NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

A GEOGRAPHIC AND FUNCTIONAL NETWORK FLOW ANALYSIS TOOL

by

Kevin M. Martin

June 2014

Thesis Co-Advisors: David L. Alderson
Rudolph Darken

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2014	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE
A GEOGRAPHIC AND FUNCTIONAL NETWORK FLOW ANALYSIS TOOL

5. FUNDING NUMBERS

6. AUTHOR(S) Kevin M. Martin

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB protocol number ____N/A____.

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

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

Critical infrastructure systems, such as water and electricity, are important for society and national defense. There is a need for network analysis tools that allow analysts to study potential scenarios to discover vulnerabilities, assess consequences, and evaluate effective solutions to overcome network weaknesses. In order to be useful, models of critical infrastructure systems need to be realistic, both geospatially and functionally. The objective of this thesis is to bridge the gap between geospatial and functional network analysis by developing a software tool that allows users to create and edit networks in a Graphical Information System (GIS) visual environment, and then also run and view the results of functional network models. Our primary contribution is to provide an easy-to-use, graphical interface in the form of a plugin that allows users, regardless of their network expertise, to create networks and exercise network flow models on them. We demonstrate the usefulness of our plugin through the analysis of a fictional case study with a realistic Internet infrastructure. We run several minimum cost flow models with simulated network attacks to assess the robustness of the network.

14. SUBJECT TERMS Network Analysis, Minimum Cost Flow, Networks, Interdiction			**15. NUMBER OF PAGES** 69
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

A GEOGRAPHIC AND FUNCTIONAL NETWORK FLOW ANALYSIS TOOL

Kevin M. Martin
Ensign, United States Navy
B.S., United States Naval Academy, 2013

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
June 2014

Author: Kevin M. Martin

Approved by: David L. Alderson
Thesis Co-Advisor

Rudolph Darken
Thesis Co-Advisor

Peter J. Denning
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Critical infrastructure systems, such as water and electricity, are important for society and national defense. There is a need for network analysis tools that allow analysts to study potential scenarios to discover vulnerabilities, assess consequences, and evaluate effective solutions to overcome network weaknesses. In order to be useful, models of critical infrastructure systems need to be realistic, both geospatially and functionally. The objective of this thesis is to bridge the gap between geospatial and functional network analysis by developing a software tool that allows users to create and edit networks in a Graphical Information System (GIS) visual environment, and then also run and view the results of functional network models. Our primary contribution is to provide an easy-to-use, graphical interface in the form of a plugin that allows users, regardless of their network expertise, to create networks and exercise network flow models on them. We demonstrate the usefulness of our plugin through the analysis of a fictional case study with a realistic Internet infrastructure. We run several minimum cost flow models with simulated network attacks to assess the robustness of the network.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

CHDS	Center for Homeland Defense and Security
CSV	Comma Separated Value
FDGC	Federal Geographic Data Committee
DHS	Department of Homeland Security
GAMS	General Algebraic Modeling System
Gbps	Gigabits per second
GUI	Graphical User Interface
GIS	Geographic Information System
PPD-21	Presidential Policy Directive 21
QGIS	Quantum GIS
XML	Extensible Markup Language

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First, I must thank Professor David Alderson. I was truly fortunate to have him as one of my advisors. I cannot express enough how grateful I am for his guidance and knowledge throughout this thesis project. He made this large undertaking into something I was able to manage and complete.

I would also like the thank Professor Rudolph Darken for his continued support and constant optimism. He kept me focused while working on this project. I hope this plugin becomes something that is used often throughout the Operations Research department.

Last, I would like to thank Erik Johnson and his team for answering my programming questions. I greatly appreciated their willingness to help me. I also really enjoyed sitting in on the weekly developer meetings.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Presidential Policy Directive 21 (PPD-21), titled "Critical Infrastructure Security and Resilience" and signed February 12, 2013 lists 16 infrastructure sectors that are vital to national security (White House 2013). These infrastructures include water, energy, transportation, banking and finance, and information and telecommunications, among others.

PPD-21 tasks the Department of Homeland Security (DHS) with increasing the security and resilience of our nation's critical infrastructures. Commercial enterprises are also interested in improving the efficiency and reliability of their private systems. Infrastructure systems are prone to many disruptions, both non-deliberate and deliberate. Non-deliberate mishaps may result from weather, technical failure, and operator error whereas deliberate acts include terrorism and vandalism. Knowing where infrastructure systems are vulnerable to deliberate and non-deliberate disruptions is exceedingly useful to increase resilience.

Most of the 16 critical infrastructure sectors are made up of systems that are networks or key components in a larger infrastructure network. The interdependencies of these network systems are so complicated that studying each critical infrastructure's resilience to disruptive events requires network analysis tools. Network analysis tools allow operators to experiment with potential scenarios to discover vulnerabilities, assess consequences, and evaluate effective solutions to overcome network weaknesses.

To be useful for critical infrastructure systems analysis, models must be realistic, both geographically and functionally. Geographic realism provides analysts with the spatial relationship between system components. Geographic representations of infrastructures are often simply maps showing the locations of physical structures. Functional realism models how the system works relying on the interactions between the components of a network. This requires detailed data. But it is often the case that real network data is kept private for security and proprietary reasons. Researchers often work around this by inferring a network's topology through experimentation or open source

data collection (e.g., Alderson et al. 2005), but the lack of realistic network topology data remains a problem nonetheless. Without the ability to run analytic techniques on real networks, these techniques are of limited value. Moreover, the lack of accessible, real network data means that there are no benchmarks for the analysis models used on simple, "toy" networks. In addition, new models cannot be validated or tested for accuracy before they are used to assess a real network. These models rely heavily on network theory but have scarcely been verified in modern applications.

There is a large focus on Geographic Information Systems (GIS) when cataloging and studying critical infrastructures (Federal Geographic Data Committee 2014). A GIS system visually represents the infrastructure on a map offering the geographic realism necessary for some types of network analysis. One problem with the exclusive use of GIS for studying critical infrastructure is that these representations are no more than drawings. The functional relationships between the images representing system components and the images representing the links between them are not represented. Although they are useful inventories of geographic data, GIS models typically cannot be used for functional network analysis, including concrete 'what if' scenarios.

To address this shortcoming, there is considerable work being done on functional models of key infrastructure networks. Researchers at the Naval Postgraduate School's Department of Operations Research and Center for Infrastructure Defense have studied critical infrastructures through the use of attacker-defender models (e.g., Brown et al. 2005, 2006; Barkley 2008; Dixon 2011; Crain 2012). These models use game theory and optimization models to find worst-case disruptions to infrastructure function, accounting for the ways in which infrastructure owners and operators try to adapt to disruptive events. Like GIS models, attack-defender models have great use, but also significant limitations. For example, applying attacker-defender analysis to an infrastructure network is not easy and typically requires domain expertise to build, execute and interpret. In addition, these models typically require expensive, proprietary software to be used. They manipulate problem data using simple text files for input and output, but they historically have offered little visualization of either input or output.

The objective of this thesis is to bridge the gap between geographic and functional network analysis. We aim to develop a tool that allows users to create and edit networks in a GIS visual environment and also allow them to run and view the results of functional models. Our goal is to "close the loop" between the network creation environment provided by GIS tools and the functional analysis tools provided by network models (see Figure 1).

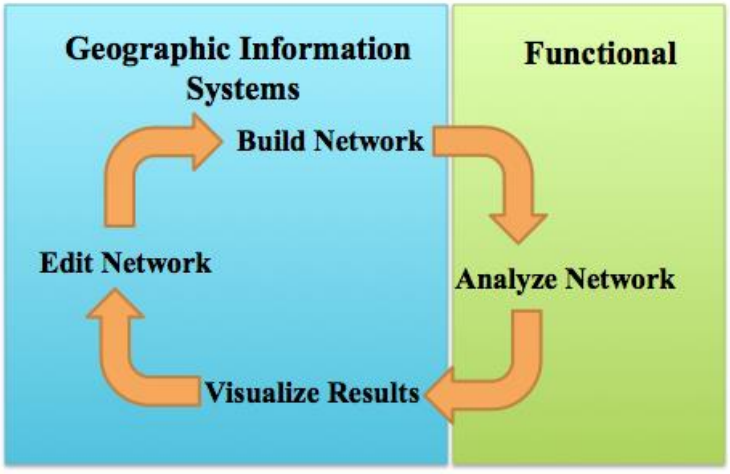

Figure 1. Our tool completes the loop between geographic and functional network model analysis.

We demonstrate the usefulness of this framework by creating a sample network and running realistic scenarios on it. Specifically, we model a fiber optic communications network in the Dystopia virtual environment, which is a fictitious example of a GIS environment. This network is functionally realistic, in the sense that it adheres to the technological constraints at work for real systems and it carries network traffic for realistic populations. Using a functional model, we assess the impact of losing one or more network connections, and we also evaluate the potential benefit of specific investments in component hardening, capacity expansion, or new network construction.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND AND RELATED WORK

We start by reviewing previous research in Geographic Information Systems (GIS) modeling and functional network analysis. Then we summarize recent work attempting to improve the limitations associated with each of these.

A. GEOGRAPHIC INFORMATION SYSTEMS TOOLS AND DYSTOPIA

The Federal Geographic Data Committee (FGDC) is an interagency organization that promotes the coordinated development, use, sharing, and dissemination of geospatial data. Geospatial information technology provides homeland security decision-makers with important information to handle disruptions that include natural disasters, terrorist attacks, and sabotage. The FGDC lists several major benefits of GIS data for homeland security applications: detection of weaknesses, preparedness for incidents, prevention of threats and attacks, protection against failure, and more effective response and recovery. (Federal Geographic Data Committee 2006)

Dystopia (Center for Homeland Defense and Security 2013) is a collection of geospecific metadata that creates a comprehensive virtual world. It was developed by the Naval Postgraduate School's Center for Homeland Defense and Security (CHDS) to be a realistic, flexible scenario environment for educational exercises. These scenarios task students with finding solutions to the issues posed in the exercise. Dystopia is not a game itself, but a context for the games to be run. By definition, a dystopia is "a place where bad things happen"—a fitting name for a place where the scenarios are often full of destruction. As shown in Figure 2, Dystopia is an island annotated with detailed spatial GIS information. It contains many population centers divided by a national border.

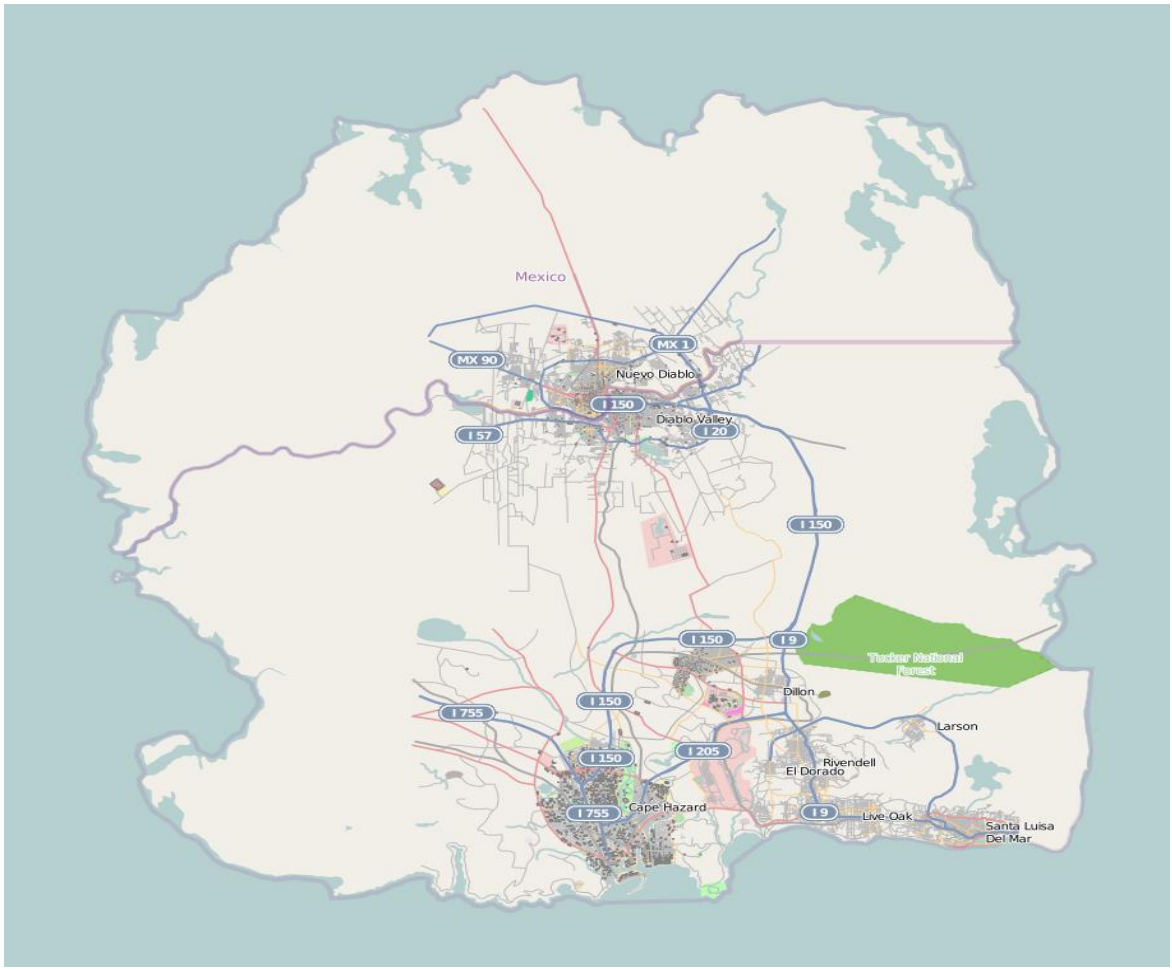

Figure 2. Screenshot from Dystopia's web interface hosted by the Naval Postgraduate School's Center for Homeland Defense and Security. (Center for Homeland Defense 2013)

Dystopia contains several GIS layers that include key infrastructure systems such as oil, power, water, important points of interest (e.g., commercial, education, government, military), and transportation infrastructure (e.g., railways, roads, and airports).

Dystopia's usefulness stems from its realistic detail, while still being a fictitious world. This allows Dystopia to be used to practice real world scenarios without the complications of using real world data. The limitation of Dystopia is that it is merely a map with images on it. Zooming in reveals detailed roads and buildings, but they are not represented by anything other than the lines drawn on the screen. This thesis work

attempts to overcome this fundamental GIS limitation creating additional data layers to support the creation and analysis of functional models.

B. MODELS OF INFRASTRUCTURE FUNCTION

Functional models allow network analysts to represent their networks as mathematical objects that can be used to run simulations. These simulations give insight to network efficiency, redundancy, and security—exactly the attributes that network operators constantly review. One class of functional models is based on the use of network flows as described by Ahuuja, Magnanti, and Orlin (1993). There exists a variety of network flow models, such as minimum cost flow, shortest path, maximum flow, assignment, transportation, circulation, and multicommodity flow. This thesis focuses on minimum cost network flow problems. Like the name suggests, a minimum cost flow problem seeks the least costly way to transfer a commodity from one node to another, taking into consideration the cost and capacities of each arc that lies in between. Minimum cost models can be applied to any commodity distribution service, making these models extremely useful.

Throughout this thesis, a network will be described by a set of nodes and a set of directed arcs. *Nodes* are the connection points, and *arcs* are the one-way links between them. Consider an electric power system as an example; power plants can be represented by nodes and power lines can be represented by arcs. Networks are more than simple mathematical graphs because nodes and arcs have attributes associated with them. These attributes are customizable and provide the flexibility for network models to be applied to most any infrastructure network. Using again the example of the electric power grid, power station nodes may have the following attributes: power output capacity (watts), cost (dollars/year), fuel type (coal/nuclear), etc. Power line arcs may also have several attributes: length (miles), voltage (volts), current (amperes), etc. These attributes are different than the node and arc attributes in the water sector and those in the information technology sector.

Attacker-defender models combine game theory with network flow models to provide worst-case modeling scenarios. They assume an intelligent attacker, able to

choose a point of weakness to attack. The defender tries to minimize the effects of these attacks through security measures such as network hardening and redundancies. We look at previous work applying attacker-defender models to communication networks in order to give insight into extending the GIS's additional functionality for vulnerability analysis.

Barkley (2008) uses an attacker-defender model to analyze communication networks based on the Internet Protocol (IP). Assuming an IP infrastructure, Barkley considers point-to-point traffic or "flows" and looks at both extremes of packet routing: shortest path and maximum flow. Traffic on a shortest path infrastructure will travel the smallest distance (fewest hops, lowest "mileage") from point 'A' to 'B'. In contrast, a maximum flow problem takes advantage of all available paths and capacities to transfer flows from 'A' to 'B'. Barkley compares both routing implementations and the optimal interdiction to most significantly reduce the flow rate.

Crain (2012) assesses the robustness of the present day undersea cable infrastructure. He considers the redundancies of more than 220 real cable systems, and he uses attacker-defender models on these networks to test redundancy and look for weaknesses. Crain uses *gravity models* (see Alderson et al. 2006 for a review) to generate a realistic traffic matrix. Gravity models in a networking sense are analogous to the law of gravitation in physics. Physics states that the larger the mass of an object, the higher the force of gravity on surrounding objects. Using a city's population analogously to mass we are able to estimate the traffic demand across a link.

Most network models consider only a single infrastructure in isolation. For example, a water system might be represented as a network of pipes (arcs) and pumps (nodes). These models often do not consider the interdependencies between infrastructures. Water pumps require electricity; therefore, the model analyzing the robustness of the water system should have some notion of the robustness of the electrical system that powers it. Dixon (2011) focuses on these interdependencies and the application of attacker-defender models to interdependent networks. Although this thesis work does not include details for specific interdependent layers, we used Dixon's work to lay the groundwork for the format and layer requirements.

Functional models can be extremely helpful to analyze networks for weaknesses; however, they traditionally have suffered from several limitations. First, functional models often lack the geographical relationships within networks found in GIS models. Second, many functional models rely on local computation using expensive third-party software licenses, complicating the already intricate workflow. Last, these models are not modular themselves making it difficult to transfer the output of one model to the input of another. This lack of modularity also prevents model creators from adapting existing models to more complex applications. Model makers must then create each model entirely themselves.

C. COMPLETING THE LOOP

Previous work has looked into simplifying the application of functional models on networks. Gun (2013) worked extensively with the General Algebraic Modeling System (GAMS 2010) computation engine. His goal was to create a cloud implementation for the GAMS software where users could interact with a web interface rather than a program running locally. This approach is similar to the goal of this thesis: automating the mathematical model in a simple user interface. Gun's major focus was model integrity checking and visualizations. The importance of model integrity checks is paramount, as this process has become extremely modular. There are many potential breakpoints if each process isn't verified properly. Gun's idea of visualization is different from the visualization expected in this thesis. While Gun expected immediate visualization of the GAMS output, he considered only simple line, bar and pie charts. While these visualizations can be a nice way to see changes in data, viewing the effects of a model actually *on* the network is often more useful. For example, we consider the use of color gradients for a particular attribute to be reflected on the network for simple and easy visualization.

This thesis work relies heavily on Quantum GIS (QGIS 2012) to model our simulated networks. QGIS is a free, open source Geographic Information System software suite. We ran version 1.8.0-Lisboa for plugin compatibility reasons.

THIS PAGE INTENTIONALLY LEFT BLANK

III. ARCHITECTURE AND IMPLEMENTATION

The primary contribution of this thesis is the design and development of an extension to an existing plugin for Quantum GIS that allows users to run functional models on networks displayed in a GIS editor. Our modified plugin combines the benefits of the geographic information contained in the GIS metadata and the functional relationships between the attributes that make up the nodes and arcs of the network. This chapter first describes the network flow model we use to validate our workflow, and then it explains how our plugin uses the model's output to visualize the results.

A. NETWORK FLOW MODEL

Our proof of concept network analysis uses a minimum-cost network flow model described in Ahjua et al. (1993). This formulation is described in depth in Crain (2012), and for the most part we treat it as a black box. The model is run using GAMS and solved by CPLEX (ILOG 2007). The model requires an input network with specific attributes associated to the nodes and arcs. Nodes must have unique names, while arcs are described in terms of a "head", "tail", "capacity", and "cost". The model also requires a traffic matrix specifying the amount of traffic traveling from each node to every other node. This information is supplied to GAMS using comma-separated-value (CSV) files.

Consider the simple network illustrated in Figure 3 as our example case.

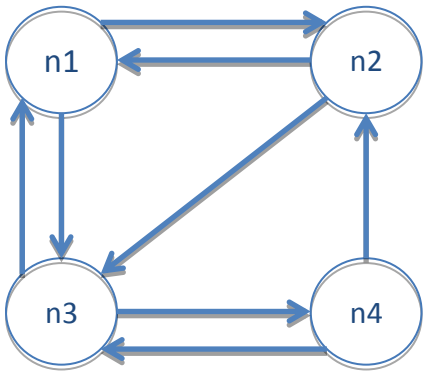

Figure 3. A simple network to run the network flow model.

For simplicity, we assume that every arc in this network has a capacity of 6 and a cost of 1. Table 1 shows the traffic matrix that is used. The units of flow in this simple example are notional.

End-To-End Traffic Demands

	n1	n2	n3	n4
n1	-8	2	2	4
n2	2	-8	4	2
n3	2	4	-8	2
n4	4	2	2	-8

Table 1. Traffic matrix for the network.

Table 2 shows the flows that minimize the routing cost while adhering to arc capacities. This information is returned after running the flow model on this simple network.

tail	head	flow	destination
n2	n1	6	n1
n3	n1	2	n1
n4	n2	4	n1
n1	n2	6	n2
n3	n1	4	n2
n4	n2	2	n2
n1	n3	2	n3
n2	n3	4	n3
n4	n3	2	n3
n1	n3	2	n4
n2	n3	2	n4
n3	n4	6	n4

Table 2. The output of the network flow model on the simple network in Figure 3 based on the traffic demands in Table 1.

Even for this simple case, the results in Table 2 are not easy to understand; we see the lack of visualization of the results for our network. Instead of reading the flow results from a table we would like to see the flows on the network itself, as in Figure 4. Our plugin aggregates this output, and updates the attributes within the visualization tool to show changes of flow.

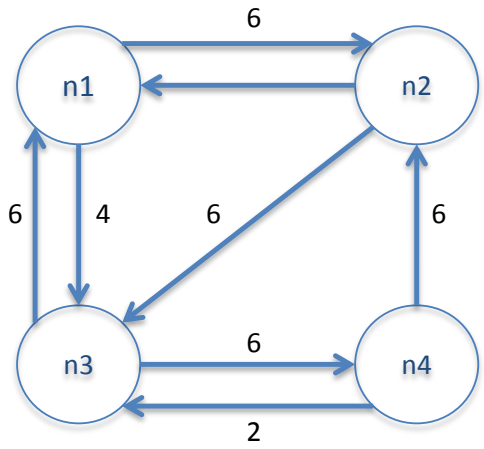

Figure 4. Our simple model redrawn to show network flow in a more intuitive manner.

B. PLUGIN FUNCTIONALITY

We extend the Quantum GIS ArcMaker plugin developed by Johnson et al. (2013) to include support for network simulation allowing users to run mathematic models on layers within QGIS. Figure 5 shows the extended plugin. The ArcMaker plugin allows users to create nodes and arcs that are logically connected. Moving a node also redraws all of the arcs connected to that node in order to stay connected to it. This is a feature that QGIS is currently missing, but a necessity when creating and editing complicated networks.

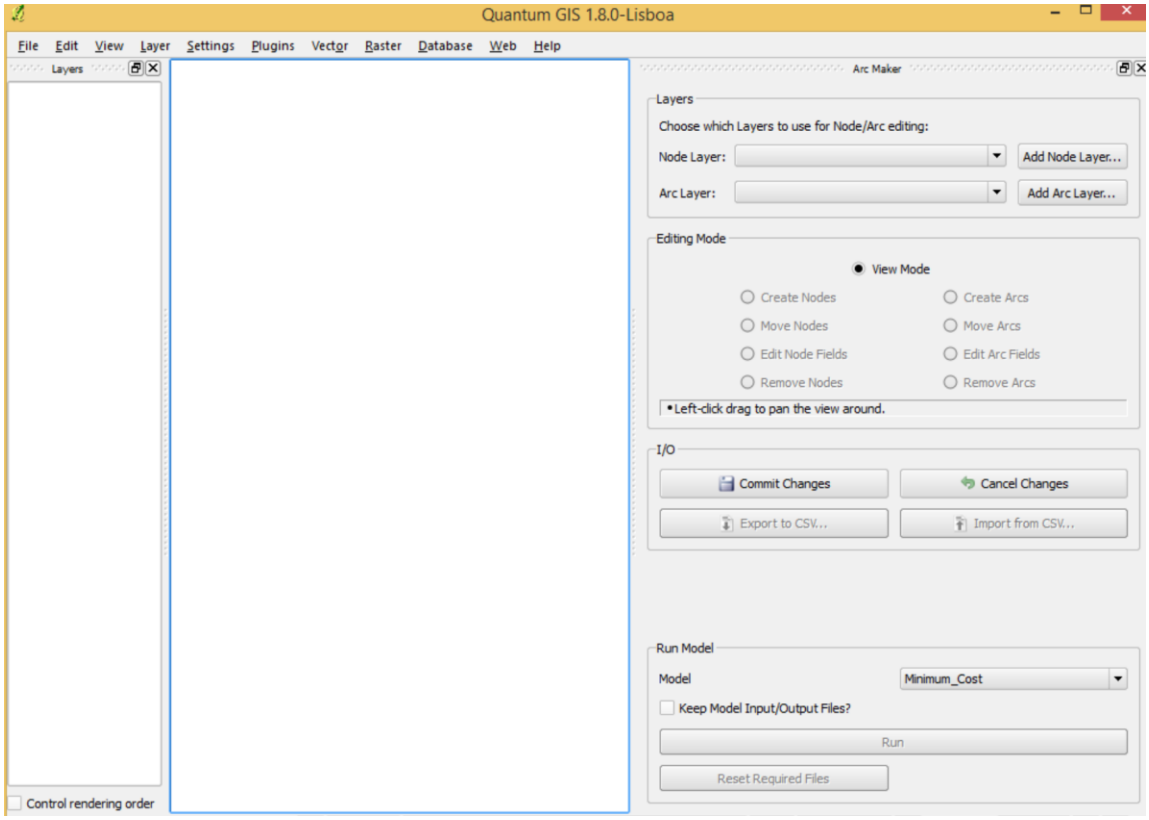

Figure 5. Quantum GIS with the ArcMaker plugin.

Figure 6 shows the additional functionality to the ArcMaker plugin. The plugin is very simple from the user's perspective. We allow the user to choose from one of the installed network models from the dropdown menu, and then run it on the currently selected node and arc layer. Choosing to keep the input and output files prevents the plugin from deleting them upon uploading the data back into QGIS. This gives the user a chance to debug unusual behavior. In the case of the network flow model, the traffic matrix is a required file that is not produced by QGIS. When running the model, the plugin will prompt the user to load one of these files. The plugin will continue to use this file until the user selects "Reset Required Files". This gives the user the chance to choose another file in the case that the traffic matrix changed.

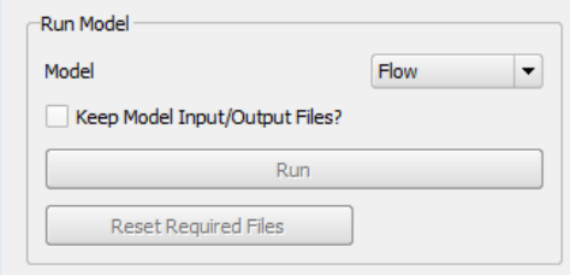

Figure 6. Additional functionality added to the Arc Maker plugin

Once the run button is clicked, the plugin starts the loop illustrated in Figure 7. Simply, we turn the QGIS layer attributes into GAMS input, run the selected model through GAMS, and then reload the model's output as the new attributes in QGIS.

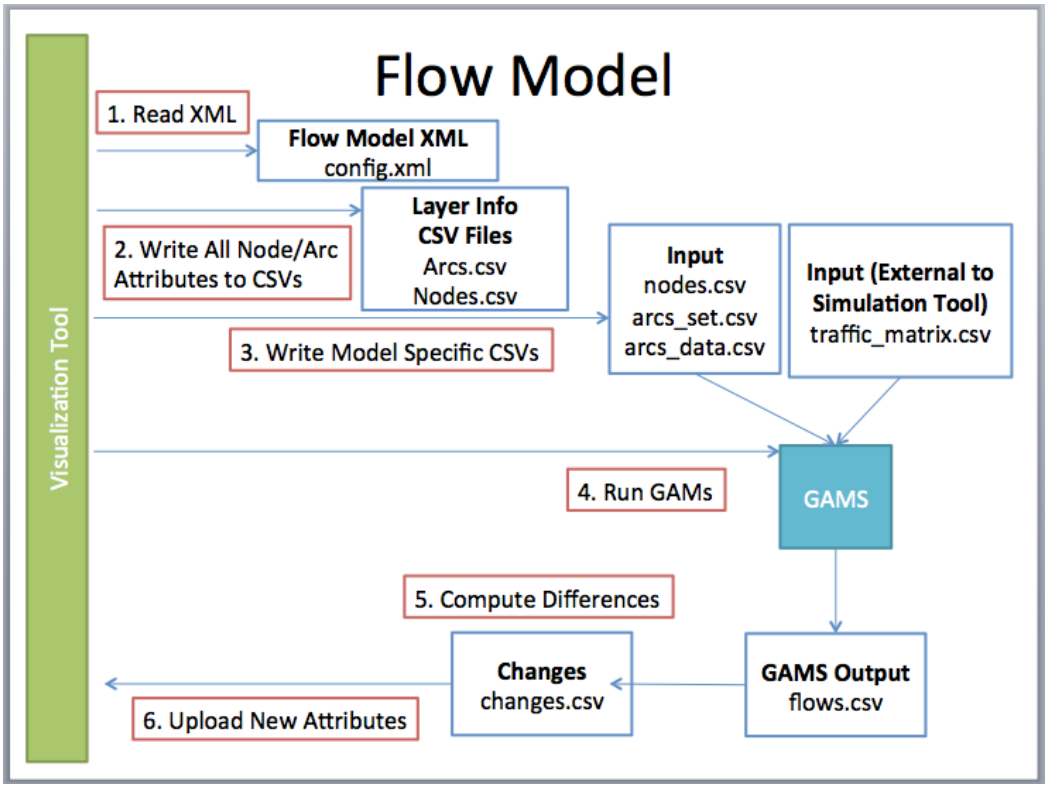

Figure 7. Work flow of running the flow model on a QGIS layer.

Work flow of running the flow model on a QGIS layer. Rather than working on the data in an Excel spreadsheet, users can now edit all of the layer information in the QGIS

graphical user interface (GUI). The GAMS box is the computation engine for the specified model. In this thesis, we treat this process as a black box to which we feed input and receive output. However, there are many complicated steps described in Crain (2012).

C. PLUGIN IMPLEMENTATION

We extend the ArcMaker plugin by inserting our own Python methods into the source code. Here is a description of each of the important steps throughout the flow model shown in Figure 7:

1. The first step is to read a configuration file, an Extensible Markup Language (XML) file that contains all of the necessary information describing the input and output of the model. Each model has its own configuration file inside the ArcMaker plugin in a directory with the model's name. The configuration file contains the filename of the GAMS model, the filenames and necessary attributes for the GAMS input, and a description of the GAMS output. Parsing this XML configuration file allows the plugin to create the input files with the correct format. Missing one comma or new line can cause the plugin to crash. An example configuration file is in the Appendix.
2. When the Run button is selected, all of the attributes from the chosen node and arc layer are written to Comma Separated Value (CSV) files. CSV is used because GAMS takes its input and writes its output in the CSV format. The plugin contains the active node and arc layer in the top two dropdown menus. The node and arc layer's attributes are written to separate CSV files. The attributes in these files will be used to create the resulting input files for GAMS.
3. Based on the configuration file, the model-specific input files are created. We iterate through the input files contained in the configuration file, writing them as we go using the attributes in the output of step 2.
4. With all of the input files in place, GAMS is executed. GAMS reads the CSV files as input and generates CSV files as output. The configuration file contains all of the information needed to read the output.
5. The output of GAMS is compared against the original input. We iterate through the GAMS output and the original attributes contained in step 2's output. Comparing these attributes, we create a "changes" file to summarize the output.
6. Within QGIS, we upload the new attributes from the changes file. Depending on the visualization style settings, the changes can be viewed instantly if the style is set to reflect a gradient on the attribute that the model changed.

D. PLUGIN USEFULNESS

The plugin overcomes the limitations of GIS and functional network analysis by combining the benefits of each. There is now a standalone environment for creating and editing networks, as well as running network flow models on them. This connection eliminates the need for users to run models using Excel files. Instead, the plugin handles the data conversion between QGIS and GAMS.

The plugin is intended to be simple and robust. There are integrity checks for every step to prevent a user from trying something that will not work or cause an error. Since we treat GAMS as a black box, there is a strong potential for errors if passed the wrong input. We monitor the GAMS output closely and pass as much information as possible back to the user in the event there is a problem. This puts a lot of the responsibility on the GAMS model creators to develop models that are robust to crashes.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. CASE STUDY: FIBER OPTIC COMMUNICATIONS BACKBONE IN DYSTOPIA

One of the major goals of this thesis is to create a realistic Internet infrastructure to test the plugin functionality discussed in the previous chapter. Even with its many detailed layers, Dystopia is missing a cyber infrastructure. We aim to draw a realistic, backbone Internet infrastructure on Dystopia on which to run our network flow models.

A. METHODOLOGY

1. Network Design Considerations

Figure 8 shows a plausible fiber optic communication network for Dystopia. It contains 18 nodes and 24 bidirectional arcs that collectively represent hundreds of miles of backbone fiber cable, both over land and undersea.

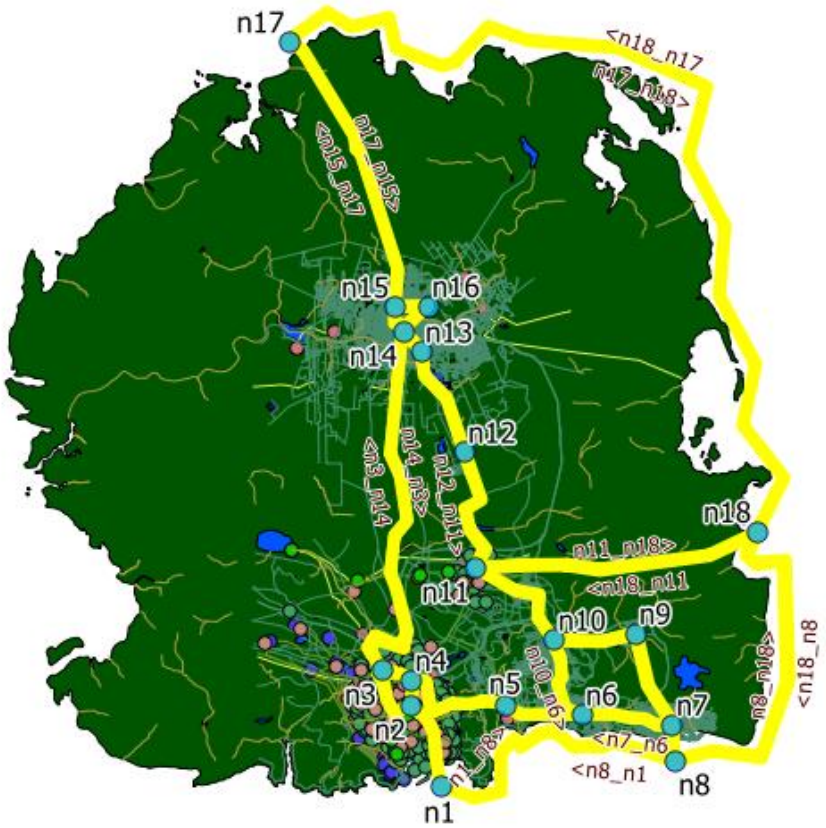

Figure 8. A simple fiber-optic backbone network for Dystopia.

Its design reflects an approximate economic cost associated with laying this fiber down, as links are very expensive. Fewer and shorter links save Internet Server Providers (ISPs) money, while many longer links provide more robust, resilient networks. Balancing redundancy and monetary expense is a tenet held closely for all ISPs. This Internet infrastructure focuses solely on the network core. To achieve this, we closely followed the first principles approach discussed in Alderson et al. (2005) describing the tradeoff between bandwidth and number of links. As noted in that work, routers nearest the core of a network tend to have fewer connections, but faster throughput (Alderson et al. 2005). This network also conforms to the general principles of network design identified by Topology Zoo and the BRITE Topology generator (Byers et al. 2014; Bowden 2013). Both projects aim to accurately map the network topologies behind the Internet.

We currently show only the core routers and the backbone links that connect them. With this example, this project provides a general proof of concept that could be extended to create a very detailed, low-level network within one city. The network's granularity could even be so detailed to show individual users as nodes connected to the network.

Fiber optic cables can be bidirectional meaning they simultaneously send data in both directions. If being used in a single direction, fiber cables almost certainly exist adjacent to a parallel link (Strachan 2005). We represent this in Quantum GIS (QGIS) using two arcs (one in either direction) between nodes. Figure 9 shows two arcs drawn closely to represent bidirectional flow between two nodes.

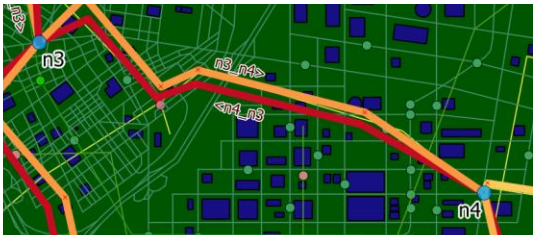

Figure 9. Bidirectional representation of arcs in Dystopia.

2. Network Attributes

With the network drawn as layers on Dystopia in Quantum GIS (QGIS), our next step is to analyze it using the network flow model discussed in Chapter III, Section C. Before we can run our analysis, we need to specify the input data for the problem.

a. Traffic Matrix

First, we must specify a traffic matrix for the amount of traffic traveling from one node to every other node. We represent this traffic in a two-way table based on a gravity model (Alderson et al 2006). Our gravity model is based on the estimated population surrounding each node. Since Dystopia is completely fictional, there is little information about city population. Therefore, we assign city populations based on Dystopia's total population of more than 400,000 people and a few specified city populations. We approximate the populations around the other nodes based on the infrastructure surrounding them. Residential areas receive higher populations than commercial areas.

After assigning populations, we consider the number of Internet users. Dystopia has a national border splitting the island into a northern and southern region. Some of Dystopia's scenarios consider the southern country to be the United States and the northern country to be Mexico. Using data from the World Bank, we treat only 38.4% of the population in Mexican cities as Internet users and 81% of the population around nodes in the United States as Internet users (The World Bank 2014). We recognize that these percentages may be low, as they take into account the entire country of Mexico and the United States, rather than the percentage of Internet users within cities. We accept this oversimplification, as this scenario is really only a proof of concept and the numbers are fictional anyway.

Based on the number of Internet users per node, we estimate the amount of traffic each node receives. We use statistics released by AT&T (AT&T 2014) and Comcast XFINITY (Comcast 2014) to roughly estimate the traffic demanded by each Internet user each month. We assume an estimated 21 GB of data demanded each month based on an assumed 30 days per month. These steps give a figure for GB of traffic per day for each node. Table 3 shows the progression from population at each node to GB of traffic per

day on each node. The figure denotes undersea nodes as blue, nodes in the United States as red, and nodes in Mexico as green. Undersea nodes have a zero population and therefore have a zero demand for data.

Node																	
n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
Population (Thousands)																	
0	70	120	100	10	25	55	0	15	35	55	5	15	50	40	30	0	0
Internet Users (Thousands)																	
0	57	97	81	8	20	45	0	12	28	45	4	12	19	15	12	0	0
Daily Traffic (GB/Day)																	
0	40	68	57	6	14	31	0	9	20	31	3	9	13	11	8	0	0

Table 3. Assumed population, number of Internet users, and daily traffic demand per node.

We multiply the traffic per day for each node by every other node's traffic demand. Table 4 below shows the resulting traffic matrix that the flow model will use to create our network flows.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n2	0	-11076	2701	2251	226	563	1238	0	338	788	1238	113	338	534	427	321	0	0
n3	0	2701	-17052	3858	386	965	2122	0	579	1351	2122	193	579	915	732	549	0	0
n4	0	2251	3858	-14857	322	804	1769	0	483	1126	1769	161	483	763	610	458	0	0
n5	0	226	386	322	-1781	81	177	0	49	113	177	17	49	77	61	46	0	0
n6	0	563	965	804	81	-4323	443	0	121	282	443	41	121	191	153	115	0	0
n7	0	1238	2122	1769	177	443	-8970	0	266	619	973	89	266	420	336	252	0	0
n8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n9	0	338	579	483	49	121	266	0	-2645	169	266	25	73	115	92	69	0	0
n10	0	788	1351	1126	113	282	619	0	169	-5935	619	57	169	267	214	161	0	0
n11	0	1238	2122	1769	177	443	973	0	266	619	-8970	89	266	420	336	252	0	0
n12	0	113	193	161	17	41	89	0	25	57	89	-903	25	39	31	23	0	0
n13	0	338	579	483	49	121	266	0	73	169	266	25	-2645	115	92	69	0	0
n14	0	534	915	763	77	191	420	0	115	267	420	39	115	-4110	145	109	0	0
n15	0	427	732	610	61	153	336	0	92	214	336	31	92	145	-3316	87	0	0
n16	0	321	549	458	46	115	252	0	69	161	252	23	69	109	87	-2511	0	0
n17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 4. Notional traffic matrix based on assumed populations. The values in this matrix represent relative demands. Absolute demand is scaled by network capacity.

Our traffic matrix's gravity model is a simplification of actual network traffic demand. Data content providers like Akamai (Akamai 2014) and Google (Google 2014) place large server farms and data caches near population centers to reduce the need for

heavy backbone traffic. *Wired.com* reports that 25% of North American's Internet Traffic is served by Google (2013). These local caches support huge data streams. Servers also have very asymmetric traffic flow, as the traffic outbound is often several times larger than the request traffic; this is not represented in our traffic matrix but could be incorporated. Another simplification is the time of day. Network flow changes drastically during the progression of the day. Network analysts are most interested in studying peak traffic, as this is when they would see the greatest number of bottlenecks in their routing. According to iMediaConnections (Harlin 2012), this peak happens at four o'clock in the afternoon.

b. Arc Capacity

For simplicity, we assume all arcs have the same capacity. We calculate the capacity in Gigabytes per day, as this is the unit we use for our traffic demand. According to Network Dictionary, Optical Carrier 192 (OC-192) was the most common fiber used for backbones by large Internet Service Providers in 2006. OC-192 has a capacity of 9.6 Gigabits per second (Gbps) (Dong 2007). After unit conversion, our capacity is 103,680 Gigabytes per day.

c. Arc Cost

Since arc cost is the notional "price" of traversing an arc, it can vary greatly on the networked infrastructure. Since we are modeling IP traffic traveling over fiber optic cables, it is the latency between nodes that determines the cost. Network latency is caused by delays in routers, while the transmission time of data through cabling is often negligible. Therefore, when determining the cost of each arc, we only count the number of "hops" over routers. Thus we make each arc cost 1, regardless of the length of the link connecting the two nodes.

B. NETWORK FLOW ANALYSIS

Our next step is to apply the minimum-cost flow model to our network. Again, we use the minimum-cost flow model from Crain (2012). We consider several scenarios, each with one or more variations. First, we consider a base case to show the network

under normal conditions. Second, we look at the case where an arc is removed. Third, we look at the case where a particular node sees a large increase in traffic demand. Last, we look at the case where a new arc is added. In each of these network manipulations, we are interested in how the flow of traffic is affected and redirected to compensate for the changes.

1. Base Case

We begin our analysis by studying normal traffic through an undisturbed network to give us a base case to compare the more interesting cases. Figure 10 shows our network in QGIS after running the minimum cost network flow model with data defined in the previous section. We use QGIS's style feature to classify and color code the arcs based on their flow attribute. Throughout this thesis, we use a graduated color scheme where dark red arcs represents those with the highest flows, and lighter, yellow arcs represent those with smaller flows. Applying this style makes it simple to visualize the flows, relative to each other.

Figure 10. Dystopia's normal network flows with a graduated color scheme after running the minimum cost network flow model with our traffic matrix and arc attributes.

The relative flows shown above are exactly what we expect for the GAMS input. We see higher traffic flow between the population centers around nodes n2, n3, n4 in the south, and nodes n13, n14, n15, n16 in the north. Table 5 represents this flow information as a matrix with units Gigabyte per day (GB/day). Empty cells show there is no arc between those two nodes.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1		2814						6371										
n2	6371		4434	3828														
n3		7991		9727										10410				
n4		3828	12848		11244													
n5				14365		10691												
n6					13812		2389			7236								
n7						5769		2325	3051									
n8	2814						6889											1987
n9							1867			2989								
n10						6977			1805		5862							
n11										4419		4137						3529
n12											4022		4404					
n13												4289		5209				
n14			10846										5094		2251	1690		
n15														2251		821	1391	
n16														2011	500			
n17															1712			1391
n18								2994			2201						1712	

Table 5. Node to node network flows for Dystopia's undisturbed IP network.

We compare the results of future scenarios against these flows in order to understand how the network is redirecting traffic. Comparing the average link utilization across all arcs as well as the amount of dropped traffic reveals an easy comparison for our network disturbances. We calculate link utilization by dividing the average flow over all links by the capacity of one link. The average link utilization for the base case is 47.74 percent with no dropped traffic.

2. Deleted Arc

We make the assumption that an interdicted arc will completely disrupt that link's flow in both directions. Our minimum network flow model is instrumented to find the minimum set of flows, even in the case where some arcs are interdicted. We use GAMS to enumerate each possible interdicted arc, and to compute the minimum-cost solution for each. We do this for one and two arc failures seeking, as an adversary would, the maximum disruption.

a. Single Interdictions

Table 6 shows the results for single interdictions sorted by severity. While most single failures result in no traffic loss, several large impacts on traffic flow. Specifically,

there are four arcs (n1-n2, n1-n8, n3-n14, n4-n5) that yield the most dropped traffic. A fifth arc (n5-n6) leads to nearly as much dropped traffic.

Tail	Head	Dropped Traffic
n1	n2	194040
n1	n8	194040
n3	n14	194040
n4	n5	194040
n5	n6	191952
n10	n11	41616
n7	n8	41616
n8	n18	23496
n11	n12	0
n11	n18	0
n12	n13	0
n13	n14	0
n14	n15	0
n14	n16	0
n15	n17	0
n15	n16	0
n17	n18	0
n2	n3	0
n2	n4	0
n3	n4	0
n6	n10	0
n6	n7	0
n7	n9	0
n9	n10	0

Table 6. The amount of dropped traffic after a single interdiction.

We show two results from Table 6 in more detail. We start by visualizing the effects of deleting arc (n3, n14). We "delete" the arc by setting its capacity to zero in our QGIS editor. This is functionally equivalent to deleting the arc; however, it prevents the user from having to redraw the arc to bring it back online.

Figure 11 shows how the traffic flow is redirected after losing the crucial connection from the southern city (n3) to the northern city (n14).

Figure 11. Dystopia's network flows after a single interdiction between nodes n3 and n14.

Notice the increased traffic rate through the undersea cables traveling on the eastern coast. The traffic between n1 and n2 nearly quadrupled. Table 7 shows the new flow matrix in GB/day. The highlighted zeros show the absence of flow traversing the interdicted links. This single interdiction raised the average link utilization to 63.04 percent and caused 194,040 GB of traffic to be dropped each day.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1		11468						14169										
n2	14169		7791	5659														
n3		10492		6560										0				
n4		5659	9261		11196													
n5				13897		11109												
n6					13810		2764			7842								
n7						5151		3120	3222									
n8	11468						6766											11436
n9							1963			3248								
n10						8156			1989		6823							
n11										5878		4199						6874
n12											4429		3532					
n13												3762		1439				
n14			0										1669		2900	201		
n15														2900		2310	7586	
n16														431	2080			
n17															7816			7586
n18								12381			5699						7816	

Table 7. Node to node network flows in Dystopia after a single interdiction between nodes n3 and n14.

In the same way, we delete the arc between nodes n4 and n5. Figure 12 shows the resulting network flow as displayed in QGIS. Similar to the previous case, we observe traffic rerouting around the disconnected arc through the undersea cables connecting nodes n1 and n8.

Figure 12. Dystopia's network flows after a single interdiction between nodes n4 and n5.

This disruption results in the flow matrix displayed in Table 8. Once again, the highlighted cells represent the arcs that have been deleted.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1		15136						13590										
n2	13590		8906	8524														
n3		8486		6333										12195				
n4		7398	7459		0													
n5				0		1781												
n6					1781		5049			2036								
n7						4455		12127	4138									
n8	15136						13225											4131
n9							2446			3198								
n10						2630			1506		6887							
n11										5789		5805						6424
n12											7322		6072					
n13												7589		6779				
n14			10649										8296		2457	1529		
n15														2037		982	2116	
n16														1920	591			
n17															2087			2116
n18								6775			3809						2087	

Table 8. Node to node network flows in Dystopia after a single interdiction between nodes n4 and n5.

Deleting the arc connecting n4 to n5 causes the average link utilization to increase ten percent to 57.74 percent from the base case utilization rate and for the network to drop 194,040 GB of traffic per day.

b. Double Interdictions

Next we look at cases involving the simultaneous loss of two links. Using the same method to determine the worst-case losses, we find the worst-case double interdictions to be removing links (n1,n8) and (n4,n5) or (n1.n8) and (n3,n14).

Deleting the links (n1, n8) and (n4, n5) results in the network flows depicted in Figure 13. As we expect, the traffic demand is placed almost entirely on the arc connecting n3 and n14 as the southern city (n3) has lost its other two connections to the rest of the island. Table 9 shows the flow matrix after these two arcs are lost.

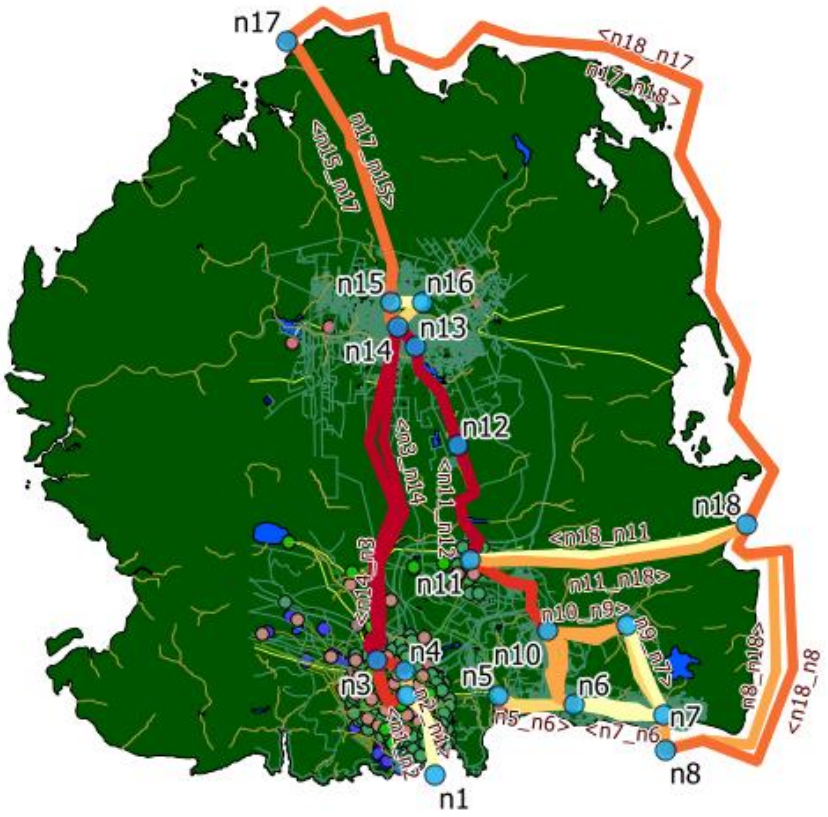

Figure 13. Dystopia's network flows after a double interdiction between nodes n1, n8 and n4, n5.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1		0						0										
n2	0		8825	2251														
n3		8825		12606										25365				
n4		2251	12606		0													
n5				0		1781												
n6					1781		995			5302								
n7						1189		6673	1950									
n8	0						7771											6673
n9							1046			3810								
n10						5108			2906		12341							
n11										11243		15392						2291
n12											16035		15659					
n13												16302		16366				
n14			25365										17009		7586	1529		
n15														7586		982	6993	
n16														2172	339			
n17															7636			6993
n18								7771			550						7636	

Table 9. Node to node network flows in Dystopia after a double interdiction between nodes n1, n8 and n4, n5.

We see huge increases on the remaining arcs. The link connecting the south (n3) to the north (n14), increases from 10,410 GB/day in the base case to 25,365 GB/day. This network configuration yields an average link utilization of 71.35 percent with 401,400 GB per day of dropped traffic.

Next we delete the links between nodes n1, n8 and nodes n3, n14. The result is shown in Figure 14. Similar to the previous interdiction, the highly populated southern city (n3) relies on one link (n4, n5) to pass traffic. Table 10 shows the resulting flow matrix from this double interdiction.

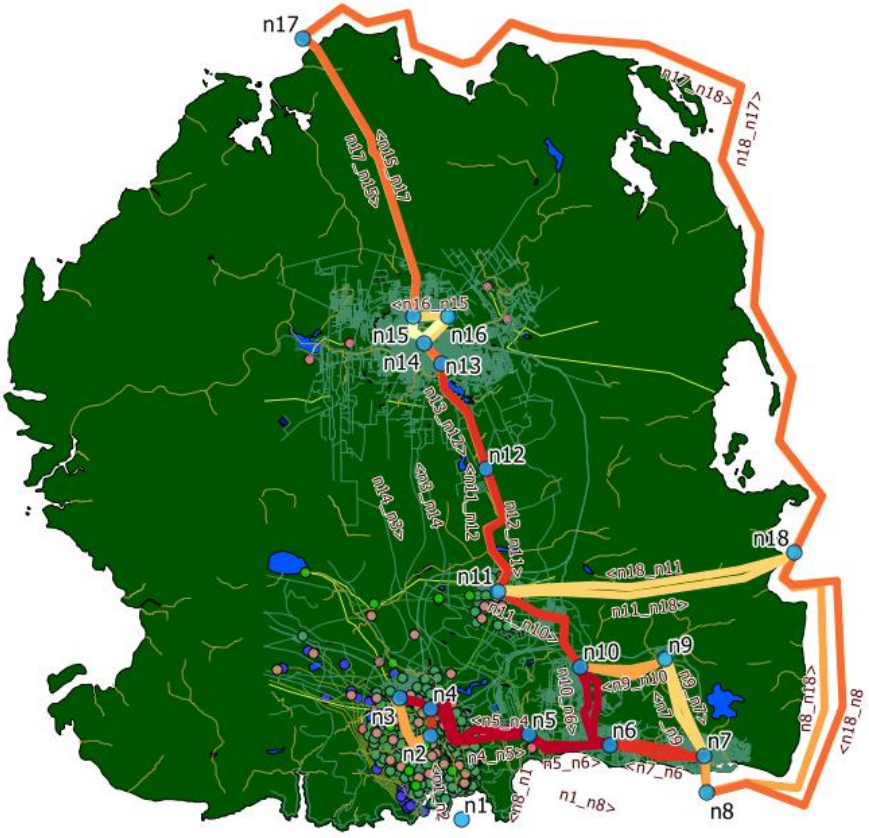

Figure 14. Dystopia's network flows after a double interdiction between nodes n1, n8 and n3, n14.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1		0						0										
n2	0		2701	8375														
n3		2701		14351										0				
n4		8375	14351		25365													
n5				25365		25278												
n6					25278		9221			15909								
n7						8483		4641	1950									
n8	0						4807											4641
n9							1046			3810								
n10						16647			2906		13618							
n11										13452		6411						2291
n12											7833		5744					
n13												7166		3651				
n14			0										5073		688	201		
n15														688		2310	4182	
n16														1623	888			
n17															5604			4182
n18								4807			703						5604	

Table 10. Node to node network flows in Dystopia after a double interdiction between nodes n1, n8 and n3, n14.

As expected, we see a huge increase on the link from node n4 to node n5. This link goes from passing 14,365 GB per day in the base case to 25,365 GB per day. This case increases the average link utilization to 70.79 percent and causes 401,400 GB per day of dropped traffic.

Both Table 9 and Table 10 show an interesting case. Notice there are zeros outside the highlighted, deleted arcs. Deleting arc (n1,n8) prevents all traffic from traversing through (n1,n2). Node n1 does not have a population because it is undersea cable station. Since it can no longer pass traffic to node n8, there is no reason for traffic to be sent to it.

We show the capability of deleting arcs using our plugin and its effects on our network in Dystopia. The process is very simple from the user's perspective. Deleting an arc consists of setting its capacity to zero, saving the changes, and then running the model. The new visualization will refresh automatically.

3. Increased Demand

In this scenario, we consider changes to our traffic matrix. First, we double the traffic demand at nodes n14, n15, and n16. Then we look at the difference after tripling the original demands in these three nodes.

Table 11 shows the new traffic matrix used in the next simulation. Notice the increases in demand under nodes n14, n15, and n16.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n2	0	-12356	2701	2251	226	563	1238	0	338	788	1238	113	338	1067	854	641	0	0
n3	0	2701	-19247	3858	386	965	2122	0	579	1351	2122	193	579	1829	1464	1098	0	0
n4	0	2251	3858	-16686	322	804	1769	0	483	1126	1769	161	483	1525	1220	915	0	0
n5	0	226	386	322	-1964	81	177	0	49	113	177	17	49	153	122	92	0	0
n6	0	563	965	804	81	-4780	443	0	121	282	443	41	121	382	305	229	0	0
n7	0	1238	2122	1769	177	443	-9975	0	266	619	973	89	266	839	671	503	0	0
n8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n9	0	338	579	483	49	121	266	0	-2919	169	266	25	73	229	183	138	0	0
n10	0	788	1351	1126	113	282	619	0	169	-6575	619	57	169	534	427	321	0	0
n11	0	1238	2122	1769	177	443	973	0	266	619	-9975	89	266	839	671	503	0	0
n12	0	113	193	161	17	41	89	0	25	57	89	-994	25	77	61	46	0	0
n13	0	338	579	483	49	121	266	0	73	169	266	25	-2919	229	183	138	0	0
n14	0	1067	1829	1525	153	382	839	0	229	534	839	77	229	-8716	579	434	0	0
n15	0	854	1464	1220	122	305	671	0	183	427	671	61	183	579	-7087	347	0	0
n16	0	641	1098	915	92	229	503	0	138	321	503	46	138	434	347	-5405	0	0
n17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 11. The traffic matrix after doubling the demand in nodes n14, n15, and n16.

Figure 15 shows the result of running the flow model on our original network with the new traffic matrix. Table 12 shows the resulting flow matrix.

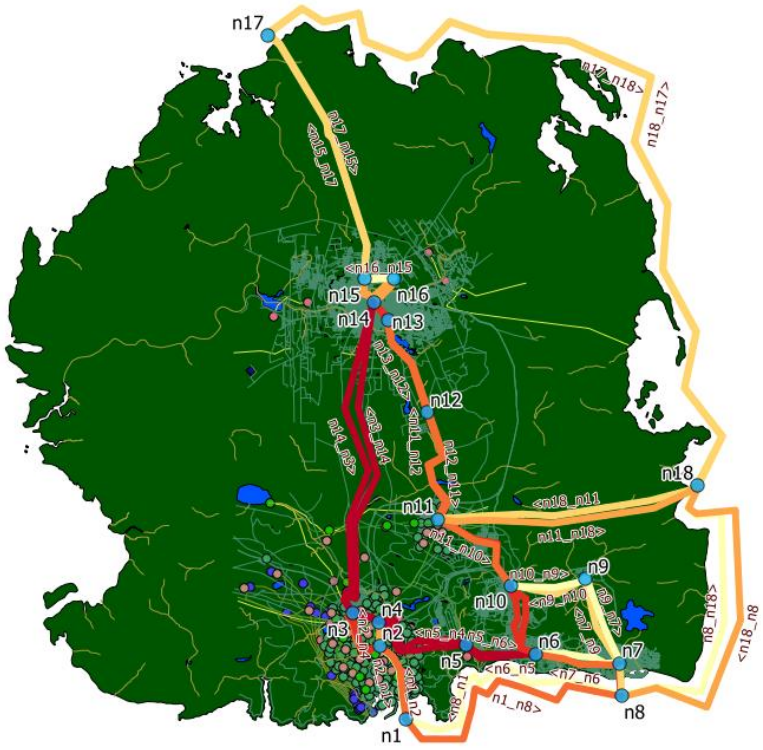

Figure 15. Dystopia's network flows after doubling the demand in nodes n14, n15, and n16.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1		2814						6354										
n2	6354		5714	3828														
n3		9254		12196										16773				
n4		3828	15736		11884													
n5				15424		11148												
n6					14688		2389			7236								
n7						6188		3071	2958									
n8	2814						7801											2733
n9							2027			3103								
n10						6977			2172		6616							
n11										5426		4937						4488
n12											5075		5295					
n13												5433		6374				
n14			16773										6512		4788	3593		
n15														4788		1812	3279	
n16														3731	1674			
n17															3417			3279
n18								3923			3160						3417	

Table 12. Node to node network flows in Dystopia doubling the demand in nodes n14, n15, and n16.

This change produces relatively small changes to the base case's flow matrix. It is apparent that the links connecting to the nodes with doubled demand did see an increase in flow, however there was not a tremendous difference. The network did see an increase in average link utilization with 58.93 percent and 109,444 GB of dropped traffic per day.

In order to see a larger difference, we now triple the original demand for nodes n14, n15, and n16. Table 13 shows the traffic matrix used in this simulation. Figure 16 shows Dystopia after running the second scenario where we increase demand.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n2	0	-13637	2701	2251	226	563	1238	0	338	788	1238	113	338	1601	1281	961	0	0
n3	0	2701	-21442	3858	386	965	2122	0	579	1351	2122	193	579	2744	2195	1647	0	0
n4	0	2251	3858	-18514	322	804	1769	0	483	1126	1769	161	483	2287	1829	1372	0	0
n5	0	226	386	322	-2147	81	177	0	49	113	177	17	49	229	183	138	0	0
n6	0	563	965	804	81	-5237	443	0	121	282	443	41	121	572	458	343	0	0
n7	0	1238	2122	1769	177	443	-10981	0	266	619	973	89	266	1258	1006	755	0	0
n8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n9	0	338	579	483	49	121	266	0	-3193	169	266	25	73	343	275	206	0	0
n10	0	788	1351	1126	113	282	619	0	169	-7216	619	57	169	801	641	481	0	0
n11	0	1238	2122	1769	177	443	973	0	266	619	-10981	89	266	1258	1006	755	0	0
n12	0	113	193	161	17	41	89	0	25	57	89	-1086	25	115	92	69	0	0
n13	0	338	579	483	49	121	266	0	73	169	266	25	-3193	343	275	206	0	0
n14	0	1601	2744	2287	229	572	1258	0	343	801	1258	115	343	-13828	1301	976	0	0
n15	0	1281	2195	1829	183	458	1006	0	275	641	1006	92	275	1301	-11323	781	0	0
n16	0	961	1647	1372	138	343	755	0	206	481	755	69	206	976	781	-8690	0	0
n17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 13. The traffic matrix after tripling the demand in nodes n14, n15, and n16.

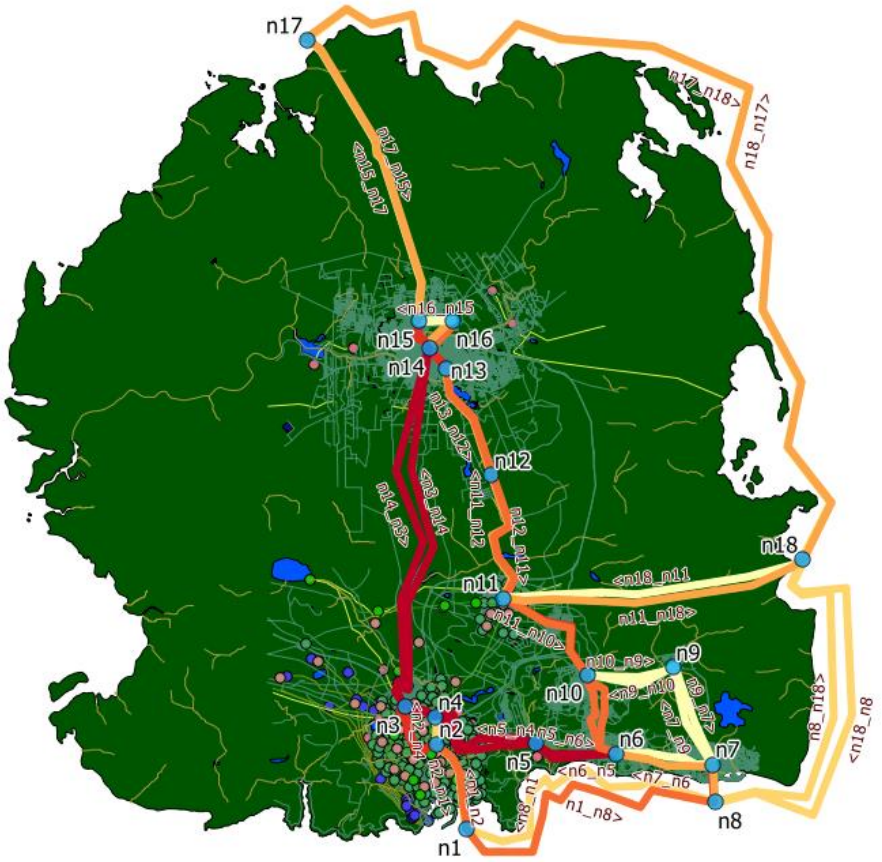

Figure 16. Dystopia's network flows after tripling the demand in nodes n14, n15, and n16.

Table 14 shows the resulting flow matrix. As expected, we see this traffic matrix change make a bigger impact throughout the entire network than in our last scenario. Average link utilization is now up to 70.48 percent. A large increase considering we only tripled the demand in three of our 18 nodes. This configuration also causes 116,280 GB of traffic to be dropped each day.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1		4072						6773										
n2	6773		8253	3828														
n3		10954		14664										23136				
n4		3828	17365		12524													
n5				15225		11605												
n6					14306		2389			7236								
n7						5349		5076	3050									
n8	4072						8899											3480
n9							2187			3217								
n10						6977			2354		7371							
n11										6249		5737						5449
n12											6698		6187					
n13												7148		7540				
n14			23136										8501		7614	5712		
n15														7614		2978	4164	
n16														6673	2017			
n17															5125			4164
n18								4602			3366						5125	

Table 14. Node to node network flows in Dystopia tripling the demand in nodes n14, n15, and n16.

4. Added Arc

As our last scenario, we consider the impact of adding an arc to the base case network. First we add an arc between nodes n3 and n10. The resulting visualization is displayed in Figure 17. Table 15 shows the new flow matrix. The highlighted cells are the new arcs and their respective flows.

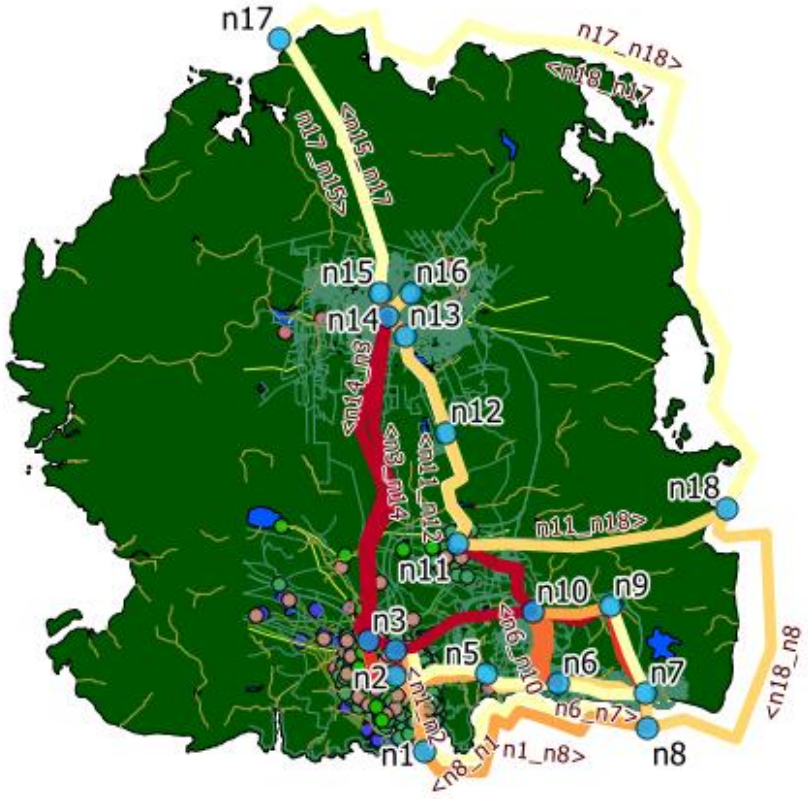

Figure 17. Dystopia's network flows after adding an arc between nodes n4 and n10.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1		1238						3360										
n2	3360		4434	5404														
n3		6443		13185										8739				
n4		5517	15421		2936					13681								
n5				1452		2366												
n6					882		2389			4645								
n7						1395		1918	5749									
n8	1238						5696											680
n9							977			7770								
n10				17514		4155			2998		8840							
n11										7411		1794						2130
n12											2159		1626					
n13												1991		2431				
n14			8512										2796		2671	1920		
n15														2557		591	1344	
n16														2172	339			
n17															1482			1344
n18								2336			336						1482	

Table 15. Node to node network flows in Dystopia after adding an arc between nodes n4 and n10.

As we expect, adding this arc decreases the network utilization. This change causes the percent utilization to drop from the base case's 47.74 percent to 39.31 percent. Obviously adding this arc did not cause dropped traffic.

Next, we add an arc from node n3 to node n15 in an effort to reduce the traffic traveling from n3 to n14. Figure 18 shows the result. Table 16 shows the result of the addition of this new arc.

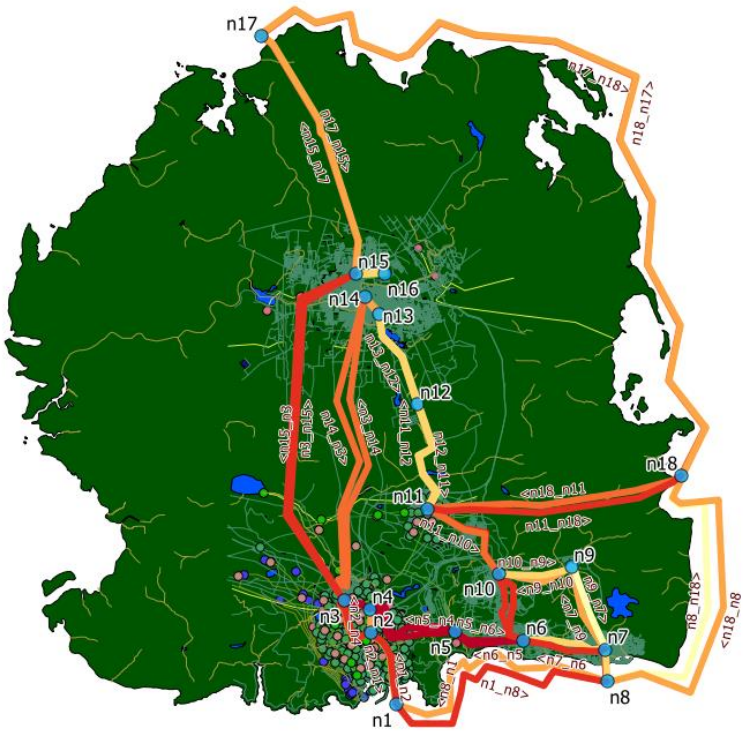

Figure 18. Dystopia's network flows after adding an arc between nodes n3 and n15.

	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
n1		2814						5515										
n2	5515		4434	3828														
n3		7135		9727										4816	5594			
n4		3828	12848		11244													
n5				14365		10691												
n6					13812		2389			7236								
n7						5769		2325	2867									
n8	2814						6705											1987
n9							1867			2989								
n10						6977			1989		5862							
n11										4603		2015						5651
n12											2084		2282					
n13												2351		3087				
n14			4557										3156		688	201		
n15			5433											268		2310	4185	
n16														431	2080			
n17															3834			4185
n18								3666			4323						3834	

Table 16. Node to node network flows in Dystopia after adding an arc between nodes n3 and n15

Similarly, we see decreases in the flow of several arcs. This addition brings the average link utilization down to 44.98 percent and does not cause any traffic to be dropped.

5. Summary of Results

We looked at three cases of network manipulations: arc deletion, change in demand, and arc addition. Figure 19 shows a summary of link utilization for each of our test cases alongside our base case. Figure 20 shows a summary of dropped traffic for each of our test cases.

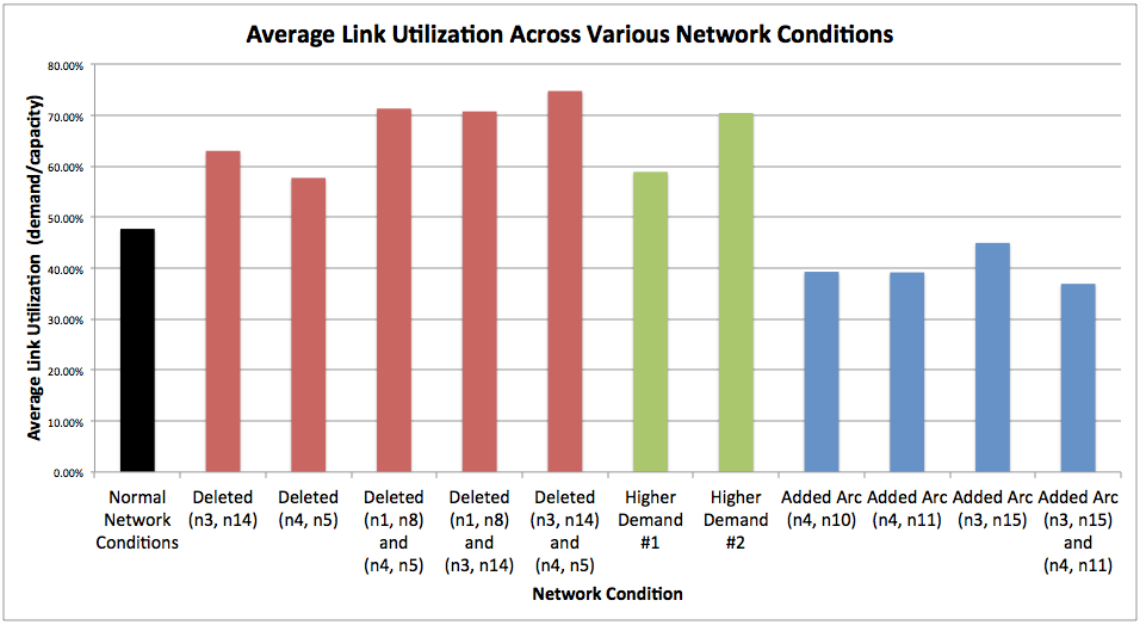

Figure 19. Summary of average link utilization across network scenarios.

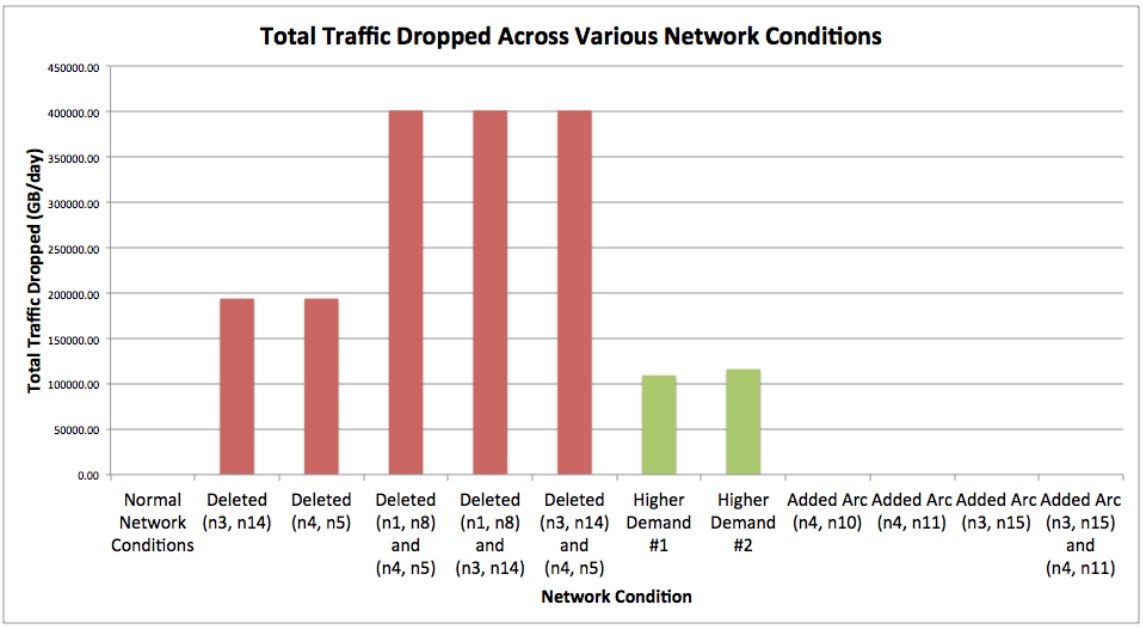

Figure 20. Summary of total traffic dropped across network scenarios.

Our addition to QGIS allows us to easily modify the network we are analyzing. After drawing a new arc, deleting an existing arc, or modifying the traffic matrix, the user simply has to select the run button to recompute all of the flow information, which is instantly reflected in the visualization.

V. CONCLUSION

A. CONTRIBUTION

Our primary contribution in this thesis is to provide an easy to use, graphical interface that allows users, regardless of their network expertise, to create networks and run models of system function on them. Doing so connects the benefits of the geographic information system's visualization with the functional modeling capabilities of a network's metadata. Users are now able to run these tests with one mouse click. This proves to be a much easier solution than the tedious task of manually generating input data and reading output data as text only. We also show the success in our attempt by applying this additional functionality to a realistic network we built in the fictional world of Dystopia.

Our case study applies the network flow model to the Internet infrastructure within Dystopia, but shows that we can just as easily apply this network building procedure to a variety of the critical infrastructures.

Not only is switching network infrastructures easy, applying network analysis to real world applications is just as simple. We create our Internet backbone network on a background layer of Dystopia, but we are equally capable of creating, for example, the high-level water infrastructure of Los Angeles on a map. The process is the same and can be applied to any networked infrastructure study.

B. FUTURE WORK

While our contributions are exciting in themselves, they open the door to many new projects. Currently, GAMS is required to be installed and licensed locally to run any GAMS model. A future step is to run a server that receives requests to run a model on files submitted to it then passes the input back to the client computer. Running the model on a server simplifies the installation process of the required tools, as GAMS would no longer be necessary.

This work analyzes networks independent of other networks. In reality, networks are interlaced with dependencies. For example, most network infrastructures rely on electricity. Therefore a water network, equipped with pumps and water stations are only effective if they have the energy to run them. Adding this functionality to the models that GAMS runs is a necessary first step before incorporating it to QGIS and our plugin.

Since we only look at a high-level backbone of the Internet layer in Dystopia, one could continue drawing the intricacies of the network. This would include the network links down streets and into buildings. One could implement access routers to provide points of presence for the citizens of Dystopia. Since Dystopia is so large and complex, adding the detail necessary for wireless routers, and individual building's access would be extremely time consuming. An alternative would be to pick a subsection of a city to implement this finer detail.

This thesis focuses on network flow models. We did not test this design on other models, but are confident it will extend easily to them assuming a configuration file can be created describing the input and output of the model. A next step here is to write configuration files for other models and test them with our plugin.

Another interesting problem relating to network modeling is wireless access points. Traditional networks have links and arcs that are static. Wireless technology allows nodes to dynamically change the network configuration. Representing this feature of wireless devices creates new challenges in simulating the behavior of these networks. We look to past work of Shankar (2008) and Nicholas (2009) could serve as starting points. We hope the network modeling technique described within this thesis serves as a step forward in those endeavors.

APPENDIX. DATA FILES

A. CONFIGURATION FILE

This Extensible Markup Language (XML) file contains all of the information needed to describe the Minimum Cost Flow model to the plugin. It contains the necessary attributes for arcs and nodes, where to find them, and how to rearrange them into the input to GAMS. It also contains the format of GAMS output for seamless transition back into QGIS for visualization.

```
<?xml version="1.0" encoding="UTF-8"?>
<OperatorModel name = "Min Cost" file = "MinCost.gms" results = "changes.csv">
	<ModelInput name = "nodes.csv" header = "0">
		<Attribute name = "node" source = "nodes_all.csv"/>
	</ModelInput>
	<ModelInput name = "arcs_data.csv" header = "1">
		<Attribute name = "tail" source = "arcs_all.csv"/>
		<Attribute name = "head" source = "arcs_all.csv"/>
		<Attribute name = "capacity" source = "arcs_all.csv"/>
		<Attribute name = "cost" source = "arcs_all.csv"/>
	</ModelInput>
	<ModelInput name = "arcs_set.csv" header = "0">
		<Attribute name = "tail" source = "arcs_all.csv"/>
		<Attribute name = "head" source = "arcs_all.csv"/>
	</ModelInput>
	<ModelInput name = "traffic_matrix.csv" header = "2" external="1"/>
	<ModelOutput name = "flows.csv" header = "1">
		<Attribute name = "tail" source = "arcs_all.csv" constant="1"/>
		<Attribute name = "head" source = "arcs_all.csv" constant="1"/>
		<Attribute name = "flow" source = "arcs_all.csv" constant="0"/>
		<Attribute name = "destination"/>
	</ModelOutput>
	<Layer>
		<Node output = "nodes_all.csv">
			<Attribute name = "node"/>
		</Node>
		<Arc output = "arcs_all.csv">
			<Atrribute name = "tail"/>
			<Attribute name = "head"/>
			<Attribute name = "capacity"/>
			<Attribute name = "cost"/>
			<Attribute name = "flow"/>
		</Arc>
	</Layer>
</OperatorModel>
```

B. INPUT FILES

The files below represent the GAMS input for running the minimum cost flow on the sample network in Figure 3.

1. arcs_set.csv

Arcs_set.csv lists the arcs that make up the network. They are listed from the "tail" node to the "head" node.

```
n1,n2
n1,n3
n2,n1
n2,n3
n3,n1
n3,n4
n4,n2
n4,n3
```

2. nodes.csv

Nodes.csv contains the list of nodes in the network. The names must be unique.

```
n1
n2
n3
n4
```

3. arcs_data.csv

Arcs_data.csv contains the attributes describing the arcs in the network.

```
tail,head,capacity,cost
n1,n2,6,1
n1,n3,6,1
n2,n1,6,1
n2,n3,6,1
n3,n1,6,1
n3,n4,6,1
n4,n2,6,1
n4,n3,6,1
```

4. traffic_matrix.csv

Traffic_matrix.csv is a two way table listing the demands between each node pair.

```
,n1,n2,n3,n4
n1,-8,2,2,4
```

n2,2,-8,4,2
n3,2,4,-8,2
n4,4,2,2,-8

C. OUTPUT FILE: FLOW.CSV

Flows.csv is the output from GAMS. We use the flow column to determine the new flow between the arc from the "tail" node to the "head" node.

tail,head,flow,destination
n2,n1,6.00,n1
n3,n1,2.00,n1
n4,n2,4.00,n1
n1,n2,6.00,n2
n3,n1,4.00,n2
n4,n2,2.00,n2
n1,n3,2.00,n3
n2,n3,4.00,n3
n4,n3,2.00,n3
n1,n3,2.00,n4
n2,n3,2.00,n4
n3,n4,6.00,n4

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Ahuja, Ravindra K., Magnanti, Thomas L., and James B. Orlin. 1993. *Network Flows: Theory, Algorithms, and Applications*. Englewood Cliffs, NJ: Prentice Hall.

Akamai Technologies. 2014. "Cloud Solutions." Accessed March 10. http://www.akamai.com/html/solutions/index.html.

Alderson, Dave, Li, Lun, Willinger, Walter, and John C. Doyle. 2005. "Understanding Internet Topology: Principles, Models, and Validation." *IEEE/ACM Transactions on Networking* 13, no. 6:1205-1218.

Alderson, David, Uhlig, Steve, Roughan, Matthew, Chang, Hyunseok, and Willinger, Walter. 2006. "The Many Facets of Internet Topology and Traffic." *Networks and Heterogeneous Media* 1, no. 4: 569–600.

AT&T. 2014. "Broadband usage FAQs—AT&T Support." Accessed March 10. http://www.att.com/esupport/article.jsp?sid=KB409045&cv=801#fbid=IZ59cq85qRg.

Comcast. 2014. "Average Internet Usage." Accessed March 10. http://customer.comcast.com/help-and-support/internet/data-usage-average-network-usage.

Barkley, Timothy R. 2008. "An Attacker-Defender Model for IP-based Networks." Master's thesis, Naval Postgraduate School.

Bowden, Rhys, Knight, Simon, Nguyen, Hung, Falkner, Nickolas, and Matthew Roughan. 2014. "The Internet Topology Zoo." The Internet Topology Zoo. Accessed March 10. http://www.topology-zoo.org/.

Byers, John, Medina, Alberto, Lakhina, Anukool, and Ibrahim Matta. 2014. "BRITE." BRITE Universal Topology Generator. Accessed March 10. http://www.cs.bu.edu/brite/.

Brown, Gerald, Carlyle, Matthew, Salmeron, Javier and Kevin Wood. 2005. Analyzing The Vulnerability of Critical Infrastructure to Attack, and Planning Defenses. *Tutorials in Operations Research: Emerging Theory, Methods, and Applications*.

———. 2006. "Defending Critical Infrastructure. *Interfaces* 36, no. 6: 530–544.

Center for Homeland Defense and Security. 2013. *Dystopia*. Accessed June 10. http://www.chds.us/?dystopia:map.

Crain, John K. 2012. "Assessing Resilience in the Global Undersea Cable Infrastructure." Master's thesis, Naval Postgraduate School.

Dixon, Cory A. 2011. “Assessing Vulnerabilities in Interdependent Infrastructures Using Attacker-Defender Models.” Master’s thesis, Naval Postgraduate School.

Dong, Jielin. 2007. *Network Dictionary*. Saratoga, CA: Javvin Technologies.

Federal Geographic Data Committee. 2014. “Homeland Security and Geographic Information Systems. Accessed May 6. http://www.fgdc.gov/library/whitepapers-reports/white-papers/homeland-security-gis.

GAMS Development Corporation. 2010. *General Algebraic Modeling System* (GAMS) 23.6. Washington, DC

Google. 2014. “Our Services.” Accessed March 10. https://www.google.com/intl/en/policies/terms/

Gun, Selcuk. 2013. “Design and Implementation of a Computation Sever for Optimization with Application to Analysis of Critical Infrastructure.” Master’s thesis, Naval Postgraduate School.

Harlin, Ky. 2012. “How Time of Day Affects Content Performance.” iMedia Connection. Accessed June 10, 2014. http://www.imediaconnection.com/content/31577.asp.

ILOG. 2007. *ILOG CPLEX 11.0 User’s Manual*. Accessed June 9, 2014. http://wwweio.upc.es/lceio/manuals/cplex-11/html/.

Johnson, Erik. 2013. *ArcMaker.* Plugin for Quantum Geographic Information System (QGIS) version 1.8.

Nicholas, Paul J. 2009, “Optimal Transmitter Placement in Wireless Mesh Networks.” Master’s thesis, Naval Postgraduate School.

Shankar, Arun. 2008, “Optimal Jammer Placement to Interdict Wireless Network Services.” Master’s thesis, Naval Postgraduate School.

Strachan, David. 2005. “Designing Fiber Optic Systems”. https://secure.connect.pbs.org/conferences/technology/2005/Presentations/Designing.Fiber.Optic.Systems.pdf (accessed May 11, 2014)

Quantum Geographic Information System. 2012. Retrieved from http://www.qgis.org/en/site/index.html.

White House, The. 2013. *Presidential Policy Directive—Critical Infrastructure Security and Resilience*. Washington, DC: White House.

World Bank. 2012. “Internet Users (per 100 People).” Accessed March 10. http://data.worldbank.org/indicator/IT.NET.USER.P2/countries?order=wbapi_data_value_2012%20wbapi_data_value%20wbapi_data_value-last&sort=desc&display=default.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California