run time required to process a network. Many of these improvements could potentially be achieved by converting the coding to a new, more robust language.

Due to the size and time required to calculate the resilience for a small geostationary satellite system, the LEO scenario was not run, even though the model, heuristic, and coding are capable of analyzing a larger system. Based upon current case study variations, it is believed that a LEO system would require in excess of 12 hours per iteration on a personal computer. This time requirement is likely acceptable for network operators and architects of multimillion dollar satellite systems. However, due to time restrictions, such a scenario was not processed in this work. Future research could analyze a LEO satellite network, making improvements to the code or heuristic where necessary, just as improvements were made to the code to improve GEO run time.

The potential future research stemming off of this thesis will serve to advance aspects of theory and application. Various topics include heuristics analysis and construction, multi-level mathematical programming, reliability theory, risk analysis, and computer programming. Other questions, topics, and methods for improvement exist in dealing with resilience. This suggests a rich area for future work.

Appendix A: Models and Heuristics

Model Index Change

Short-Term Model:	Primary Time is Short Term Interval (STI)
	Secondary Time is always 1
Medium/Long-Term Model:	Primary Time is Long Term Interval (LTI)
	Secondary Time is Short Term Interval (STI)

Indices

i, j, k	All three are to be used to denote the node.
t	Source node of the transmission
l	The network system the transmission last experienced
s	The security level of the transmission
d	Time (Primary)
∂	Time (Secondary)
δt	Duration of a single STI
a	Attack type
e	Active Defense node
r	Orbital Radius {0,1,2}

Variables

$y_{i,j,t,l,s,\partial,d}$	Flow of transmission from node t currently flowing from node i to node j with security level s and node i operating in network system l
$\partial y_{i,t,s,\partial d}$	Amount of unsatisfied demand at node i with security level s
$\text{TCactive}_{\partial,d}$	Binary variable noting whether an active TC station exists at time d,∂ (1=yes, 0=no)
$\text{Excess\_Cap}_{s,\partial,d}$	Amount of capacity of security level s in the network that remains after all demand has been filled at time d.
$\text{Excess\_Value}_{\partial,d}$	Value of remaining capacity in network at time d
$Excess\_Cap\_PCT_{\partial,d}$	Percent capacity remaining in Network at time d
Total_Net_Value	Total value of the attempted calls in the network
$\varphi_{\text{i,a,g,d}}$	Amount of resources of type g to be used against node i in attack type a $\left[0, RqRsc_{\text{i,a,g}}\right]$
$\omega_{i,a,d}$	Binary Variable denoting an attack of type a against i at time d
$\text{ADFired}_{e,a,d}$	Binary Matrix specifying if defense e,a was used at time d
$x_{i,\partial,d}$	Degradation of node i at LTI time d
$Fuel_i\left(DistPos_i\right)$	Fuel burned by node i given the distance repositioned
$\text{empty}_{i,d}$	Binary variable denoting if node i is out of fuel at time d
$RepoD_i(DistPos_i)$	Time to reposition node i given the distance repositioned
$RepoS_i$	LTI time node i begins its repositioning
$DistPos_i$	Distance repositioned by node i
$\text{Rep}_{i,d}$	Binary matrix signalling a repair for node i at time d

Parameters:

$b_{i,t,s}$	Supply/Demand at node i from source node t with security level s
$c_{i,t,s}$	Value of transmission to node i
$cap_{i,l,s}$	Capacity of node for network type l and security level s
$x_{i,\partial,d}$	Operational Status of node i at time d,∂
TC_i	Vector denoting if node i is a Tracking and Control station
$Network\_Value$	Sum of network capacity multiplied by value of security rating
ec_s	Value of excess capacity with security level s
$Conn_{i,j,\partial,d}$	Binary parameter denoting the connection between node i and j at time interval d,∂
$costO_i$	Cost of operating node i for 1 time step
$costR_{i,a}$	Cost of reparing node i from attack type a
GlobeMax	Global capacity of network
RegMax	Regional capacity of network
CostMax	Maximum possible cost resulting from a probable event
$AdvRsc_g$	Adversary Resources of type g
$RqRsc_{i,a,g}$	Required Resources of type g to eliminate node i via attack type a
$f_{i,a,d}(\varphi_{i,a,g,d}, ADFired_{e,i,a,d})$	Function computing node i effectiveness at time d based upon amount of resources used and active defenses used
$IntDur$	Duration of each time step
$TimeR_{i,d}(x_{i,a,d})$	Time to repair node i after an attack degrading to $x_{i,a,d}$
$CostR_{i,d}(x_{i,a,d})$	Cost to repair node i after an attack degrading to $x_{i,a,d}$
$TimeRpc_i$	Time to replace node i
$CostRpc_i$	Cost to replace node i
$ST_{e,a}$	Minimum threat of type a worthy of considering defense from e
$ADFP_{e,a}$	Maximum distance defense e can protect against attack type a
$ADRC_{e,a}$	Cost to use defense node e against attack type a
$ADRT_{e,a}$	Time between uses of defense e given it was used to protect against type a
$ADProt_{e,i,a}$	Binary matrix denoting if node i is protected from attack type a by defense node e
$Mobile_i$	Binary variable denoting if node i is mobile
$maxpos_i$	Maximum distance node i may be respositioned to
$minpos_i$	Minimum distance node i may be respositioned to
$maxfuel_{i,d}$	Maximum fuel available to node i at time d
$FuelUseR_i$	Amount of fuel used by node i per time step d
$footprint_{i,r}$	Radius of footprint of node i for orbital radius r
OD	Polar Orbit Direction of orbital plane at some starting time d=0 in regards to a contiguous half sphere of Earth (1 if North, -1 if South)
CN	Number of lateral orbits completed during each cycle
OT	Orbit Time
CT	Cycle Time

Short Term Model

$$\underset{\varphi}{Max}\ \mathrm{v} = \frac{Cost}{CostMax} + \frac{\sum_i \sum_s c_{i,s} \bullet \partial \mathrm{y}_{i,s}}{GlobeMax} + \frac{\sum_{j \in R} \sum_s c_{j,s} \bullet \partial \mathrm{y}_{j,s}}{\mathrm{RegMax}}$$

Subject To:

$$[1]\mathrm{Cost} = \sum_i \left(\left(\sum_a \mathrm{CostR}_{i,a}(x_{i,a}) \right) + (1 - x_i) \bullet \mathrm{CostO}_i \right) + \sum_i \sum_d \mathrm{CostR}_{\mathrm{i,d}}(x_{i,a,d}) + \sum_e \sum_a ADRC_{e,a} \left(\sum_d ADFired_{e,a} \right)$$

$$[2]\ x_{i,d} = \sum_a x_{i,a,d} \quad \forall i$$

$$[3]\ \sum_i \sum_a \varphi_{i,a,g} \le AdvRsc_g \quad \forall g$$

$$[4]\ x_{i,a,d} = f_{i,a,d}(\varphi_{i,a,g,d}, ADFired_{e,i,a,d}) \quad \forall i,a,d$$

$$[5]\sum_g \varphi_{\mathrm{i,a,g,d}} \le M \bullet \omega_{i,a,d} \quad \forall i,a,d$$

$$[6]\ \sum_{d'=d}^{d+\bar{d}} \sum_a \omega_{i,a,d'} = 1 + M(1 - \phi_i) \quad \forall i,d, \text{ where } \bar{d} = \left\lceil \frac{TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,g,d}))}{IntDur} \right\rceil$$

$$[7]\phi_{i,d} \le TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,g,d})) \quad \forall i,d$$

$$[8]\ \underset{y}{Min}\, z = \frac{1}{Total\_Net\_Value} \bullet \sum_d \sum_s \sum_j c_{j,s} \bullet \partial \mathrm{y}_{j,s,d}$$

Subject To:

(1) $\sum_j y_{(j,i),l,s,d} \le cap_{i,l,s,d} \bullet x_{i,d} \quad \forall i,l,s$

(2) $\sum_l \sum_s \sum_j y_{(j,i),l,s,d} \le cap_i \bullet x_{i,d} \quad \forall i$

(3) $-\sum_{l,i} y_{(j,i),l,s,d} - \partial \mathrm{y}_{j,s,d} = b_{j,\mathrm{s,d}} \quad \forall j,s$

(4) $\left(\sum_l \left(\sum_j y_{(j,i),l,s,d} - \sum_k y_{(i,k),l,s,d} \right) \right) \le b_{i,\mathrm{s,d}} \quad \forall i,s$

(5) $\mathrm{cap}_i \bullet Conn_{j,i,d} \ge \sum_{j,l,s} y_{(j,i),l,s,d} \ge 0 \ \forall i$

(8) $TCactive_d \le \sum_i (TC_i \bullet x_{i,d})$

(9) $Excess\_Cap_{s,d} = \sum_{i,d} \left(cap_i \bullet x_{i,d} - \sum_l \sum_j y_{(j,i),t,l,s,d} \right) \quad \forall s,\mathrm{d}$

(10) $\mathrm{x}_{i,d} \le 1 \quad \forall \mathrm{i,d}$

$$(10) Excess\_Value_d = \sum_s ec_s \bullet Excess\_Cap_{s,d}$$

$$(11) Weighted\_Excess\_Cap\_PCT_d = Excess\_Value_d \Big/ Total\_Net\_Value_d$$

$$(12) Total\_Net\_Value_d = \sum_s \sum_j c_{j,s} \bullet b_{j,s,d}$$

$$(14)\ ST_{e,a} \bullet ADFired_{e,i,a,d} \le f(\varphi_{i,a,g,d}, 0) \bullet \text{ADProt}_{e,i,a} \quad \forall e,i,a,d$$

$$(15)\ \sum_{d'=d}^{d+\hat{d}} \sum_i ADFired_{e,i,a,d} \le 1 \quad \forall e,a,d, \text{ where } \hat{d} = \frac{ADRT_{e,a}}{IntDur}$$

Medium/Long Term Model

$$\underset{\varphi}{Max}\ \mathrm{v} = \frac{Cost}{CostMax} + \frac{\sum_i \sum_t \sum_s c_{i,t,s} \bullet \partial\, \mathrm{y}_{i,t,s}}{GlobeMax} + \frac{\sum_{i \in R} \sum_t \sum_s c_{i,t,s} \bullet \partial\, \mathrm{y}_{i,t,s}}{\mathrm{RegMax}}$$

Subject To:

$$[1]\mathrm{Cost} = \sum_i \left(\left(\sum_a \mathrm{CostR}_{i,a}(x_{i,a}) \right) + (1 - x_i) \bullet \mathrm{CostO}_i \right) + \sum_i \sum_d \mathrm{CostR}_{\mathrm{i,d}}(x_{i,a,d}) + \sum_e \sum_a ADRC_{e,a} \left(\sum_d ADFired_{e,a} \right)$$

$$[2]\ x_{i,d} = \sum_a x_{i,a,d} \quad \forall i$$

$$[3]\ \sum_d \sum_i \sum_a \varphi_{i,a,d} \le AdvRsc$$

$$[4]\ x_{i,a,d} \ge \left(\phi_{i,d} \bullet f_{i,a,d}(\varphi_{i,a,d}) \right) + \left((1 - \phi_{i,d}) \bullet x_{i,a,d-1} \right) \quad \forall i, a, d$$

$$[5]\varphi_{\mathrm{i,a,d}} \le M \bullet \omega_{i,a,d} \quad \forall i, a, d$$

$$[6]\ \sum_{d'=d}^{d+\overline{d}} \sum_a \omega_{i,a,d'} = 1 + M(1 - \phi_i) \quad \forall i, d, \qquad \text{where } \overline{d} = \left\lceil \frac{TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,g,d}))}{IntDur} \right\rceil$$

$$[7]\phi_{i,d} + \mathrm{Rep}_{\,i,d + \frac{TimeR_{i,d}(f_{i,a,d}(\varphi_{i,a,d}))}{IntDur}} \le M \bullet (1 - \sum_a f_{i,a,d}(\varphi_{i,a,d})) \quad \forall i, d$$

$$[8]x_{i,\partial,d} \ge x_{i,d} \quad \forall i, \partial, d$$

$$[9]\ \underset{y}{Min}\ \ z = \frac{1}{Total\_Net\_Value} \bullet \sum_d \sum_\partial \sum_s \sum_j \left(c_{j,s} \bullet \partial\, \mathrm{y}_{j,s,\partial,d} \right)$$

Subject To:

(1) $\sum_j y_{(j,i),l,s,\partial,d} \le cap_{i,l,s,\partial,d} \bullet x_{i,\partial,d} \quad \forall i, l, s$

(2) $\sum_l \sum_s \sum_j y_{(j,i),l,s,\partial,d} \le cap_i \bullet x_{i,\partial,d} \quad \forall i$

(3) $-\sum_{l,i} y_{(j,i),l,s,\partial,d} - \partial\, \mathrm{y}_{j,s,\partial,d} = b_{j,\mathrm{s},\partial,d} \quad \forall j, s$

(4) $\left(\sum_l \left(\sum_j y_{(j,i),l,s,\partial,d} - \sum_k y_{(i,k),l,s,\partial,d} \right) \right) \le b_{i,\mathrm{s},\partial,d} \quad \forall i, s$

(5) $\mathrm{cap}_i \bullet Conn_{j,i,\partial,d} \ge \sum_{l,s} y_{(j,i),l,s,\partial,d} \ge 0 \ \forall j, i, \partial, d$

(8) $TCactive_{\partial,d} \le \sum_i (TC_i \bullet x_{i,\partial,d})$

(9) $Excess\_Cap_{s,\partial,d} = \sum_{i,d} \left(cap_i \bullet x_{i,\partial,d} - \sum_l \sum_j y_{(j,i),t,l,s,\partial,d} \right) \quad \forall s, \partial, d$

(10) $x_{i,\partial,d} \le 1 \quad \forall \mathrm{i},\partial,d$

(11) $Excess\_Value_d = \sum_{\partial}\sum_{s} ec_s \bullet Excess\_Cap_{s,\partial,d}$

(12) $Weighted\_Excess\_Cap\_PCT_d = \frac{1}{Total\_Net\_Value_d} \bullet \sum_{\partial} Excess\_Value_{\partial,d}$

(13) $Total\_Net\_Value_d = \sum_{\partial}\sum_{s}\sum_{i} c_{i,s} \bullet b_{i,s,\partial,d}$

(15) $ST_{e,a} \bullet ADFired_{e,i,a,\partial} \le f(\varphi_{i,a,g,d}, 0) \bullet \mathrm{ADProt}_{e,i,a} \quad \forall e,i,a,d$

(16) $\sum_{d'=d}^{d+\hat{d}} \sum_{i} ADFired_{e,i,a,d} \le 1 \quad \forall e,a,d \quad \text{where } \hat{d} = \left\lceil \frac{ADRT_{e,a}}{IntDur} \right\rceil$

(17) $x_{i,\partial,d} \le \mathrm{State}_{\mathrm{i,d}} \bullet (\mathrm{State}_{\mathrm{i,d}} - 1) \quad \forall \mathrm{i},\partial,d$

(18) $\mathrm{POS}\Phi_{i,\partial,d} = \mathrm{Sin}[(\delta t \bullet \partial \bullet 360 / OT) - 90 \bullet (OD - 1) + OD \bullet \mathrm{Arcsin}(\Phi_{i,d} / 90)] \bullet 90 \quad \forall i,\partial,d$

(19) $\mathrm{POS}\Theta_{i,\partial,d} = \Theta_{i,d} + 180 - 360 \bullet (\delta t \bullet \partial / (CT / CN)) + 360 \bullet Wrap_{i,\delta,d} \quad \forall i,\partial,d$

(20) $0 \le \mathrm{POS}\Theta_{i,\delta,d} \le 360 \quad \forall i,\partial,d$ (21) $\mathrm{Wrap}_{i,\delta,d} = \{0,1\} \quad \forall i,\partial,d$

(22) $\mathrm{HA}_{i,j,\partial,d} = Sin^2[(\mathrm{POS}\Phi_{i,\partial,d} - \mathrm{POS}\Phi_{j,\partial,d}) \bullet 0.5] +$
$+ Cos[\mathrm{POS}\Phi_{i,\partial,d}] \bullet Cos[\mathrm{POS}\Phi_{j,\partial,d}] * Sin^2[(\mathrm{POS}\Theta_{i,\partial,d} - \mathrm{POS}\Theta_{j,\partial,d}) * 0.5] \quad \forall \mathrm{i,j},\partial,d$

(23) $dist_{i,j,\partial,d} = \mathrm{Atan2}\left(\sqrt{1 - \mathrm{HA}_{i,j,\partial,d}}, \sqrt{\mathrm{HA}_{i,j,\partial,d}}\right) \quad \forall i,j,\partial,d$

(24) $dist_{i,j,\partial,d} \le \mathrm{footprint}_{i,r_j} \bullet ChckI_{i,j,\partial,d} + M \bullet (1 - ChckI_{i,j,\partial,d}) \quad \forall \mathrm{i,j},\partial,d$

(25) $dist_{i,j,\partial,d} \le \mathrm{footprint}_{j,r_i} \bullet ChckJ_{i,j,\partial,d} + M \bullet (1 - ChckJ_{i,j,\partial,d}) \quad \forall \mathrm{i,j},\partial,d$

(26) $\mathrm{Conn}_{i,j,\partial,d} \le 0.5 \bullet (ChckI_{i,j,\partial,d} + ChckJ_{i,j,\partial,d}) \quad \forall \mathrm{i,j},\partial,d$

(27) $\mathrm{maxfuel}_{i,d} = \mathrm{maxfuel}_{i,d-1} - FuelUseR_i - D_{i,d} \bullet Fuel_i(DistPos_i) \quad \forall \mathrm{i},d$

(28) $\sum_{d} \mathrm{D}_{i,d} \le Mobile_i \quad \forall \mathrm{i}$

(29) $\mathrm{maxfuel}_{i,d} \le M \bullet empty_{i,d} \quad \forall \mathrm{i},d$

(30) $\sum_{\partial} \mathrm{x}_{i,\partial,d} \le empty_{i,d} \quad \forall \mathrm{i},d$

(31) $\mathrm{State}_{i,d} \le M \bullet x_{i,d} \quad \forall i,d$

(32) $\mathrm{R}_{i,d} \le \mathrm{R}_{i,d-1} + D_{i,d} \quad \forall i,d$

(33) $\sum_{d} \mathrm{R}_{i,d} \ge \mathrm{RepoD}_i(DistPos_i) \quad \forall i$

(34) $\mathrm{State}_{i,d} \le 3 - 2 \bullet R_{i,d} \quad \forall i,d$

(35) $\mathrm{DistPos}_i = dist_{(i,1),(i,|d|)} \quad \forall i$

(36) $\mathrm{minpos}_i \bullet \mathrm{D}_{i,d} \le dist_{(i,d-1),(i,d)} \le \mathrm{maxpos}_i \bullet \mathrm{D}_{i,d} \quad \forall i,d$

(37) $x_{i,d} \le x_{i,d} + \mathrm{Rep}_{i,d} \quad \forall i,d$ (38) $\mathrm{Rep}_{i,d} = \{0,1\} \quad \forall i,d$

Time Dependent Set Covering Problem Heuristic

{1} $If \sum_{i \in I_m} \left(C_{i,\partial',d'} \right) < \left| I_m \right|$ then

Select $\left\{ i \mid \text{Req}_{i,\partial',d'} * C_{i,\partial',d'} \le \left(\sum_j Conn_{j,i,\partial',d'} * \text{x}_{j,\partial',d'} \right), i \in I_m \right\}$

Select $\left\{ \text{j} \mid \min_j \left(\text{dist(j,i)-footprint}_j > 0 \right), j \notin J \right\}$ and J=J+{j}

{2} Calculate expected position of all mobile nodes except for j

[Constraints (18) – (20)]

{3} $\max_{\Theta\Phi} \left\{ \sum_{k \in I_m, \partial, d'} \left(w_k C_{k,\partial,d'} \right) \mid POS_{j,0,d'} = (\Theta, \Phi) \right\}$

Subject to:

(22) $\text{HA}_{i,j,\partial,d'} = Sin^2[(\text{POS}\Phi_{i,\delta,d'} - \text{POS}\Phi_{j,\delta,d'}) \bullet 0.5] +$

$+Cos[\text{POS}\Phi_{i,\delta,d'}] \bullet Cos[\text{POS}\Phi_{j,\delta,d'}] \bullet Sin^2[(\text{POS}\Theta_{i,\delta,d'} - \text{POS}\Theta_{j,\delta,d'}) \bullet 0.5] \quad \forall \text{i},\partial$

(23) $dist_{i,j,\partial,d'} = \text{Atan2}\left(\sqrt{1 - \text{HA}_{i,j,\partial,d'}}, \sqrt{\text{HA}_{i,j,\partial,d'}} \right) \quad \forall i, \partial$

(24) $dist_{i,j,\partial,d'} \le \text{footprint}_{i,r_j} \bullet ChckI_{i,j,\partial,d} + M \bullet (1 - ChckI_{i,j,\partial,d'}) \quad \forall \text{i},\partial$

(25) $dist_{i,j,\partial,d'} \le \text{footprint}_{j,r_i} \bullet ChckJ_{i,j,\partial,d} + M \bullet (1 - ChckJ_{i,j,\partial,d'}) \quad \forall \text{i},\partial$

(26) $\text{Conn}_{i,j,\partial,d'} \le 0.5 \bullet (ChckI_{i,j,\partial,d'} + ChckJ_{i,j,\partial,d'}) \quad \forall \text{i},\partial$

(37) $\text{Req}_{i,\partial,d'} \bullet C_{i,\partial,d'} \le \left(\sum_j Conn_{j,i,\partial,d'} \bullet \text{x}_{j,\partial,d'} \right) \quad \forall i, \partial$

(new) $\text{Conn}_{i,j,\partial',d} = 1$

{4} $If \ Fuel_j \left[\text{dist(j,}(\Theta\Phi)) \right] > \text{maxfuel}_{j,d} - FuelUseR_j \bullet (d_{\max} - d')$

Then J=J-{j}, J'=J'+{j}, and return to {1}

{5} Calculate time for movement subject to constraints and Record Connections from {3}

[constraints (32) – (34)] simplify to

set $\text{State}_{j,d} = 1$ for all $d' \le d < d' + \text{RepoD}_j(DistPos_j)$

[constraints (22) – (26)]

{6}*C*alculate Value function $\sum_d z_d$ subject to linear program:

$$\underset{y}{Min}\ z_d = \frac{1}{Total\_Net\_Value}\sum_{\partial}\sum_{s}\sum_{t}\sum_{i} c_{i,t,s} \bullet \partial\, \mathrm{y}_{i,t,s,\partial}$$

Subject To:

(1.1) $\sum_{t,j} y_{(i,j),t,l,s,\partial,d} \le cap_{i,l,s,\partial,d} \bullet x_{i,\partial,d} \quad \forall i,l,s,\partial,d$

(1.2) $\sum_{t,j} y_{(j,i),t,l,s,\partial,d} \le cap_{i,l,s,\partial,d} \bullet x_{i,\partial,d} \quad \forall i,l,s,\partial,d$

(2) $\sum_{l}\sum_{s}\sum_{j} y_{(j,i),t,l,s,\partial,d} \le cap_i \bullet x_{i,\partial,d} \quad \forall i,\partial,d$

(3) $-\sum_{l}\sum_{k} y_{(t,k),t,l,s,\partial,d} \ge b_{t,t,s,\partial} \bullet \mathrm{x}_{t,\partial,d}$ (Supply node t) $\forall t,s,\partial,d$

(4) $\left(\sum_{l}\left(\sum_{j} y_{(j,i),t,l,s,\partial,d} - \sum_{k} y_{(i,k),t,l,s,\partial,d}\right)\right) + \partial y_{i,t,s,\partial,d} = b_{i,t,\mathrm{s},\partial,d} \quad \forall i,t,s,\partial,d$

(5) $\mathrm{cap}_i \bullet Conn_{j,i,\partial,d} \ge \sum_{j,t,l,s} y_{(j,i),t,l,s,\partial,d} \ge 0 \ \forall i,\partial,d$ (6) $\partial y_{i,t,s,\partial,d} \ge 0 \ \forall i,t,s,\partial,d$

(7) $\sum_{t,s} \partial y_{i,t,s,\partial,d} \le ((\sum_{i}(TC_i \bullet x_{i,\partial,d})) - 1) \bullet (cap_i \bullet (1 - TC_t)) \bullet TCactive_{\partial,d} +$

$+ (cap_i \bullet (x_{i,\partial,d} + (1 - TCactive_{\partial}))) \quad \forall i,\partial,d$

(8) $TCactive_{\partial,d} \le \sum_{i}(TC_i \bullet x_{i,\partial,d}) \quad \forall \partial,d$

(10) $\mathrm{x}_{i,\partial,d} \le 1 \ \forall \mathrm{i},\partial,d$

{7}*If* $(\partial < \partial_{\max})$ and $\left(J + J',\ \mathrm{J}' \ne \varnothing \text{ contains all mobile nodes or } \sum_{i \in I_m, d}\left(C_{i,\partial',d}\right) \ge d_{\max} * \left|I_m\right|\right)$,

then set ∂'=∂'+1 and return to step {1}

{8}*If* $J + J'$ contains all mobile nodes, then set J=J'=∅ and proceed to next d' such that there exists a node i where $\left|\mathrm{x}_{i \in I_m, d'} - \mathrm{x}_{i \in I_m, d'-1}\right| > 0$

Return to step {1}

Repeat until d'=$\mathrm{d}_{\max}$

Heuristic One

1) Set B to the amount of resources, z=0, ε be an arbitrarily small number, m be a sufficiently large integer, and v'=$\max_i(v_i)$

2) For every i,d combination, set $x_{i,d'} = 0$, (or expected with no incident) for all $d' \geq d$, and record resulting Attacker objective as value $v_{i,d}'$

3) Reset all $x_{i,d}$ to initial values and set $x_{i,d}' = x_{i,d}$, $x_{i,a,d}' = 0$

4) Set $p_{i,d} = \frac{v_{i,d}}{\sum_{i=0}^{n} v_{i,d}}$ and $p_{i,d}' = p_{i,d}$ for each i,d

5) Set $p_{i,d} = \frac{p_{i,d}}{\sum_{i=0}^{n} p_{i,d}}$

6) If $\sum_a \omega_{i,a,d} = |A|$, then set $p_{i,d} = 0$ and return to step 5.

7) If $x_{i,d}' = 1$, then set $p_{i,d} = 0$ and return to step 5.

8) Select a random number $x \in [0,1]$

9) Set $\varphi = \begin{cases} 0 & \text{if } p_{i,d} = 0 \text{ or } p_{i,d} < \max_d\{p_{i,d}\} \\ Min_d\left\{\left(\min_{a \mid \omega_{i,a,d}=0}\{f_{i,a,d}^{-1}(1) - f_{i,a,d}^{-1}(x_{i,d}')\}\right), B\right\} & \text{if } x \in [\sum_{j=1}^{i} p_{j,d}, p_{i,d} + \sum_{j=1}^{i} p_{j,d}) \\ 0 & \text{Otherwise} \end{cases}$

10) $x_{i,a,d} = f_{i,a,d}(\varphi_{i,a,d} = \varphi + f_{i,a,d}^{-1}(x_{i,d}'))$ for each a

11) If $x_{i,a,d} = \max_{a \mid \omega_{i,a,d}=0}\{x_{i,a,d}\}$ and no other $x_{i,a',d} = \max_{a \mid \omega_{i,a,d}=0}\{x_{i,a,d}\}, a' \neq a,$ exists, then set $x_{i,a,d}' = x_{i,a,d}$. If there exists $x_{i,a',d} = x_{i,a,d} = \max_{a \mid \omega_{i,a,d}=0}\{x_{i,a,d}\}, a' \neq a,$ then set $x_{i,a,d}' = x_{i,a,d}$ such that $TimeR_{i,d}(x_{i,a,d}) = \max_a\{TimeR_{i,d}(x_{i,a,d})\}$, and set $\omega_{i,a,d} = 1$.

12) Set $x_{i,d}' = \sum_a x_{i,a,d}', B = B - \varphi, p_{i,d} = 0$

13) If B>0, then return to step 5, else continue

14) Calculate resulting Attacker objective value and denote value as v.

15) If $1+\varepsilon > \frac{v}{v'} > 1-\varepsilon$ then z=z+1

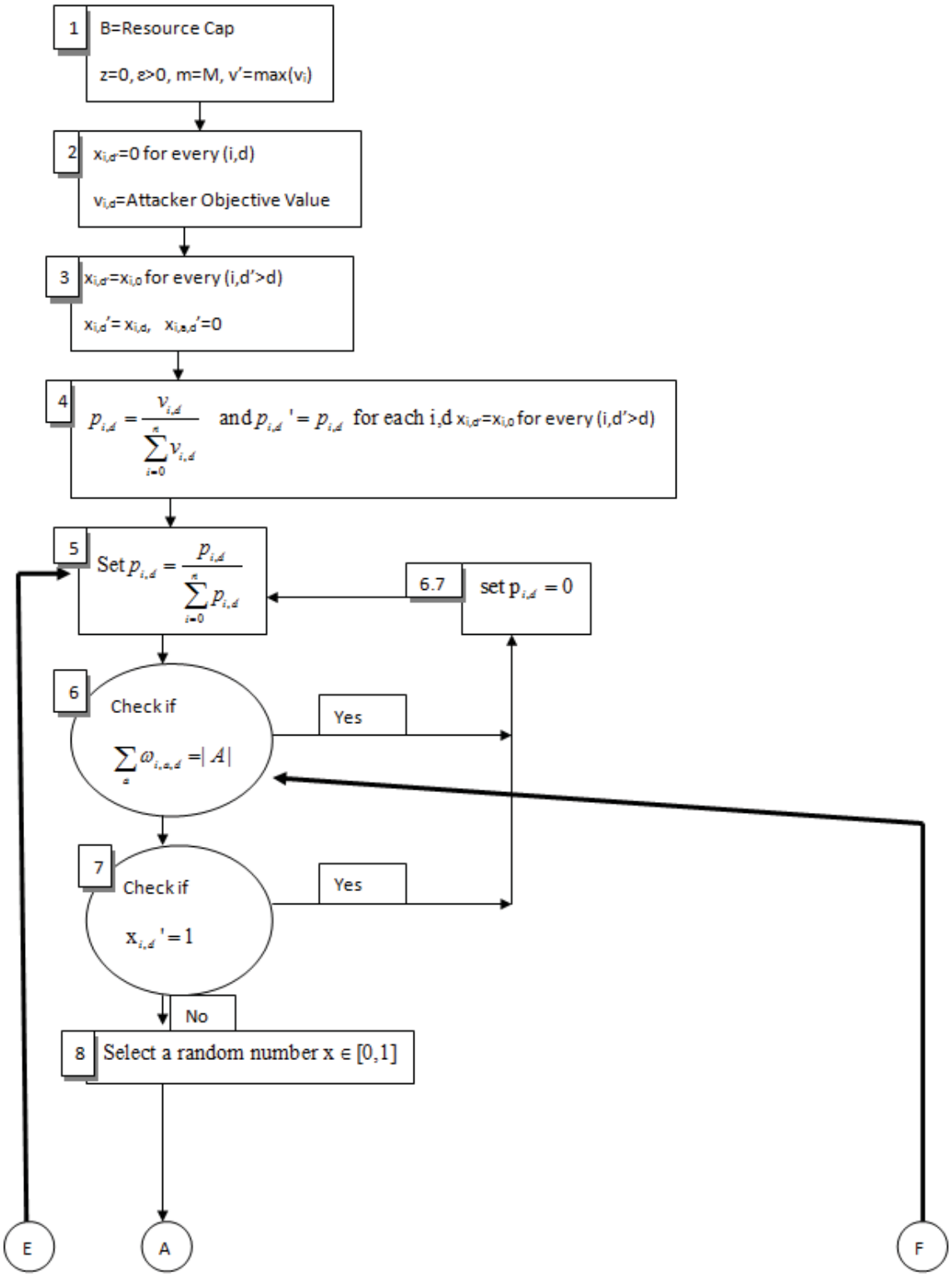
16) If v>v', then set v'=v and $X_{i,a,d} = x_{i,a,d}' \ \forall \ i,a,d.$

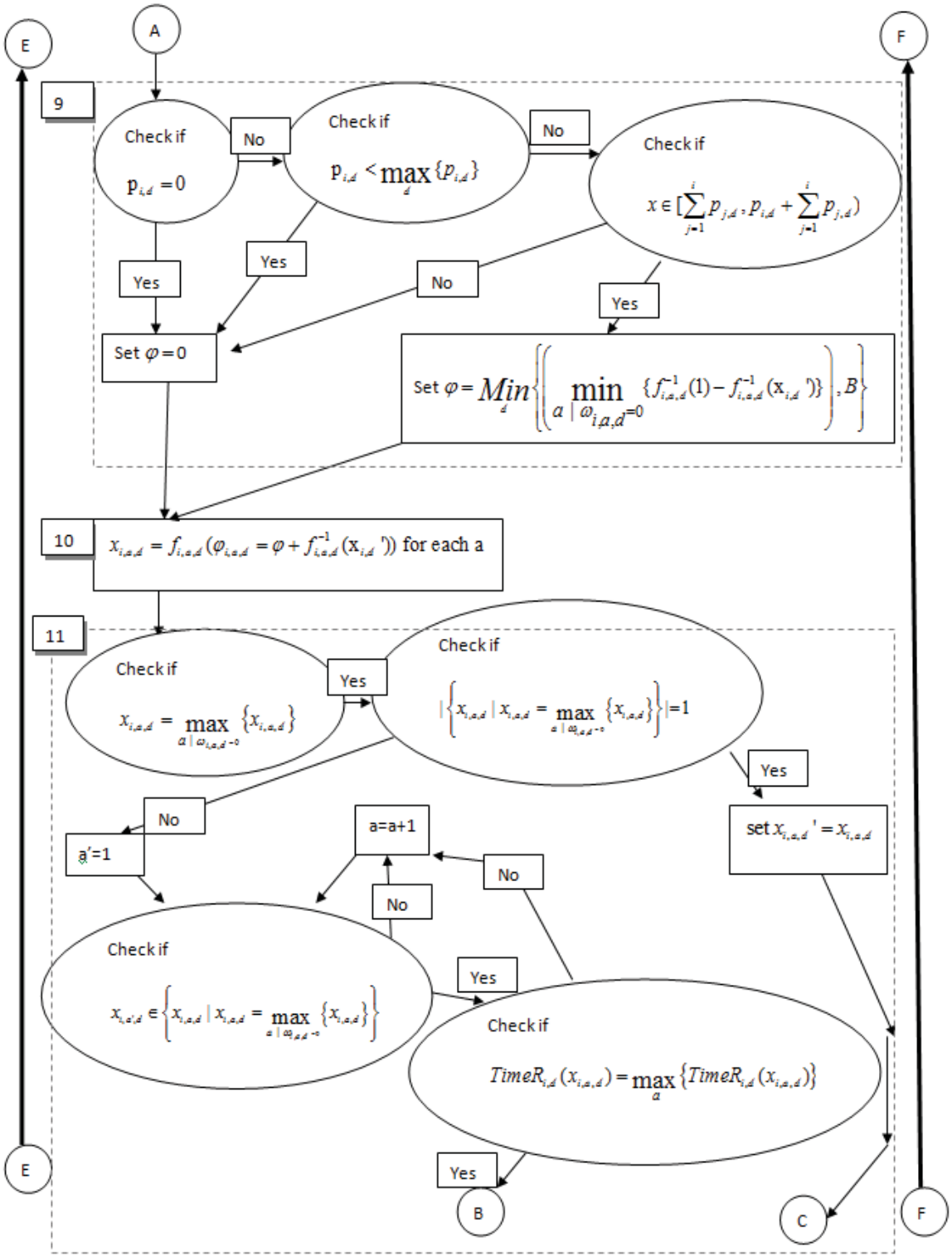
17) If z>m then terminate, else continue

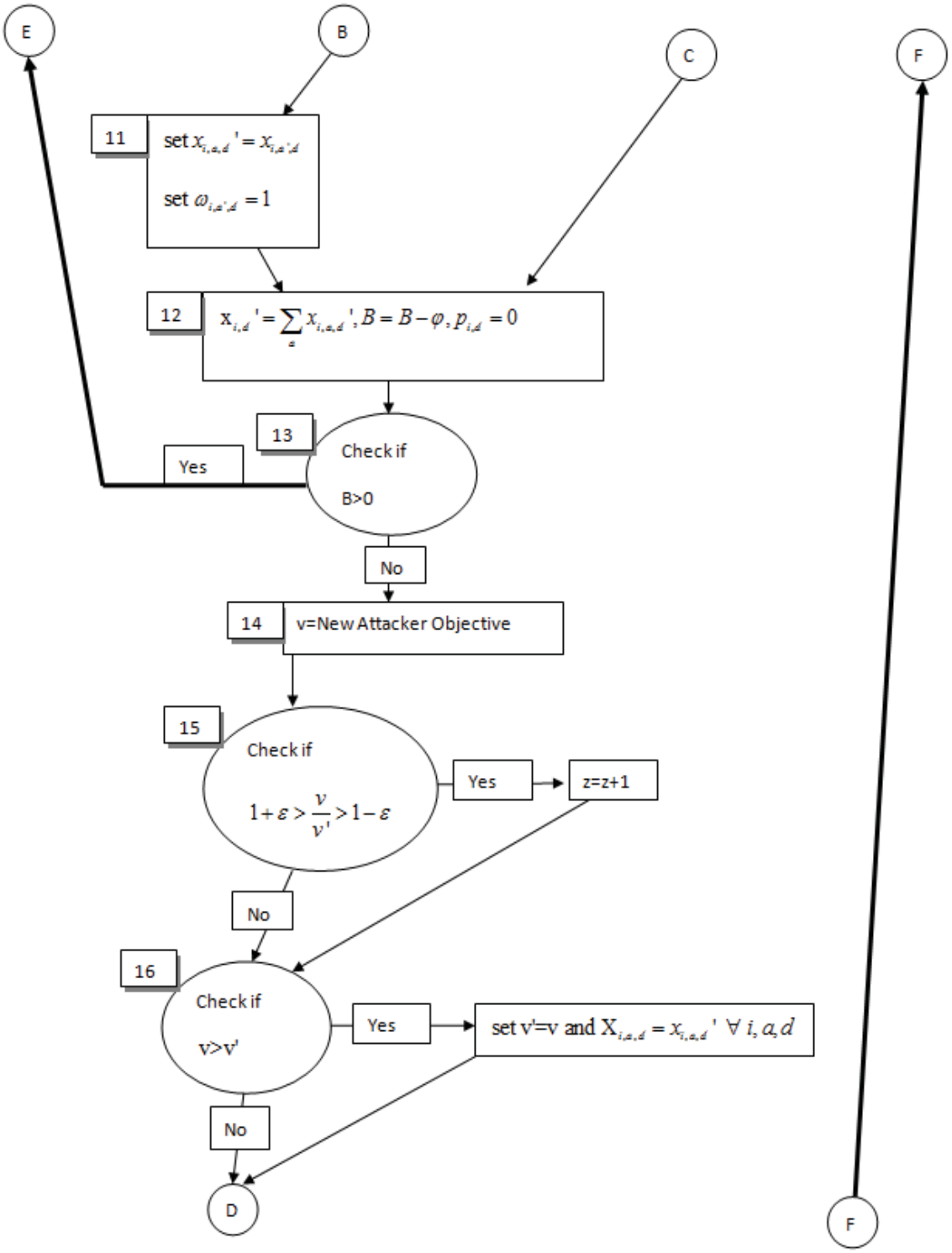
18) Set $p_{i,d} = 0.5 \bullet \frac{p_{i,d}'}{\sum_{i=0}^{n} p_{i,d}'} + 0.5 \bullet \frac{x_{i,d}'}{\sum_{i=1}^{n} x_{i,d}'}$

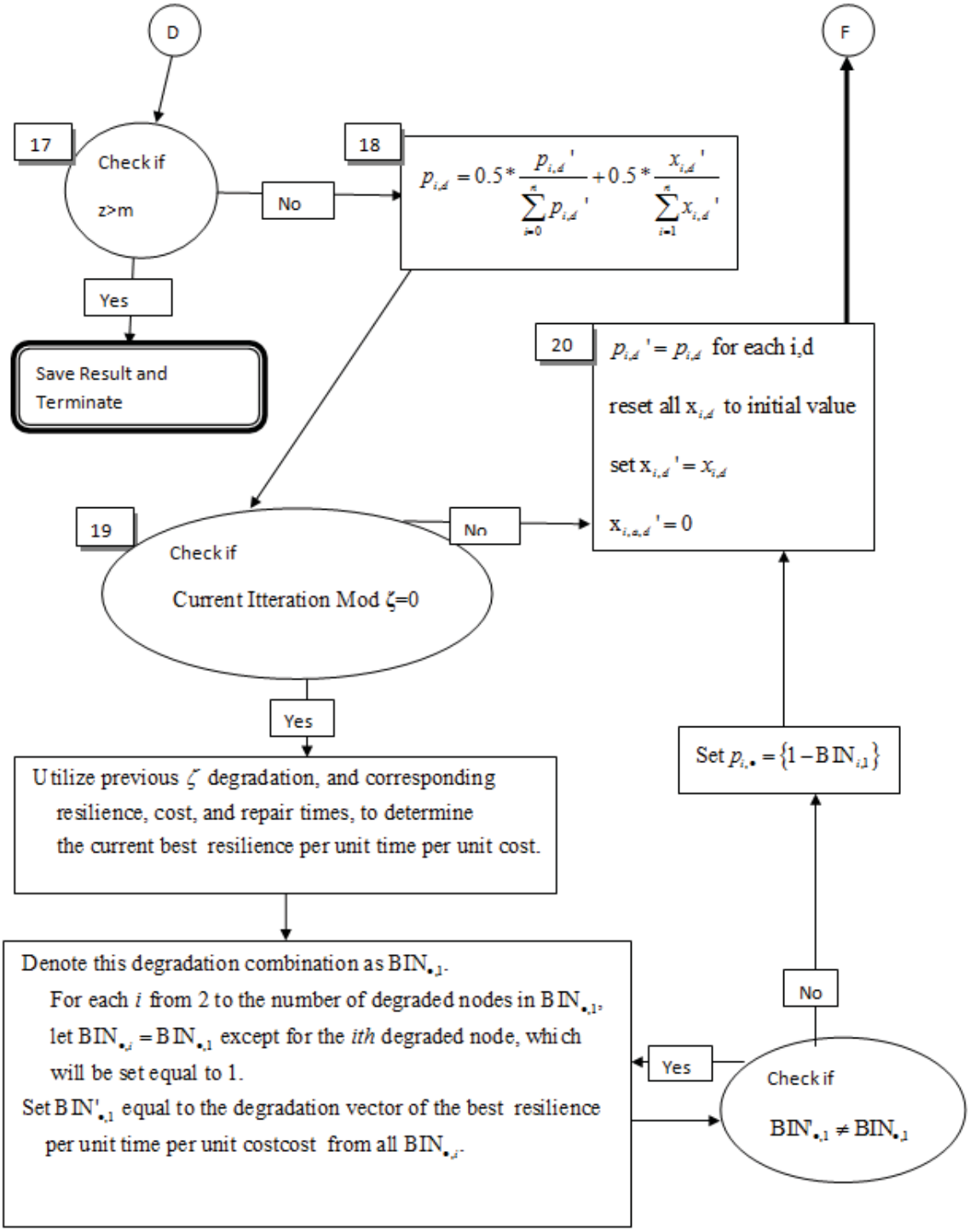
19) Set $p_{i,d}' = p_{i,d}$ for each i,d, reset all $x_{i,d}$ to initial values,
and set $x_{i,d}' = x_{i,d}$, $x_{i,a,d}' = 0$ and return to step 6.

Heuristic One Flow Chart




E
A
F
9
Check if
$p_{i,d}=0$
No
Check if
$p_{i,d} < \max_{d}\{p_{i,d}\}$
No
Check if
$x \in [\sum_{j=1}^{i} p_{j,d}, p_{i,d} + \sum_{j=1}^{i} p_{j,d})$
Yes
Yes
No
Yes
Set $\varphi = 0$
Set $\varphi = Min_{d}\left\{\left(\min_{a \mid \omega_{i,a,d}=0}\{f^{-1}_{i,a,d}(1) - f^{-1}_{i,a,d}(x_{i,d}')\}\right), B\right\}$
10
$x_{i,a,d} = f_{i,a,d}(\varphi_{i,a,d} = \varphi + f^{-1}_{i,a,d}(x_{i,d}'))$ for each a
11
Check if
$x_{i,a,d} = \max_{a \mid \omega_{i,a,d}=0}\{x_{i,a,d}\}$
Yes
Check if
$\left|\left\{x_{i,a,d} \mid x_{i,a,d} = \max_{a \mid \omega_{i,a,d}=0}\{x_{i,a,d}\}\right\}\right|=1$
Yes
No
set $x_{i,a,d}' = x_{i,a,d}$
a'=1
a=a+1
No
No
Check if
$x_{i,a',d} \in \left\{x_{i,a,d} \mid x_{i,a,d} = \max_{a \mid \omega_{i,a,d}=0}\{x_{i,a,d}\}\right\}$
Yes
Check if
$TimeR_{i,d}(x_{i,a,d}) = \max_{a}\{TimeR_{i,d}(x_{i,a,d})\}$
Yes
E
B
C
F


E
B
C
F
11
set $x_{i,a,d}' = x_{i,a',d}$
set $\omega_{i,a',d} = 1$
12
$x_{i,d}' = \sum_a x_{i,a,d}', B = B - \varphi, p_{i,d} = 0$
13
Check if
B>0
Yes
No
14
v=New Attacker Objective
15
Check if
$1+\varepsilon > \frac{v}{v'} > 1-\varepsilon$
Yes
z=z+1
No
16
Check if
v>v'
Yes
set v'=v and $X_{i,a,d} = x_{i,a,d}' \ \forall i,a,d$
No
D
F


D
17
Check if
z>m
No
18
$p_{i,d} = 0.5 * \frac{p_{i,d}'}{\sum_{i=0}^{n} p_{i,d}'} + 0.5 * \frac{x_{i,d}'}{\sum_{i=1}^{n} x_{i,d}'}$
Yes
Save Result and
Terminate
F
20
$p_{i,d}' = p_{i,d}$ for each i,d
reset all $x_{i,d}$ to initial value
set $x_{i,d}' = x_{i,d}$
$x_{i,a,d}' = 0$
19
Check if
Current Itteration Mod ζ=0
No
Yes
Set $p_{i,\bullet} = \{1 - BIN_{i,1}\}$
Utilize previous ζ degradation, and corresponding
resilience, cost, and repair times, to determine
the current best resilience per unit time per unit cost.
Denote this degradation combination as $BIN_{\bullet,1}$.
For each *i* from 2 to the number of degraded nodes in $BIN_{\bullet,1}$,
let $BIN_{\bullet,i} = BIN_{\bullet,1}$ except for the *ith* degraded node, which
will be set equal to 1.
Set $BIN'_{\bullet,1}$ equal to the degradation vector of the best resilience
per unit time per unit costcost from all $BIN_{\bullet,i}$.
No
Yes
Check if
$BIN'_{\bullet,1} \neq BIN_{\bullet,1}$

Appendix B: Case Study and Variation Parameters

Node Name	Satellite (Y/N)	Operating (Y/N)	Spare (Y/N)
TTAC-1 (Svalbard, Norway)	N	Y	N
TTAC-2 (Fairbanks, AK)	N	Y	N
TTAC-3 (Vancouver, Canada)	N	Y	N
TTAC-4 (Toronto, Canada)	N	Y	N
TTAC-5 (Reykjavik, Iceland)	N	Y	N
Sat Network Ops Center (Leesburg, Virginia)	N	Y	N
Commercial Gateway (Tempe, Arizona)	N	Y	N
DoD Gateway (Wahiawa, HI)	N	Y	N
HQ (Bethesda, MD)	N	Y	N
GEOSAT1	Y	Y	N
GEOSAT2	Y	Y	N
GEOSAT3	Y	Y	N
GEOSAT4	Y	Y	N
GEOSAT5	Y	Y	N
GEOSAT6	Y	Y	N

Node Name	Target (Y/N)	InterSat Link (km)	Up/Downlink Radius(km)
Telemetry Tracking and C/C (TTAC)-1 (Svalbard, Norway)	Y	N/A	12000
TTAC-2 (Fairbanks, AK)	Y	N/A	12000
TTAC-3 (est Vancouver, Canada)	Y	N/A	12000
TTAC-4 (est Toronto, Canada)	Y	N/A	12000
TTAC-5 (est Reykjavik, Iceland)	Y	N/A	12000
Sat Network Ops Center (SNOC) (Leesburg, Virginia)	Y	N/A	12000
Commercial Gateway (Tempe, Arizona)	N	N/A	12000
DoD Gateway (Wahiawa, HI)	N	N/A	12000
HQ (Bethesda, MD)	N	N/A	12000
GEOSAT1	Y	90000	7768
GEOSAT2	Y	90000	7768
GEOSAT3	Y	90000	7768
GEOSAT4	Y	90000	7768
GEOSAT5	Y	90000	7768
GEOSAT6	Y	90000	7768

	Location at Time=0 (in degrees)	
Node Name	Latitude (NS)	Longitude (EW)
TTAC-1 (Svalbard, Norway)	80.238166	12.447236
TTAC-2 (Fairbanks, AK)	66.8350185	-149.65307
TTAC-3 (Vancouver, Canada)	51.25	-126.1
TTAC-4 (Toronto, Canada)	45.652527	-82.381961
TTAC-5 (Reykjavik, Iceland)	66.1333	-24.9333
Sat Network Ops Center (Leesburg, Virginia)	41.252181	-80.744541
Commercial Gateway (Tempe, Arizona)	35.414842	-114.909319
DoD Gateway (Wahiawa, HI)	23.502574	-161.022938
HQ (Bethesda, MD)	40.98472	-80.09472
GEOSAT1	0	0
GEOSAT2	0	120
GEOSAT3	0	-120
GEOSAT4	0	60
GEOSAT5	0	180
GEOSAT6	0	-60

Node Name	Resources Needed to Permanently Interdict Node ($US Thousands)	Sink (Y/N)
TTAC-1 (Svalbard, Norway)	1410	N
TTAC-2 (Fairbanks, AK)	1410	N
TTAC-3 (Vancouver, Canada)	1410	N
TTAC-4 (Toronto, Canada)	1410	N
TTAC-5 (Reykjavik, Iceland)	1410	N
Sat Network Ops Center (Leesburg, Virginia)	1410	N
Commercial Gateway (Tempe, Arizona)	1410	Y
DoD Gateway (Wahiawa, HI)	1410	Y
HQ (Bethesda, MD)	1410	N
GEOSAT1	180000	N
GEOSAT2	180000	N
GEOSAT3	180000	N
GEOSAT4	180000	N
GEOSAT5	180000	N
GEOSAT6	180000	N

Degradation Function

Function1: Sharp Spike at End $D=(x)^{15}$

Function2: Gentle Spike at End $D= (x)^{5}$

Function3: S Curve with Early High $D=3*(x)^{2}-2*(x)^{3}$

Function4: S Curve with Right Translated Point of Inversion $D= (1/3)*(5*x^{2}-2*x^{5})$

Function5: Linear $D=x$

Repair Function

Function 1: Full Rebuild Time Always

Function 2: Recovery Time Linearly Related to Degradation

Node Name	Degradation Function	Capacity (Users)	Repair Function	Rebuild (Cycles)
TTAC-1 (Svalbard, Norway)	2	172000	2	42
TTAC-2 (Fairbanks, AK)	2	172000	2	42
TTAC-3 (Vancouver, Canada)	2	172000	2	42
TTAC-4 (Toronto, Canada)	2	172000	2	42
TTAC-5 (Reykjavik, Iceland)	2	172000	2	42
Sat Network Ops Center (Leesburg, Virginia)	2	172000	2	42
Commercial Gateway (Tempe, Arizona)	2	172000	2	42
DoD Gateway (Wahiawa, HI)	4	172000	2	42
HQ (Bethesda, MD)	2	172000	2	42
GEOSAT1	3	20000	1	182
GEOSAT2	3	20000	1	182
GEOSAT3	3	20000	1	182
GEOSAT4	3	20000	1	182
GEOSAT5	3	20000	1	182
GEOSAT6	3	20000	1	182

Node Name	Mobile (Y/N)	Max Fuel	Fuel Used per Deg. Shift
TTAC-1 (Svalbard, Norway)	N	1	1
TTAC-2 (Fairbanks, AK)	N	1	1
TTAC-3 (Vancouver, Canada)	N	1	1
TTAC-4 (Toronto, Canada)	N	1	1
TTAC-5 (Reykjavik, Iceland)	N	1	1
Sat Network Ops Center (Leesburg, Virginia)	N	1	1
Commercial Gateway (Tempe, Arizona)	N	1	1
DoD Gateway (Wahiawa, HI)	N	1	1
HQ (Bethesda, MD)	N	1	1
GEOSAT1	Y	5113	340
GEOSAT2	Y	5113	340
GEOSAT3	Y	5113	340
GEOSAT4	Y	5113	340
GEOSAT5	Y	5113	340
GEOSAT6	Y	5113	340

Node Name	Fuel Used per Cycle	Reposition Deg per Cycle	Min Cycles Operational Post Move
TTAC-1 (Svalbard, Norway)	0	1	1
TTAC-2 (Fairbanks, AK)	0	1	1
TTAC-3 (Vancouver, Canada)	0	1	1
TTAC-4 (Toronto, Canada)	0	1	1
TTAC-5 (Reykjavik, Iceland)	0	1	1
Sat Network Ops Center (Leesburg, Virginia)	0	1	1
Commercial Gateway (Tempe, Arizona)	0	1	1
DoD Gateway (Wahiawa, HI)	0	1	1
HQ (Bethesda, MD)	0	1	1
GEOSAT1	1	2.4	42
GEOSAT2	1	2.4	42
GEOSAT3	1	2.4	42
GEOSAT4	1	2.4	42
GEOSAT5	1	2.4	42
GEOSAT6	1	2.4	42

Variation Changes

Baseline 6	Degrees		
Node Name	Latitude (NS)	Longitude (EW)	Rebuild
TTAC-1 (Svalbard, Norway)	80.238166	12.447236	42
TTAC-2 (Fairbanks, AK)	66.8350185	-149.65307	42
TTAC-3 (Vancouver, Canada)	51.25	-126.1	42
TTAC-4 (Toronto, Canada)	45.652527	-82.381961	42
TTAC-5 (Reykjavik, Iceland)	66.1333	-24.9333	42
Sat Network Ops Center (Leesburg, Virginia)	41.252181	-80.744541	42
Commercial Gateway (Tempe, Arizona)	35.414842	-114.909319	42
DoD Gateway (Wahiawa, HI)	23.502574	-161.022938	42
HQ (Bethesda, MD)	40.98472	-80.09472	42
GEOSAT1	0	-120	182
GEOSAT2	0	-60	182
GEOSAT3	0	0	182
GEOSAT4	0	60	182
GEOSAT5	0	120	182
GEOSAT6	0	180	182

Shifted	Degrees		
Node Name	Latitude (NS)	Longitude (EW)	Rebuild
TTAC-1 (Svalbard, Norway)	80.238166	12.447236	42
TTAC-2 (Fairbanks, AK)	66.8350185	-149.65307	42
TTAC-3 (Vancouver, Canada)	51.25	-126.1	42
TTAC-4 (Toronto, Canada)	45.652527	-82.381961	42
TTAC-5 (Reykjavik, Iceland)	66.1333	-24.9333	42
Sat Network Ops Center (Leesburg, Virginia)	41.252181	-80.744541	42
Commercial Gateway (Tempe, Arizona)	35.414842	-114.909319	42
DoD Gateway (Wahiawa, HI)	23.502574	-161.022938	42
HQ (Bethesda, MD)	40.98472	-80.09472	42
GEOSAT1	0	-150	182
GEOSAT2	0	-90	182
GEOSAT3	0	-30	182
GEOSAT4	0	30	182
GEOSAT5	0	90	182
GEOSAT6	0	150	182

Repair	Degrees		
Node Name	Latitude (NS)	Longitude (EW)	Rebuild
TTAC-1 (Svalbard, Norway)	80.238166	12.447236	21
TTAC-2 (Fairbanks, AK)	66.8350185	-149.65307	21
TTAC-3 (Vancouver, Canada)	51.25	-126.1	21
TTAC-4 (Toronto, Canada)	45.652527	-82.381961	21
TTAC-5 (Reykjavik, Iceland)	66.1333	-24.9333	21
Sat Network Ops Center (Leesburg, Virginia)	41.252181	-80.744541	21
Commercial Gateway (Tempe, Arizona)	35.414842	-114.909319	21
DoD Gateway (Wahiawa, HI)	23.502574	-161.022938	21
HQ (Bethesda, MD)	40.98472	-80.09472	21
GEOSAT1	0	-120	91
GEOSAT2	0	-60	91
GEOSAT3	0	0	91
GEOSAT4	0	60	91
GEOSAT5	0	120	91
GEOSAT6	0	180	91

Reduced 4	Degrees		
Node Name	Latitude (NS)	Longitude (EW)	Rebuild
TTAC-1 (Svalbard, Norway)	80.238166	12.447236	42
TTAC-2 (Fairbanks, AK)	66.8350185	-149.65307	42
TTAC-3 (Vancouver, Canada)	51.25	-126.1	42
TTAC-4 (Toronto, Canada)	45.652527	-82.381961	42
TTAC-5 (Reykjavik, Iceland)	66.1333	-24.9333	42
Sat Network Ops Center (Leesburg, Virginia)	41.252181	-80.744541	42
Commercial Gateway (Tempe, Arizona)	35.414842	-114.909319	42
DoD Gateway (Wahiawa, HI)	23.502574	-161.022938	42
HQ (Bethesda, MD)	40.98472	-80.09472	42
GEOSAT1	0	-180	182
GEOSAT2	0	-90	182
GEOSAT3	0	0	182
GEOSAT4	0	90	182

Increased 8	Degrees		
Node Name	Latitude (NS)	Longitude (EW)	Rebuild
TTAC-1 (Svalbard, Norway)	80.238166	12.447236	42
TTAC-2 (Fairbanks, AK)	66.8350185	-149.65307	42
TTAC-3 (Vancouver, Canada)	51.25	-126.1	42
TTAC-4 (Toronto, Canada)	45.652527	-82.381961	42
TTAC-5 (Reykjavik, Iceland)	66.1333	-24.9333	42
Sat Network Ops Center (Leesburg, Virginia)	41.252181	-80.744541	42
Commercial Gateway (Tempe, Arizona)	35.414842	-114.909319	42
DoD Gateway (Wahiawa, HI)	23.502574	-161.022938	42
HQ (Bethesda, MD)	40.98472	-80.09472	42
GEOSAT1	0	-180	182
GEOSAT2	0	-135	182
GEOSAT3	0	-90	182
GEOSAT4	0	-45	182
GEOSAT5	0	0	182
GEOSAT6	0	45	182
GEOSAT7	0	90	182
GEOSAT8	0	135	182

Bibliography

"0303110F Def Satellite Comm Sys". FY98 USAF Military Space RDDS. 1999. http://www.fas.org/spp/military/budget/peds_98f/0303110f.htm (accessed 27 March 2013)

Alderson, D.L., & Brown, G.G., Carlyle, W.M., Wood, R.K. "Solving Defender-Attacker-Defender Models for Infrastructure Defense". *Operations Research, Computing, and Homeland Defense.* INFORMS 2011, Hanover, MD, pp. 28-49. http://faculty.nps.edu/dlalders/. (Accessed 8 October 2013)

Argonne National Laboratory. "RI Dashboard". http://www.dis.anl.gov/images/ri_dashboard_RI4.jpg. (Accessed 3 February 2014)

Arroyo, Jose M. & Francisco Galiana. "On the Solution of the Bilevel Programming Formulation of the Terrorist Threat Problem." *IEEE Transactions on Power Systems*: Volume 20, Number 2. May 2005

Bard, Jonathan. "Practical Bilevel Optimization: Algorithms and Applications". Springer, 1998.

Brown, Gerald & Matthew Carlyle, Javier Salmeron, Kevin Wood. "Analyzing the Vulnerability of Critical Infrastructure to Attack and Planning Defenses". *Tutorials in Operations Research, INFORMS.* 2005.

Brown, Gerald & Matthew Carlyle, Javier Salmeron, Kevin Wood. "Defending Critical Infrastructure". Interfaces Vol. 36, No. 6, November 2006.

Cain, Fraser. "Circumference of the Earth". Universe Today. March 2009. http://www.universetoday.com/26461/circumference-of-the-earth/ (accessed 16 March 2013)

Calvete, Herminia, & Carmen Gale, Maria-Jose Oliveros. "Bilevel model for production-distribution planning solved by using ant colony optimization". Computer and Operations Research, 38:1, pages 320-327. January 2011. http://www.sciencedirect.com/science/article/pii/S0305054810001206 (accessed 24 July 2013)

Carlson, L. & G. Bassett, W. Buehrin, M. Collins, S. Folga, B. Haffenden, F. Petit, J. Phillips, D. Verner, and R. Whitfield. "Resilience: Theory and Applications". Argonne National Laboratory, Decision and Information Sciences Division. January 2012.

Cheruku, Dharma Raj. "Satellite Communication". I. K. International Pvt Ltd. 2010.

CIP Vigilance. "Power Grid-Interactive Simulation". http://ciip.wordpress.com/tag/power-grid/ (accessed 18 March 2013)

Clark, Stephen. "Sizing up America's Place in the Global Launch Industry." Spaceflight Now. November 24, 2013. http://spaceflightnow.com/falcon9/007/131124commercial/ (Accessed 4 February 2014)

"Commercial Communications Satellites". The Exploration of Space. http://www.century-of-flight.net/Aviation%20history/space/Commercial%20Communications%20Satellites.htm (accessed 25 March 2013)

Committee on Appropriations. "Department of Defense Appropriations Bill, 2012". House of Representatives. 2011. http://appropriations.house.gov/uploadedfiles/fy_2012_defense_full_committee_report.pdf (accessed 27 March 2013)

Corporate Staff. "AEHF Satellite Program Advances". Aerospace. January 2013. http://www.aerospace.org/2013/01/31/aehf-satellite-program-advances/ (accessed 15 March 2013)

Costlow, Terry. "DOD builds foundation for future SATCOM". Defense Systems. February 25, 2013. http://defensesystems.com/articles/2013/02/25/special-report-dod-future-satellite-communications.aspx (accessed 18 March 2013)

Cimellaro, Gian & Andrei Reinhorn, Michel Bruneau. "Framework for analytical quantification of disaster resilience". Engineering Structures. Vol. 32, pp3639-3649. 2010.

Cushman, John. "U.S. Satellite Plans Falter, Imperiling Data on Storms." New York Times. October 26, 2012.

Defense Information Systems Agency (DISA). "Fiscal Year 2011 Budget Estimates". Department of Defense. February 2010. http://comptroller.defense.gov/defbudget/fy2011/budget_justification/pdfs/01_Operation_and_Maintenance/O_M_VOL_1_PARTS/DISA_FY11.pdf (accessed 27 March 2013)

Department of Homeland Security (DHS). "National strategy for homeland security". 2002

DIA. "Network Services Telecommunications Service Level Agreement". Customer Services Division. November 2012. http://www.disa.mil/Services/Network-Services/Service-Level-Agreement (accessed 28 March 2013)

DISA. "COMSATCOM Scoop: Volume 5". Department of Defense. January 2013.

DISA. "COMSATCOM Services". Department of Defense. http://www.disa.mil/Services/SATCOM/Comsatcom-Services (accessed 3 April 2013)

DISA. "Satellite Communications". Department of Defense. http://www.disa.mil/Services/SATCOM (accessed 28 March 2013)

DISA. "Telecommunications Service Level Agreement (SLA)". Department of Defense. November 2012. http://www.disa.mil/Services/Network-Services/Service-Level-Agreement (accessed 17 May 2013)

DHS. "National Infrastructure Protection Plan: 2006". Diane Publishing. June 20, 2006.

DHS. "DHS Risk Lexicon." Risk Steering Committee. September 2008 http://www.dhs.gov/xlibrary/assets/dhs_risk_lexicon.pdf *(*Accessed 30 Apr 2013*)*

DHS. "National Infrastructure Protection Plan. Partnering to Enhance Protection and Resiliency." January 2009.

Dictionary.com. "Resilience". *Collins English Dictionary - Complete & Unabridged 10th Edition*. Source location: HarperCollins Publishers. http://dictionary.reference.com/browse/resilience. (accessed 15 March 2013)

Department of Defense (DoD). "National Security Space Strategy: Unclassified Summary". January 2011. http://www.defense.gov/home/features/2011/0111_nsss/docs/NationalSecuritySpaceStrategyUnclassifiedSummary_Jan2011.pdf (accessed 11 July 2013)

"Fact Sheet: Resilience of Space Capabilities." January, 2011. http://www.defense.gov/home/features/2011/0111_nsss/docs/DoD%20Fact%20Sheet%20-%20Resilience.pdf (Accessed 12 March 2014)

Encyclopedia Astronautica. "Iridium". Astronautix. http://www.astronautix.com/project/iridium.htm#chrono (accessed 25 March 2013)

EO 13010. "Critical Infrastructure Protection." Executive Order. The White House. http://www.fas.org/irp/offdocs/eo13010.htm July 1996. (accessed 15 March 2013)

EO 13636. "Improving Critical Infrastructure Cybersecurity." Executive Order. The White House. http://www.fas.org/irp/offdocs/eo13010.htm February 2013. (Accessed 15 March 2013)

Eremenko, P. & C. Roberts & O. Brown. "Cost Benefit Analysis of a Notional Fractionated SATCOM Architecture." AIAA International Communications Satellite Systems Conference, 11-14 June 2006. http://arc.aiaa.org/doi/abs/10.2514/6.2006-5328 (Accessed 30 July)

Eshel, Tamir. "Raytheon Awarded $321 Million for the Production of 24 SM-3 Block 1B Missiles." Defense Updated. March 30, 2011. http://defense-update.com/20110330_sm3_block1b.html (Accessed 4 February 2014)

FAS. "DSCS-3". Military Space Programs. 1999. http://www.fas.org/spp/military/program/com/dscs_3.htm (Accessed 27 March 2013)

Galdorisi, George. "U.S. Navy Missile Defense: Operation Burnt Frost." Defense Media Network. May 18, 2013. http://www.defensemedianetwork.com/stories/u-s-navy-missile-defense-operation-burnt-frost/ (Accessed 3 December 2013)

Geo-orbit. "Geostationary, LEO, MEO, HEO Orbits". http://www.geo-orbit.org/sizepgs/geodef.html (Accessed 23 October 2013)

Global Horizons Team. "Global Horizons: United States Air Force Global Science and Technology Vision." AF/ST TR 13-01. 21 June 2013. http://www.defenseinnovationmarketplace.mil/resources/GlobalHorizonsFINALREPORT6-26-13.pdf . (Acessed 23 October 2013)

Graham, William. "ULA Atlas V successfully deploys new GPS satellite." NASA Spaceflight.com. May 15, 2013. http://www.nasaspaceflight.com/2013/05/ula-atlas-v-launch-new-gps-satellite/ (Accessed 20 August 2013)

Haimes, Y.Y. ""Risk Modeling, Assessment, and Management: Third Edition." John Wiley and Sons Inc. Publishing. 2009.

Hill, Edward & Clair, Wial, Wolman, Atkins, Blumenthal, Ficenec, Friedhoff. "Economic Shocks and Regional Economic Resilience". George Washington, Urban Institute. May, 2010.

Holling, C.S. "Resilience and stability of ecological systems." Annual Review of Ecology and Systematics 4. 1973

Ignizio, James. "Goal Programming and Extensions". Lexington Books, University of Michigan. 1976.

Iridium. "The Global Network: Satellite Constellation". July 2012. http://www.iridium.com/About/IridiumGlobalNetwork/SatelliteConstellation.aspx (Accessed 255 March 2013)

"Iridium Satellite Phones". Global Satellite Communications. http://www.globalsatellitecommunications.com/iridium (Accessed 25 March 2013)

Joint Publication 3-3 "Joint Interdiction." Department of Defense. 14 October 2011.

Jennings, Barbara & Eric Vugrin & Deborah Belasich. "Resilience Certification For Commerical Buildings: A Study of Stakeholder Perspectives." Springer Environment Systems and Decisions: Volume 33, Issue 2, pages 184-194. June 2013 http://link.springer.com/article/10.1007%2Fs10669-013-9440-y (Accessed 30 July 2013)

Keil, Todd M. "Enhancing Ctirical Infrastructure Resilience". Infrastructure Protection. 22 December, 2010. http://www.dhs.gov/blog/2010/12/22/enhancing-critical-infrastructure-resilience (Accessed 15 March 2013)

Keohane, Jonathan & Gail Rohrbach. "Ask an Astrophysicist". Imagine the Universe, NASA. April 1997. http://imagine.gsfc.nasa.gov/docs/ask_astro/answers/970408d.html (Accessed 25 March 2013)

Klibi, Walid & Alain Martel. "Modeling Approaches for the Design of Resilient Supply Networks under Disruptions". Cirrelt. November 2010. http://www.sciencedirect.com/science/article/pii/S0925527311004580 (Accessed 15 March 2013)

Kolmogorov, Andrei & Sergei Fomin. "Introductory Real Analysis." Courier Dover Publications. New York. 1975.

Kumar, K & V. Joshi. "An attitude control approach for compensation of satellite drift in elliptic synchronous orbits." Acta Astronautica, Vol. 8, No. 7, pp. 719-731, 1981. http://ac.els-cdn.com/0094576581900138/1-s2.0-0094576581900138-main.pdf?_tid=acb1acca-09aa-11e3-b6d7-00000aab0f02&acdnat=1377011619_c27c2ca2344af69fa3286239ecadaf92 (Accessed 20 August 2013)

Lewis, Ted G. "Critical Infrastructure Protection in Homeland Security". John Wiley & Songs. April 2006.

Lim, Churlzu & J. Smith. "Algorithms for discrete and continuous multicommodity flow network interdiction problems." IIE Transactions, 39:1, 15-26. April 2007. http://dx.doi.org/10.1080/07408170600729192. (Accessed 24 July 2013)

Lockheed Martin. "Double Buy for AEHF Satellites." Lockheed Martin. 2012. http://www.lockheedmartin.com/us/news/features/2012/communications-satellites.html (Accessed 4 February 2014)

Luthar, S.S & Cicchetti, D. "The construct of resilience: Implications for interventions and social policies." Development and Psychopathology, 12. 2000

"Maximized Network Availability With Diverse VSAT Routing". Encore Networks. http://www.encorenetworks.com/app_note/app_note_diverse_vsat_routing.htm (Accessed 15 March 2013)

Meckling, William. "Communications Satellites: Supplemental Information on the Cost Estimates Given in Research Memorandum RM-2709-NASA." RAND. June 30, 1961. http://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM2778.pdf. (Accessed 5 August 2013)

Miller, Robert D. "Computing the Area of a Spherical Polygon." Graphic Gems: Section 11.4. 2008. http://www.iut-arles.up.univ-mrs.fr/raffin.r/myspip/fichiers_static/lpin/reflexions2008/Computing%20the%20Area%20of%20a%20Spherical%20Polygon_Graphics%20Gems%20IV_pp147.pdf (Accessed 13 March 2014)

Moore, James Thomas. "Extensions to the Multilevel Programming Problem." University of Texas at Austin, May 1988.

Moskowitz, Clara. "US Headed for 'Perfect Storm' in Space, Air Force General Says." SPACE.com http://www.space.com/20586-military-space-reliance-perfect-storm.html 9 April 2013. (Accessed 17 April 2013)

National Geographic. "Earth's Atmosphere". http://science.nationalgeographic.com/science/earth/earths-atmosphere/ (Accessed 15 March 2013)

National Institute of Standards and Technology (NIST) "Common Vulnerability Scoring System Version 2 Calculator". National Vulnerability Database. http://nvd.nist.gov/cvss.cfm?calculator&version=2 (Accessed 18 March 2013)

National Security Telecommunications Advisory Committee (NSTAC). "NSTAC Secure Government Communications Scoping Report". January 2013. http://www.ncs.gov/nstac/nstac_publications.html (Accessed 18 March 2013)

NSTAC. "NSTAC Report to the President on Commercial Satellite Communications Mission Assurance". November 2009. http://www.ncs.gov/nstac/nstac_publications.html (Accessed 18 March 2013)

National Telecommunications and Information Administration (NTIA). "FY 2011 Budget as Presented to Congress". U.S. Department of Commerce. February 2010. http://www.osec.doc.gov/bmi/budget/11CJ/NTIA%20FY2011%20Congressional%20Budget.pdf (Accessed 27 March 2013)

Nykamp, D. Q. "Network definition." *Math Insight*. http://mathinsight.org/network_definition (Accessed 5 August 2013)

Office of the Press Secretary. "PPD-21 Critical Infrastructure Security and REsilience." Presidential Policy Directive. The White House. http://www.whitehouse.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil (Accessed 15 March 2013)

Office of the Under Secretary of Defense (Comptroller). "Fiscal Year 2012 Budget Estimates". Department of Defense Efficiency Initiatives. http://comptroller.defense.gov/defbudget/fy2012/FY2012_Efficiency_Justification_Book.pdf (Accessed 27 March 2013)

Office of the Under Secretary of Defense (Comptroller). "Defense Working Capital Fund: Defense-Wide Fiscal Year 2013 Budget Estimates". Department of Defense. February 2012. http://comptroller.defense.gov/defbudget/fy2013/budget_justification/pdfs/06_Defense_Working_Capital_Fund/PB_13_DWWCF_Operating_Budget.pdf (Accessed 27 March 2013)

Patera, Russell P.. "General Method for Calculating Satellite Collision Probability", Journal of Guidance, Control, and Dynamics, Vol. 24, No. 4 (2001), pp. 716-722. http://arc.aiaa.org/doi/abs/10.2514/2.4771?journalCode=jgcd. (Accessed 20 August 2013)

Peat, Chris. "AEHF 1 (USA 214) -Orbit". Heavens Above. February 2013. http://www.heavens-above.com/orbit.aspx?satid=36868 (Accessed 15 March 2013)

Polaris Project. "Types of Orbits". Iowa State University. http://www.polaris.iastate.edu/EveningStar/Unit4/unit4_sub3.htm (Accessed 25 March 2013)

Powel, Robert W. "DoD Hands Qwest $100M Contract". Telecom Ramblings. April 2011. http://www.telecomramblings.com/2011/04/dod-hands-qwest-100m-contract/ (Accessed 28 March 2013)

Pratt, Stephen & Richard Raines, Carl Fossa, Michael Temple. "An Operational and Performance Overview of the IRIDIUM Low Earth Orbit Satellite System." *IEEE Communications Surveys*: Second Quarter, 1999. http://www.comsoc.org/pubs/surveys (Accessed 19 April 2013)

Press Center. "Global Star Announces Successful Second Launch of Six New Satellites". Globalstar Inc. July 2011. http://www.globalstar.com/en/index.php?cid=7010&pressId=681 (Accessed 25 March 2013)

Reibeek, Holli. "Catalog of Earth Satellite Orbits". NASA. September 2009. http://earthobservatory.nasa.gov/Features/OrbitsCatalog/ (Accessed 25 March 2013)

Romanosky, Sasha & Peter Mell, Karen Scarfone. "A Complete Guide to the Common Vulnerability Scoring System Version 2.0". First. http://www.first.org/cvss/cvss-guide (Accessed 18 March 2013)

Royden, H. L. & P. M. Fitzpatrick. "Real Analysis: 4th Edition." Pearson Eduction Inc, Boston, MA. 2010.

Salmeron, Javier & Kevin Wood, Ross Baldick. "Analysis of Electric Grid Security Under Terrorist Threat". *IEEE Transactions on Power Systems, Vol. 19.* May 2004.

Salmeron, Javier & Kevin Wood, Ross Baldick. "Worst Case Interdiction Analysis of Large Scale Electric Power Grids". *IEEE Transactions on Power Systems, Vol. 24.* February 2009.

SatBeams. "List of Satellites at Geostationary Orbit". SatBeams. March 2013. http://www.satbeams.com/satellites (Accessed 25 March 2013)

Satellite Servicing Capabilities Office (SSCO). "Robotic Refueling Mission." Nasa. http://ssco.gsfc.nasa.gov/robotic_refueling_mission.html. (Accessed 5 August 2013)

"Satellite Network." *ATIS Telecom Glossary 2007*. ATIS. http://www.expertglossary.com/definition/. (Accessed 15 March 2013)

Science Clarified. "Earth's Interior". http://www.scienceclarified.com/Di-El/Earth-s-Interior.html (Accessed 15 March 2013)

Sladen, Rod. "Iridium Constellation Status". Rod Sladen Satellite Watching. November 22, 2012. http://www.rod.sladen.org.uk/iridium.htm (Accessed 13 August 2013)

Solnon, Christine. "Ant Colony Optimization and Constraint Programming." *John Wiley and Sons Inc*., Hoboken, NJ. 2010.

Sullivan, Laura. "Why Resilience Matters." ActionAid. http://www.actionaidusa.org/eu/2012/10/why-resilience-matters. October, 2012 (Accessed 15 March 2013)

Szpankowski, Wojiech. "Average Case Analysis of Algorithms on Sequences: Chap 3." Department of Computer Science, Purdue University. http://www.cs.purdue.edu/homes/spa/book.html. December, 2001. (Accessed 24 June 2013)

Taleb, Nassim Nicholas. "The Black Swan: Second Edition: The Impact of the Highly Improbable Fragility." Random House LLC, May 11, 2010.

Tatum, Steve. "Lockheed Martin-Built Milstar Satellite Surpasses 10-year On Orbit Design Life". Lockheed Martin. February 2012. http://www.lockheedmartin.com/us/news/press-releases/2012/february/0210-ss-milstar.html (Accessed 15 March 2013)

Tatum, Steve. "Second Advanced EHF Military Communications Satellite Built By Lockheed Martin Launched Successfully for the US Air Force". Lockheed Martin. May 2012. http://www.lockheedmartin.com/us/news/press-releases/2012/may/0504-ss-aehf-2.html (Accessed 15 March 2013)

United States Government Accountability Office (GAO). "Critical Infrastructure Protection." October 2012. *www.gao.gov/assets/650/649705.pdf* (Accessed 15 March 2013)

Vugrin, Eric D. &, Drake E. Warren, Mark A. Ehlen, R. Chris Camphouse. "A Framework for Assessing the Resilience of Infrastructure and Economic Systems". *Sustainable and Resilient Critical Infrastructure Systems*. Springer. 2010.

Weinberger, Sharon. "Cruise Missiles: The Million-Dollar Weapon." Huffington Post: Business. March 25, 2011. http://www.huffingtonpost.com/2011/03/25/cruise-missiles-missile_n_840365.html (Accessed 4 February 2014)

Werner, Markus & Axel Jahn, Erich Litz, Axel Bottcher. "Analysis of System Parameters for LEO/ICO Satellite Communication Networks." *IEEE Journal on Selected Areas in Communications*. Vol. 13, No. 2, pp 371-381. February 1995. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=345881 (Accessed 13 November 2013)

Whalen, David J. "Communications Satellites: Making the Global Village Possible". NASA. http://history.nasa.gov/satcomhistory.html (Accessed 25 March 2013)

Wright, David & Laura Grego, Lisbeth Gronlund. "The Physics of Space Security". 2005. http://www.amacad.org/publications/Physics_of_Space_Security.pdf (Accessed 25 March 2013)

Zobel, Christopher & Lara Khansa. "Characterizing Multi-event disaster resilience". *Computers & Operations Research*. Vol. 42, pp83-94. 2014.

Vita

Second Lieutenant Jonathan S. Turner graduated from Alvarado High School in Alvarado, Texas in 2008. He entered undergraduate studies at Texas State University-San Marcos where he graduated with a Bachelor of Science degree in Mathematics in May 2011. He then went on to earn his Masters of Science in Mathematics in August 2012. He was commissioned through the Detachment 840 AFROTC at Texas State University.

His first assignment was at Wright Patterson AFB where he entered the Graduate School of Engineering and Management, Air Force Institute of Technology, as an Operations Research Master's student. Upon graduation, he is assigned to Rome Labs, Rome, NY.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
27 Mar 2014	Master's Thesis	Aug-2012 - Mar-2014

4. TITLE AND SUBTITLE

A Methodology For Measuring Resilience in a Satellite-Based Communication Network

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Turner, Jonathan S, 2dLt

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
Wright-Patterson AFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-14-M-31

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Headquarters Air Force A-9 Information Systems
James Muccio, GS-15, DAF, Chief, Integrated Space Analysis
1570 Air Force Pentagon Room 4E214
Washington, DC 20330-1570
James.Muccio@pentagon.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

HQ USAF/A9IS

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution Statement A: Approved For Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States."

14. ABSTRACT

According to Presidential Policy Directive 21, increasing resilience of critical infrastructures is not only desired, but United States policy. Communications infrastructures are one such critical infrastructure.
The purpose of this research is to develop a methodology for measuring resilience in satellite communication systems for use as a key criterion in the selection and acquisition of new satellite architectures, in accordance with the National Security Space Strategy. The base methodology utilized in this thesis is Extreme Event Modeling implemented through the use of Bi-Level Programming with monotonically nonlinear continuous and mixed integer variables. This model differs from previous efforts applied to other critical infrastructures in that it captures the temporal component associated with multiple events, as well as the repairs, or reconstitution, of infrastructure components. Furthermore, a heuristic based upon a ratio of impact to cost and local searches is developed to solve the resulting continuous bi-level problem.

15. SUBJECT TERMS

Resilience, Bi-Level, Attacker-Defender, Risk, Robust

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Richard F. Deckro, AFIT/ENS
U	U	U	UU or SAR	259	19b. TELEPHONE NUMBER *(Include area code)* (937) 785-3636 x4325 Richard.Deckro@Afit.edu