3-26-2015

Determining the Most Vital Arcs within a Multi-Mode Communication Network Using Set-Based Measures

Christopher A. Hergenreter

Determining the Most Vital Arcs Within a Multi-Mode Communication Network Using Set-Based Measures

THESIS

MARCH 2015

Christopher A. Hergenreter, Capt, USAF

AFIT-ENS-MS-15-M-131

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-15-M-131

DETERMINING THE MOST VITAL ARCS WITHIN A MULTI-MODE COMMUNICATION NETWORK USING SET-BASED MEASURES

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

Christopher A. Hergenreter, B.A.

Capt, USAF

MARCH 2015

AFIT-ENS-MS-15-M-131

DETERMINING THE MOST VITAL ARCS WITHIN A MULTI-MODE COMMUNICATION NETWORK USING SET-BASED MEASURES

Christopher A. Hergenreter, B.A.
Capt, USAF

Committee Membership:

Dr. Sarah G. Nurre
Chair

Dr. Jeffery D. Weir
Member

AFIT-ENS-MS-15-M-131

Abstract

Technology has dramatically changed the way the military has disseminated information over the last fifty years. The Air Force has adapted to the change by operating a network with various ways to disseminate information. The Air Operating Center (AOC) is a large contributor to disseminating information in the Air Force. When the standard mode of sending information is disrupted, the AOC seeks both alternative ways available to send information and long term approaches to decrease vulnerability of its standard procedures. In this thesis, we seek to identify and quantify the most vital components within a multi-mode communications network via a combination of a set-based efficiency and set-based cost efficiency measures that utilize the all pairs shortest path (APSP) problem and minimum cost flow (MCF) problem. We capture the phenomenon that network components must work together to provide flow by examining how the network performs when sets of arcs are disrupted. We run 125 different computational experiments examining varying degrees of damage experienced by the network. From these results, we deduce insights into the characteristics of the most vital arcs in a multi-mode communication network which can inform future fortification decisions.

Key words: Air Operating Center (AOC), Air Tasking Order (ATO), All Pairs Shortest Path (APSP), Minimum Cost Flow (MCF), Set-Based Efficiency, Set-Based Cost Efficiency, Network, Most Vital, Communication Network

I would like to thank my lovely wife for all her constant support, encouragement, and allowing me to spend the days and nights needed to work through my masters. Additionally, I thank my youngens for waking up every morning to say good bye and always welcoming me home with hugs and kisses.

Acknowledgements

I would like thank my thesis advisor, Dr. Sarah Nurre for her expertise, guidance, and unending patience during my time at AFIT. I would also like to thank Capt Andrew Colson for his expertise and always being available to assist.

Lastly, I would like to thank all my peers that helped me and spent countless hours in the COA as we forged on to graduate.

Christopher A. Hergenreter

Table of Contents

List of Figures

List of Tables

DETERMINING THE MOST VITAL ARCS WITHIN A MULTI-MODE COMMUNICATION NETWORK USING SET-BASED MEASURES

I. INTRODUCTION

Since the beginning, military units have always sent information from one point to another, for example when a headquarters sends orders to the front lines. Currently, the Air Force has a network that has various ways to transfer information to and from different destinations. When this network is disrupted or destroyed the Air Force needs to determine the best modes to distribute information. Herein, we model the dissemination of information using a network, examine methods for determining the best dissemination mode when faced with disruption, and quantify the most vital arcs. Subsequently, the main contributions of this work are discussed.

1.1 Background

Technology has dramatically changed the way the military sends information over the last fifty years. One cause for this change is the desire to increase the security of the way in which information is sent; increased security ensures that the information being sent arrives in the correct format and in a time that still makes the information relevant to its intended destination. When the standard mode of distributing information is disrupted, military personnel look for other means to send information. Deciding which is best has increased in difficulty since there are many avenues through which information can be sent, and there are multiple factors which need to be considered when assessing the utility of the dissemination mode, including security, risk, and timeliness. The Air Operating Center (AOC) is a main Air Force contributor of disseminating information and they are most concerned about security and timeliness of the information being sent.

The AOC has an important role for the Air Force as well as the Joint environment. Joint Doctrine explains the AOC with the following definition: "[the AOC] provides the capability to plan, coordinate, allocate, task, execute, monitor, and assess the activities of assigned or attached forces" [1]. It is comprised of individuals from all the services and is considered a fully integrated command center.

The air tasking cycle that the AOC uses to produce an Air Tasking Order (ATO) can be seen in Figure 1. Traditionally the air tasking cycle is a 72-hour iterative process. The ATO production team is always working on the next day's ATO. Each component (Weaponeering and Allocation, Target Development, Objectives, Effects and Guidance) completes its portion of the ATO days prior to ATO production. The ATO production team is responsible for constructing, publishing, and disseminating the ATO and updates to the Special Instructions (SPINS) that accompany an ATO [1]. The ATO and SPINS along with the Air Operations Directive (AOD), and Airspace Command Order (ACO) "provide operational and tactical direction at appropriate levels of detail" [1]. These documents need to be sent securely and in a timely matter. The Joint Publication 3-30 states "a timely ATO is critical" [1].

The mode in which the ATO production team disseminates information may be disrupted by a variety of ways such as enemy or weather. Since the ATO production team faces these threats, it needs to understand and be prepared for such events by identifying alternative dissemination modes and long-term approaches to decrease the vulnerability of its standard procedures. To model this problem, a network composed of nodes and arcs is utilized. In this context nodes can represent those locations where information originates, waypoints, and locations needing information. Arcs represent ways information can be transmitted from one location to another. These

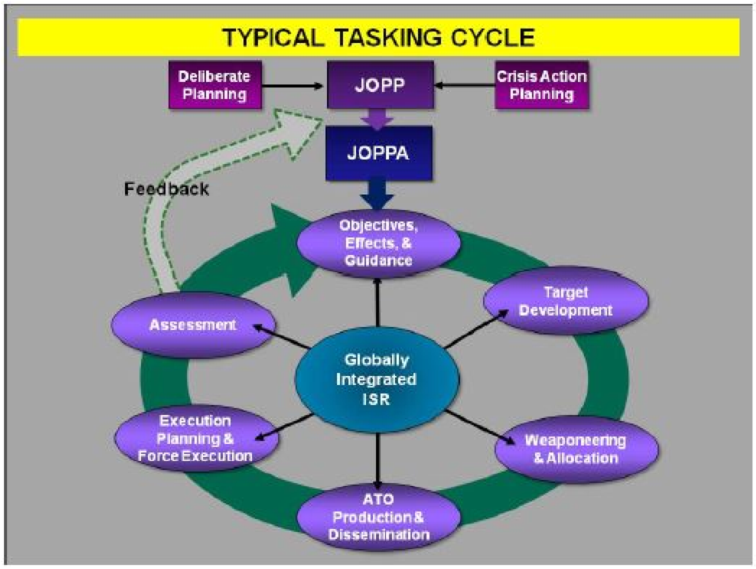


Figure 1. AOC's Air Tasking Cycle [1]

arcs have an associated cost parameter quantifying security, traversal time, distance and level of disruption. Flow through the network represents the ATO and SPINS[1]. By solving a minimum cost flow problem (MCF) on a network and an all pairs shortest path (APSP) problem [2], we are able to determine (i) if paths exist in the network allowing all demands to receive the ATO and (ii) if so, the least cost solution. We solve both a MCF and APSP problems for different levels of disruption to the network, denoted as scenarios. The last step in the scope of this thesis is to use the results from MCF and APSP problems in two vitality equations to quantify the vitality of network components[2] thus creating a rank ordering. The rank ordered list can then be used to perform a risk analysis on whether or not the vital components are vulnerable. We combine these two measures in a culminating vitality score. Once the vulnerabilities are identified it can drive the location of where to fortify with redundant components in the network, thus increasing the resiliency of the network [2].

A decision support system (DSS) was created in Microsoft Office Excel 2007 to

[1]Henceforth, the use of ATO represents both ATO and SPINS
[2]Components can be arcs and/or nodes

execute all algorithms, calculations, and output the results. The results produced by the DSS will aid a decision maker to make informed decisions and identify where a communication network's vital arcs and vulnerabilities are located and where redundant arcs should be placed to increase the network's resiliency.

Due to the sensitive nature of an actual Air Force communication network, we model and analyze an artificial multi-mode communication network. We perform computational experiments on this artificial network and then compare, analyze, and present the results produced.

1.2 Main Contributions and Summary

The five main contributions this research are: (i) a model for a multi-mode communication network, (ii) determination of the best way to send information by solving MCF and APSP problems, (iii) identifying the most vital components of a network utilizing scenarios and calculations which capture how sets of arcs interact, (iv) validation of the model and method through computational testing on a communication network, and (v) the development of a DSS which will help Air Force leaders to make informed decisions on the future of ATO dissemination.

The thesis proceeds as follows, Chapter II presents relevant work related to the application and methodology. Chapter III sets forth the formal mathematical problem statement, model, and methodology. The results of computational tests on realistic networks are presented in Chapter IV, including policy insights regarding the vitality of network components. In Chapter V, we conclude and propose promising and necessary avenues for future work.

II. Literature Review

In this chapter, we discuss literature relevant to the application and methodological approach. We explore the intricacies of the shortest path problem, specifically the all pairs shortest path problem (APSP), and the minimum cost flow problem (MCF). In addition to introducing the MCF and APSP, we discuss how they are relevant to modeling a multi-mode communication network. Further, we examine how many have defined and quantified resilient and vulnerable networks, identify most vital components, and what has been explored in the realm of ATO dissemination. Lastly we summarize aspects from the literature which we utilize to solve the problem.

2.1 Shortest Path

A shortest path problem identifies first if a path exist in a network from a specific source node to a specific sink node and secondly which path is the shortest to traverse from the two nodes [2]. The shortest path problem can be solved many different ways some examples are: Dijkstra's algorithm, Dijkstra's two tree algorithm, or the generic label-correcting algorithm [2, 3]. An extension of the shortest path problem is the APSP problem which identifies the shortest path between all pairs of nodes [2]. Dijkstra's algorithm can be iteratively solved to find solutions to the APSP problem. Further the Floyd-Warshall algorithm [2] can also be utilized to solve the APSP problem.

2.2 Minimum Cost Flow Problem

Although similar to the shortest path problem, a MCF problem is used to identify the cheapest option to send something through a network in order to meet a demand within the network while not exceeding the capacity on any arc [2]. Where shortest path problem is determining the one path that is the shortest from the source node to a sink node, MCF problem is solving to determine which path or paths is the cheapest to take from all source nodes to all demand nodes. Just as there are many ways to find the shortest path there are many ways to solve the MCF problem [2, 4, 5, 6]. We utilize the negative cycle canceling algorithm outlined in Ahuja et al. [2]. MCF has been used to solve transportation, distribution, and scheduling problems [2]. Some of the more recent applications of MCF have been used in scheduling information technology projects and assigning projects to human resources [7]. Dewil et al. [8] formulate a maximum covering/patrol routing problem as a MCF problem. Rao et al. [9] also use MCF in an attempt to minimize electricity cost. The minimum cost flow problem's versatility and ease of coding lends itself useful in many applications.

2.3 Resilient Networks

Using either APSP or MCF problems can lead to determine whether or not a network is resilient to an attack, natural disaster, or other instances where a network would be disrupted. Resiliency has been discussed in regards to many different networks to include transportation networks [10], community networks [11], series parallel networks [12], social ecological networks [13], and power networks [14]. A broad definition of network resiliency is how well a network performs under duress. In this section, we continue exploring how different application areas define and address the concepts of resiliency when modeled on a network.

The resilience of a network, according to Colbourn [12], is the measure of the network's reliability. He suggests using the expected number of node pairs that communicate as a way to measure reliability, thus providing a statistic for the resilience of a network. The connectivity or ability to communicate between pairs of nodes can be extracted from the solution of an APSP problem. Colbourn presents an algorithm that computes the resilience of a series parallel network. Sterbenz et al. [15] define resilience by how well the network provides the demanded service while disrupted. Their analytical approach to evaluating network resilience includes a two-dimensional state space defined by the operational state of the network and the services delivered. Different scenarios are simulated on the network resulting in a point in the two-dimensional space for each scenario. The combination of all points is denoted as the resilience trajectory. The resilience of the network is then quantified by taking the area under the resilience trajectory.

O'Rourke [11] defines resilience by how well the network "bounces" back after experiencing a disruption. For a community network he uses a simple resilience factor which is labeled as R. This factor is a measure of the networks resilience using the expected loss in quality or probability of failure over the period of time it takes for the network to recover. Barker et al. [16] define resilience almost verbatim as O'Rourke. Both papers use time to quantify the resilience of a network. Barker et al. [16] write that resilience can be measured by the time dependent ratio of recovery over loss. They state that there are four parts to resilience: reliability, vulnerability, survivability, and recoverability. The focus, when finding the resilience of a network, is on the vulnerability and recoverability. These two aspects over time are used to develop a resilience-based component importance measure (CIM). Additionally, Barker et al. [16] use a 20-node and 30-link network to demonstrate their component importance measure on a variety of different disruptive events.

2.4 Vulnerable Networks

Similar and often complimentary to network resilience, network vulnerability has been discussed in regards to many different types of networks, such as computer networks [17], power networks [18, 19], and community networks [11].

The more robust a network is, the less vulnerable the network will be when it is attacked [20]. Dekker et al. [20] defines robustness as a network's ability to continue performing while under or after an attack and if the topology of the graph is robust it will be due to the alternate paths within the network. The authors state that there are two ways in graph theory that can measure connectivity in a graph: node connectivity and arc connectivity. They argue that node connectivity is more useful in performing this connectivity metric. Node connectivity is defined in their paper as the smallest number of nodes that if removed would disconnect or create a single node graph.

Others have used a centrality index to aide in structural vulnerability analysis [18]. Dwivedi et al. [18] uses the centrality definition set forth by Freeman et al. [21]. Freeman et al. [21] uses a betweenness-based measure of centrality, which means that any component's centrality degree is based on how often it is on the shortest path between all components to all other components within a network. Armed with this centrality index, Dwivedi et al. [18] ranks the network components (specifically arcs) on their centrality index based on the portion of flow they carry through the network. Those arcs with the highest centrality value are deemed most vital. Knowing which arcs have the highest centrality index based on maximum flow through the network allows for one to pinpoint the most vital components. Thus knowing where the network is vulnerable.

2.5 Most Vital Network Components

Ventresca and Aleman [22] provide a good overview of the ways to determine the most vital components in a network. The authors define robustness as the ability for a network to continue operating while being attacked and list four ways to measure robustness; graph entropy, largest component, efficiency, and pairwise connectivity.

They then go on to mention six measures' of centrality which attempt to quantify the importance of a node; betweenness, closeness, degree, PageRank, Kleinberg's authority score, and leverage. All four robustness measures are used and compared after two attacking strategies are executed on a network; (i) randomly choosing one of the six centrality measures at each iteration and (ii) greedy choice which chooses one of the six centrality measure that will cause the largest drop in robustness. Over all the iterations there was little difference between the four robustness measures results.

The most vital arc in a network is defined by Ball and Golden [23] as the set of arcs that if simultaneously removed would have the greatest increase on the shortest distance between two specific nodes. For a transportation network, Novak et al. [24] present a similar approach to Ball and Golden which they call Network Robustness Index (NRI). This index is used to find the most vital arcs. The NRI approach evaluates the change in cost between networks with and without disturbances. They used three hypothetical models to compare the new approach to the traditional volume/capacity (V/C) ratio approach [25]. If V/C has a value larger than one then congestion is expected along that arc. The authors show how these two methods identify different sets of arcs that are most vital, and it is argued that NRI presents a better picture of the system-wide impacts opposed to the narrow localized V/C approach.

Wollmer [26], when referring to rail systems, defined the vital arc as the arc that would decrease the network's throughput the greatest. He determines the most vital set of arcs by taking the dual of a planar network and solves the shortest route problem thus identifying the set which is the minimum cut of a maximum flow problem. He states that his algorithm could be used to find where a transportation network would be sensitive to road closures or traffic accidents that tend to cause a decrease in the maximum flow through the network or where roads could be added to increase the maximum flow. Ratliff et al. [27] define the most vital arcs similarly to Wollmer [26] as the set of arcs that if removed simultaneously would cause the greatest decrease in throughput. They developed a procedure to find these set of arcs by sequentially modifying the network making cuts resulting in a minimum cut. It is recommended that this procedure be used with logistics or communications networks, where there may be an entity wanting to defend the network and another entity trying to interdict.

Latora et al. [28] define the critical components of a network as those that when removed cause the largest change between the efficiency of the network before and after the removal. The efficiency measure used is the same one referenced in Ventresca and Aleman [22]. This metric is the sum over all pairs of nodes, the inverse of the shortest path value between the pairs. As mentioned, this calculation is done before and after the removal of a component. Thus the importance of the component is the difference of efficiency before removal and after removal. Then those components with the highest change in efficiency are the most vital. To show versatility they apply their method to both a communication network and a terrorist network. We use this efficiency measure, but instead of removing one component we remove a set of components. It is the interaction of sets of network components that allow a network to perform. Thus, the concentration on sets represents an attempt to capture the importance of how the network responds rather than one single component.

Alderson et al. [29] argues that currently there is no set of all encompassing rules that find the most vital components when the components work together to produce a capacity or throughput. They state that it really depends on which set of arcs are damaged and takes analysis of the problem to determine which rule or set of rules best fit the problem at hand. They provide examples and ways to assess the critical components and the worst-case set of components. The examples range from as small as seven nodes and 9 arcs to the 1950s Soviet Train Network which has 50 nodes and approximately 90 arcs. Using this concept that network vitality is network specific, we focus on communication networks, such as an ATO dissemination network.

2.6 ATO Dissemination

With the exception of Allen [30], little has been written or researched about the dissemination of the ATO. Allen seeks to answer two main questions, the first being whether or not the current method of dissemination gets the job done. The second question is focused on the pros and cons of additional methods. He compares options that are used to disseminate the ATO in the joint environment (air gap, Multi-Level Security, web/Global Broadcast Service (GBS)) and states that web/GBS is the most suitable option to accomplish the mission. However, he does state that further research should be done prior to selecting one method over others.

2.7 Summary

Herein we provided an overview of APSP and MCF, both of which we use to quantify the vital components of a multi-mode communication network. Further, we examine previous quantifications of vulnerability and vitality which both motivate and compliment my approach. We proceed by providing the methodology we used for finding the most vital components in a network.

III. Methodology

We provide in this chapter the model and methodology utilized for determining the most vital components of a multi-mode communication network, such as an ATO dissemination network. First, we outline the mathematical notation describing the network, we put forth the formulation and algorithms for solving the all pairs shortest path (APSP) problem and the minimum cost flow (MCF) problem. Using the solutions of APSP and MCF, we describe the calculation for determining the vital network components which we denote as the set-based efficiency and the set-based cost efficiency measures. Combining these two measures, we set forth an arc vitality ranking, therby indicating the most vital arcs. We begin by representing the multi-mode communication network as a network described in the next section.

3.1 Mathematical Notation and Formulations

Let $G = (N, A)$ be a network defined by node set N and arc set A. Each arc $(i, j) \in A$ has a cost c_{ij}, capacity u_{ij}, and level of disruption $\beta_{ij} = 1$ if available and 0 otherwise. Cost can represent a variety of realistic entities, including time, security, and distance. A lower cost value indicates a greater incentive to utilize the arc (e.g. higher level of confidence in security). The length of the shortest path from node i to node j is denoted d_{ij}. Define s as the source where flow originates and τ as the sink where flow terminates. The decision variable x_{ij} represents the flow on each arc $(i, j) \in A$. For each node $i \in N$, define b_i to equal the supply (positive), demand (negative), and 0 otherwise.

3.2 Minimum Cost Flow Problem

On this network, we determine the MCF through the network. The MCF problem is considered the most fundamental of all network flow problems [2]. The problem can easily be defined as trying "to determine a least cost option of moving a commodity through a network in order to satisfy demands at certain nodes from available supplies at other nodes" [2]. The minimum cost flow which we solve for is directly dependent on the set of available arcs in the network (i.e. those with $\beta_{ij} = 1$). Thus, we extend the traditional MCF linear programming formulation [2] to include the availability of arcs.

$$\text{Minimize} \sum_{(i,j)\in A} c_{ij} x_{ij} \tag{1a}$$

$$\text{s.t.} \sum_{\{j:(i,j)\in A\}} x_{ij} - \sum_{\{j:(j,i)\in A\}} x_{ji} = b_i \qquad \forall\, i \in N \tag{1b}$$

$$0 \le x_{ij} \le u_{ij}\beta_{ij} \qquad \forall\, (i,j) \in A \tag{1c}$$

Equation (1a) sets forth the traditional MCF objective which multiplies the cost for each arc by its flow. Constraint (1b) is the mass balance constraint which ensures that the total flow leaving node i is not more than what is entering node i and at the same time is still meeting the demand, b_i. Our alteration to the formulation appears in Constraint (1c). This constraint limits the flow on each arc by its capacity if it is available and 0, otherwise.

This formulation can be solved to optimality using a commercial solver. However, there are exact algorithms for solving a MCF problem which are appropriate for the extension to the problem presented. The cycle-canceling algorithm [2] is one, which we outline in Algorithm 1.

Algorithm 1 Cycle-Canceling Algorithm for Solving MCF [2]

1: Input: $G = (N, A)$, for each arc : {cost=c_{ij}, capacity=u_{ij}}, and each node: {supply or demand=b_i}
2: Determine a feasible flow x_{ij} on the network satisfying all b_i values (e.g. maximum flow)
3: Create residual network $G' = (N', A')$ where $N' = N, A' = A \cup \{(j, i) : (i, j) \in A$ and $x_{ij} > 0\}$,
4: Calculate residual capacity

$$r_{ij} = \begin{cases} u_{ij} - x_{ij}, & \forall\ (i,j) \in A \\ x_{ij}, & \forall\ (i,j) \in A' \setminus A. \end{cases}$$

5: Set $c_{ij} = -c_{ij}\ \forall\ (i,j) \in A' \setminus A$
6: **while** $G' = (N', A')$ contains a negative cycle, C **do**
7: Let: $\gamma = \min\limits_{(i,j) \in C} r_{ij}$.
8: Update flow values on G'

$$x_{ij} = \begin{cases} x_{ij} + \gamma, & \text{if } (i,j) \in C \cap A \\ x_{ij} - \gamma, & \text{if } (i,j) \in C \text{ and } (j,i) \in C \cap A' \setminus A \\ x_{ij}, & \text{otherwise.} \end{cases}$$

9: Update G' as in Step 3 and Step 4.
10: **end while**
11: Output the minimum cost flow value and solution.

In this algorithm, first a feasible flow is found which satisfies all of the demand. Ford-Fulkerson's maximum flow algorithm [2] is one which can find a feasible flow. The cycle-canceling algorithm continues by rerouting this flow through the identification of negative cycles. A negative cycle in the residual network indicates that flow can be rerouted, still maintaining the satisfaction of demand, in a cheaper way. The algorithm terminates when no negative cycles remain.

A benefit to using the cycle-canceling algorithm as opposed to solving the linear programming formulation is that it does not require the use of a commercial solver. Thus, it can be implemented in any programming language. As will be discussed more thoroughly in Chapter IV, we have implemented the cycle-canceling algorithm in VBA within Microsoft Excel 2007. This provides a familiar platform for users managing the application of ATO dissemination discussed in Chapter I.

3.3 All Pairs Shortest Path Problem

On the network set forth in Section 3.1, we also solve for the shortest path between all pairs of nodes. The shortest path problem is considered to be the simplest of all network flow problems [2], which seeks to determine the shortest length from a set origin to a set destination. This problem can be solved utilizing the MCF linear programming formulation outlined in Equations (1a)-(1c). With this formulation, by setting $b_s = 1$, $b_\tau = -1$, and $b_i = 0 \; \forall i \in N \setminus \{s, \tau\}$, we can find the shortest path from origin s to destination τ. For the APSP problem, we seek to determine not the shortest path from one origin to one destination, but the shortest path between all pairs of nodes. As follows, we iteratively solve the shortest path problem by changing the origin and destination.

As was the case for the MCF problem, the shortest path problem can also be solved via alternative exact algorithms not requiring commercial solvers. Dijkstra's algorithm is one such algorithm [2]. Dijkstra's algorithm is useful because it not only finds the shortest path from s to τ but finds the shortest path from s to all other nodes within the network. It accomplishes this by scanning out from s and maintaining a distance label for each node which acts as an upper bound on the shortest path to that node [2]. Throughout the execution of the algorithm, it maintains two groups, one group that has been permanently labeled and the other group is temporarily labeled. The distance label for all permanent nodes is the shortest distance from s to that node. This algorithm continues until all nodes are permanently labeled, resulting in the shortest distance from s to all other nodes. Hence, we can determine a solution to the APSP problem by solving $O(|N|)$ shortest path problems, one for each node in the network, by changing the origin. Because of this result and not requiring the use of a commercial solver, we have implemented the iterative solving of Dijkstra's algorithm to determine a solution to the APSP problem.

3.4 Vitality Measures

Using the solutions from the APSP problem and MCF problem, we calculate the vitality of arcs within the network. The two vitality measures proposed herein are motivated by the work of Latora et al. [28]. The traditional efficiency measure defined by Latora et al. [28] is presented in Equations (2) and (3). In Equation (2) the efficiency of the network is calculated by taking the average of the inverse of the length of the shortest path between all pairs of nodes. This calculation is utilized in Equation (3) which computes the change in the efficiency of the network when a network component i (arc or node) is removed. Latora et al. [28] states that in contrast to calculating the change in efficiency when a single component is removed, it can also be calculated when a set of components are removed. The examination of sets of components is particularly important in networks, as it is the combination and interaction of network components which allows a network to perform (e.g. deliver flow to demand nodes). However, even with the examination of sets of components it is desirable to attribute a vitality score to individual components, thereby informing fortification decisions.

As such, we utilize the efficiency measure outlined in Equations (2) and (3) with sets of arcs removed. By examining many scenarios of arcs which are removed, we then attribute a vitality score back to each arc.

$$E(G) = \frac{1}{|N|(|N|-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}} \tag{2}$$

$$\Delta E = E(G) - E(G - component_i) \quad i = 1, ..., A, \tag{3}$$

Define K as the set of possible scenarios considered. For each $k \in K$, define arc set A^k as the set of available arcs, specifically $A^k = \{A|\beta_{ij}^k = 1\}$ where $\beta_{ij}^k = 1$ when arc is available in scenario k. From this we define $G^k = (N, A^k)$ as the network under scenario k. Without loss of generality, we focus on the availability of arcs, as nodes can be equivalently represented as arcs. Define w^k to represent the weight of scenario k. This weight may convey the likelihood of scenario occurring or the scenario's importance. With this we define the new set-based efficiency measure denoted e_{ij} for all arcs $(i, j) \in A$. Directly influenced by Equation (3) from Latora et al. [28], we calculate the change in network efficiency for scenario k, denoted ΔE^k, via Equation (4). Combining w^k, β_{ij}^k, and ΔE^k we introduce the set-based efficiency measure in Equation (5) where $\sum_{k=1}^{|K|}(1 - \beta_{ij}^k)$ calculates the number of scenarios with which arc (i, j) is disrupted. Hence, the efficiency of arc (i, j), e_{ij} is calculated by taking the average weighted change in efficiency of those scenarios with which arc (i, j) is disrupted.

$$\Delta E^k = E(G) - E(G^k) \tag{4}$$

$$e_{ij} = \frac{1}{\sum_{k=1}^{|K|}(1 - \beta_{ij}^k)} \sum_{k=1}^{|K|}(1 - \beta_{ij}^k)\Delta E^k w^k \tag{5}$$

The calculation of the first efficiency measure was based on the length of the shortest path between all pairs of nodes. For the second vitality measure we present, the calculation is based on the minimum cost flow. We continue the examination of sets of arcs which leads to an individual arc based calculation. Define ΔC^k as the difference in the minimum cost flow between network G^k and G, as shown in Equation (6).

$$\Delta C^k = \sum_{(i,j)\in A^k} c_{ij}x_{ij}^k - \sum_{(i,j)\in A} c_{ij}x_{ij} \tag{6}$$

As was done in Equation (5) for the set-based efficiency measure, we combine w^k, β_{ij}^k, and ΔC^k in the new set-based cost efficiency measure, h_{ij}, set forth in Equation (7). This equation calculates h_{ij} as the average weighted change in the minimum cost flow for the scenarios in which $\beta_{ij}^k = 0$.

$$h_{ij} = \frac{1}{\sum_{k=1}^{|K|}(1-\beta_{ij}^k)} \sum_{k=1}^{|K|}(1-\beta_{ij}^k)\Delta C^k w^k \tag{7}$$

3.5 Determining Vitality

For each vitality measure, e_{ij} and h_{ij}, a ranked list is created where r_{ij}^e and r_{ij}^h represents the ranking of arc $(i,j) \in A$ for set-based efficiency and set-based cost efficiency, respectively. A lower ranking indicates arc (i,j) had a greater e_{ij} and h_{ij} value. This higher value indicates it was apart of scenarios which led to the greatest decrease in network performance.

To produce a ranked list of the most vital arcs we combine both metrics. To ensure the combination of both measures does not consider one measure more important than another, we normalize both between 0 and 1. Define $\hat{e_{ij}}$ and $\hat{h_{ij}}$ as the normalized set-based efficiency and set-based cost efficiency, respectively. The normalization is calculated via Equations (8) and (9).

$$\hat{e_{ij}} = \frac{e_{ij} - \min\limits_{(i,j)\in A} e_{ij}}{\max\limits_{(i,j)\in A} e_{ij} - \min\limits_{(i,j)\in A} e_{ij}} \tag{8}$$

$$\hat{h_{ij}} = \frac{h_{ij} - \min\limits_{(i,j)\in A} h_{ij}}{\max\limits_{(i,j)\in A} h_{ij} - \min\limits_{(i,j)\in A} h_{ij}} \tag{9}$$

Using these calculations, we define a combined vitality measure v_{ij} for all $(i, j) \in A$ presented in Equation (10). Which adds the normalized set-based efficiency and set-based cost efficiency measures.

$$v_{ij} = \hat{e_{ij}} + \hat{h_{ij}} \tag{10}$$

The v_{ij} values are utilized to create a ranked list. This ranking is our final vitality score. Those with a lower ranking indicate the most vital arcs.

In the next chapter we validate the use of v_{ij}, our vitality measure, in an artificial multi-mode communication network and discuss the results and analysis from 125 computational experiments.

IV. Computational Results

In this section we describe the network structure and demonstrate how to use the minimum cost flow (MCF) problem and all pairs shortest path (APSP) problem to inform the development of set-based cost efficiency and set-based efficiency vitality measures. Additionally, we validate these vitality measures explained in Chapter III and show how they directly feed the calculation to identify vital components for a communications network by performing computational experiments.

4.1 Network

For the computational experiments we consider a multi-mode artificial communication network with $|N| = 52$ and $|A| = 752$, see Figure 2 for an illustration of the network. Each node $i \in \{1, 2\}$ is connected to all nodes $j \in \{3, \ldots, 27\}$ by four parallel arcs (i, j) with four distinct costs 1, 2, 3, or 4. Additionally, each node $i \in \{1, 2\}$ is connected to all nodes $j \in \{28, \ldots, 52\}$ by three parallel arcs (i, j) with three distinct costs 1, 2, and 4. Define nodes 1 and 2 as supply nodes and nodes $j \in \{3, \ldots, 52\}$ as demand nodes. In this network, arcs connect the nodes in $\{3, \ldots, 27\}$ together and those nodes in $\{28, \ldots, 52\}$ together. Each node $i \in \{3, \ldots, 27\}$ is connected to five randomly selected distinct nodes $j \in \{3, \ldots, 27\}$, where $i \neq j$ via two parallel arcs (i, j). For each arc (i, j), a random cost of 1, 2, 3, or 4 is generated and set equal to c_{ij}. Similarly, each node $i \in \{28, \ldots, 52\}$ is connected to three randomly selected distinct nodes $j \in \{28, \ldots, 52\}$, where $i \neq j$ via two parallel arcs (i, j). These arcs are also assigned a random cost of 1, 2, 3, or 4. We augment the network with a dummy node $i = 53$ to ensure a feasible MCF can be achieved. This node is connected to all demand nodes $j \in \{3, \ldots, 52\}$ via arcs with an arbitrarily high cost set to 100. This high cost ensures that the other options in the network are more desirable. Additional augmentation is done on the network with the addition of a super source and super sink defined as $s = 54$ and $\tau = 55$, respectively. The super source has $b_s = 50$ and super sink has $b_\tau = -50$ indicating the supply and demand to meet all 50 demand nodes, respectively. The super source s is connected by one arc (s, i) to each node $i \in \{1, 2, 53\}$ with a capacity $u_{si} = 50$ and cost $c_{si} = 0$. The super sink is connected via one arc (j, τ) from each node $j \in \{3, ..., 52\}$ with a capacity $u_{j\tau} = 1$ and cost $c_{i\tau} = 0$. Knowing that no more than 50 will flow through the network based on $b_s = 50$, we place a non-limiting constraint on each arc (i, j) by setting the capacity $u_{ij} = 50$ for all arcs (i, j) with $i \in \{1, 2, 53\}$ and $j \in \{3, \ldots, 52\}$. The non-limiting constraint with $u_{ij} = 50$ is also applied to the sets of arcs (i, j) for nodes $i, j \in \{3, \ldots, 27\}$ with $i \neq j$ and $i, j \in \{28, \ldots, 52\}$ with $i \neq j$. Arcs (i, j) where $i \in \{1, 2\}$ and $j \in \{3, \ldots, 52\}$ are classified as "Type 1" and arcs (i, j) are classified "Type 2" when $i, j \in \{3, \ldots, 52\}$ with $i \neq j$. This classification is a way to separate the topology of the network for added analysis. Type 1 arcs directly connect the supply nodes to the demand nodes and Type 2 arcs connect demand nodes together.

Figure 2 provides a visual of the network. Due to the overabundance of arc redundancy, some arcs in the network are not shown in the figure.

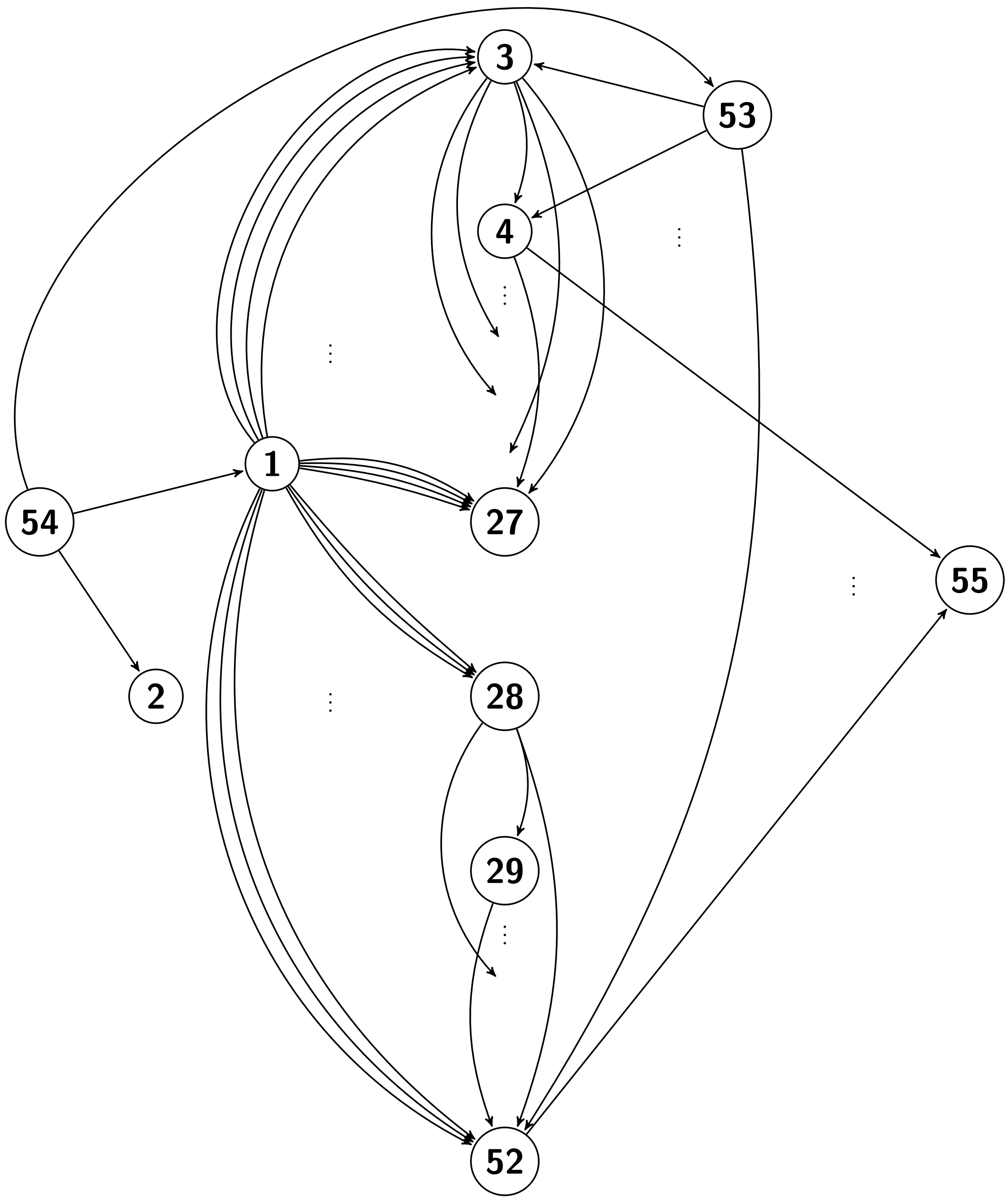


Figure 2. We show a graphical representation of G. Present are arcs from the super source node 54 to supply nodes 1, 2, and dummy node 53. The arcs that originate from node 2 mimic those that originate from node 1. In the set of nodes $\{3, \ldots, 27\}$ there are two parallel arcs that connect each node with five other nodes in the set. In the set of nodes $\{28, \ldots, 52\}$ there are two parallel arcs that connect each node with three other nodes in the set. Node 53 is connected to all the nodes in the set $\{3, \ldots, 52\}$ by one arc. Each node in the set $\{3, \ldots, 52\}$ is connected by one arc to node 55.

4.2 Scenarios

For the computational experiments, we consider five different degrees of damage, specifically when 10%, 20%, 30%, 40%, and 50% of the arcs are damaged. For each of these five damage percentages, we create twenty five random instances by uniformly at random selecting the appropriate number of arcs to be damaged. This random selection is conducted using the Microsoft Excel random function and results in 125 different scenarios.

These computational experiments are conducted in a decision support system (DSS) built in Microsoft Office Excel 2007 with the Visual Basic for Applications environment. The DSS inputs a network, solves for both the MCF and APSP solutions, computes both vitality measures, and produces a final ranked order list of the most vital arcs within the network. The results shown hereafter were performed on a desktop with 3.40 GHz AMD A4-5300B processor and 8.00 GB of RAM.

4.3 Implementation

Both the network described in Section 4.1 and the scenarios from Section 4.2 in the form of Excel files are used as input into the DSS mentioned in Section 4.2. The DSS then calculates ΔE^k and ΔC^k for each damage scenario $k \in K$ according to Equations (3) and (6). These values are then used to calculate the two arc vitality measures: set-based efficiency (e_{ij}) and set-based cost efficiency (h_{ij}). Table 1 shows the values associated with each percentage of damage: the total arcs in the network; total number of damaged arcs; and scenario weight w^k value used in Equation (5) and (7). The weight (w^k) utilized for each damaged percentage is calculated as the inverse of the number of damaged arcs. With this calculation, the impact felt by a damage scenario is equally distributed between all arcs in that particular scenario.

Table 1. Values associated with each degree of damage

Percentage	Total Arcs	Total Damaged	Weight
0%	752	0	0
10%	752	75	1/75
20%	752	151	1/151
30%	752	226	1/226
40%	752	301	1/301
50%	752	376	1/376

Once the DSS calculates e_{ij} and h_{ij} it continues and solves for Equations (8), (9), and (10), $\hat{e_{ij}}$, $\hat{h_{ij}}$, and v_{ij}, respectively. We then rank order v_{ij}, resulting in the final output, which a sample of this output is shown in Table 2. Interpreting Table 2 beginning from the left most column going towards the right most column the table displays the number of each arc, where the arc originates (i), ends (j), cost (c_{ij}), the arc's category (CAT) as Type 1 or Type 2, the number of times the arc is disrupted, in the 125 scenarios the value of both the normalized set-based efficiency measure ($\hat{e_{ij}}$), $\hat{e_{ij}}$ rank, normalized set-based cost efficiency measure ($\hat{h_{ij}}$), $\hat{h_{ij}}$ rank, the summation of $\hat{e_{ij}}$ and $\hat{h_{ij}}$, (v_{ij}), and lastly the ranking associated with arc (i, j) according to v_{ij}.

Table 2. Sample DSS output showing characteristics of each arc and final $\hat{e_{ij}}, \hat{h_{ij}}$, and v_{ij} values and rankings.

Arc #	i	j	c_{ij}	CAT	# of times Disrupted	$\widehat{e_{ij}}$	$\widehat{e_{ij}}$ Rank	$\widehat{h_{ij}}$	$\widehat{h_{ij}}$ Rank	v_{ij}	v_{ij} Rank
1	1	2	4	Type 1	39	0.67	218	0.60	183	1.27	167
2	2	1	4	Type 1	34	0.66	133	0.31	633	0.97	378
3	1	3	1	Type 1	38	0.42	566	0.61	158	1.03	360
4	1	3	2	Type 1	36	0.41	459	0.38	537	0.79	518
5	1	3	3	Type 1	42	0.42	506	0.44	429	0.87	481
6	1	3	4	Type 1	45	0.25	639	0.46	397	0.72	555
7	2	3	1	Type 1	37	0.49	96	0.63	128	1.13	97
8	2	3	2	Type 1	38	0.22	456	0.29	653	0.51	587
9	2	3	3	Type 1	32	0.06	522	0.36	580	0.41	574
10	2	3	4	Type 1	37	0.61	11	0.67	93	1.28	21
11	1	4	1	Type 1	35	0.69	201	0.77	31	1.47	68
12	1	4	2	Type 1	37	0.44	469	0.51	319	0.94	403
13	1	4	3	Type 1	37	0.43	558	0.58	199	1.01	390
14	1	4	4	Type 1	38	0.39	216	0.39	523	0.78	357
15	2	4	1	Type 1	48	0.54	300	0.42	464	0.96	385

4.4 Results

In this section we describe what analysis was done on the results and output from the DSS after the implementation outlined in Section 4.3. We begin by analyzing the change of efficiency and cost efficiency at each degree of disruption. This analysis allows us to quantify how the network reacts at each degree of disruption. Along the horizontal axis in Figure 3 and Figure 4 is the percentage of arcs disrupted and the vertical axis displays the average change across each percentage. We found that the average change in the APSP (ΔE^k) and average change in MCF (ΔC^k) values increased as the percentage of disruption increased. We note in Figure 3 that the rate of ΔE^k appears linear. However, in Figure 4 we note that as 40% and 50% of the arcs are damaged the ΔC^k is more dramatic than 10% and 20%, indicating a nonlinear trend. The insight drawn from this is that a defender's cost efficiency will decrease at a greater rate after 30% of the network is disrupted rather than when the network is disrupted 30% or less.

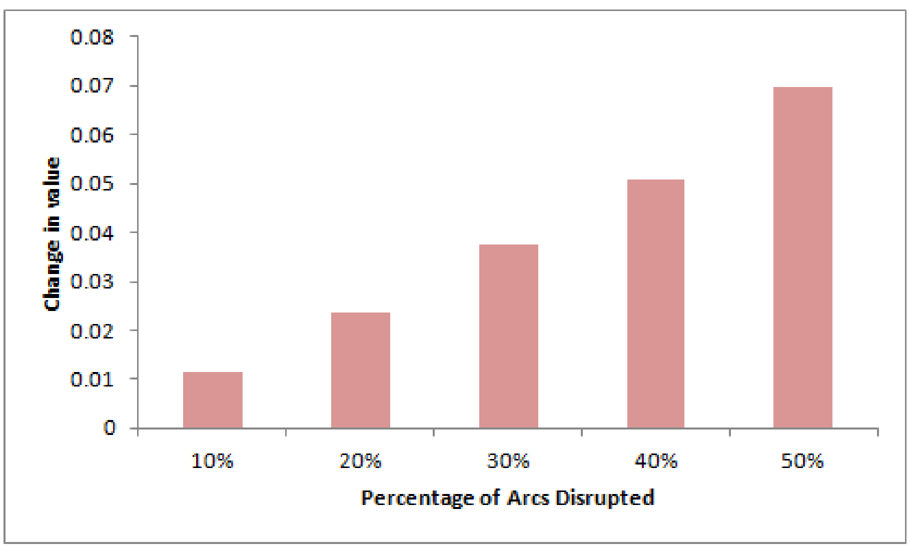


Figure 3. The average ΔE^k value as the percentage of disrupted arcs increases.

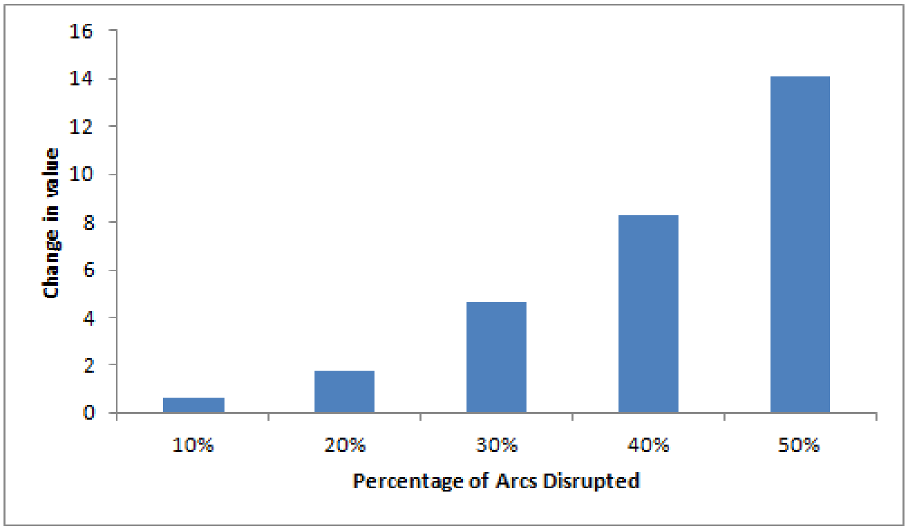


Figure 4. The average ΔC^k value as the percentage of disrupted arcs increases.

.

The next set of analysis is utilized to assess the vitality of arcs in the network. In Figure 5, we display the e_{ij} ranking vs h_{ij} ranking for each arc. On the horizontal axis is h_{ij} rank from 0 to 752 and on the vertical axis is the e_{ij} rank from 0 to 752. For each metric, 0 is considered most vital, indicating that arcs near the origin are considered most vital in both metrics. Type 1 arcs are displayed as a circle and Type 2 arcs are displayed as a addition sign. For each scenario, we randomly generated which arcs are disrupted based on the degree of disruption. The color indicates how many times an arc was disrupted in the 125 scenarios. Upon initial inspection, we notice that there is a linear relationship between e_{ij} and h_{ij}. Additionally, there is a higher concentration of dark blue points near the figure's origin and a concentration of dark red points near the top most right corner. This phenomenon is due to the random generation of scenarios and equal weighting for each scenario. These two together create cases where some arcs are only disrupted during the higher percentage scenarios (the dark blue points near the origin), when the change in efficiency and cost efficiency is the greatest. Thus, giving them a higher v_{ij} value than those arcs that are disrupted in many scenarios including the 10% scenarios (the dark red points in top most right corner) when the change in efficiency and cost efficiency is least. Without an appropriate weight assigned to each scenario it can skew the results to misrepresent which arcs are most vital.

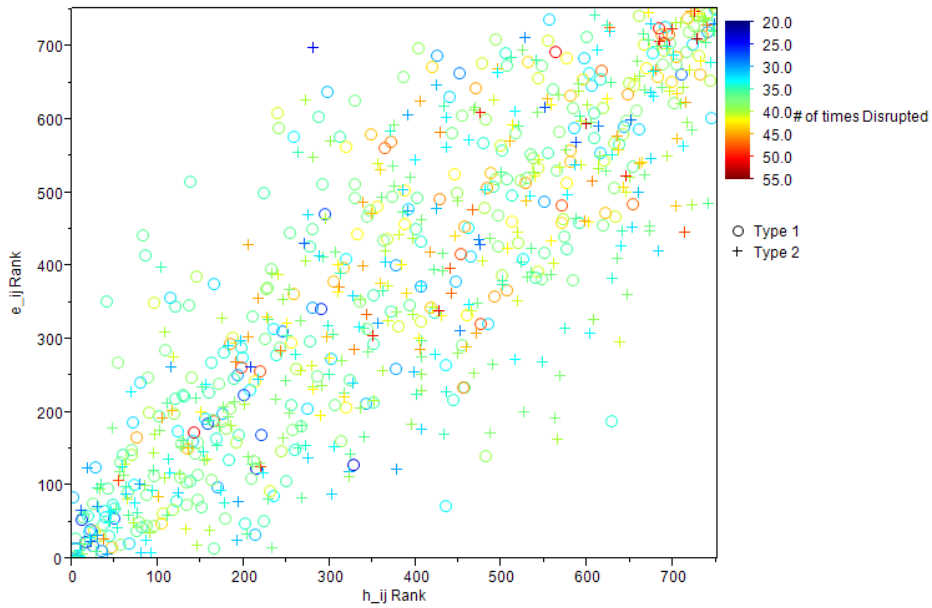


Figure 5. The e_{ij} and h_{ij} rank for all arcs within the network with a $w^k = 1$ applied to each scenario.

To address the misrepresentation of which arcs are most vital, we applied a weighting as shown in Table 1, rather than an equal weighting. This weighting change leads to different e_{ij} and h_{ij} values, thus changing the ranking. In Figure 6, similarly to Figure 5, we display the $\hat{e_{ij}}$ ranking vs $\hat{h_{ij}}$ ranking for each arc with these weights applied. Once the weight was applied the linear relationship which was seen before is no longer evident, highlighting what is most important between the two metrics.

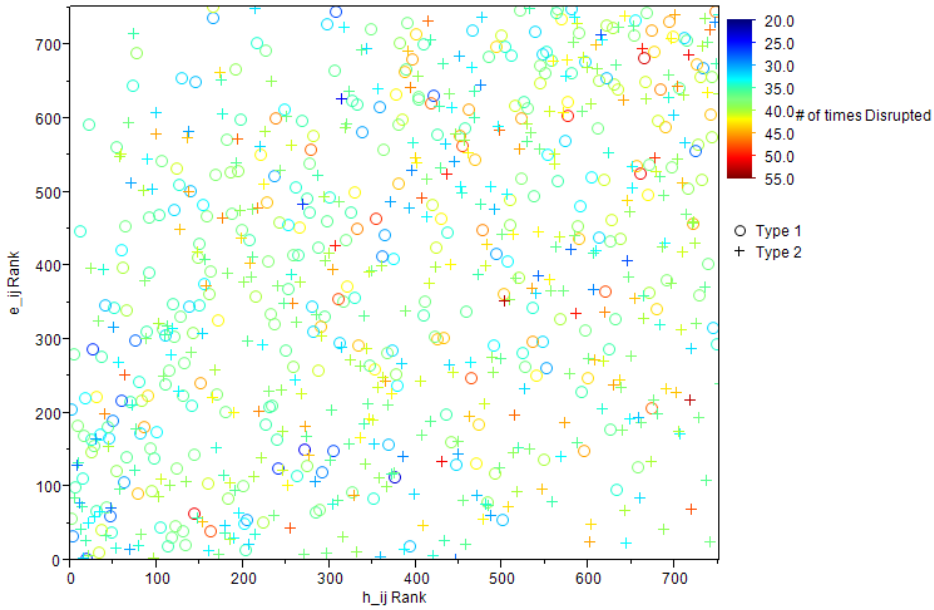


Figure 6. All the arcs within the network after a weight is applied to each scenario.

Table 3, 4, and 5 display a rank order list of the top 75 most vital arcs in the network according to v_{ij}. In addition to the rank according to v_{ij}, Table 3, 4, and 5 displays the following: the arc number associated with arc (i, j), the originating location (node i), destination (node j), the cost to flow down arc (i, j), the type of arc, the number of times the arc is disrupted over the 125 scenarios, $\hat{e_{ij}}$ value, the rank of arc (i, j) according to $\hat{e_{ij}}$, $\hat{h_{ij}}$ value, the rank of arc (i, j) according to $\hat{h_{ij}}$, the summation of $\hat{e_{ij}}$ and $\hat{h_{ij}}$ with v_{ij}, and each arc's rank according to v_{ij}.

The three Tables 3, 4, and 5, highlight four varying characteristics of the top 75 most vital arcs. The first being that the cost for each arc varies from 1, 2, 3, and 4 throughout all 75 arcs. The second varying characteristic is that half of the most vital arcs according to v_{ij} were Type 1, implying that the type of arc did not drive whether or not it was in the top 75. The number of times each arc was disrupted varied as well which was our desire by applying the appropriate weight. We wanted to ensure the number of times an arc was disrupted did not positively or negatively affect the results by favoring either a more disrupted arc or a less disrupted arc. The last characteristic to note is that the ranking for both $\hat{e_{ij}}$ and $\hat{h_{ij}}$ are not exact for any of the top 75 arcs and some are quite different. For instance, $\hat{e_{ij}}$ ranks arc number 271 as 183 and $\hat{h_{ij}}$ ranks it the 8th most vital. This can be seen in the reverse with arc number 741. However, there are some arcs where they both are closer to the same $\hat{e_{ij}}$ and $\hat{h_{ij}}$ rank, as in the case for arc 27 or 478. As expected, those arcs that are closer in rank between $\hat{e_{ij}}$ and $\hat{h_{ij}}$ rise to the top in v_{ij} ranking. However, there were only 30 arcs that both $\hat{e_{ij}}$ and $\hat{h_{ij}}$ identified as being in the top 75 most vital arcs.

Table 3. A rank ordered table of the top 75 arcs after applying a weight, according to v_{ij} ranking: 1-30

Arc #	i	j	c_{ij}	CAT	# of times	$\widehat{e_{ij}}$	$\widehat{e_{ij}}$ Rank	$\widehat{h_{ij}}$	$\widehat{h_{ij}}$ Rank	v_{ij}	v_{ij} Rank
27	1	6	1	Type 1	26	1.00	1	0.83	19	1.83	1
478	15	17	3	Type 2	32	0.92	4	0.84	15	1.76	2
284	2	41	1	Type 1	29	0.75	33	0.97	3	1.72	3
65	2	10	3	Type 1	38	0.71	56	0.98	2	1.69	4
657	37	28	1	Type 2	39	0.85	6	0.82	21	1.67	5
497	17	20	4	Type 2	35	0.77	28	0.83	17	1.60	6
611	29	33	1	Type 2	40	0.93	3	0.67	96	1.60	7
290	2	42	1	Type 1	41	0.83	11	0.77	33	1.59	8
521	19	15	3	Type 2	35	0.73	43	0.85	12	1.58	9
317	1	47	1	Type 1	31	0.57	205	1.00	1	1.57	10
639	34	35	1	Type 2	36	0.67	85	0.88	4	1.55	11
475	15	23	2	Type 2	33	0.72	52	0.83	20	1.54	12
439	11	19	4	Type 2	34	0.81	17	0.73	52	1.54	13
625	31	51	2	Type 2	30	0.68	78	0.86	10	1.54	14
450	12	17	2	Type 2	35	0.69	73	0.85	13	1.54	15
280	2	40	4	Type 1	35	0.66	99	0.87	6	1.52	16
737	50	39	1	Type 2	30	0.82	15	0.70	68	1.51	17
701	44	40	3	Type 2	36	0.75	34	0.75	41	1.50	18
402	7	17	2	Type 2	33	0.71	59	0.79	28	1.50	19
206	2	28	1	Type 1	41	0.73	41	0.76	36	1.49	20
654	36	49	2	Type 2	28	0.63	130	0.86	7	1.49	21
183	2	25	1	Type 1	34	0.65	110	0.84	14	1.49	22
651	36	51	1	Type 2	35	0.63	133	0.86	9	1.49	23
505	18	25	1	Type 2	34	0.80	20	0.69	79	1.49	24
25	2	5	3	Type 1	34	0.74	38	0.74	48	1.47	25
11	1	4	1	Type 1	35	0.69	66	0.77	31	1.47	26
341	1	51	1	Type 1	35	0.80	18	0.66	113	1.46	27
271	1	39	4	Type 1	38	0.59	183	0.86	8	1.45	28
480	15	7	3	Type 2	33	0.69	67	0.76	34	1.45	29
741	51	34	1	Type 2	37	0.85	8	0.60	172	1.45	30

Table 4. A rank ordered table of the top 75 arcs after applying a weight, according to v_{ij} ranking: 31-60

Arc #	i	j	c_{ij}	CAT	# of times	$\widehat{e_{ij}}$	$\widehat{e_{ij}}$ Rank	$\widehat{h_{ij}}$	$\widehat{h_{ij}}$ Rank	v_{ij}	v_{ij} Rank
53	1	9	3	Type 1	28	0.71	60	0.74	47	1.45	31
288	1	42	2	Type 1	37	0.69	71	0.75	42	1.44	32
689	42	52	1	Type 2	33	0.82	13	0.61	159	1.44	33
155	1	22	1	Type 1	37	0.60	169	0.84	16	1.44	34
748	52	31	2	Type 2	38	0.75	35	0.69	82	1.43	35
586	26	14	2	Type 2	28	0.69	72	0.75	46	1.43	36
143	2	20	1	Type 1	37	0.80	21	0.63	134	1.43	37
181	1	25	3	Type 1	40	0.62	146	0.81	25	1.42	38
237	2	33	2	Type 1	34	0.60	163	0.81	24	1.41	39
91	1	14	1	Type 1	37	0.61	155	0.80	27	1.40	40
79	2	12	1	Type 1	37	0.75	31	0.65	118	1.40	41
26	2	5	4	Type 1	35	0.82	14	0.57	203	1.39	42
272	2	39	1	Type 1	33	0.57	220	0.83	18	1.39	43
740	50	33	1	Type 2	28	0.96	2	0.44	446	1.39	44
151	2	21	1	Type 1	36	0.52	279	0.87	5	1.39	45
338	2	50	1	Type 1	34	0.73	45	0.66	105	1.39	46
562	23	15	4	Type 2	30	0.60	165	0.78	29	1.38	47
360	3	20	3	Type 2	37	0.85	9	0.53	270	1.38	48
68	1	11	2	Type 1	39	0.64	121	0.73	53	1.37	49
729	49	37	1	Type 2	36	0.66	98	0.71	57	1.37	50
727	48	50	1	Type 2	32	0.80	22	0.57	206	1.37	51
313	1	46	4	Type 1	37	0.73	39	0.63	130	1.37	52
134	1	19	4	Type 1	38	0.73	46	0.64	122	1.37	53
609	29	36	1	Type 2	40	0.77	29	0.59	185	1.36	54
388	6	8	2	Type 2	37	0.84	10	0.52	296	1.36	55
294	1	43	2	Type 1	33	0.60	171	0.76	35	1.36	56
552	22	7	4	Type 2	33	0.77	30	0.59	188	1.36	57
749	52	48	4	Type 2	40	0.61	156	0.75	43	1.35	58
156	1	22	2	Type 1	30	0.65	105	0.70	63	1.35	59
225	2	31	2	Type 1	44	0.66	90	0.69	78	1.35	60

Table 5. A rank ordered table of the top 75 arcs after applying a weight, according to v_{ij} ranking: 61-75

Arc #	i	j	c_{ij}	CAT	# of times	$\widehat{e_{ij}}$	$\widehat{e_{ij}}$ Rank	$\widehat{h_{ij}}$	$\widehat{h_{ij}}$ Rank	v_{ij}	v_{ij} Rank
746	51	35	4	Type 2	31	0.86	5	0.49	358	1.35	61
80	2	12	2	Type 1	32	0.60	166	0.75	45	1.35	62
149	1	21	3	Type 1	48	0.73	40	0.61	163	1.34	63
103	2	15	1	Type 1	42	0.56	222	0.78	30	1.34	64
177	2	24	3	Type 1	36	0.59	180	0.75	44	1.34	65
485	16	18	2	Type 2	35	0.74	37	0.60	179	1.34	66
426	10	21	2	Type 2	44	0.72	53	0.62	152	1.33	67
461	13	22	3	Type 2	40	0.67	87	0.66	101	1.33	68
195	1	27	1	Type 1	40	0.66	93	0.67	95	1.33	69
716	46	29	2	Type 2	46	0.58	199	0.75	39	1.33	70
646	35	48	1	Type 2	35	0.71	58	0.62	147	1.33	71
679	40	38	2	Type 2	35	0.64	122	0.69	74	1.33	72
622	31	45	1	Type 2	39	0.83	12	0.50	331	1.33	73
139	1	20	1	Type 1	36	0.63	139	0.70	66	1.33	74
196	1	27	2	Type 1	29	0.59	190	0.74	49	1.32	75

Figure 7 displays Table 3, 4, and 5 in a graphical form giving a different insight into the results. Both axises are from 0 to 450 with the horizontal axis being h_{ij} ranking and the vertical axis being e_{ij} ranking of the top 75 most vital arcs. The colors represent the number of times an arc was disrupted and the circles are Type 1 arcs and an addition sign is Type 2 arcs. Based on the figure there is an even distribution of the number of times an arc is disrupted as well as where the arc is located within the topology of the network. Note that there are arcs in the top 75 that are ranked significantly higher by e_{ij} than h_{ij} and vice versa. The outlying arcs that are ranked higher by e_{ij} than h_{ij} are Type 2 and those outlying arcs that are ranked higher by h_{ij} than e_{ij} are Type 1 arcs. When we looked at the top 75 arcs according to e_{ij} there were only 25 of the 75 that were Type 1 and when we looked at the top 75 according to h_{ij} it was a near split with 42 of the 75 that were Type 1. This highlights the difference between the two metrics, that they are not mirror images of the other and the importance of including both metrics.

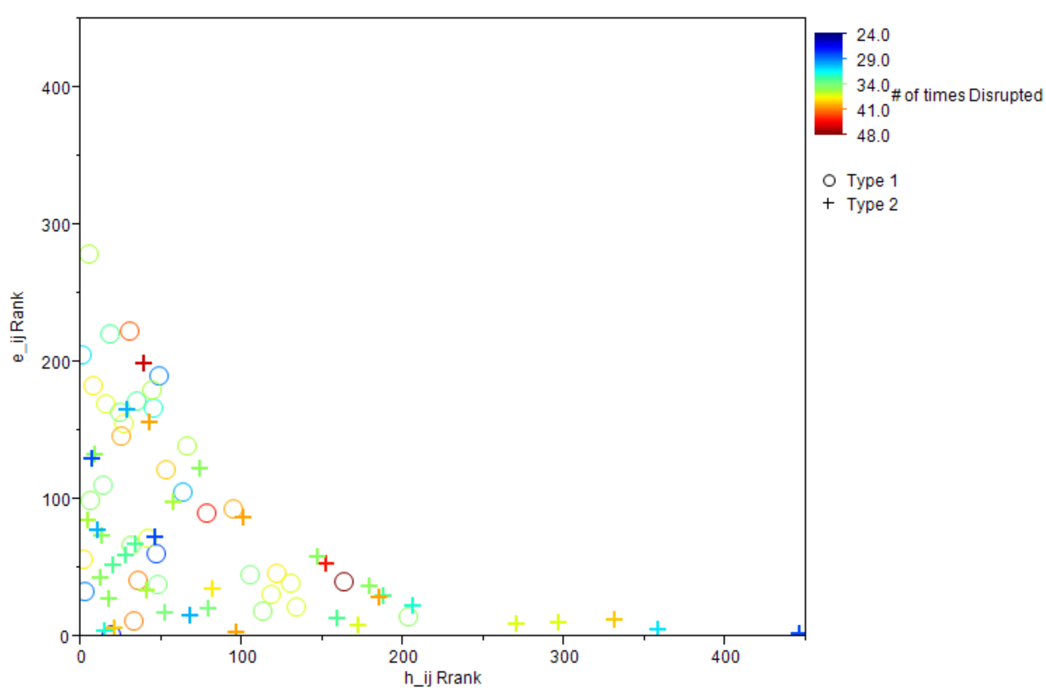


Figure 7. The top 75 arcs according to v_{ij} displayed graphically.

To verify that v_{ij} was correctly calculating which arcs are most vital, we ran another set of scenarios and instead of random disruption we took a greedy approach. A greedy approach represents an informed decision maker who knows which arcs are most vital to the network and chooses a percentage to disrupt. To implement a greedy approach, we selected a percentage of the top most vital arcs according to v_{ij} to disrupt. We chose five percentages 10%, 20%, 30%, 40%, and 50% which allowed us to compare with those scenarios that were randomly disrupted. Figure 8 displays the degree of disruption along the horizontal axis and the average change in value on the vertical axis. The colors represent ΔE^k and ΔC^k for both a random disruption and a greedy disruption. Applying the greedy approach resulted in both ΔE^k and ΔC^k values higher then when the network was disrupted randomly and this was the case for each percentage damaged. This provided validation that the metrics correctly identified are the most vital arcs in the network.

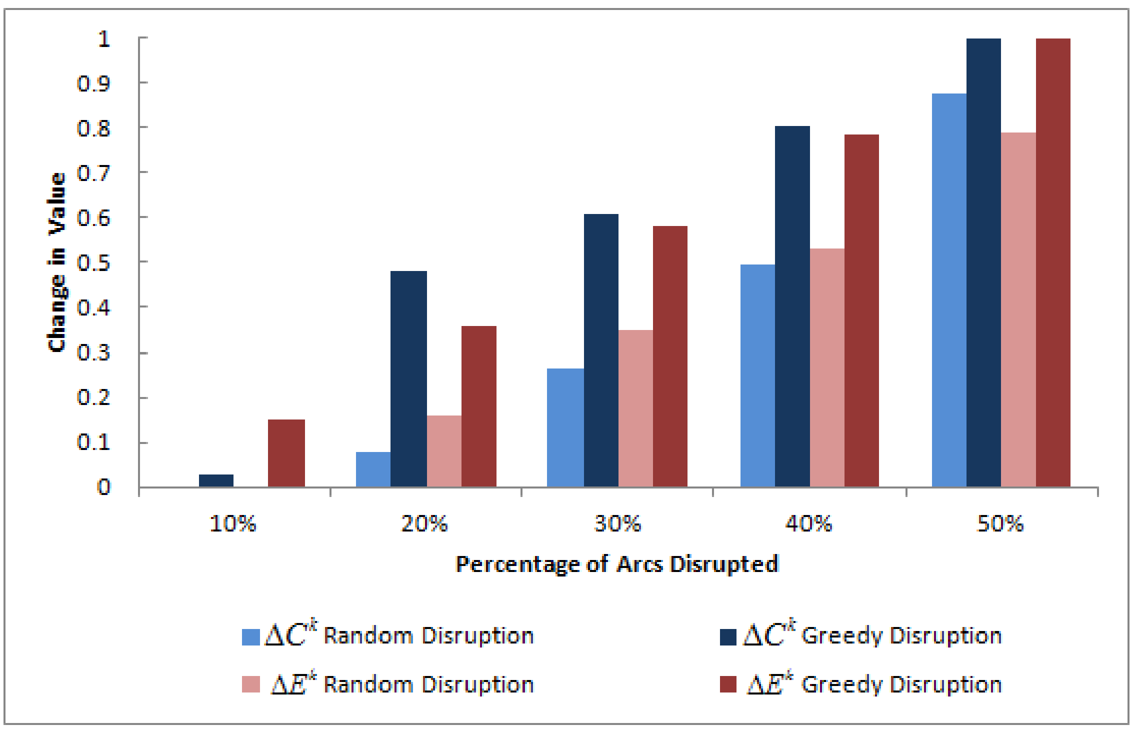


Figure 8. A bar graph displaying the average change in value at each degree of disruption while comparing random disruption approach to a greedy disruption approach for both ΔE^k and ΔC^k

Summary. In this chapter we applied the methodology from Chapter III to a multi-mode communication network in order to identify the most vital arcs within the network. We did this by using a new vitality metric and proceeded to validate the metric by comparing the 125 randomly generated computational tests to five greedy computational tests. These tests not only validated the model but also assigned a value to each arc which allowed us to rank order all the arcs within the network.

V. Conclusion

The distribution of information, specifically through a communication network is key to ensure the Air Force is successful during training, day to day operations, and most importantly during conflict. One of the main contributors of the Air Force for disseminating information is the Air Operating Center (AOC). The number one objective of the AOC is to enable the Air Force to be successful in it's mission. This objective can only be achieved by successfully disseminating information. If the AOC's communication network is significantly disrupted it could threaten their success and thus the success of the Air Force mission. The threat to a failed mission highlights the importance of knowing the network's topology, knowing how information will be sent in times of disruption, and identify it's vulnerabilities. Identifying a network's vulnerabilities begins with knowing what is important within the network. The objective of this paper was to identify the most vital components within a multi-mode communications network.

To meet this objective, we quantify the vital components within a network via a combination of a set-based efficiency and set-based cost efficiency measures. With both of these measures, we identify that a network operates to its full potential when arcs correctly interact together. Thus, we examine the network performance when sets of arcs are disrupted in the network, which we denote as scenarios. We expand the traditional efficiency measure (see Latora et al. [28]) and calculate the change in the network performance, specifically all pairs shortest path value, for each scenario. This value is then attributed to all arcs which were disrupted in the scenario, thus indicating that each arc contributed to the degradation in the network performance. The same concept is applied to a set-based cost efficiency measure. We calculate the change in the minimum cost flow for the disrupted network under each scenario in comparison to the best minimum cost flow. This value is then attributed to the arcs in each scenario. We combine these two arc-based values to form a set-based vitality metric. Based on this metric we rank order the most vital arcs within a network.

We validated the set-based vitality metric by first creating a model to replicate a multi-mode communications network. Using this model, we used minimum cost flow (MCF) and all pairs shortest path (APSP) problems to determine the best way to disseminate information. The values produced from solving the MCF and APSP problems were then used in our two new set-based vitality measures. The examination of the interaction and validation of the model was done by performing 125 computational experiments on different weighted scenarios. To capture the characteristics of both metrics they were normalized and summed creating what we define as our set-based vitality metric. This metric produced a value that achieves the end goal of providing a decision maker with a single rank ordered list of the arcs that are most vital to the network.

After completing all computational tests, we analyzed the results and reported our findings. The first finding was the degree of disruption had an impact on both the change in efficiency (ΔE^k) which had a linear change and the change in cost efficiency (ΔC^k) which had a greater percentage of change when the degree of disruption was greater than 30% disruption than when the disruption is less than 30%. The second finding was the impact of not applying an appropriate weight to scenarios and how it may skew the ranking of which arcs are most vital. It was also apparent that whether the arc was Type 1 or 2 did not exclude it from being in the top 75 most vital arcs. The final finding came from the validation of the vitality metric, v_{ij}. We compared the change in value for ΔE^k and ΔC^k when the disruption was random and when the disruption was the most vital arcs. This comparison was done for each degree of disruption and in every case for both ΔE^k and ΔC^k the change in value was greater when the most vital arcs were disrupted.

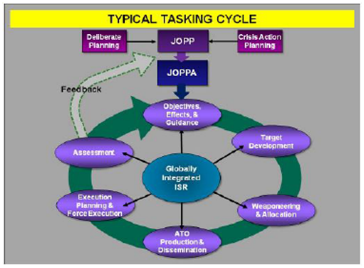
For future study, we propose that the methodology proposed in Chapter III and demonstrated in Chapter IV should be applied to a real world communication network as well as other network types. A shortcoming with this methodology is that it may not be applicable to all networks. Research needs to be conducted to determine the types of networks the methodology could be applied toward. Additional research should also be done on applying weights to the different metrics in the event one of the metrics is more important based on the network evaluated. Experiments should be conducted to evaluate whether or not other metrics could be incorporated to determine most vital components, such as graph entropy and largest component.

Determining the Most Vital Arcs Within a Multi-Mode Communication Network Using Set-Based Measures

Capt Christopher Hergenreter
Advisor: Dr. Sarah G. Nurre, PhD
Reader: Dr. Jeffery D. Weir, PhD
Department of Operational Sciences (ENS)
Air Force Institute of Technology

Background:

- In the last fifty years the way Air Force disseminates information has changed
- This change is to increase the speed and the security of the information disseminated
- AOC: constrained by time and security



Contributions:

- Modelled multi-mode communication network
- Created a DSS to do the following:
  - Determined which mode to disseminate information when the network is disrupted
  - Identified the most vital arcs by capturing how arcs interact
- Validated Vitality Metric

Literature Review:

- All Pairs Shortest Path Problem (APSP)
- Minimum Cost Flow Problem (MCF)
- Efficiency Equation

AIR FORCE INSTITUTE OF TECHNOLOGY

Approach:

- Model a multi-communication network
- Solve for APSP problem
- Calculate set-based efficiency
- Solve MCF problem
- Calculate set-based cost efficiency
- Normalize both efficiency values
- Sum normalized efficiency values for final vitality metric
- Rank order arcs according to vitality metric

Set-Based Efficiency:

$$E(G) = \frac{1}{|N|(|N|-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}$$

$$\Delta E^k = E(G) - E(G^k)$$

$$e_{ij} = \frac{1}{\sum_{k=1}^{|K|} (1-\beta_{ij}^k)} \sum_{k=1}^{|K|} (1-\beta_{ij}^k) \Delta E^k w^k$$

Set-Based Cost Efficiency:

$$\Delta C^k = \sum_{(i,j)\in A^k} c_{ij} x_{ij}^k - \sum_{(i,j)\in A} c_{ij} x_{ij}$$

$$h_{ij} = \frac{1}{\sum_{k=1}^{|K|} (1-\beta_{ij}^k)} \sum_{k=1}^{|K|} (1-\beta_{ij}^k) \Delta C^k w^k$$

Normalized Efficiency Values:

$$\widehat{e_{ij}} = \frac{e_{ij} - \min_{(i,j)\in A} e_{ij}}{\max_{(i,j)\in A} e_{ij} - \min_{(i,j)\in A} e_{ij}}$$

$$\widehat{h_{ij}} = \frac{h_{ij} - \min_{(i,j)\in A} h_{ij}}{\max_{(i,j)\in A} h_{ij} - \min_{(i,j)\in A} h_{ij}}$$

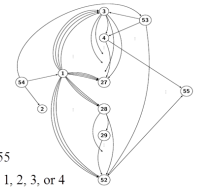
Final Vitality Metric

$$v_{ij} = e_{ij} + h_{ij}$$

Network:

- $|N| = 52$
- $|A| = 752$
- $s = 54$
- $\tau = 55$
- dummy = 53
- $b_s = 50$
- $b_\tau = -50$
- $u_{ij} = 50$, where $j \neq 55$
- c_{ij} = varies between 1, 2, 3, or 4

where $i \neq 54$ or $j \neq 55$



Scenarios:

- 125 Randomly generated scenarios

Percentage	Total Arcs	Total Damaged	Weight
0%	752	0	0
10%	752	75	1/75
20%	752	151	1/151
30%	752	226	1/226
40%	752	301	1/301
50%	752	376	1/376

Table 1. Values associated with each degree of damage

Change in Efficiency:

- Linear change for efficiency
- Non-Linear change for cost efficiency

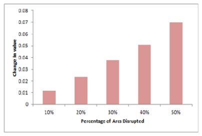


Figure 3. The average ΔE^k value as the percentage of disrupted arcs increases.

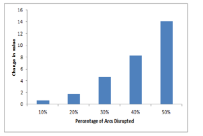


Figure 4. The average ΔC^k value as the percentage of disrupted arcs increases.

Most Vital:

- Top 75 most vital arcs according to the final vitality metric
- Type 1 and Type 2 arcs surfaced as most vital

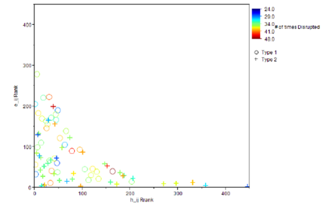


Figure 7. The top 75 arcs according to v_{ij} displayed graphically.

Random Disruption vs. Greedy Disruption:

- Greedy Disruption caused a higher rate of change for both efficiency measures

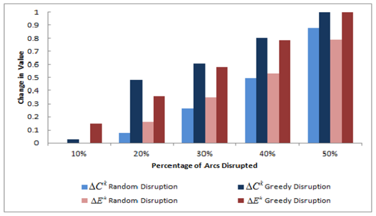


Figure 8. A bar graph displaying the average change in value at each degree of disruption while comparing random disruption approach to a greedy disruption approach for both ΔE^k and ΔC^k

Conclusions:

- Degree of disruption caused change in efficiency
- The type of arc did not affect whether or not it was most vital
- Vitality metric identified most vital arcs

Bibliography

1. U.S. Joint Chiefs of Staff, *Joint Mobilization, Joint Publication 3-30*. Washington, DC, Feburary 2014. `http://www.dtic.mil/doctrine/new_pubs/jp3_30.pdf`.

2. R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows: Theory, algorithms, and applications*. Englewood Cliffs, New Jersey: Prentice-Hall, 1993.

3. G. A. Klunder and H. N. Post, "The shortest path problem on large-scale real-road networks," *Networks*, vol. 48, no. 4, pp. 182–194, 2006.

4. A. Goldberg and R. Tarjan, "Solving minimum-cost flow problems by successive approximation," in *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pp. 7–18, ACM, 1987.

5. Y. Dinic, "An algorithm for the solution of the max-flow problem with the polynomial estimation," *Doklady Akademii Nauk SSSR*, vol. 194, no. 4, pp. 1277–1280, 1970.

6. J. B. Orlin, "A faster strongly polynomial minimum cost flow algorithm," *Operations research*, vol. 41, no. 2, pp. 338–350, 1993.

7. R. Kolisch and C. Heimerl, "An efficient metaheuristic for integrated scheduling and staffing IT projects based on a generalized minimum cost flow network," *Naval Research Logistics*, vol. 59, no. 2, pp. 111–127, 2012.

8. R. Dewil, P. Vansteenwegen, D. Cattrysse, and D. Van Oudheusden, "A minimum cost network flow model for the maximum covering and patrol routing problem," *Technical Report*, 2014. Available at: `https://lirias.kuleuven.be/bitstream/123456789/469354/1/patrolroutingworkingpaper.pdf`.

9. L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing electricity cost: Optimization of distributed internet data centers in a multi-electricity-market environment," in *Proceedings from IEEE INFOCOM*, pp. 1–9, 2010.

10. W. Ip and Q. Wang, "Resilience and friability of transportation networks: Evaluation, analysis and optimization," *IEEE Systems Journal*, vol. 5, no. 2, pp. 189–198, 2011.

11. T. D. O'Rourke, "Critical infrastructure, interdependencies, and resilience," *The Bridge*, vol. 37, no. 1, pp. 22–29, 2007.

12. C. J. Colbourn, "Network resilience," *SIAM Journal on Algebraic Discrete Methods*, vol. 8, no. 3, pp. 404–409, 1987.

13. M. A. Janssen, Ö. Bodin, J. M. Anderies, T. Elmqvist, H. Ernstson, R. R. McAllister, P. Olsson, and P. Ryan, “Toward a network perspective of the study of resilience in social-ecological systems,” *Ecology and Society*, vol. 11, no. 1, p. 15, 2006.

14. M. Shinozuka and S. E. Chang, “Evaluating the disaster resilience of power networks and grids,” in *Modeling spatial and economic impacts of disasters*, pp. 289–310, Springer, 2004.

15. J. P. G. Sterbenz, E. K. Cetinkaya, M. A. Hameed, A. Jabbar, S. Qian, and J. P. Rohrer, “Evaluation of network resilience, survivability, and disruption tolerance: analysis, topology generation, simulation, and experimentation,” *Telecommunication systems*, vol. 52, no. 2, pp. 705–736, 2013.

16. K. Barker, J. E. Ramirez-Marques, and C. Rocco, “Resilience-based network component importance measures,” *Reliability Engineering and System Safety*, vol. 117, pp. 89–97, 2013.

17. H. R. Shahriari, R. Sadoddin, R. Jalili, R. Zakeri, and A. R. Omidian, “Network vulnerability analysis through vulnerability take-grant model (VTG),” in *Information and Communications Security*, pp. 256–268, Springer, 2005.

18. A. Dwivedi, X. Yu, and P. Sokolowski, “Analyzing power network vulnerability with maximum flow based centrality approach,” in *8th IEEE International Conference on Industrial Informatics (INDIN)*, pp. 336–341, 2010.

19. Å. J. Holmgren, “Using graph models to analyze the vulnerability of electric power networks,” *Risk analysis*, vol. 26, no. 4, pp. 955–969, 2006.

20. A. H. Dekker and B. D. Colbert, “Network robustness and graph topology,” in *Proceedings of the 27th Australasian conference on Computer science-Volume 26*, pp. 359–368, Australian Computer Society, Inc., 2004.

21. L. C. Freeman, S. P. Borgatti, and D. R. White, “Centrality in valued graphs: A measure of betweenness based on network flow,” *Social networks*, vol. 13, no. 2, pp. 141–154, 1991.

22. M. Ventresca and D. Aleman, “Network robustness versus multi-strategy sequential attack,” *Journal of Complex Networks*, 2014.

23. M. O. Ball, B. L. Golden, and R. V. Vohra, “Finding the most vital arcs in a network,” *Operations Research Letters*, vol. 8, no. 2, pp. 73–76, 1989.

24. D. M. Scott, D. C. Novak, L. Aultman-Hall, and F. Guo, “Network robustness index: A new method for identifying critical links and evaluating the performance of transportation networks,” *Journal of Transport Geography*, vol. 14, no. 3, pp. 215–227, 2006.

25. Y. Dheenadayalu, B. Wolshon, and C. Wilmot, "Analysis of link capacity estimation methods for urban planning models," *Journal of transportation engineering*, vol. 130, no. 5, pp. 568–575, 2004.

26. R. D. Wollmer, "Some methods for determining the most vital link in a railway network," *RAND Memorandum*, 1963. `http://www.rand.org/pubs/research_memoranda/RM3321.html`.

27. D. H. Ratliff, T. G. Sicilia, and S. H. Lubore, "Finding the n most vital links in flow networks," *Management Science*, vol. 21, no. 5, pp. 531–539, 1975.

28. V. Latora and M. Marchiori, "How the science of complex networks can help developing strategies against terrorism," *Chaos, solitons & fractals*, vol. 20, no. 1, pp. 69–75, 2004.

29. D. L. Alderson, G. G. Brown, M. W. Carlyle, and L. Anthony Cox, "Sometimes there is no "most-vital" arc: Assessing and improving the operational resilience of systems," *Military Operations Research*, vol. 18, no. 1, pp. 21–37, 2013.

30. C. T. Allen, "Air tasking order dissemination: Does it get the job done?," 2002. `http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm`.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
26–03–2015	Master's Thesis	SEP 2013 — MAR 2015

4. TITLE AND SUBTITLE

Determining the Most Vital Arcs Within
a Multi-Mode Communication Network Using Set-Based Measures

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Hergenreter, Christopher A., Captain, USAF

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/ENS)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-MS-15-M-131

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

Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

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

Technology has dramatically changed the way the military has disseminated information over the last fifty years. The Air Force has adapted to the change by operating a network with various ways to disseminate information. The Air Operating Center (AOC) is a large contributor to disseminating information in the Air Force. When the standard mode of sending information is disrupted, the AOC seeks both alternative ways available to send information and long term approaches to decrease vulnerability of its standard procedures. In this thesis, we seek to identify and quantify the most vital components within a multi-mode communications network via a combination of a set-based efficiency and set-based cost efficiency measures that utilize the all pairs shortest path (APSP) problem and minimum cost flow (MCF) problem. We capture the phenomenon that network components must work together to provide flow by examining how the network performs when sets of arcs are disrupted. We run 125 different computational experiments examining varying degrees of damage experienced by the network. From these results, we deduce insights into the characteristics of the most vital arcs in a multi-mode communication network which can inform future fortification decisions.

15. SUBJECT TERMS

Air Operating Center (AOC), Air Tasking Order (ATO), All Pairs Shortest Path (APSP), Minimum Cost Flow (MCF), Set-Based Efficiency, Set-Based Cost Efficiency, Network, Most Vital, Communication Network

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. Sarah G. Nurre (ENS)
					19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	51	(937) 255-3636 x4319 sarah.nurre@afit.edu