AFIT/GOR/ENS/94M-12

(1)

ASSESSING THE VULNERABILITY OF MULTI-COMMODITY NETWORKS WITH FAILING COMPONENTS

THESIS

Alan R. Robinson
Captain, USAF

AFIT/GOR/ENS/94M-12

Approved for public release; distribution unlimited

AFIT/GOR/ENS/94M-12

ASSESSING THE VULNERABILITY OF MULTI-COMMODITY NETWORKS WITH FAILING COMPONENTS

THESIS

Presented to the Faculty of the Graduate School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Operations Research

Alan R. Robinson, B.A. and M.S.

Captain, USAF

March, 1994

Approved for public release; distribution unlimited

THESIS APPROVAL

STUDENT: Alan R. Robinson, Captain, USAF CLASS: GOR-94M

THESIS TITLE: Assessing the Vulnerability of Multi-Commodity Networks with Failing Components

DEFENSE DATE: 23 February 1994

Committee:	Name/Title/Department	Signature
Advisor	YUPO CHAN Professor of Operations Research Department of Operational Sciences	Yupo Chan
Reader	DENNIS C. DIETZ, Lt Col, USAF Assistant Professor of Operations Research Department of Operational Sciences	Dennis C. Dietz

Preface

This research investigated an analytical approach to assess flow disturbance, or "compromise," in multi-commodity communication networks where network arcs are subject to random failure. This effort is the result of six months of research in the fields of network reliability, origin-destination matrix estimation, and compromise programming.

The successful completion of this thesis can be attributed to many individuals. First, I owe many thanks to my thesis advisor, Dr. Yupo Chan, for his expert technical guidance and his insightful suggestions. Second, I must thank my thesis reader, LtCol Dennis Dietz, for keeping the edits at a minimum and for keeping me honest as well. Third, I thank my thesis sponsors for proposing the research project and for making the time to answer many, many questions. Finally, I thank my wife, Laura, both for her invaluable advice as the resident WordPerfect expert and for ensuring that, at all times, I knew where my priorities lie.

Alan R. Robinson

Table of Contents

Page

List of Figures

List of Tables

Table		Page

AFIT/GOR/ENS/94M-12

Abstract

This research proposes an analytical approach for assessing flow disturbance, or "compromise," based on limited sampling of arc flow information in multi-commodity, or multiple origin-destination (O-D), networks with failing arcs. There were three objectives established for this research. The first objective was to bound the expected flow, given the arcs fail with certain probabilities, which was accomplished by reviewing current approaches for single-commodity networks and extending the results to the multi-commodity case. The second objective was to determine the best placement of flow monitors to obtain the most accurate estimates of O-D pair volumes. This was accomplished using a multi-criteria approach for defining and evaluating all possible monitor placement strategies satisfying monitor availability. The O-D pair volumes were estimated using the l_p-norm metric for varied levels of p. The final objective was to define a compromise metric providing confident assessments on the occurrence of "compromise." This was accomplished using simple regression techniques to generate confidence intervals around the expected flow for each O-D pair. The approach proposed in this research is provided as an initial look into "compromise" assessment based on limited network information.

ASSESSING THE VULNERABILITY OF MULTI-COMMODITY NETWORKS WITH FAILING COMPONENTS

I. Introduction

This chapter begins with a brief background on the study of vulnerability of communication networks. Next, the research problem and objectives are stated, followed by the assumptions which are carried throughout the study. Also, the scope of the research is discussed, both for the current study as well as for potential follow-on efforts. Finally, a brief outline of the different sections of this study is presented.

1.1 Background

Considerable research over the past several decades has been accomplished in the field of communication network design where the objective is the efficient design of network structures maximizing transmission throughput at minimal cost [1, 8, 16, 37]. The level of research in this field does not appear to be dwindling given the increasing complexity of global communication systems and the achieved efficiencies in large network solution algorithms.

The efficiency of a communication network design can be assessed using various measures (e.g. throughput, cost, transmission delay, reliability, and vulnerability) [22, 25, 35]. Focusing on vulnerability measurement, researchers have generally defined vulnerability as either the effort needed to disconnect a network [3, 6, 21] or "the effort required to disrupt the maximum amount of traffic of the network by disconnecting a minimum number of links" [19]. In essence, research has focused on developing efficient analytical and simulation models to pin-point the weakest part (or "Achilles Heel") of a network. Obvious beneficiaries of this research include communication carriers and the military community who have used the results to better design functional, reliable, and survivable communication networks. An important observation is that the research has been primarily geared toward assisting *designers* of communication networks.

Vulnerability measurement of a communication network from a *user* perspective appears somewhat less developed. The user perspective, as perceived for this study, is interested in assessing the vulnerability of its transmissions within an operational communication network. A potentially useful measure is the vulnerability of communication transmissions to unexpected disruptions (e.g. unexpected rerouting or tampering) [13]. Whereas the designer focuses on identifying and improving the weakest link(s), the user (who must assume a more passive role given the network configuration already includes the weak links) must focus on identifying transmission disruptions. Given the apparent lack of a precedent in the literature, the following definition of vulnerability is used throughout this study: ***vulnerability*** is a measure of risk of the susceptibility of communication traffic (or flow), between a specified origin and destination, to disruptions over and beyond "expected" disruptions due to carrier management of traffic. Any observed disruptions exceeding "expected" bounds define a compromised network flow configuration.

1.2 Research Problem

Generally stated, the problem is one of determining if a user can assess the vulnerability of its communications to disruptions. The approach is to place a limited number of costly monitors at specific points within a network to monitor communication flow patterns, and based on sample measures of flow deviations (i.e. between expected and observed flows) assess the vulnerability of the user's communications [13]. The problem itself is twofold:

(1) separate the "unexpected" disturbance (i.e. the disturbance we want to detect) from the existing, or built-in, disturbance due to normal operations; and

(2) determine the placement of monitors such that the sampled information provides the most accurate representation of the network flow pattern.

The problem is further complicated in that network components (i.e. nodes and links) are not totally reliable and will fail randomly, and multiple users within the same network are competing for common assets which may lead to congestion and rerouting.

1.3 Research Objectives.

The following objectives are identified to examine and assess the feasibility of the above stated problem:

<u>Objective #1:</u> Establish a steady-state flow through a multi-commodity

communication network with failing components that minimizes carrier costs and satisfies user throughput demands. This steady-state system represents the expected flow pattern in an uncompromised scenario.

Objective #2: Determine monitor placement strategies that maximize the likelihood that the origin-destination (O-D) demands estimated from partial sampling of network components accurately represent the actual O-D demands in the original network.

Objective #3: Define a metric for determining if a flow disturbance, or "compromise," has occured, and evaluate the compromise assessment for varying O-D estimation perspectives.

1.4 Assumptions

The following assumptions are carried throughout this study:

- The basic network configuration is a multi-path network with multiple O-D pairs (or commodities) representing the users of the system.
- The network structure assumes a circuit-switched design rather than a packet-switched design; in other words, once a route is established it remains fixed or dedicated.
- The routing algorithm that determines the path(s) taken by each O-D pair is a simple minimum path algorithm (e.g. minimum distance).
- The flow control algorithm that limits traffic on arcs to avoid excessive congestion is a function of constant arc capacities.

- Flow disturbance, or "compromise," is exhibited as an increase in flow along a particular O-D pair's path(s) [13].
- Network components are either in a state of operation or failure with known probabilities (given that nodes can be modeled as arcs, the problem is further simplified by assuming all nodes are 100% reliable).

1.5 Scope of Research

1.5.1 Current Study. This study is partitioned into four distinct problems: first, measure the expected flow pattern in a multi-commodity network subject to random arc failures; second, determine the optimal location of flow monitors within the network subject to conflicting objectives; third, measure the estimated O-D demands based on a sampling or subset of arc flow observations; and fourth, define a measurement tool to enable probabilistic assessments of flow "compromise." These problems can be solved either analytically or by simulation. This study uses an analytical approach to solving these problems using mathematical programming techniques. The advantages of this approach are that an analytical solution can yield as useful a result as a more complicated simulation, especially under the simplified assumptions of this study, and the feasibility of the approach can be readily assessable from this analytic result.

The network structure itself represents a circuit-switched design rather than a packet-switched design as assumed earlier. A circuit-switched network reduces a significant amount of complexity by establishing virtual routes (fixed routes) and is the more common network structure used by carriers for voice communication. A packet-switched network introduces a great deal more variability since transmission routing is dynamic and is designed primarily for data communication.

1.5.2 Follow-on Studies. Evidently, the scope of the current study is at best limited from a practical perspective given the simplifying assumptions. Its purpose, nevertheless, is crucial for determining the overall feasibility of the stated approach. Assuming the results of this study are favorable, the scope of the follow-on research can be readily expanded to encompass the realities and complexities of real systems.

The next logical step is perhaps to either develop a more complex analytical model or develop a discrete-event simulation model of a realistic communication network, based on either a circuit-switched or packet-switched design. This model can be progressively enhanced to incorporate (1) larger sized networks, (2) variable arc capacities as a function of time, (3) continuous arc reliability functions (to model the more realistic case of partial versus total failure), (4) preferential route assignments (based on some message priority heuristic), and (5) improved network performance measures based on (i) average traffic loads every hour (rather than the limiting overall expected value used in this study), or (ii) each individual call (or entity) accessing the system.

1.6 Summary

This chapter presented an overview of the focus of this study. It included a brief discussion on past studies of vulnerability measurement, a suggested interpretation of the term *vulnerability* as used in this study, the problem statement, and the objectives of the study. Also included were the assumptions which mainly served to simplify the problem so it could be handled using an analytical solution approach. Finally, the scope of the research was presented for both the current study and for potential follow-on efforts.

Chapter II, Literature Review, presents a comprehensive review of the literature as it regards the measurement of expected flow in multi-commodity networks with failing arcs, the location of flow monitors based on potentially conflicting objectives, and O-D demand estimation based on limited sampling of network flows.

Chapter III, Model Formulation, describes the different models developed to meet the objectives stated in this chapter and presents an overall solution algorithm.

Chapter IV, Compromise Measurement, describes the compromise metric used to assess flow disturbance.

Chapter V, Case Study, introduces the case network and applies the overall solution algorithm to a set of possible monitor location strategies. This chapter also evaluates the performance of these strategies in assessing compromise.

Chapter VI, Conclusions and Recommendations, provides a summary of the approach and its conclusions, and recommends appropriate extensions for follow-on efforts.

II. Literature Review

This chapter presents a review of the literature applicable to the estimation of flow in stochastic networks and the assessment of compromise using multiple-criteria decision making (MCDM) principles. First, we present notation along with a basic review of network flow models. Next, we extend this review to networks with failing components and present existing estimation methods. The next section provides an overview of MCDM principles and demonstrates how they apply to monitor location problems. Finally, we review current methods for estimating O-D demands from partial network information.

2.1 Minimal Cost Network Flow Representation

2.1.1 Network Parameters. Let $G(N, A)$ be a directed graph, or network, defined by a set of nodes N, $\{i = 1, 2, \ldots, m\}$, and a set of arcs $A = \{(i, j), (k, l), \ldots, (s, t)\}$. The network consists of m distinct nodes and n directed arcs collectively referred to as network components. With each node i in G is associated an external flow parameter b_i which is positive, negative or zero depending on whether node i is, respectively, a source (supply), sink (demand), or intermediate node. With each arc(i, j), directed from node i to node j, is associated both a flow capacity parameter, u_{ij}, which constrains the maximum feasible flow along arc(i, j), and a unit flow cost parameter, c_{ij}.

In a telecommunication network the nodes may represent users and switching centers within the network and the arcs may represent communication facilities such as access lines and trunks interconnecting the switching centers [33: 3]. The external flow parameter b may represent the total number of simultaneous calls, or flow units, input to the system from a specific user node. The arc capacities may represent the maximum number of simultaneous calls handled by each communication facility and expressed in the same unit measurement as the external flow. Finally, the arc cost parameter may represent various criteria. For instance, the cost function may be expressed as delay in traversing an arc, as arc distance (e.g. miles), or as a communication facility usage charge [35: 362]. Since it was assumed in the previous chapter that the flow routing algorithm was based on the shortest path, arc costs will be expressed as distance measures.

2.1.2 Network Flow Variables. The objective in a minimum cost network flow problem (MCNFP) is to determine the optimal value of flow variables that satisfy both the flow conservation and arc capacity constraints at minimal cost [4: 420]. The flow variables can be expressed as either arc or path flow variables. The choice of modeling either flow variable dictates the form of the incidence matrix H used to represent the network structure. If we use the arc flow convention we must represent the network in a m x n node-arc incidence matrix where each element h_{ik} of H is defined as +1 if arc i is directed ***at*** node k, -1 if arc i is directed ***from*** node k, and 0 otherwise. By constrast, if we use the path flow convention we must represent the network as a n x p arc-path incidence matrix where p represents the number of possible paths connecting the source and sink nodes, and where each element h_{ij} of H is defined as +1 if arc i lies on path j,

$\{j = 1, \ldots, p\}$, and 0 otherwise.

To assist in deciding which convention to use we make use of the assumption stated in Chapter I that the system modeled is a voice network. In most cases, the manner in which communication facilities, or arcs, are shared among users in ordinary telephone calls is determined by the circuit switching technique. This technique establishes a dedicated path through the network connecting the source (origin) node and sink (destination) node for the entire duration of the call [33: 2-3; 35: 95-97]. This process is one reason why the arc-path formulation for the MCNFP is used.

2.1.3 Network Flow Model. In its simplest from the MCNFP is formulated as a single-commodity flow model since only one type of flow is required between one origin node and one destination node, or one origin-destination (O-D) pair. In many situations, however, more elaborate flow models are required to capture the complexities of practical networks. Examples include flowing multiple commodities, or flow types, along common arcs, and modeling multiple origin and/or multiple destination nodes. A comprehensive review of multi-commodity, nonsimultaneous, and multiterminal network flow models is presented in [25: 315-317]. The class of multi-commodity network flow models is of particular interest in this research for it allows communication between many distinct O-D pairs to occur simultaneously throughout the network.

The general deterministic minimum cost multi-commodity network flow (MCMCNF) problem formulated on path flow variables is presented in Figure 2.1.

Figure 2.1. Deterministic MCMCNF Model

Objective Function

Minimize

$$Z = \sum_{k \in K} \sum_{p \in P^k} c_p^k f_p^k$$

subject to

$$\sum_{p \in P^k} f_p^k = b^k \qquad \forall\ k \in K$$

$$\sum_{k \in K} \sum_{p \in P^k} h_{ij}^{kp} f_p^k \le u_{ij} \qquad \forall\ (i,j) \in A$$

$$f_p^k \ge 0 \qquad \forall\ k \in K\ ,\ p \in P^k$$

where A = set of directed arcs in the network
K = set of distinct O-D pairs (or commodities)
P^k = set of paths connecting O-D pair k
b^k = total flow input to the network by O-D pair k
c_p^k = path cost equal to the sum of all arc costs lying on path $p \in P^k$
f_p^k = path flow on the p^{th} path of the set of candidate paths P^k
h_{ij}^{kp} = 1 if arc(i, j) lies on path $p \in P^k$; 0 otherwise
u_{ij} = flow capacity of arc(i, j)

The objective function is similar to that of the MCNF problem (single-commodity) except for the subset of paths which must be considered for each O-D pair. The first constraint represents the set of flow conservation constraints ensuring the sum of the path flow variables for each O-D pair equals the flow volume input to the network by the O-D pair. The second constraint represents the set of arc capacity constraints ensuring the sum of the path flow variables for all O-D pairs sharing a specific arc does not exceed the arc's capacity.

2.1.4 Solution Algorithms. Unlike single-commodity network flow problems, the capacitated MCMCNF problem does not have a completely special structure allowing it to be solved efficiently using the network simplex method. It is characterized by its set of "nice" constraints, represented by the flow conservation constraints partitioned into block angular and network structures, and "general" or "complicating" constraints, represented by the arc capacity constraints. Problems with this structure can be solved using the Dantzig-Wolfe decomposition algorithm [4: 320-349].

MCMCNF problems in the node-arc incidence matrix formulation can be solved fairly efficiently using commercial software packages such as SAS Netflow, which uses the Dantzig-Wolfe decomposition techniques; however, problems of this form which model specific O-D pairs can yield very large linear programs resulting in potential computational difficulties [15: 671]. Farvolden, et al [15], suggest the arc-path formulation based on the observation that not all paths need be enumerated since a smaller subset of paths will dominate the majority.

2.2 *Network Models With Failing Components*

2.2.1. Component Reliability. As described earlier, network components are represented by nodes and arcs, which in this research are assumed to be in either a complete state of operation or failure with probability r and $(1 - r)$ respectively [1: 1080; 37]. A network with this characteristic is sometimes referred to as a *stochastic binary system* [3: 154; 34: 101]. For simplicity, and without loss of generality, we assume the nodes are perfectly reliable. As demonstrated in [37: 10-11] a node subject to random failure can be modeled as two nodes connected by a "dummy arc."

It is further assumed that the arc failures are independent of each other [3, 34, 37] yielding, for each path $p \in P^k$ connecting a specific O-D pair k, a path survival probability R_p^k represented by the product of the individual arc survival probabilities connected in series from origin to destination [8, 33: 43]:

$$R_p^k = \prod_{(i,j)\in A_p^k} r_{ij}$$

2.2.2 Expected Minimum Cost Network Flow. Determining the exact flow through a network subject to random component failures, whether it be a minimum cost or maximum flow, single- or multiple-commodity problem, "is classified as NP-hard which simply means that the computational effort grows exponentially with the number of stochastic components" [37]. Yim [37] describes the computational effort required to compute the exact value of the expected maximum flow through a small, single-commodity network, which can be generalized for the MCMCNF problem. This example clearly illustrates the computational complexity inherent to networks of interesting size.

The expected value of the MCMCNF problem formulated on path flow variables is presented in Figure 2.2 [28: 766; 34: 8].

2.2.3 Solution Approaches. In this section we present a few approaches to solving for the exact value of flow through a network. The form of the objective function and network is not necessarily pertinent to this discussion.

Shier [34: 101-117] proposes an approximation to the exact value for a binary system which "involves generating a relatively small number of [failure] states that encompass a relatively large proportion of the total probability." The method is one of "generating the [failure] states of a system in order of nonincreasing probability" such that a "maximum coverage of the state space (in terms of probability) [is] obtained for a specified number of generated states." For example, in a network consisting of 25 arcs with *equal* survival probabilities at 0.90, Shier's method reduces the failure-state space to be considered from 2^{25} = 33,554,432 states to a mere 15,100 states (or 0.045 percent) to achieve a specified coverage probability of 0.90. His results suggest that the proportion of states needed to be evaluated significantly decreases as arc survivability rates increase; for a communication network, it is common that arc survivability rates tend to unity [28, 34].

Figure 2.2. Expected Value of MCMCNF

Objective Function

$$\text{E(MCMCNF)} = \sum_{s \in S} Z(s)P(s)$$

where S = set representing all failure states s, $\{s = 1, \ldots, 2^{n'}\}$, where any given state represents a combination of operating and failed arcs

n' = number of arcs having survival probability strictly less than 1 ($n' \leq n$)

$Z(s)$ = objective function value of the deterministic MCMCF problem when in state s

$P(s)$ = probability of failure state s and is obtained as follows:

$$\prod_{(i,j) \in A_s^o} r_{ij} \prod_{(i,j) \in A_s^f} (1 - r_{ij})$$

where A_s^o = subset of arcs in A operating in state s

A_s^f = subset of arcs in A failed in state s

Based on the last observation that arc reliabilities tend to unity in practical communication networks, a second approach assumes that it is highly unlikely that more than one arc fail simultaneously, assuming they fail independently of each other [25]. Failure-state enumeration for this method is reduced to simply evaluating the $n' + 1$ combinations of failure states with at most one failed arc.

A third approach was implemented by Bailey [2] who actually simulated a single-commodity, maximum flow network to obtain the expected value of the maximum flow. Currently, an efficient solution algorithm for determining the exact value of the expected flow, short of simulation, does not appear forthcoming [37: 13].

2.2.4 Bounds For The MCMCNF Problem. Prior research of binary systems has mainly focused on the development of lower and upper bounds for the single-commodity, maximum flow case [8, 28, 37, 38]. Aneja and Nair [1] did extend their work to encompass multi-commodity networks, but the focus was to develop an algorithm to compute the expected maximum flow. The common purpose of this prior research has been the development of relatively tight bounds on the expected maximum flow through a network irresepective of any demand constraints. Our research, besides assuming a minimum cost objective which can easily be transformed to a maximization problem, adds complexity by requiring all demands between O-D pairs be satisfied. Essentially, the model resembles a multi-commodity transportation problem; however, the addition of arc reliabilities requires us to resort to bounding the expected minimum cost flow in the network. Proposed bounds are presented in Chapter III.

2.3 *Multiple Criteria Decision Making*

Most realistic decision problems can best be modeled as multiple criteria decisions [14, 30, 39]. An important characteristic of these models is that the criteria, or stated objectives of the decision maker (DM), are often in conflict. Given this conflict, it is usually not possible to obtain a single optimal solution such that all criteria are simultaneously optimized. More realistically, a compromise solution must be arrived at which satisfies the DM's revealed preference structure. This preference structure can be either stated up-front in the form of goals and solved independently of the DM, such as goal programming, or it can be applied interactively where the DM can influence the direction of the solution algorithm to obtain the best solution [39].

These approaches have been used to solve a variety of multi-criteria decision problems, the most prominent relevant to this course of study being facility location. The next sections will illustrate the multi-criteria location problem in greater depth.

2.3.1 Multi-Criteria Location Problems. Location problems, generally referred to as facility location problems, "concern the location of facilities to serve clients economically" [26: 7]. The problem is essentially one of determining the optimal placement of facilities such that the cost of placing a facility at a specific site is minimized while client demand for facility service is satisfied [26]. The decision variable denoting facility location is modeled as a binary variable $x_{ij} = 1$ if a facility is located at site ij (or in this context, arc (i, j), and $x_{ij} = 0$ otherwise. The demand variable, which is not considered in our approach, is generally modeled as a non-negative, continuous variable.

Examples of single-criteria locational problems are numerous. A recent study [5], for example, focused on optimally locating "discretionary service facilities" (e.g. automatic teller machines and gasoline service stations) so as to maximize the amount of intercepted flow, or consumers. However, the extent of the research for locational decisions involving multiple criteria appears somewhat less developed. According to some researchers [29, 30] facility location problems are inherently multi-criteria problems where a typical set of conflicting criteria would include cost (minimize) and service (maximize).

The general multi-criteria problem can be expressed mathematically as follows [29, 39]:

$$\text{maximize } f_i(x), \quad i = 1 \ldots q$$

$$\text{subject to} \qquad g_j(x) \leq 0, \quad j = 1 \ldots m$$

where $x \in$ X and X represents the decision space, $f_i(x)$ represents the criterion functions and $g_j(x)$ the constraints. Assuming the criterion functions are conflicting in nature, it may be difficult, if not impossible, to obtain a single optimal solution to all criteria. In this case, we search for the best possible solution, also referred to as efficient solutions based on the Pareto preference concept of "more is better", for each criterion function. By definition, a decision x^0 is a Pareto optimal or efficient solution for the set of criteria f_i if there exists no $x \in$ X such that $f_i(x) \geq f_i(x^0)$ for all i [29: 84; 39: 22].

2.3.2 Complexity of the 0-1 Multi-Criteria Problem. The difficulty in solving the above problem when expressed in the context of a location problem is twofold: (1) given the decision variables are binary 0-1 variables, as the number of possible sites (or arcs) increases, the complexity of the branch-and-bound enumeration algorithm grows exponentially [5: 203; 26: 114]; and (2) given that every *feasible* solution to the binary problem is an extreme point [26: 457], "using parametric analysis of the convex combination problem [is] of little use in the search for all efficient solutions" [29: 84].

The first difficulty indicates that this class of problems is considered NP-hard and, as such, can be very difficult to solve for medium to large problems (or networks). The fact that the number of possible solutions grows exponentially as problem size grows linearly is a well-known characteristic of the binary problem and many solution approaches have been proposed including implicit enumeration and cutting plane algorithms [24: 41-55, 81-86; 26: 456-464].

The second difficulty is inherent to the multi-criteria binary problem. In the general linear case, where the decision variables are continuous, we can apply Geoffrion's results [18] to determine the complete set of properly efficient solutions (i.e. set of efficient solutions minus the subset of improperly efficient solutions which allow unbounded tradeoffs [39: 29]) by parametric analysis of the following problem with respect to λ:

$$\text{maximize } \textstyle\sum_i \lambda_i f_i(x)$$

$$\text{subject to} \quad x \in X, \quad \lambda \geq 0$$

Ross and Soland [30: 313] note that the set of efficient and properly efficient solutions are identical when either the solution set is finite or the criteria functions are linear and the constraint set X is a linear polytope. For this research, then, we can generalize the case to simply the set of efficient solutions.

Continuing the discussion for the continuous variable case, parametric analysis of the above problem implies that the efficient solutions can be obtained by maximizing various convex combinations of the criterion functions. However, for the binary variable case, the decision space is discrete and the convexity assumption no longer applies [29]. Therefore, to determine the set of efficient solutions for the binary multi-criteria problem other solution approaches must be investigated.

2.3.3 Solution Approaches. Current solution approaches for the deterministic, multi-criteria problem are classified as methods involving value functions, efficient solution sets, and interactive algorithms [30: 308]. We introduce new notation in the following discussion where Y represents the outcome space, $y = f(x)$ represents the vector of q criterion functions introduced earlier, and y^1 represents the outcome for decision, or alternative, 1.

2.3.3.1 Value Functions. The use of value functions to solve the multi-criteria location problem assumes prior knowledge of the DM's underlying preference structure. A DM's value function, $v(y)$, is defined such that the DM prefers y^1 to y^2 if and only if $v(y^1) > v(y^2)$ [30: 312; 39: 96]. The optimal solution is obtained from the following problem [30: 312]:

$$\begin{aligned} &\text{maximize } v(f(x)) \\ &\text{subject to} \quad x \in X \end{aligned}$$

The difficulty with this approach is determining the value function v. Its construction is usually based on interviews with the DM and may require repeated sessions to ensure a consistent model is developed [39: 122].

2.3.3.2 Efficient Frontier. The efficient frontier is defined as the set of efficient solutions. Unlike the value function approach, the analyst needs little information regarding the DM's preference structure to determine the efficient frontier [30: 313]. In this case, it may be possible for the DM to determine the optimal solution simply by choosing one from the set presented to him.

The difficulty with this approach, as discussed earlier, is the discreteness of the feasible set. Rasmussen [29] presents a comparative study of solution methods used to establish a complete, or even partial, set of efficient solutions for the binary, multi-criteria problem. Yu [39] presents another set of solution methods using compromise programming to obtain the set of compromise, or efficient, solutions. This latter approach is implemented in this research and fully presented in Chapter III.

2.3.3.3 Interactive Algorithms. Interactive algorithms, unlike the prior two approaches, involve the DM and analyst working together in breaking down the solution space iteratively incorporating the DM's preferences. Ross and Soland [30: 313-314] warn that although this type of approach can usually lead to a quicker solution, some of the existing algorithms "will not, in practice, yield a final choice which is an efficient solution."

The obvious difficulty with this approach is having the luxury of a DM who will interact with the analyst. An interactive approach was not considered for this research, therefore a description of algorithms will not be presented. Refer to [12, 14, 29: 84-92; 30: 317-318; 39: 325-333] for more information on interactive algorithms as applied to multi-criteria problems.

2.4 Origin-Destination Matrix Estimation

A typical problem which occurs frequently in urban planning is the estimation of origin-destination (O-D) matrices based on limited information regarding the particular transportation networks under consideration [10, 11, 17, 32]. Proposed O-D estimation methods applied to transportation networks are extended in this research to simplified communication networks given the inherent similarities in network structures and parameters.

An O-D demand was defined earlier as the parameter b^k representing the total flow, or external flow, contributed to the network by O-D pair k. The O-D matrix, D, is defined as the matrix representing the proportion of flow b^k connecting O-D pair k on specific arcs [32: 441] and is referred to in later chapters as a routing matrix. The purpose of the estimation process, then, is to obtain an accurate estimate of the target demands b for all O-D pairs, denoted as the vector F, using as little information as possible from the underlying network [10: 1; 11: 27].

For the networks considered in this research, the only information available to the analyst is flow totals on particular arcs, namely those arcs where flow monitors are located. Some studies [10, 11] have considered introducing additional available information into the estimation process, such as trip distribution and "turning movement" information, to increase estimation accuracy without increasing the need for more flow sampling; although the results appear favorable, this additional information is not assumed known in this research.

The O-D estimation problem can be represented mathematically as the set of linear equations [10: 2]:

$$\mathrm{D\ F = V}$$

where D is an n'' x t routing matrix, where n'' is the number of arcs sampled ($n'' \leq n$) and t is the number of distinct O-D pairs. The entries of D, denoted as $d_{ij}^{\,k}$, represent the proportion of external flow b^k routed along arc (i, j). Since the networks considered can assume multi-path assignments for each O-D pair, the entries $d_{ij}^{\,k}$ can take any value between 0 and 1. In addition, F is a t x 1 vector representing the estimated demands, or estimated external flows, for all O-D pairs where each element of the vector is denoted as F^k. Finally, V is a n'' x 1 vector representing the observed flow totals on the sampled arcs.

A survey of optimization methods for obtaining the estimated O-D matrix is presented in [10, 11, 32]. The procedure, in general, is the inverse of obtaining the minimum cost network flow. In other words, given V, which is usually the variable in traditional network flow problems, the objective is to determine the estimated O-D demands, F, which is usually a given parameter in network problems. The concept of matrix inversion is the basis for the following optimization methods: Generalized Inverse, Entropy Maximization, and L_p-Norm Minimization.

2.4.1 Generalized Inverse Method. The generalized inverse method consists of an objective formulated as a quadratic penalty function measuring the deviation between the observed and estimated flow values on sampled arcs. The objective is minimized over all sampled arcs resulting in a generalized least-squares estimator given by:

$$\text{minimize } \Sigma_{ij} \left(V_{ij} - \Sigma_k d_{ij}^k F^k\right)^2$$

A peculiarity of this approach is presented in [10, 11] where the authors note that the dot-product of F, given by $\Sigma_k (F^k)^2$, is minimized also, resulting in a set of equalized O-D estimates, given by $F^k = \Sigma_k F^k / n''$, which does not necessarily minimize the least-squares estimator.

2.4.2 Entropy Maximization Method. The entropy maximization method uses an entropy penalty function which "has a theoretical justification based on principles of information theory" [32: 449]. A simplified form of the entropy maximization formulation is presented in Figure 2.3 [10: 4].

It is demonstrated in [23] that obtaining a satisfactory estimate of the O-D demands is indeed difficult given the non-linearity and lack of strict concavity of the objective function. In fact, a unique solution is not guaranteed. Further, it is shown that for a given $F_{tot} = \Sigma_k F^k$, the entropy maximization method yields a set of equalized O-D demand estimates identical to the generalized inverse method [10].

Figure 2.3. Entropy Maximization Model

Objective Function

Maximize

$$Z_{Entropy} = -\sum_k F^k(\log F^k - 1)$$

subject to

$$\sum_k d_{ij}^k F^k = V_{ij} \qquad \forall\ (i,j) \in A$$

$$F^k \geq 0 \qquad \forall\ k \in K$$

2.4.3 L_p-Norm Minimization Method. In [10] the authors propose an algorithm to improve the O-D estimation process based on the class of l_p-norm deviational measures. Besides simplifying the computational aspect of the problem, the use of l_p-norms allows the analyst to model deviations between target demands, b^k, and estimated demands, F^k, incorporating limited information regarding the DM's preference structure [39: 66-71]. The l_p-norm formulation is presented in Figure 2.4 [10: 6-7; 39: 69]. The function $r(p)$, $1 \leq p \leq \infty$, is interpreted as a regret function using distance measure to denote the level of *regret* using F^k instead of b^k. The parameter λ_k represents weight values used to assess varying levels of importance to specific O-D pairs. In the general case all weights are equal. Finally, the exponent p is used to define the DM's risk preference. When $p = 1$, the regret function $r(1)$ minimizes the sum of the absolute deviations. When $p = 2$, $r(2)$ is the least-squares solution similar in function to the generalized inverse approach. When $p = \infty$, $r(\infty)$, also known as the Tchebyschev norm, minimizes the maximum deviation.

Results identified in [10: 7-8, 20-21] using the l_p-norm metric appear to indicate that as p tends to ∞, for a fixed F_{tot}, the solution tends to represent an equalized set of O-D estimates similar to results obtained using the two previous methods. However, the authors demonstrate that the l_∞-norm, specifically, "'tracks' the [target] O-D values instead of simply equalizing the grand sum [F_{tot}] -- which is a highly desirable property" [10: 8].

Figure 2.4. L_p-Norm Minimization Models

Objective Functions

for $1 \le p < \infty$:

$$r(p) = \min \left[\sum_k \lambda_k^p |F^k - b^k|^p\right]^{\frac{1}{p}}$$

for $p = \infty$:

$$r(\infty) = \min\left[\max_k |F^k - b^k|\right]$$

subject to

$$\sum_k d_{ij}^k F^k = V_{ij} \qquad \forall\ (i,j) \in A$$

$$\sum_k \lambda_k = 1$$

$$F^k \ge 0 \qquad \lambda_k^p \ge 0$$

2.5 Summary

Communication networks, factoring in simplifying assumptions regarding network parameters, can be adequately modeled as multi-commodity networks. Therefore, determining the optimal flow through a network satisfying all commodity demands at minimum cost can be solved using a deterministic, multi-commodity, minimum cost network flow model. However, when arcs are subject to failure, we must resort to obtaining the expected flow through the network. Since obtaining a solution to this problem is computationally infeasible, we must develop bounds on the expected flow representing a range of feasible flows capturing the exact flow.

Once flow patterns are determined for both bounds we can determine an optimal placement of flow monitors by applying multi-criteria decision making principles to location modeling. Formulating the location problem as a weighted compromise program gives us flexibility to evaluate various location strategies with regards to the accuracy of the O-D estimates obtained and the compromise assessment itself.

Finally, having selected various location strategies, we formulate the O-D estimation model as a range of l_p-norm functions factoring in both the sampled arc flows and the DM's preference structure.

III. Model Formulation

This chapter presents a description of the modeling approach and associated mathematical programs. First, we summarize the problem and the objectives the overall model is designed to achieve. The next three sections describe the models developed: network flow bounds, monitor location strategy, and O-D demand estimation models. Next, we present a comprehensive step-by-step procedure which incorporates the above models. Finally, we review the software package used in this research.

3.1 Problem Summary

The problem is stated from a user perspective where a particular user of a communication network is interested in determining if its communication transmissions have been "compromised" by outside influences. For this research, a "compromise" scenario is exhibited as an increase in flow along specific paths and modeled by increasing a particular O-D pair's external flow. The user has the option of placing a limited number of flow sampling monitors on arcs within the network with the purpose of obtaining arc flow information to estimate O-D pair external flow values. O-D estimation is initially performed for a network under "normal" conditions where the estimates are used to establish a compromise measure. Subsequent O-D estimates are obtained from periodic flow sampling and compared to the compromise measure to determine if a "compromise" condition exists.

The overall model is designed to satisfy three specific objectives where each objective is represented by a specific sub-model. These objectives are summarized below:

(1) Given that arcs have reliability rates associated with them, define a lower and upper bound on the expected minimum cost flow through a network satisfying all user demands.

(2) Develop a monitor location model and present a strategy maximizing the likelihood that the O-D demands estimated from partial sampling of network components accurately represent the target O-D demands in the actual network.

(3) Develop an O-D demand estimation model and a corresponding compromise metric.

3.2 Minimum Cost Multi-Commodity Network Flow Models

3.2.1 Deterministic and Expected Network Flow Models. The first phase of modeling this problem is formulating the basic network model upon which all future assessments will be made. In its simplest form the network is totally deterministic where all network components are perfectly reliable. A mathematical model for the perfectly reliable network is presented in Chapter II.

In our problem the network is defined as a binary system [34] in the sense that an individual arc(i, j) can either be up or down with probability r_{ij} or ($1 - r_{ij}$), respectively. As mentioned earlier, solving for the exact minimum cost flow for networks of interesting size when reliability is introduced can be very difficult. This difficulty is illustrated in [37] and a corresponding mathematical model for the expected minimum cost flow is presented in Chapter II.

3.2.2 Bounds for the Expected Network Flow. Although obtaining an exact value for the expected flow may be computationally infeasible, it is possible to obtain lower and upper bounds on the expected flow thus defining for each arc and O-D pair a range of "expected" flow behavior. The bounds established for this research are partly based on results forwarded by [1, 8] for the single-commodity maximum flow network. The assertion for this research is that the bounds formulated below indeed capture the expected flow. Some properties related to these bounds are presented in Appendix A.

3.2.2.1 Lower Bound Formulation. The lower bound on the expected flow represents the best case from the user perspective where total cost is less than expected cost. The objective function represents the total cost (or distance) of the system and is formulated on arc-path decision variables f_p^k. The network flow is subject to several constraint sets. The first set of constraints requires all O-D pair demands to be satisfied and is identical to the deterministic constraint set. The second set constrains maximum arc utilization to the expected capacity of the arc, for all arcs. This aspect of the formulation is adapted from [1, 8, 38] where expected capacity was used to establish an upper bound on maximum flow in a single-commodity network. The final constraint set establishes non-negativity requirements for the decision variables. The mathematical model is presented in Figure 3.1.

Figure 3.1. Lower Bound Network Flow Model

$$\underline{\text{minimize}} \qquad Z_{LB} = \sum_{k} \sum_{p \in P^k} c_p^k f_p^k$$

$$\text{s.t.} \qquad \sum_{p \in P^k} f_p^k = b^k \qquad \forall\, k \in K$$

$$\sum_{k} \sum_{p \in P^k} h_{ij}^{kp} f_p^k \le r_{ij} u_{ij} \qquad \forall\, (i,j) \in A$$

$$f_p^k \ge 0 \qquad \forall\, k \in K\,,\, p \in P^k$$

where A = set of directed arcs in the network
K = set of distinct O-D pairs (or commodities)
P^k = set of paths connecting O-D pair k
b^k = total flow input to the network by O-D pair k
c_p^k = path cost equal to sum of arc costs lying on path $p \in P^k$
f_p^k = path flow on the p^{th} path of the set of candidate paths P^k
h_{ij}^{kp} = 1 if arc(i, j) lies on path $p \in P^k$; 0 otherwise
r_{ij} = reliability of arc(i, j)
u_{ij} = flow capacity of arc(i, j)

and where the expected arc capacity, e(u_{ij}), is equal to $r_{ij}\ u_{ij}$

3.2.2.2 Upper Bound Formulation. The upper bound on the expected flow represents a worse case scenario from the user perspective where total cost is greater than expected cost. The objective function represents the total cost (or distance) of the system and is identical to the lower bound objective. The network flow is subject to several constraint sets. The first set of constraints requires all O-D pair demands to be satisfied; however, each path variable has an associated loss parameter if at least one arc in the path is not totally reliable. This potential loss of flow may lead to an influx of slack external flow at some or all O-D pair origin nodes. Consequently, system cost increases as slack external flow increases. This aspect of the formulation is also adapted from [1, 8, 38] where the loss parameter was used as an objective function coefficient to establish a lower bound on maximum flow in a single-commodity network.

The second set constrains maximum arc utilization to the original capacity of the arc, for all arcs, and is identical to the deterministic constraint set. The final constraint set establishes non-negativity requirements for the decision variables. The mathematical model is presented in Figure 3.2.

3.3 *Multi-Criteria Monitor Location Model*

Once the flow patterns are established for both bounds, the next step is to determine which subset of arcs within the network will maximize the likelihood of obtaining flow information from all users while deploying the least number of monitors. This subproblem is modeled as a multi-criteria 0-1 integer program and solved using compromise programming to obtain the best monitor location strategy common to both

Figure 3.2. Upper Bound Network Flow Model

$$\underline{\text{minimize}} \qquad Z_{UB} = \sum_{k} \sum_{p \in P^k} c_p^k f_p^k$$

$$\text{s.t.} \qquad \sum_{p \in P^k} R_p^k f_p^k = b^k \qquad \forall\, k \in K$$

$$\sum_{k} \sum_{p \in P^k} h_{ij}^{kp} f_p^k \le u_{ij} \qquad \forall\, (i,j) \in A$$

$$f_p^k \ge 0 \qquad \forall\, k \in K\,,\, p \in P^k$$

where A = set of directed arcs in the network
K = set of distinct O-D pairs (or commodities)
P^k = set of paths connecting O-D pair k
R_p^k = reliability of path p for O-D pair k
b^k = total flow input to the network by O-D pair k
c_p^k = path cost equal to sum of arc costs lying on path $p \in P^k$
f_p^k = path flow on the p^{th} path of the set of candidate paths P^k
h_{ij}^{kp} = 1 if arc(i, j) lies on path $p \in P^k$; 0 otherwise
u_{ij} = flow capacity of arc(i, j)

and where path reliabilities are obtained as follows:

$$R_p^k = \prod_{(i,j) \in A_p^k} r_{ij}$$

network flow patterns (i.e. lower and upper bounds).

3.3.1 Criteria Functions. For this research we choose two criterion functions, among many possible functions, to represent an arbitrary decision maker's (DM) preferences regarding where the monitors should be placed in the network such that a flow disturbance is intercepted. A basic assumption related to criterion selection is that no prior knowledge is given regarding specific "hot spots" or particularly vulnerable arcs; instead, we assume that a compromise flow is simply routed along the least costly path(s). The two proposed criterion functions are described below and mathematically presented in Figure 3.3:

(1) Criterion Function 1 (CF1) - Maximize arc reliability to cost ratio, $v_{ij} = r_{ij} / c_{ij}$, where it is expected that the majority of the flow will traverse the arcs with the greatest ratio. Note that v_{ij} increases for increasing r_{ij} or decreasing c_{ij} ($c_{ij} \geq 1$).

(2) Criterion Function 2 (CF2) - Maximize arc flow where we want to place monitors on arcs with the maximum proportion of flow. CF2 may appear at first glance to be equivalent to CF1, but it is possible for both criteria to conflict on certain arcs since the second criteria, unlike the first criteria, is indirectly a function of arc capacities. In other words, a highly reliable, low cost, and low capacity arc may lead to conflict between both criteria.

The arc consideration parameter w_{ij} included in both criterion functions is presented in Section 3.3.3.3, Location Model Simplifications.

3.3.2 Decision Space. The decision variable is modeled as a binary 0-1 variable,

$x_{ij}^{k'}$, which takes on a value of 1 if a monitor is placed on arc(i, j) and 0 otherwise, and where k' is an index denoting the possible combinations of users on arcs of the network. This index will grow exponentially as the number of distinct O-D pairs increases, where the maximum number of combinations contained within the set is

$$\sum_{k=1}^{t} \binom{t}{k} = 2^t - 1$$

by the binomial theorem [31: 10]. Fortunately, this number can be significantly reduced by evaluating the network flow patterns and determining a routing table which identifies the specific user combinations for each arc. Therefore, the maximum number of user combinations is reduced to no greater than the number of active arcs (i.e. arcs carrying flow) in the network.

The decision space X for this problem consists of four proposed constraints sets. These sets are described below and mathematically presented in Figure 3.3:

(1) Redundancy Avoidance - This set of constraints limits the number of monitors located on arcs where the user combinations are identical to no more than one monitor. This is justified because locating two or more monitors covering the same combination of users yields two or more redundant arc flow constraints in the O-D estimation model (see section 3.4. O-D Demand Estimation Models).

(2) Monitor Goal - This is an equation which either constrains the decision to fielding a maximum of M monitors or is relaxed to determine the maximum number of monitors required.

(3) Minimum User Coverage - This set of constraints requires that every user,

Figure 3.3. Monitor Location Model For Each Network Bound

Criterion Functions

$$CF1 = y_1 = \max \sum_i \sum_j w_{ij}\, v_{ij}\, x_{ij}^{k'}$$

$$CF2 = y_2 = \max \sum_i \sum_j w_{ij}\, f_{ij}\, x_{ij}^{k'}$$

subject to

$$\sum_i \sum_j x_{ij}^{k'} \le 1 \qquad \forall\, k'$$

$$\sum_{k'} \sum_i \sum_j x_{ij}^{k'} \le M$$

$$\sum_i \sum_j x_{ij}^{k'} \ge 1 \qquad \forall\, k \in k'$$

$$x_{ij}^{k'} \in (0,1)$$

where w_{ij} = arc consideration coefficient = 1 if $u_{ij} - \Sigma_k \Sigma_{p \in Pk} f_p^k > 0$; 0 otherwise

f_{ij} = average proportion of flow on arc (i, j) for all users of arc $(i, j) = (\Sigma_k \Sigma_{p \in Pk}\, h_{ij}^{kp} f_p^k) / (\Sigma_k \Sigma_{p \in Pk}\, f_p^k)$

or O-D pair, be accounted for by the monitor location strategy; in other words, we desire flow information on all users of the network. This requirement may be relaxed if some users are of no consequence in the assessment of compromise.

(4) Binary Requirement - The decision variable is constrained to either 0 or 1.

3.3.3 Solution Approach.

3.3.3.1 Compromise Programming. To solve this bi-criterion problem, we use compromise programming techniques to obtain the best compromise solution to both criterion functions. The objective function for the compromise program is commonly refered to as a regret function $r(y)$ where the objective is to minimize the "*regret* of using y instead of obtaining the *ideal* point y^*," which can be modeled as a distance measure [39: 68]. Yu [39: 69] presents a class of l_p-norms which can be used to model the DM's preference structure with regards to the measurement of regret. Gershon [20: 245, 248] presents a generalized version of the l_p-norm introducing weights and normalization which permit us to interpret and compare the compromise solutions obtained by a weighted solution method [14: 164; 39] as weighted distances reflecting a percent shortfall from the ideal solution. Using this method we can determine which monitor location strategy will best achieve the DM's stated criteria. The compromise program for the monitor location model is presented in Figure 3.4, where the regret function is generalized for $1 \leq p < \infty$; however, in this research we only evaluate compromise solutions based on absolute deviations ($p = 1$).

Figure 3.4. Monitor Location Model Compromise Program

Compromise Function

$$r(y;\lambda,p=1) = \min \sum_{l=1}^{q} \lambda_l \left(\frac{y_l^* - y_l}{y_l^*} \right)$$

subject to

$$x \in X$$

where

$$\Lambda = \left\{ \lambda \in R^l \mid \lambda \geq 0, \sum_{l=1}^{q} \lambda_l = 1 \right\}$$

3.3.3.2 Arc Selection Strategy. The above solution approach results in possible monitor location strategies for each of the network bound configurations when viewed separately. However, a realistic scenario requires a single overall placement of monitors. In order to achieve this it is necessary to design a common monitor location strategy to both network bounds. Two simple approaches come to mind for selecting a common set of arcs: the first approach is to only place monitors on the common arcs shared by both network bounds (i.e. the intersection of both sets), and the second approach is to place monitors on all the arcs selected in both bounds (i.e. the union of both sets). The first approach is advantageous in that fewer arcs are likely to be selected, which reduces the cost of placing monitors; however, it is possible that arc selection in both bounds yield individual strategies sufficiently diverse that when considered together yield no arc selection strategy at all. On the other hand, the second approach may appear more costly yet guarantee a sufficient number of monitors are placed to collect the flow information required for O-D estimation. Further, if cost is a consideration, the value M used to obtain the compromise solutions for each bound can be incrementally decreased and evaluated. Given this flexibility, the second approach is used in this research.

The algorithm for the monitor location problem is summarized below:

Step 1: Solve for the compromise solution by minimizing $r(y; p)$ for the lower bound network flow.

Step 2: Repeat Step 1 for the upper bound network flow.

Step 3: Define an overall monitor location strategy as the union of the sets obtained in Steps 1 and 2.

We can extend this arc selection strategy to the situation where a maximum number of monitors, say M', can be fielded. For example, if no more than 3 monitors can be fielded ($M' = 3$), then Steps 1 and 2 above are run where $M = 3$, noting that y_l^* $\{l = 1, 2\}$ is also obtained for $M = 3$ to ensure consistency. In Step 3 a common strategy is defined where the number of selected arcs ranges from 3 (best case) to 6 (worse case). If the best case is achieved then stop; otherwise, if the number of selected arcs is greater than the maximum requirement then either eliminate the additional arcs based on some ranking scheme or repeat Steps 1 and 2 for $M = 2$, noting that y_l^* is also obtained for $M = 2$. Step 3 results in a common strategy of 2 to 4 arcs where either a strategy is selected or the process continues until the number of common arcs is less than or equal to M'.

3.3.3.3 Location Model Simplifications. The monitor location model considered in this research can be simplified by introducing an arc consideration coefficient, w_{ij}, and modifying the compromise program objective function in Figure 3.4.

First, the arc consideration parameter, w_{ij}, is defined as a 0-1 objective function coefficient which assumes the value 1 if an arc (i, j) has remaining capacity after an optimal flow is obtained, and the value 0 otherwise (i.e. if the arc is at full capacity). The reason this parameter is included in the model is particular to the earlier definition of "compromise" or flow disturbance: an increase in flow along a specific path or subset of paths connecting an O-D pair. If an arc is expected to be 100% utilized in a normal situation, then introducing a compromise flow will not affect the status of this arc.

Therefore, considering it for possible placement of a monitor is not productive. On the other hand, if an arc is under-utilized then compromise flow may or may not manifest itself on this arc and, therefore, this arc should be considered for possible monitor placement. The parameter w_{ij} is obtained by computing the following expression for each arc in A:

$$u_{ij} - \sum_{k \in K} \sum_{p \in P^k} h_{ij}^{kp} f_p^k$$

If the above expression is positive, or greater than a specified tolerance, then $w_{ij} = 1$ and arc(i, j) is considered; else if the expression equals zero, then $w_{ij} = 0$.

The next two simplifications occur in the formulation of the compromise program objective function $r(y;\ \lambda,\ p)$ for $p = 1$. First, the function r as stated represents the percent shortfall (when multiplied by 100) from the ideal solution $(y_1^*,\ y_2^*)$. As illustrated in Figure 3.5 (a), we can simplify the model by optimizing the function r', which represents the percent coverage (when multiplied by 100) of the ideal, and where $r = 1 - r'$.

Second, a single parameter α_{ij}, for fixed weights λ, is calculated for each arc, which consolidates both criteria function coefficients and restates the objective function in terms of a single variable, $x_{ij}^{k'}$. The parameter α_{ij} can also be evaluated as the function $\alpha_{ij}(\lambda)$ for varying weights and is illustrated in Figure 3.5 (b). The revised monitor location compromise model is presented in Figure 3.6.

Figure 3.5 (a). Simplification Of Compromise Program Objective Function

Compromise Function Reformulation for p = 1 and λ fixed:

$$r = \min \sum_{l=1}^{q} \lambda_l \left(\frac{y_l^* - y_l}{y_l^*} \right)$$

$$r = \min \left(\sum_{l=1}^{q} \lambda_l - \sum_{l=1}^{q} \lambda_l \left(\frac{y_l}{y_l^*} \right) \right)$$

$$r = 1 - \max \sum_{l=1}^{q} \lambda_l \left(\frac{y_l}{y_l^*} \right)$$

where the function to be optimized is r':

$$r' = \max \sum_{l=1}^{q} \lambda_l \left(\frac{y_l}{y_l^*} \right)$$

Figure 3.5 (b). Simplification Of Compromise Program Objective Function Coefficients

Expanding the bi-criteria function r':

$$r' = \max \left(\lambda_1 \left(\frac{\sum_{i,j} w_{ij} v_{ij} x_{ij}^{k'}}{y_1^*} \right) + \lambda_2 \left(\frac{\sum_{i,j} w_{ij} f_{ij} x_{ij}^{k'}}{y_2^*} \right) \right)$$

$$r' = \max \sum_{i,j} w_{ij} \left(\frac{\lambda_1 v_{ij}}{y_1^*} + \frac{\lambda_2 f_{ij}}{y_2^*} \right) x_{ij}^{k'}$$

we obtain the simplified function for variable λ:

$$\alpha_{ij}(\lambda) = w_{ij} \left(\frac{\lambda_1 v_{ij}}{y_1^*} + \frac{\lambda_2 f_{ij}}{y_2^*} \right)$$

Figure 3.6. Revised Monitor Location Model Compromise Program

Compromise Function

$$r'(y;\lambda,p{=}1) = \max \sum_{i,j} \alpha_{ij}(\lambda)\, x_{ij}^{k'}$$

subject to

$$x \in X$$

$$\lambda \in \Lambda$$

where X = decision space for monitor placement
Λ = decision space for criterion weighting

3.4 Origin-Destination Demand Estimation Models

The final model is designed to estimate the O-D demands, or external flows, input to the network by sampling flow information on the arcs where the monitors are fielded. This estimation process is initially conducted for both network bounds assuming the flow represents "normal" flow through the system. The O-D estimates are then used in defining a compromise metric, or confidence interval for "normal" flow behavior, presented in Chapter IV. This estimation process is repeated in subsequent time periods to obtain new O-D estimates which are used to assess "compromise" by comparing the new estimates to the established confidence interval.

3.4.1 L_p-Norm Minimization Model. This subproblem is modeled using the class of l_p-norm deviational measures presented in Chapter II. The general L_p model where p is bounded by 1 and infinity, and where each O-D pair is weighted, is provided in Figure 3.7. For this research we only investigate the effects of estimating O-D demands for p = 1, 2 and infinity.

3.4.2 Goal Programming Solution Approach. The general l_p-norm model is operationalized as a goal program due to possible numerical problems when optimizing absolute value functions using General Algebraic Modeling System (GAMS) software (reference Section 3.6 for discussion of software) [7: 92]. The revised model is as presented in [39: 84-89] where the absolute value functions are eliminated by introducing non-negative deviational variables, dp^k and dm^k, for each O-D pair k. The goal program

Figure 3.7. Generalized L_p-Norm Minimization Model

Objective Functions

Minimize (for $1 \leq p < \infty$)

$$l(p) = \left[\sum_k \lambda_k^p |F^k - b^k|^p \right]^{\frac{1}{p}}$$

Minimize (for $p = \infty$)

$$l(\infty) = \max_k |F^k - b^k|$$

subject to

$$\sum_k d_{ij}^k F^k = V_{ij} \qquad \forall\ (i,j) \in S$$

$$\sum_k \lambda_k = 1$$

$$F^k \geq 0 \qquad \lambda_k^p \geq 0$$

where S = set of arcs selected for monitor placement
b^k = target O-D demand for user k
d_{ij}^k = proportion of user k flow on arc(i, j), obtained from routing matrix
F^k = estimated O-D demand for user k
V_{ij} = observed flow on arc (i, j)

using equal weights for all O-D pairs is represented in Figures 3.8 (a) and (b)

3.5 Solution Algorithm

This section presents an algorithm for solving the problem under consideration which incorporates the models presented above. The algorithm first evaluates a network under "normal" conditions to establish a monitor location strategy and develop a compromise metric (presented in Chapter IV, Compromise Measurement) which defines a threshold for "normal" O-D demand variation based on O-D estimates. The algorithm then evaluates the same network under "compromise" conditions where revised O-D estimates are compared to previously established thresholds to assess compromise. The algorithm is summarized below as a sequence of steps:

Step 1: Once the network parameters (i.e. O-D pair target demands, arc reliabilities, capacities and costs, and possible path sets) are defined for the "normal" scenario, solve for the optimal minimum expected cost for each network bound which satisfies all O-D demands. Define a routing matrix representing the proportion of user flow on each arc for both bounds.

Step 2: Using the arc flow values obtained in Step 1, solve for a common monitor location strategy to both network bounds for specified M and fixed weighting of criterion functions (reference arc selection algorithm in Section 3.3.3.2).

Step 3: Using the arc flow values for the specific arcs having monitors, obtain the O-D demand estimates for $p = 1, 2$ and ∞.

Figure 3.8 (a) O-D Estimation Goal Programming Model ($1 \le p < \infty$)

Objective Function

Minimize

$$l(p) = \sum_k (dm^k + dp^k)^p$$

subject to

$$F^k + dm^k - dp^k = b^k \qquad \forall\, k$$

$$\sum_k d_{ij}^k F^k = V_{ij} \qquad \forall\, (i,j) \in S$$

$$F^k \ge 0 \qquad \forall\, k$$

Figure 3.8 (b) O-D Estimation Goal Programming Model ($p = \infty$)

Objective Function

Minimize

$$l(\infty) = W$$

subject to

$$W - dm^k - dp^k \geq 0 \qquad \forall\, k$$

$$F^k + dm^k - dp^k = b^k \qquad \forall\, k$$

$$\sum_k d_{ij}^k F^k = V_{ij} \qquad \forall\, (i,j) \in S$$

$$F^k \geq 0 \qquad \forall\, k$$

Step 4: Based on the O-D estimates in Step 3, define the compromise metric for each level of p (see Chapter IV).

Step 5: Create a "compromise" scenario by introducing an increase in flow at an arbitrary O-D pair and repeat Step 1 to obtain the revised arc flow values for each network bound.

Step 6: Using the arc flow values for the specific arcs having monitors, obtain the revised O-D demand estimates for $p = 1, 2$ and ∞ as in Step 3.

Step 7: Compare results in Step 6 to the compromise metric defined in Step 4.

3.6 *Software*

The solution algorithm is programmed primarily using GAMS. GAMS is a commercially available software package that provides a high-level language simplifying the development and solution of large-scale and complex models. The development of a model is performed independently of the solution algorithm, which allows for quick and easy modifications. The solution is obtained through accompanying solvers such as MINOS 5 (linear and non-linear programs) and ZOOM (mixed-integer programs) [7: 3, 105]. GAMS Version 2.20 is the VMS version used for this research.

3.7 *Summary*

This chapter reviewed the various mathematical models used in this research. They include lower and upper bounds on the expected minimum cost network flow, a compromise monitor location solution based on multiple criteria, and O-D estimates based on the class of L_p-norm functions. Also presented was a step-by-step algorithm for solving the problem of compromise assessment for a given network, where the compromise metric itself is developed in the next chapter. Finally, the software tool used for programming the various models was identified.

IV. Compromise Measurement

This chapter presents a modeling approach for developing a compromise metric (Step 4 of the solution algorithm in Chapter III) used for assessing flow disturbance, or "compromise." The problem of compromise measurement is first presented followed by a description of the modeling approach, which identifies the assumed regression model, the different sources of potential variability, and the compromise metric itself. A final section is included which presents an alternative method for assessing "compromise" where the O-D sample size is small and the normality assumption may not hold.

4.1 The Measurement Problem

This aspect of the modeling approach is considered separately given the importance of the measurement problem. The original problem assumes that "compromise" occurs randomly from the user perspective and therefore the assessment of "compromise" from only partial information of a network involves uncertainty. For this reason we develop a compromise metric as a means of determining if a "compromise" has occurred and locating the specific O-D pair contributing to the apparent flow increase. This metric is defined in terms of confidence intervals to allow us to express our assessment of "compromise" as a confidence statement. The intervals themselves are a function of the expected variability due to "normal" network behavior (i.e. the expected flow pattern under "normal" conditions) which we assume can be readily obtained.

The use of confidence intervals to bound O-D flows assumes the O-D flows are normally distributed. We use the law of large numbers to justify this assumption where in more realistic networks we would expect the number of O-D pairs to be on the order of n^2 for fully connected networks, where n represents the number of nodes [31: 31].

A numerical example of the measurement problem is presented as part of the case study in Chapter V, Section 5.4.1, Step 4.

4.2 *Modeling Approach*

The model proposed in this section is used to explain the variability due to flow differences in both network bounds and to construct appropriate confidence bounds around "normal" flow behavior. The O-D pairs in each network bound are characterized by the specific paths selected and the flow values on the arcs along those paths. To understand the flow behavior between both bounds, we analyze the sources of variance, namely the path selection differences and the arc flow differences.

To understand this variance we develop a statistical relation between the dependent variable Y and the independent variable X of the form

$$Y = f(X)$$

where $Y = (Y_1, Y_2, \ldots, Y_t)$ is a vector of O-D pair volumes represented as the difference of the lower bound from the upper bound O-D external flow, and $X = (X_1, X_2, \ldots, X_t)$ is a vector of path differences representing the number of distinct path selection differences between the lower and upper bound network flow patterns for each O-D pair.

The variable Y_k could be modeled as the O-D pair external flows for lower bound

($b^k{}_{LB}$) and upper bound ($b^k{}_{UB}$) separately, but is consolidated as a deviation measure ($b^k{}_{UB}$ - $b^k{}_{LB}$) for simplicity. Further, Y_k is assumed to be normally distributed, as stated earlier, with mean E(Y_k) and equal variance σ^2 for all k.

The independent variable X_k is selected to help explain the potential variability in Y_k due to path selection preferences in both network bounds. The proposed model presented in the next section accounts for this variability. The arc flow variance, however, is not accounted for by this model and is therefore used to construct the confidence bounds.

4.2.1 Regression Model. It is assumed that a simple first-order model, linear in the regression parameters and the independent variable, adequately describes the relationship between Y and X, and is of the form

$$Y_k = \beta_0 + \beta_1 X_k + \varepsilon_k$$

where β_0 and β_1 are the regression parameters, and ε_k represents the deviation of Y_k from its mean value E(Y_k) and is normally distributed with mean E(ε_k) = 0 and variance $\sigma^2\{\varepsilon_k\} = \sigma^2$ [27: 31]. See [27: 38-43] for a discussion of methods to obtain the estimated regression function:

$$\hat{Y}_k = b_0 + b_1 X_k$$

4.2.2 Analysis of Variance.

4.2.2.1 Sum of Squares. The analysis of variance approach to regression analysis is useful in understanding the sources of variability in the model. As presented in [27: 87 - 93], the total variation in the observations, or O-D deviations, is measured as the total sum of squares, SSTO, which is composed of the sum of the regression sum of squares, SSR, and the error sum of squares, SSE, where

$$\text{SSTO} = \sum_{k=1}^{t} (Y_k - \bar{Y})^2, \qquad \bar{Y} = \frac{\sum_k Y_k}{t}$$

$$\text{SSR} = \sum_{k=1}^{t} (\hat{Y}_k - \bar{Y})^2$$

where SSR represents the squared deviation of the fitted regression around the overall mean, and

$$\text{SSE} = \sum_{k=1}^{t} (Y_k - \hat{Y}_k)^2$$

where SSE represents the variation of the O-D deviations around the fitted regression line.

4.2.2.2 Mean Squares. The associated mean squares for SSR and SSE, when divided by their respective degrees of freedom, are as follows [27: 91]:

$$\text{MSR} = \frac{\text{SSR}}{1} = \text{SSR}$$

$$\text{MSE} = \frac{\text{SSE}}{t - 2}$$

where MSR now represents the total variation in the model explained by the estimated regression function. In other words, MSR denotes the variability due to different path selections explained by the model. The variance of interest, however, is MSE which represents the unexplained error and will be used to help build the subsequent confidence intervals.

4.2.2.3 O-D Estimation Error. A third source of variability not accounted for by the analysis of variance approach is the error due to O-D estimation: the deviation between the target, b^k, and estimated, F^k, O-D pair external flow for each network bound. Similarly, this variance can be expressed as the deviation between the target O-D deviation, Y^k, and the estimated O-D deviation, D_k, where $D_k = (F^k_{UB} - F^k_{LB})$. This variance, also referred to as an accuracy measure [9: 370], is denoted as MSE_{est} where:

$$\text{MSE}_{est} = \frac{\sum_{k=1}^{t} (D_k - Y_k)^2}{t}$$

4.2.3 Compromise Metric. There are 2 forms of the compromise metric proposed for this research: individual confidence intervals ("individual metric") for each O-D pair and an overall joint confidence interval ("joint metric").

4.2.3.1 Individual O-D Confidence Intervals. Constructing a confidence interval around each O-D pair serves the dual purpose of identifying a "compromise" and its source. The objective is to estimate for each O-D pair the mean O-D deviation denoted as $E(Y_k)$ for specific k. Since this value is generally not known, we use the point estimator $\hat{Y}_k$ defined in the earlier regression model for specific X_k.

An initial confidence interval for $E(Y_k)$ is developed using the t distribution (not to confuse with the scalar t used to denote the number of O-D pairs) [27: 77-78]. The $1 - \alpha$ confidence limits are:

$$\hat{Y}_k \pm t(1-\alpha/2;t-2)s_{\hat{Y}_k}$$

where

$$s^2_{\hat{Y}_k} = \text{MSE}\left(\frac{1}{t} + \frac{(X_k-\bar{X})^2}{\sum_k (X_k-\bar{X})^2}\right)$$

A final confidence interval for $E(Y_k)$ incorporates the mean error due to the O-D estimation process and is of the form:

$$\hat{Y}_k \pm \left(t(1-\alpha/2;t-2)s_{\hat{Y}_k} + \sqrt{\text{MSE}_{\text{est}}}\right)$$

4.2.3.2 Joint Confidence Interval. A joint metric is considered to assess the sensitivity of the model to mean O-D volumes with no expectation of being able to determine the specific source of compromise. The joint metric is of the form:

$$\bar{Y} \pm \left(t(1-\alpha/2;t-2)\sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{\text{est}}} \right)$$

where the mean of the O-D pair deviations Y_k is equivalent to the mean of the fitted O-D pair deviations $\hat{Y}_k$ [27: 49].

4.3 Alternative Compromise Metric

An alternative method for assessing "compromise" is proposed in this section as an extension of this research. The method, as described by Upton and Fingleton [36: 170-75], uses the Moran statistic *I* to determine if correlation exists between the values of each O-D pair. A set of O-D pairs is considered correlated if their exists a connection or similarity between the values of each O-D pair.

This method is useful for the case where the number of locations, or O-D pairs, considered is small. Although the distribution of *I* is assumed approximately normally distributed for a large number of sampled locations, the method allows for probabilistic assessments of the observed statistic *I* for smaller sample sizes [36: 171].

The stated hypotheses for this test are:

Null Hypothesis: No correlation present (i.e. no similarity exists between the source and destination node flow values for each O-D pair)

Alternative Hypothesis: Correlations present (i.e. a similarity does exist between the source and destination node flow values for each O-D pair)

In the context of our research, it is proposed that rejecting the null hypothesis is equivalent to stating that no "compromise" has occurred given that the deviation between source and destination flows for each O-D pair is sufficiently close to be deemed similar. On the contrary, failing to reject the null hypothesis indicates that the deviations are sufficiently large to be deemed unsimilar. This last statement suggests that a possible "compromise" has occurred given that a "compromise" scenario is exhibited as an increase in flow at a particular O-D pair's source node.

The test statistic I is presented below in the authors' original notation and related to the context of our research. Note that i denotes the source node and j denotes the destination node.

$$I = \frac{n}{S_o} \frac{\sum_i \sum_j W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where n = number of locations, or nodes, sampled = $2t$

x_i = value of flows at source nodes = D_k^i

x_j = value of flows at destination nodes = $D_k^j = 0$ given that flow demands are unchanged from lower to upper bound

$\bar{x}$ = average of all node values x_i and x_j

W_{ij} = matrix representing the O-D pairs where $W_{ij} = 1$ if nodes i and j denote an O-D pair, and $W_{ij} = 0$ otherwise

$S_o = \sum_i \sum_j W_{ij} \quad (i \neq j)$

Furthermore, we define the matrix of elements Y_{ij} where

$$Y_{ij} = (x_i - \bar{x})(x_j - \bar{x})$$

Given the lack of appropriate probability tables for small sample size problems, we make use of this proposed method which defines the cumulative distribution function (CDF) for I as a reference curve against which observed values of the test statistic I can be compared to assess the significance of the observation. The CDF for I is obtained using a randomization process of the matrix Y where (1) the x_i and x_j values assume the estimated O-D deviations in the "normal" scenario, and (2) $n!$ Y matrices are obtained by random permutations of the n rows and n columns [36: 152]. It is under this randomization process that the authors claim the distribution of I to be approximately normally distributed with specified mean and variance [36: 171]. The CDF for I is defined by solving I for each specific random matrix Y and obtaining a frequency count of the number of Y matrices resulting in a specific I value. The resulting CDF is a frequency distribution of I in the "normal" scenario.

To assess "compromise," we would compute the I statistic where x_i and x_j assume the estimated deviations in a "compromise" scenario. To assess the significance of the resulting I value we compare it to the CDF for I. If the observed value is extreme (e.g. only in a few occasions is the observed value equalled or exceeded) then we would conclude that correlation is present; in other words, we would conclude that "compromise" has not occurred with certain significance. Contrarily, if the observed value is exceeded by a specified percentage of observations, then we would conclude that "compromise" has occurred.

4.4 Summary

This chapter reviewed the proposed metric for determining if a "compromise" situation exists based on the comparison of estimated O-D external flow values to pre-established confidence intervals encompassing "normal" network behavior. This chapter also presented, as an extension, an alternative test statistic for measuring "compromise" for small O-D sample sizes. The next chapter applies the models presented in Chapter III to a test case network and provides, as part of this case network, a numerical example of the method presented in this chapter.

V. Case Study

This chapter applies the solution algorithm presented in Chapter III to a test case network. The first section briefly restates the research objectives and includes a description of the case network and all related network parameters. The second section illustrates Step 1 of the solution algorithm presented in Section 3.5, where a lower and upper bound network flow pattern is obtained. The third section illustrates Step 2, monitor location strategy, and describes the two sets of experiments to be conducted. The final two sections illustrate the remaining five steps for both experiments.

5.1 Network Description

The purpose of this chapter is to apply the solution algorithm on a case network and analyze its performance, keeping in mind the stated research objectives of Chapter I which are summarized below:

Objective #1: Obtain a network flow that bounds the expected minimum cost network flow.

Objective #2: Obtain a monitor location strategy that maximizes the accuracy of estimating the true O-D external flows in both "normal" and "compromised" network configurations.

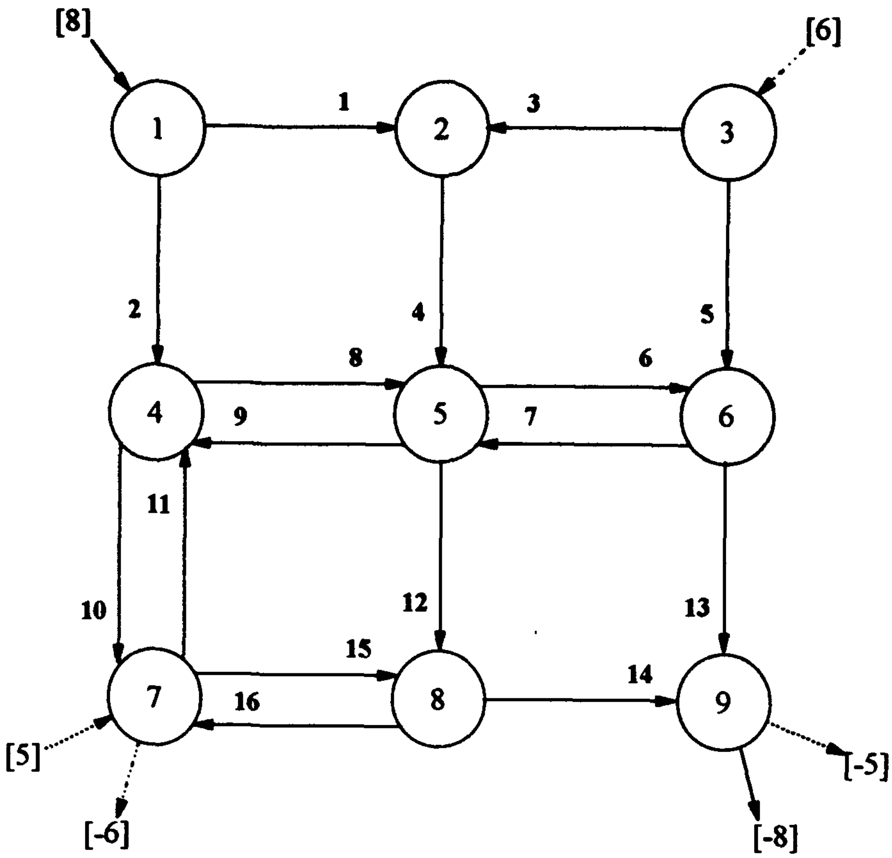
Objective #3: Define the compromise metric and evaluate the compromise assessment based on O-D estimation preferences for p = 1, 2, and ∞.

The case network is a three-commodity (or user) network with multiple unreliable arcs as defined in Figure 5.1. The network arc parameters, O-D parameters, and O-D path sets for the original "normal" configuration are provided in Tables 5.1, 5.2, and 5.3, respectively. Note that the arc notation (i, j) used in the previous models is simplified from this point on to the single arc index i (refer to Figure 5.1).

Table 5.1 Network Arc Parameters

i	from	to	r_i	u_i	c_i
1	1	2	1	7	2
2	1	4	.95	5	1
3	3	2	1	5	2
4	2	5	1	10	3
5	3	6	.98	4	1
6	5	6	.98	8	3
7	6	5	1	4	3
8	4	5	.7	10	2
9	5	4	.8	5	2
10	4	7	1	3	1
11	7	4	1	5	1
12	5	8	1	9	3
13	6	9	1	10	3
14	8	9	.98	9	2
15	7	8	.95	6	2
16	8	7	.95	4	2

Figure 5.1 Three-Commodity Case Network



\# arc index i, $i = 1 .. 16$

[ ] O-D Pair Volume

Table 5.2 Network O-D Pair Demands

O-D Pair k	Node Pair	Demand b^k
1	(1, 9)	8
2	(3, 7)	6
3	(7, 9)	5

Table 5.3 O-D Pair Path Sets

k	p	Arcs	R_p^k	c_p^k
1	1	1-4-6-13	.98	11
	2	1-4-12-14	.98	10
	3	2-8-6-13	.6517	9
	4	2-8-12-14	.6517	8
2	5	3-4-9-10	.8	8
	6	3-4-12-16	.95	10
	7	5-7-9-10	.784	7
	8	5-7-12-16	.931	9
3	9	11-8-6-13	.686	9
	10	15-14	.931	4

5.2 Network Bounds

In this section the results are presented for Step 1 of the solution algorithm as stated below:

<u>Step 1</u>: *Obtain the optimal minimum cost network flow for both bounds representing the "normal" scenario.* The GAMS models for both bounds are included in Appendices B.1 and B.2, and the optimal flow patterns displayed in Appendix B.3. The

optimal objective function values are $Z_{LB}^* = 144.76$ and $Z_{UB}^* = 169.2$. The resulting target O-D path flows and external flows are summarized in Tables 5.4 and 5.5.

Table 5.4 Path Flows f_p^k for both Network Bounds

k	p	f_p^k (LB)	f_p^k (UB)
1	1	0	3.371
	2	3.25	3.629
	3	4.18	1.749
	4	.57	0
2	5	0	0
	6	2.08	2.86
	7	3	3
	8	.92	1
3	9	0	0
	10	5	5.371

Table 5.5 Target O-D External Flow

k	b^k (LB)	b^k (UB)
1	8	8.749
2	6	6.86
3	5	5.371

From the flow patterns in Appendix B.3, we can develop the flow routing matrix for each network bound representing the proportion of user flow on each arc. This matrix is shown below in Table 5.6 for both bounds:

Table 5.6 Flow Routing Matrix

	Lower Bound			Upper Bound		
i	$k = 1$	$k = 2$	$k = 3$	$k = 1$	$k = 2$	$k = 3$
1	0.41	-	-	0.80	-	-
2	0.59	-	-	0.20	-	-
3	-	0.35	-	-	0.42	-
4	0.41	0.35	-	0.80	0.42	-
5	-	0.65	-	-	0.58	-
6	0.52	-	-	0.59	-	-
7	-	0.65	-	-	0.58	-
8	0.59	-	-	0.20	-	-
9	-	0.50	-	-	0.44	-
10	-	0.50	-	-	0.44	-
11	n/a	n/a	n/a	n/a	n/a	n/a
12	0.48	0.50	-	0.42	0.56	-
13	0.52	-	-	0.59	-	-
14	0.48	-	1.00	0.42	-	1.00
15	-	-	1.00	-	-	1.00
16	-	0.50	-	-	0.56	-

5.3 *Monitor Location Strategies*

In this section the results are presented for Step 2 of the solution algorithm:

Step 2: *Obtain a common monitor location strategy to both bounds for specified* M. For this case network we set $M = 2$ to ensure a worse case strategy of no more than four arcs. We also propose to explore the set of compromise solutions which are also efficient in both network bounds.

5.3.1 Set of Efficient Solutions. Obtaining the set of efficient solutions allows us to investigate different location strategies for various weighting of the two criterion functions. By incrementally varying the criteria weights, we obtain a set of efficient solutions (i.e. strategies) for each bound. The common strategies are thus defined as the union of the efficient solutions, or strategies, and are presented in Table 5.7.

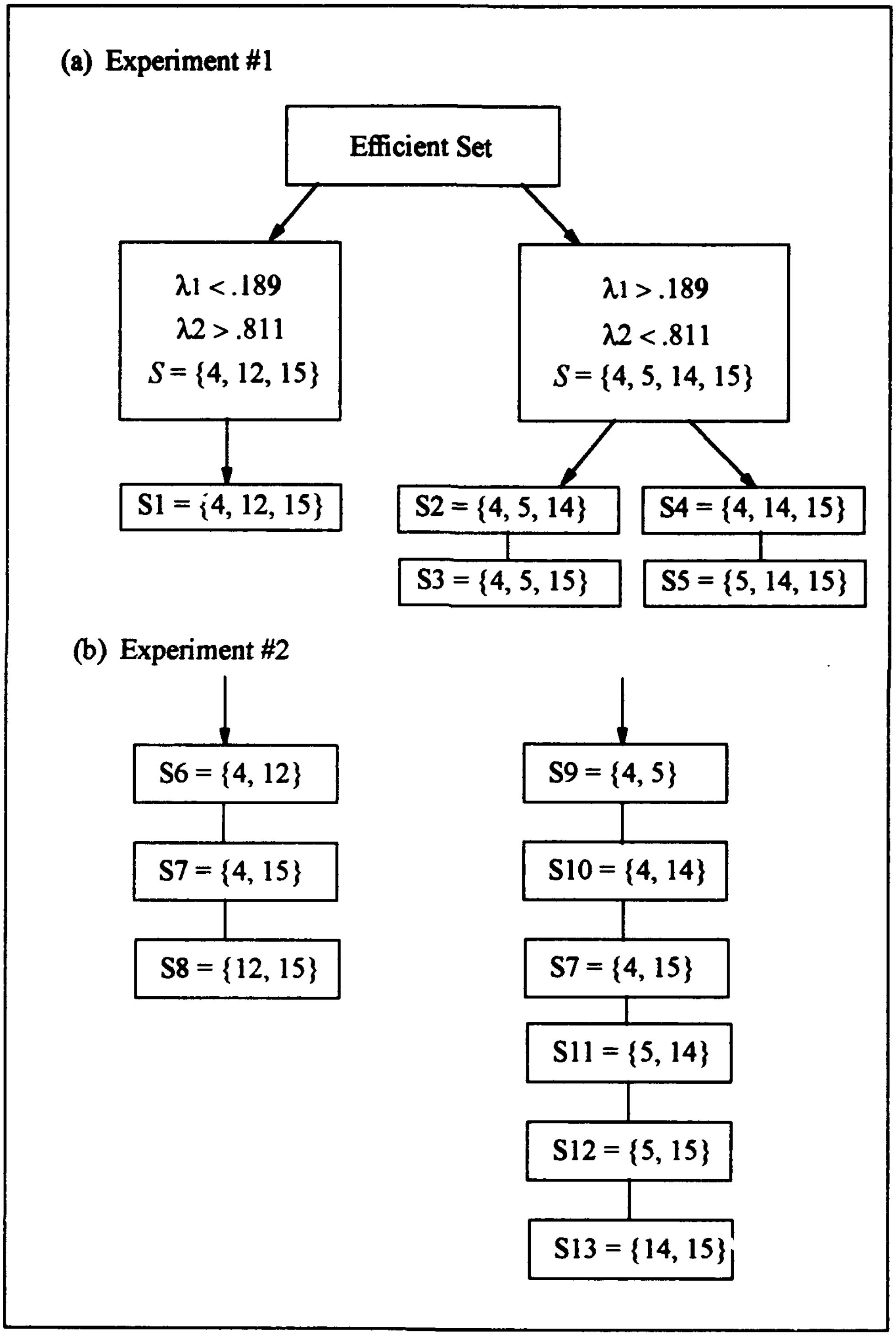
Table 5.7 Set of Efficient Solutions (both Bounds)

Weights		Bound Strategy	Common Strategy, S
$\lambda_1 < 0.189$	Lower Bound	{12, 15}	{4, 12, 15}
	Upper Bound	{4, 15}	
$\lambda_1 > 0.189$	Lower Bound	{5, 14}	{4, 5, 14, 15}
	Upper Bound	{4, 15}	

For $M = 2$ the optimal values for criterion function 1 are $y_1{*} = 1.47$ (lower bound) and $y_1{*} = 1.487$ (upper bound), and for criterion function 2 $y_2{*} = 0.808$ (lower bound) and $y_2{*} = 1.632$ (upper bound). The lower bound network yielded two distinct efficient solutions whereas the upper bound efficient solution is unique. The GAMS location models for both bounds are included in Appendices C.1 and C.2.

5.3.2 Experimental Design. Two separate experiments are conducted based on the common strategies identified above, which exhaustively evaluate all possible 3-arc and 2-arc combinations. The first experiment assumes three monitors are fielded (or 20% sampling of active arcs) and investigates the performance of each 3-arc strategy. The second experiment extends the investigation to all possible 2-arc strategies (or 13.33% sampling of active arcs). The possible location strategies for Experiments #1 and #2 are presented in Figure 5.2 (a) and (b).

Figure 5.2 Experimental Design



The next two sections explore the results for all strategies within each experiment. To illustrate the remaining steps of the solution algorithm, the first strategy for Experiment #1 is developed in greater detail. All subsequent strategies are presented in summary form.

5.4 Experiment #1

5.4.1 Location Strategy S1. Strategy S1 consists of three monitors placed on arcs {4, 12, 15}.

<u>Step 3:</u> *Obtain the O-D pair external flow estimates for both bounds representing the "normal" scenario for p = 1, 2 and infinity.* First, we note that since we are sampling from three arcs, the constraint set for the O-D estimation model is fully determined (i.e. 3x3) therefore the O-D estimates are unique for all $p \geq 1$. When we sample from 2-arc strategies the relationships for different norm exponents become more evident. The constraint set is defined by the arc flow values and the routing matrix for the specific arcs sampled as illustrated for S1 in Table 5.8 for both bounds.

Table 5.8 O-D Estimation Constraint Set for S1

i	Lower Bound, d_i^k			V_i	Upper Bound, d_i^k			V_i
	$k = 1$	$k = 2$	$k = 3$		$k = 1$	$k = 2$	$k = 3$	
4	0.41	0.35	-	5.33	0.80	0.42	-	9.86
12	0.48	0.50	-	6.82	0.42	0.56	-	7.489
15	-	-	1.00	5	-	-	1.00	5.371

The GAMS O-D estimation model for S1 is included in Appendix D.1 and the resulting estimates are in Appendix D.2.

Step 4: *Define the compromise metric.* Using the approach proposed in Chapter IV, we define individual and joint confidence intervals around each O-D pair's expected flow and overall mean flow for the "normal" scenario. In the general case where we have a large number of O-D pairs, a DM would be interested in assessing "compromise" with a certain degree of confidence. This confidence is reflected in our confidence intervals using the appropriate t-statistic from look-up tables. In this case study, however, given the small sample size we simply use a scaling factor of two to conservatively widen our intervals. As a result, confidence statements related to this case study are inappropriate and the results themselves inconclusive; however, the goal is to gain sufficient insight to assess the performance of the different metrics. An alternative approach was proposed in Chapter IV which may allow for probabilistic assessments of "compromise" for small samples. The applicability of this approach is left as an extension of this research.

Initial confidence intervals are developed based on the target flows. As presented in Chapter IV, we define the variables $Y_k = (b^k_{\mathrm{UB}} - b^k_{\mathrm{LB}}) = (0.749, 0.86, 0.371)$ and $X_k = (2, 0, 0)$, where user 1 (k = 1) has two distinctly different path selections, and users 2 and 3 (k = 2 and 3) have none. Using classical regression methods, we fit a simple linear model with regression coefficients $\beta_0 = 0.616$ and $\beta_1 = 0.067$. Partitioning the sum of squares, we obtain an SSE = MSE = 0.12, which is used to calculate the variance term for the fitted value $\hat{Y}_k$. The results for the initial confidence intervals (i.e. not including the O-D estimation error) are presented in Appendix D.3. Note that these intervals remain fixed for all strategies.

These intervals are slightly widened upon introducing the O-D estimation error denoted as MSE_{est}. First, we define for strategy S1 the deviational variable $D_k = (F^k_{UB} - F^k_{LB}) = (1.235, 0.384, 0.371)$ which represents the deviation in the O-D estimates for the "normal" scenario. The error is then computed as the mean squared deviation of the O-D estimated deviation D_k and the O-D target deviation Y_k. This error is added to the confidence bounds defined earlier to obtain the individual and joint metrics (see Appendix D.2), which are summarized below in Table 5.9:

Table 5.9 Compromise Metrics for S1

Strategy	Individual Confidence Intervals (CIs)			Joint CIs
	$k = 1$	$k = 2$	$k = 3$	
S1 = {4, 12, 15}	(-.335, 1.833)	(-.266, 1.497)	(-.266, 1.497)	(-.132, 1.452)

Note that the individual intervals for users 2 and 3 are identical since both users have no path selection differences between lower and upper bound, whereas user 1 has two distinct path differences.

Step 5: *Introduce a "compromise" scenario.* For this case network we define "compromise" as a single flow unit increase in the first O-D pair's demand ($k = 1$), from 8 to 9 units. To obtain the revised arc flow values we solve the network bound models in Appendices B.1 and B.2 with the revised O-D demand parameter. The "compromise" network patterns are included in Appendix B.3 where the thicker arcs represent the "compromised" paths for both bounds. This step is unchanged for all other strategies.

Step 6: *Obtain the revised O-D pair external flow estimates for both bounds* *representing the "compromise" scenario for p = 1, 2 and infinity.* Similarly to the process in Step 3, we solve the O-D estimation model in Appendix D.1 with appropriate changes to affected arcs. For this strategy, and all subsequent strategies, the only affected arc is $i = 4$ in the lower bound pattern where V_4 increases from 5.33 to 6.33 units. The revised O-D estimates are included in Appendix D.2.

Step 7: *Assess compromise.* This final step requires the comparison of the revised O-D estimates to the previously established confidence intervals. Since the confidence intervals bound the expected deviation of upper to lower bound flow, the "compromise" estimates need to be similarly expressed as deviation measures. These results are included in Appendix D.2 and summarized in Table 5.10.

Table 5.10 "Compromise" Estimates for S1

Strategy	$k = 1$	$k = 2$	$k = 3$	Joint
S1	-1.568	.811	.371	-.129

From these results we would conclude that, from an individual user basis, user 1 is the source of "compromise," whereas from a joint user basis, we would conclude that no "compromise" has occurred. Note that an actual confidence statement would be inappropriate as stated earlier.

5.4.2 Summary Results. The results for the remaining four strategies, namely Steps 3, 4, 6 and 7, are included in Appendices E.1 through E.4. The overall results for Experiment #1 (i.e. confidence intervals and "compromise" values) are presented in Table 5.11. Table 5.12 summarizes the results by displaying for the individual confidence

Table 5.11 Experiment #1 Summary Results

Efficient Solutions	Strategy	Individual Confidence Intervals			Joint CIs
		k = 1	k = 2	k = 3	
$\lambda_1 < .189$	S1 = {4, 12, 15}	(-.335, 1.833) -1.568	(-.266, 1.497) .811	(-.266, 1.497) .371	(-.132, 1.452) -.129
$\lambda_1 > .189$	S2 = {4, 5, 14}	(-.017, 1.515) -1.587	(.052, 1.179) .866	(.052, 1.179) 1.464	(.186, 1.134) .248
	S3 = {4, 5, 15}	(-.002, 1.5) -1.587	(.067, 1.164) .866	(.067, 1.164) .371	(.201, 1.119) -.117
	S4 = {4, 14, 15}	(-.093, 1.591) .682	(-.024, 1.255) -1.745	(-.024, 1.255) .371	(.11, 1.21) -.231
	S5 = {5, 14, 15}	(.019, 1.479) .682	(.088, 1.143) .866	(.088, 1.143) .371	(.222, 1.098) .64

Table 5.12 Experiment #1 Summary Results (cont)

Efficient Solutions	Strategy	Individual CIs	Joint CIs
$\lambda_1 < .189$	S1 = {4, 12, 15}	k = 1	NO
$\lambda_1 > .189$	S2 = {4, 5, 14}	k = 1 & 3	NO
	S3 = {4, 5, 15}	k = 1	YES
	S4 = {4, 14, 15}	k = 2	YES
	S5 = {5, 14, 15}	None	NO

intervals the source(s) of potential "compromise", and for the joint confidence intervals, whether a "compromise" has occurred ("YES") or not ("NO").

The key questions to be answered in this experiment are (1) how well the location model performs in selecting the more susceptible arcs to "compromise," and (2) how accurate the proposed compromise metrics perform in assessing "compromise" and its source in a network known to be compromised.

Of the candidate arcs selected for monitor placement, only $i = 4$ incurred an increase in flow. In other words, of the six arcs carrying the increased flow between both bounds (see Appendix B.3), our location model managed to select only one, resulting in a 16.7% selection rate for this particular network. It is not necessarily the intention of this research to construct a more accurate location model, however the apparent weakness is acknowledged and discussed in the recommendations section of the final chapter.

Of the five strategies considered in this experiment, four contain $i = 4$ and are therefore of interest. Strategy S5 does not contain $i = 4$ and does not detect "compromise," which is an expected result. The first four strategies (S1 - S4) do detect a source of "compromise" when the compromise metric is defined as the individual metric. S1 through S3 correctly identify user 1 as a potential source of "compromise" whereas S4 = {4, 14, 15} flatly identifies the wrong source. The reason for this error is that arcs $i = 14$ and 15 fix users 1 and 3 to "normal" flow levels forcing user 2 to compensate for the flow increase on $i = 4$. S2 = {4, 5, 14}, which identifies user 1, also identifies user 3. This uncertainty is caused by user 2 being fixed to "normal" levels on $i = 5$ forcing user 1 to account for the flow increase on $i = 4$. It also forces user 3 to compensate for user 1's increase to maintain the arc constraint equality on $i = 14$.

The individual metric approach yields three of five strategies correctly identifying user 1 as a potential source of "compromise." The key characteristic of the three successful strategies is that none of the arc constraint sets fix user 1 to a "normal" level, unlike S4. This observation suggests that better prediction capabilities may be achieved from a combination of arc samples serving a multitude of varied users concurrently, and from sampling a greater number of "compromised" arcs where the objective is to isolate the single user common to all "compromised" arcs.

The joint metric approach to assessing "compromise" yields mixed results. When $i = 4$ was not included (i.e. S5), the assessment consistently agreed with the individual assessment. However, when $i = 4$ was included only two of the four remaining strategies yielded a correct assessment. Note, however, that the mean compromise estimates for S1 and S2 are relatively close to the lower confidence bound which suggests that a more rigorous approach (i.e. larger sample size) may lead to a more favorable result.

Overall, it is difficult to make a distinction between the performance of both metrics given the size of the O-D sample; however, we can observe for this case that the joint metric is less sensitive to detecting "compromise" and the individual metric provides the added benefit of identifying the potential source. So, for 3-arc strategies the individual metric provides potentially greater insight into the true behavior of the network and should be the preferred metric.

5.5 *Experiment #2*

This experiment investigates the performance of all 2-arc combinations resulting in nine separate strategies of which one is redundant. The results for the eight strategies, namely Steps 3, 4, 6 and 7, are included in Appendices F.1 through F.8. The compromise metrics and "compromise" values for all strategies and p-values are displayed in Table 5.13 and summarized, as in Experiment #1, in Table 5.14.

Having addressed the location model limitations, the key question to be answered in this experiment is how accurate the compromise metrics perform in assessing "compromise" and its source as a function of the norm exponent p.

Of the eight strategies considered four contain $i = 4$ (i.e. S6, S7, S9 and S10). The remaining four strategies do not contain $i = 4$ and do not detect "compromise," which is an expected result. Of the four strategies of interest, two (i.e. S6 and S9) only cover two of the three users and therefore provide no insight on the effects of p on assessment accuracy. In both S6 and S9, user 3 is not covered leading to results similar to removing user 3 and its flow from the network. The estimates are therefore unique for all $p \geq 1$. The results for S7 and S10 are discussed next for both metrics at each level of p.

When $p = 1$, the deviational norm assumes a totally compensatory preference structure: the l_1-norm objective, which views all criteria (i.e. the deviation between each O-D's target and estimated volume) on an equal basis, minimizes the overall deviation by emphasizing a single criteria's loss with an equivalent gain spread over the remaining criteria. In other words, the l_1-norm models the preference structure of a DM bent on identifying a single source. This result is clearly evident for S7 and S10 using the

Table 5.13 Experiment #2 Summary Results ($\lambda_1 < .189$)

Efficient Solution	Strategy	p	Individual Confidence Intervals $k = 1$	$k = 2$	$k = 3$	Joint CIs
$\lambda_1 < .189$	S6 = {4, 12}[a]	1,2,∞	(-.335, 1.833) -1.568	(-.266, 1.497) .811	(-.266, 1.497) .371	(-.132, .66) -.129
	S7 = {4, 15}[b]	1	(.002, 1.496) -1.594	(.071, 1.16) .86	(.071, 1.16) .371	(.205, 1.115) -.121
		2	(.016, 1.482) -.611	(.085, 1.146) -.294	(.085, 1.146) .371	(.219, 1.101) -.178
		∞	(.017, 1.481) -.518	(.086, 1.145) -.407	(.086, 1.145) .371	(.221, 1.099) -.185
	S8 = {12, 15}	1	(.052, 1.446) .749	(.121, 1.11) .851	(.121, 1.11) .371	(.256, 1.064) .657
		2	(.051, 1.447) .746	(.12, 1.111) .85	(.12, 1.111) .371	(.255, 1.065) .656
		∞	(.051, 1.447) .741	(.12, 1.111) .852	(.12, 1.111) .371	(.254, 1.066) .655

a. strategy does not cover user 3 ($k = 3$)
b. strategy also exists for $\lambda_1 > .189$
c. strategy does not cover user 1 ($k = 1$)
d. strategy does not cover user 2 ($k = 2$)

Table 5.13 Experiment #2 Summary Results ($\lambda_1 > .189$)

Efficient Solution	Strategy	p	Individual Confidence Intervals			Joint CIs
			$k = 1$	$k = 2$	$k = 3$	
$\lambda_1 > .189$	S9 = {4, 5}[a]	1,2,∞	(-.002, 1.5) -1.587	(.067, 1.164) .866	(.067, 1.164) .371	(.201, 1.119) -.117
	S10 = {4, 14}	1	(-.001, 1.499) .724	(.068, 1.163) -1.854	(.068, 1.163) .356	(.202, 1.118) -.258
		2	(.006, 1.492) -.492	(.075, 1.156) -.431	(.075, 1.156) .94	(.209, 1.111) .006
		∞	(.002, 1.496) -.533	(.071, 1.16) -.377	(.071, 1.16) .959	(.205, 1.115) .016
	S11 = {5, 14}	1	(.042, 1.456) .749	(.111, 1.12) .866	(.111, 1.12) .345	(.245, 1.075) .653
		2	(.043, 1.455) .741	(.112, 1.119) .866	(.112, 1.119) .348	(.246, 1.074) .652
		∞	(.041, 1.457) .728	(.111, 1.12) .866	(.111, 1.12) .354	(.245, 1.075) .649
	S12 ={5,15}[c]	1,2,∞	(.054, 1.444) .749	(.123, 1.108) .866	(.123, 1.108) .371	(.257, 1.063) .662
	S13={14,15}[d]	1,2,∞	(.019, 1.479) .682	(.088, 1.143) .86	(.088, 1.143) .371	(.222, 1.098) .638

Table 5.14 Experiment #2 Summary Results (cont)

Efficient Solutions	Strategy	p	Individual CIs	Joint CIs
$\lambda_1 < .189$	S6 = {4, 12}	1, 2, ∞	$k = 1$	NO
	S7 = {4, 15}	1	$k = 1$	YES
		2	$k = 1$ & 2	YES
		∞	$k = 1$ & 2	YES
	S8 = {12, 15}	1, 2, ∞	none	NO
$\lambda_1 > .189$	S9 = {4, 5}	1, 2, ∞	$k = 1$	YES
	S10 = {4, 14}	1	$k = 2$	YES
		2	$k = 1$ & 2	YES
		∞	$k = 1$ & 2	YES
	S11 = {5, 14}	1, 2, ∞	none	NO
	S12 = {5, 15}	1, 2, ∞	none	NO
	S13 = {14, 15}	1, 2, ∞	none	NO

individual metric where S7 singles out user 1 and S10 user 2. Whether one user or another is identified in this process is very dependent on the combination of users on the arcs sampled and where the actual "compromise" occurs. For instance, the distinction between S7 and S10 leading to conflicting assessments is that for S10 = {4, 14}, user 1 is common to both a "compromised" arc (i = 4) and a "normal" arc (i = 14), while user 2 is singled out on i = 4. For S7 = {4, 15}, this is not the case. Also, note that the joint metric in this case is consistent with the individual metric assessments.

When p = 2, the deviational norm assumes a preference structure resembling the least-squares method: the l_2-norm objective minimizes the sum of the squared deviations. The DM in this case is not necessarily focused on identifying a single source, but rather desires to minimize the potential contribution of O-Ds with relatively small deviations to help single out the potential source(s). For this case network using the individual metric the two sources identified for S7 and S10 are users 1 and 2, both common to i = 4. The most noticeable change from p = 1 to 2 is the interaction effect between user 1 and 2, which causes the "compromise" values to significantly converge or equalize. The joint metric again consistently agrees with the individual metric.

When $p = \infty$, the deviational norm assumes a totally noncompensatory preference structure: the l_∞-norm objective emphasizes the largest, dominating deviation. The DM in this case is interested in minimizing the maximum deviation regardless of the O-D generating the deviation. This process ignores the remaining O-Ds which results in a system attaining an equilibrium state. The changes in the "compromise" values from

$p = 2$ to ∞ are relatively small. The individual metric maintains users 1 and 2 as potential sources and the joint metric concurs that a "compromise" has occurred. In fact, in both strategies user 1 tends to be the furthest from its respective lower confidence bounds which may be considered relevant information in some ranking scheme to isolate a single source.

Ignoring the strategies not including $i = 4$, which consistently assess no "compromise" and are likely to be a testament of the weakness of the location model rather than the compromise metrics, the four strategies of interest yield the following results:

(a) for $p = 1$: three of four strategies using the individual metric correctly identify user 1 as the *definite* source, and three of four strategies using the joint metric concur that a "compromise" has occurred;

(b) for $p \geq 2$: four of four strategies using the individual metric correctly identify user 1 as a *potential* source, and three of four strategies using the joint metric concur that a "compromise" has occurred.

These results indicate that (1) a possible tradeoff exists between the number of strategies accurately identifying the potential source and the number of sources to be considered, and (2) the individual metric is as sensitive to "compromise" detection as the joint metric.

Introducing risk preference it would appear from the results that a DM with a totally compensatory risk preference structure would be the least conservative. In other words, the risk of incorrectly assessing the source of "compromise" is increased, yet the decision itself is limited to fewer (in this case only one) candidates. As a DM tends toward a total noncompensatory risk preference structure ($p \rightarrow \infty$), the DM becomes more conservative. In other words, the risk of incorrectly assessing the potential source is reduced, however, at the expense of sorting through more candidate sources which introduces additional uncertainty and risk.

The final observations involve O-D estimation errors and their effects on "compromise" assessment. In general, and not surprising given the small size of the network, the estimation process applied to the "normal" scenario produced very accurate estimates with the exception of a couple strategies that, although producing relatively accurate estimates, affected the final assessment.

In Experiment #1, the estimation error, denoted as MSE_{est}, was negligible for strategies S2 through S5 (all ≤ 0.023); however, for S1 = {4, 12, 15}, $MSE_{est} = 0.154$ primarily caused by O-D estimation error in the lower bound estimates. The outcome using the individual metric was unaffected by the error, however the outcome was particularly sensitive using the joint metric. In this case, a "compromise" was not identified when it should have been, and more accurate estimates would have yielded the correct result.

In Experiment #2, we observe a similar result. With the exception of strategy S6 = {4, 12} for $p \geq 1$, the estimation error was consistently less than 0.004. The error for S6 was 0.154 for the same reason as S1 and results in the same outcome.

Overall, we can observe that the estimation error in the "normal" scenario appears to decrease in Experiment #2 where we are sampling from fewer arcs (see Figure 5.3). This is explained by the reduction of the arc constraint set which permits the O-D estimation model to achieve a better solution. We note that a possible tradeoff exists for this situation where the DM must tradeoff between sampling from fewer arcs, which improves the "normal" O-D estimates and tightens the interval bounds, and sampling from more arcs, which increases the probability of intercepting a "compromised" arc.

In the "compromise" scenario we observe a similar trend. However, it is noted that an increase in the "compromise" estimation error (i.e. the average squared deviation between the "compromise" O-D estimates and the target O-Ds) denotes an increased sensitivity of the arc sampling strategy to detecting "compromise," and is therefore desirable. From Figure 5.3, it appears that a larger number of arcs sampled leads to this desired property. This observation supports our intuition that a larger number of sampled arcs should provide a more complete picture of the true state of the network.

Finally, evaluating the effects of changes in the l_p-norm exponent p, we observe that the estimation model, for Experiment #2, is more sensitive to detecting "compromise" (i.e. greater "compromise" estimation error) as the model tends toward total compensation, or $p = 1$ (see Figure 5.4). This supports the earlier observation that the l_1-norm tends to single out an individual source. Given the results for $p = 1$ yield potential inaccurate "compromise" assessments, it appears a tradeoff may exist between an increase in model sensitivity (i.e. greater "compromise" estimation error) and model accuracy (i.e. identifying the correct source).

Figure 5.3 O-D Estimation Error as a Function of Number of Arcs Sampled

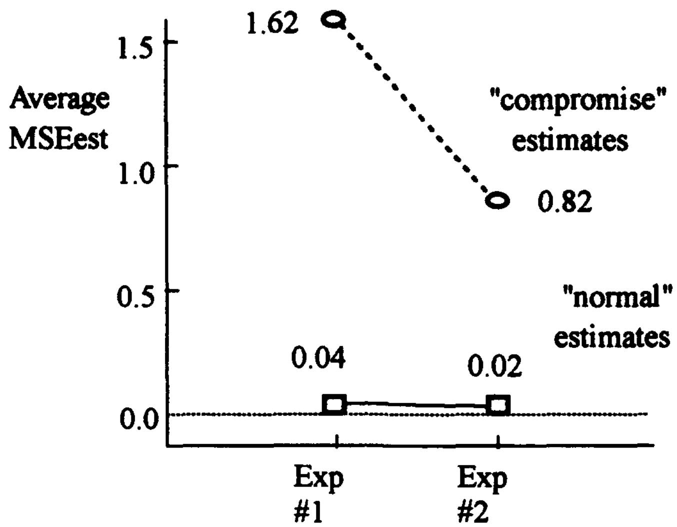
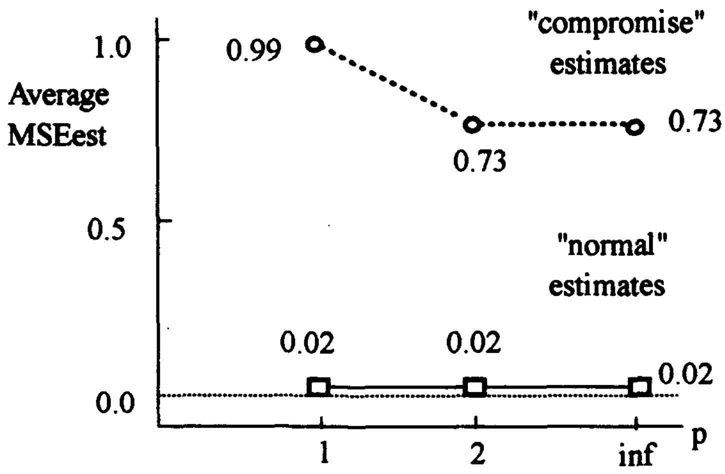


Figure 5.4 O-D Estimation Error as a Function of *lp*-Norm Exponent *p*



VI. Conclusions and Recommendations

This chapter presents the summary of the research approach, its conclusions, and recommendations for future research.

6.1 Summary

The goal of this research was to develop an analytical model for assessing the vulnerability of a multi-commodity, or multi-user, network subject to arc failures. Given that a network achieved a "compromised" state by increasing the flow along a particular O-D pair's path(s), the proposed method for assessing vulnerability was to strategically field flow sensing monitors along specific arcs of the network and to use the sampled information to help construct an image of the original network. More specifically, the image sought was an estimate of the O-D pair volumes input to the network since a "compromised" network would exhibit an increase in O-D volume. The overall objective was to evaluate whether or not the model was capable of accurately assessing "compromise" (with certain confidence), and if so, determine if it could identify the O-D pair source.

The approach consisted of developing a sequence of three separate models and a compromise metric used to make the final assessment. The first model involved bounding the flow through the network given that arcs failed totally with certain probabilities. It is known from prior research that solving for the exact flow in such a network can be intractable given the large number of possible failure states that would need to evaluated separately. Therefore, the first model establishes a lower and upper bound on the exact, or expected, flow through the network by solving simpler linear programs. The second model consisted of locating the specific arcs on which to place the flow monitors subject to possibly conflicting criteria and a predefined number of monitors. This model was formulated as a weighted compromise program to obtain the set of efficient solutions, or the set of possible monitor location strategies. The third model consisted of estimating the O-D pair volumes based on the limited arc sampling information. This model was formulated using l_p-norm objective functions and solved for $p = 1$, 2 and ∞. The estimates for the "normal" case were used to help define the compromise metrics and the estimates for the "compromise" case were used to evaluate the status of the network with respect to the compromise metrics. The compromise metrics were defined as both individual confidence intervals (for each O-D pair) and joint confidence intervals (average of all O-D pairs) around the expected flow (in this case, the deviation of upper to lower O-D volume) in a "normal" network. A "compromise" situation was established if the "compromise" estimates breached the interval bounds.

The performance of the proposed metrics was evaluated using 2 distinct experiments consisting of the set of all 3-arc and 2-arc monitor location strategies analyzed for each level of the norm exponent p. Although the case network used in this research was small, important observations were made that would need to be confirmed with more rigorous experimentation.

6.2 Conclusions

Although it would be premature to draw any concrete conclusions from this research, we were able to put forth several key observations.

First, it was noted that the performance of the location model in selecting susceptible arcs was questionable even though it identified a "compromised" arc. The original objectives of this research did not encompass the development of a location model that accurately selected susceptible arcs. However, it is our contention that a more accurate location model would provide a better assessment of "compromise" by helping to single out the O-D pair common to all "compromised" arcs. Furthermore, an accurate location model may reduce the need for more monitors.

Second, it was observed that the "compromise" assessments differed as a function of *p*. Relating the assessments to a DM's risk preference structure, it was proposed that a conservative DM's preference structure could best be modeled using a totally noncompensatory approach to O-D estimation, whereas a DM willing to accept more risk would have a preference structure totally compensatory in nature. It was also shown that the latter preference structure was more sensitive to detecting "compromise" based on the increased "compromise" estimation error. Although not foolproof, this proposition (if shown to apply in larger and more realistic networks) could provide invaluable information to a DM based on his or her's revealed preference structure.

Third, it was observed that the "normal" and "compromise" estimation error generally decreased as the number of arcs sampled decreased. This led to the observation that a DM may need to reach a compromise in the number of monitors fielded given that fewer monitors increased estimation accuracy of the "normal" scenario yet decreased (1) the probability of selecting a susceptible arc, and (2) the sensitivity of the estimation model in detecting "compromise." This tradeoff is critical given the current location model capabilities, yet may be overcome with a more accurate location model.

Finally, it was determined for the network used that no perceptible difference existed in the accuracy of the individual and joint metrics, although the individual metric provided additional information regarding the actual source of the "compromise." It is expected that an individual metric would generally be preferred given the added insight it provides the DM in assessing the source of "compromise."

6.3 Recommendations

The following recommendations are proposed for future research:

1. Assuming the models and simplifying assumptions are unchanged, more rigorous experiments could be conducted to include (1) small sample testing using the proposed Moran statistic I, and (2) larger networks with significantly more O-D pairs. This work would assess the validity of the observations made in this research.

2. A better understanding of the relationships between the lower and upper bounds and the expected value of the minimum cost flow in multi-commodity networks with failing components should be achieved. Prior research has extensively studied the single-commodity, maximum flow case, yet multi-commodity networks appear to have been generally ignored. Furthermore, we believe the lower bound proposed in this research

provides a relatively tight bound on the expected value. However, the appropriateness of the proposed upper bound remains unanswered.

3. The development of a more accurate location model should be attempted. Depending on the definition of "compromise," the location model criteria, of which there may be many appropriate formulations depending on the DM's objectives, should account for how the "compromise" flow will exhibit itself. For example, in this research we assumed "compromise" flow to be an increase in "normal" flow routed along a particular path(s) according to the pre-established routing algorithm (i.e. minimum cost routing). Therefore, to increase the probability of interception, the criteria should account for the knowledge we possess regarding the routing algorithm, namely the cost of arcs and paths in the network.

4. The definition of "compromise" can be extended to include various other forms of interest such as (1) the rerouting of flow along unexpected or undesirable paths, (2) the increase of flow along a particular arc versus a path, and (3) the unexpected loss of flow along a particular arc or path.

5. A more complex analytical model can be developed for either circuit-switched or packet-switched communication networks. Some ideas for this extension were previously mentioned in Chapter I, Section 1.5.2.

References

1. Aneja, Y. P.; and Nair, K. P. K. "Multicommodity Network Flows with Probabilistic Losses." *Management Science*, Vol 28, pp. 1080-1086 (1982).

2. Bailey, Capt Thomas G. *Response Surface Analysis of Stochastic Network Performance*. MS Thesis. AFIT/GOR/ENS/88D-01. School of Engineering, Air Force Instiitute of Technology (AETC), Wright-Patterson AFB, OH, December 1988 (AD-A202561).

3. Ball, M. "Complexity of Network Reliability Computations." *Networks*, Vol 10 (1980).

4. Bazaraa, M. S., et al. *Linear Programming and Network Flows, 2nd Ed.* John Wiley & Sons, Inc. (1990).

5. Berman, O., et al. "Optimal Location of Discretionary Service Facilities." *Transportation Science*. Vol 26, No 3 (1992).

6. Boesch, F., et al. "Graphs as Models in Communication Network Vulnerability: Connectivity and Persistence." *Networks*, Vol 11 (1981).

7. Brooke, A., et al. *GAMS: A User's Guide*. The International Bank for Reconstruction and Development/The World Bank (1988).

8. Carey, M; and Hendrickson, C. "Bounds on Expected Performance of Networks with Links Subject to Failure." *Networks*. Vol 14, pp. 439-456 (1984).

9. Cascetta, et al. "Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts." *Transportation Science*. Vol 27, No 4 (1993).

10. Chan, Yupo; and Rahi, M. Yunus. "An l_p-Norm Origin-Destination Estimation Method That Minimizes Site-Specific Data Requirements." 4th Revision, 1992.

11. Chan, Y; et al. "Inferring an Origin-Destination Matrix Directly from Network Flow Sampling." *Transportation Planning and Technology*. Vol 11, pp. 27-46 (1986).

12. Current, J. R., et al. "An Interactive Approach to Identify the Best Compromise Solution for Two Objective Shortest Path Problems." *Computers Operations Research*. Vol 17, pp. 187-198 (1990).

13. DoD Briefings, 22 Sep 93 and 15 Dec 93.

14. Evren, Ramazan. "Interactive Compromise Programming." *Journal of the Operational Research Society*. Vol 38, No. 2, pp. 163-172 (1987).

15. Farvolden, J. M., et al. "A Primal Partitioning Solution for the Arc-Chain Formulation of a Multicommodity Network Flow Problem." *Operations Research*. Vol 41 (1993).

16. Gaught, Capt William Lee. *Improving Reliability in a Stochastic Communication Network*. MS Thesis. AFIT/GSO/ENS/90D-10. School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB, OH, December 1990 (AD-A230442).

17. Gedeon, C, et al. "Determining Origin-Destination Matrices and Optimal Multiproduct Flows for Freight Transportation Over Multimodal Networks." *Transportation Research*. Vol 27B, pp. 351-368 (1993).

18. Geoffrion, A. M. "Proper Efficiency and the Theory of Vector Maximization." *Journal of Mathematical Analysis and Applications*. Vol 22, pp. 618-630 (1968).

19. Gerakoulis, D. P. "A Parameter for Vulnerability Evaluation of Packet Switched Communication Networks." *IEEE Global Telecommunications Conference*. Vol 2, pp. 944-48 (1991).

20. Gershon, Mark. "The Role of Weights and Scales in the Application of Multiobjective Decision Making." *European Journal of Operations Research*. Vol 15, pp. 244-50 (1984).

21. Harary, F. *Graph Theory*. Addison-Wesley Publishing Company, Inc, 1972.

22. Hayes, J. F. *Modeling and Analysis of Computer Communications Networks*. New York: Plenum Press (1986).

23. Isensee, E. K.; and Chan, Y. "Optimizing Scout Locations in Forward Areas Along a Transportation Network." Working Paper, 1990.

24. Jensen, Paul A. *Integer Programming Notes*. OPER 620 course notes.

25. Minoux, M. "Network Synthesis and Optimum Network Design Problems: Models, Solution Methods and Applications." *Networks*, Vol 19 (1989).

26. Nemhauser, George L.; and Wolsey, Laurence A. *Integer and Combinatorial Optimization.* John Wiley & Sons, Inc. (1988).

27. Neter, J., et al. *Applied Linear Statistical Models, 3rd Ed.* Richard D. Irwin, Inc, 1990.

28. Onaga, K. "Bounds on the Average Terminal Capacity of Probabilistic Nets." *IEEE Transactions on Information Theory.* pp. 766-768 (1968).

29. Rasmussen, L. M. "Zero-One Programming with Multiple Criteria." *European Journal of Operational Research.* Vol 26, pp. 83-95 (1986).

30. Ross, G. T.; and Soland, R. M. "A Multicriteria Approach to the Location of Public Facilities." *European Journal of Operational Research.* Vol 4, pp. 307-321 (1980).

31. Ross, Sheldon. *A First Course in Probability, 2nd Ed.* Macmillan Publishing Company (1984).

32. Schneider, M. H.; and Zenios, S. A. "A Comparative Study of Algorithms for Matrix Balancing." *Operations Research.* Vol 38 (1990).

33. Sharma, Roshan L. *Network Topology Optimization.* Van Nostrand Reinhold (1990).

34. Shier, Douglas R. *Network Reliability and Algebraic Structures.* Oxford University Press, Oxford (1991).

35. Spragins, J. D., et al. *Telecommunications: Protocols and Design.* Addison-Wesley Publishing Company, Inc. (1991).

36. Upton, G.; and Fingleton, B. *Spatial Data Analysis by Example - Volume I: Point Pattern and Quantitative Data.* John Wiley & Sons (1985).

37. Yim, Capt Eugene. *Improving the Survivability of a Stochastic Communication Network.* MS Thesis. AFIT/GOR/ENS/88D-01. School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB, OH, December 1988 (AD-A202872).

38. Yim, Eugene, et al. "Exact and Approximate Improvement to the Throughput of a Stochastic Network." Working Paper, 2nd Revision, 1992.

39. Yu, Po-Lung. *Multiple-Criteria Decision Making.* Plenum Press, New York (1985)

Appendix A. *Properties of Network Flow Bounds*

Some properties regarding the relationship between the network bound models presented in Chapter III and the expected minimum cost are identified below in Properties 1 through 4. The claim for this research is that the lower and upper bound network models envelop the true or expected network cost. In other words,

$$Z^* \leq Z_{LB}^* \leq Z_{EV} \leq Z_{UB}^*$$

where Z^* is the optimal mimimum network cost when no arcs are subject to failure. For simplicity, the arc index i is used instead of (i, j) where each arc is assigned a single identifying number.

Property 1: The optimal minimum cost, multi-commodity flow in a fully reliable network, Z^*, is less than or equal to the lower bound minimum cost, multi-commodity flow in a network subject to arc failures, Z_{LB}^*, where individual arc capacities, u_i, are set to their expected arc capacities, $e(u_i)$, where $e(u_i) = r_i\, u_i$:

$$Z^* \leq Z_{LB}^*$$

The minimum cost operator will route the maximum flow along arcs with least cost limited only by arc capacities. However, unlike the totally reliable network where all arcs are assumed 100% reliable, the lower bound network further limits the amount of flow traversing any arc with reliability strictly less than 1.0. For any preferred arc i with least cost having reliability less than or equal to 1.0, $f_i\,(u_i) \geq f_i\,(e(u_i))$, where $f_i\,(.)$ represents the optimal flow on arc i subject to arc capacity constraint (.). Therefore, in order to satisfy the demand constraints, the lower bound minimum cost operator will reroute at least as much flow along less preferred, or more costly, arcs than the totally reliable network, which demonstrates the above relationship.

Property 2: The optimal minimum cost, multi-commodity flow in the lower bound network configuration, $Z_{LB}*$, is less than or equal to the *expected* minimum cost, multi-commodity flow, Z_{EV} :

$$Z_{LB}^* \leq Z_{EV}$$

It has been shown for the single-commodity, maximum flow network that the optimal maximum flow when using expected arc capacities yields a total flow greater than or equal to the expected value of the maximum flow [1, 8, 28, 38]. In other words, the expected flow when at least one of the random arc capacities, U_i, is set to its expected capacity, $e(u_i)$, is greater than or equal to the expected flow when all arc capacities remain random variables. The general relationship is given by the following inequality:

$$E\{\text{MaxFlow}(e(u_1), e(u_2), \ldots, e(u_p), \ldots, e(u_{n-m}), U_1, \ldots, U_m)\} \geq$$
$$E\{\text{MaxFlow}(e(u_1), e(u_2), \ldots, e(u_{p-1}), e(u_{p+1}), \ldots, e(u_{n-m}), U_1, \ldots, U_p, \ldots, U_m)\}$$

Generalizing this relationship to the minimum cost, single- or multi-commodity case, it is important to note that the minimum cost objective is similar to the maximum flow objective above in that the minimum cost operator forces the maximum amount of flow over the least costly arcs. If at least one of the preferred arcs has reliability strictly less than 1.0, then by the relation above, the amount of flow traversing this arc when its capacity is set to its expected capacity is greater than or equal to the expected flow traversing this arc. The difference in flow on a particular arc between the lower bound and expected value is rerouted along arcs of equal or higher cost. It follows that the expected system cost is greater than or equal to the deterministic, lower bound system cost. The general relationship in this case becomes:

$$E\{Z_{LB}(e(u_1), e(u_2), \ldots, e(u_p), \ldots, e(u_{n-m}), U_1, \ldots, U_m)\} \leq$$

$$E\{Z_{EV}(e(u_1), e(u_2), \ldots, e(u_{p-1}), e(u_{p+1}), \ldots, e(u_{n-m}), U_1, \ldots, U_p, \ldots, U_m)\}$$

Property 3: The optimal minimum cost, multi-commodity flow in the upper bound network configuration, Z_{UB}^*, where path flows are subject to loss dependent on path reliability rates, is greater than or equal to the *expected* minimum cost, multi-commodity flow, Z_{EV} :

$$Z_{EV} \leq Z_{UB}^*$$

For this relationship to be true, it must also be true that $Z_{LB}^* \leq Z_{UB}^*$. We show this latter relationship first.

(i) For any particular path through a network (say, the path with least cost) the following property applies to the lower and upper bound path flow:

$$\text{LB:} \quad f_{p,\text{LB}} \leq \min(e(u_i))$$

$$\text{UB:} \quad f_{p,\text{UB}} \leq \min(u_i)$$

for all i lying on path p. To meet the demand, b, a new flow, $f_{p'}$, along path p' not equal to p, at equal or higher cost, is included (assuming $f_p < b$) to satisfy the following demand inequalities:

$$\text{LB: } f_{p',LB} \geq b - \min(e(u_i))$$

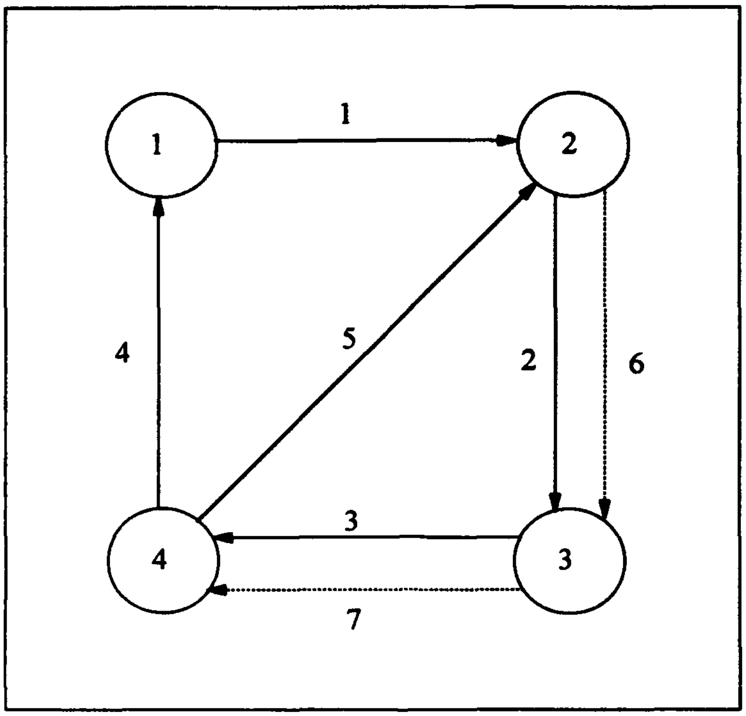
$$\text{UB: } f_{p',UB} \geq b - R_p \min(u_i)$$

where R_p is the loss parameter for path p. Now, note that $R_p \leq \min(r_i)$, for i lying on p. This implies that $R_p \min(u_i) \leq \min(e(u_i)) = \min(r_i\ u_i)$. This, in turn, implies that $f_{p',UB} \geq f_{p',LB}$. Therefore, we can expect that at least as much flow will be routed along more costly paths in the upper bound configuration as in the lower bound configuration, which shows the relationship $Z_{LB}^* \leq Z_{UB}^*$. In fact, for a path p where more than one arc has reliability strictly less than 1.0, R_p is strictly less than $\min(r_i)$; thus, $Z_{LB}^* < Z_{UB}^*$.

(ii) To demonstrate that $Z_{EV} \leq Z_{UB}^*$ we use a specific example to show where the relationship does hold. However, it is not advisable to generalize this result to all networks since no definitive relationship was found. The example network is a two - commodity network, represented in Figure A.1, where all but two arcs are perfectly reliable, resulting in $2^2 = 4$ possible failure states to examine to determine the expected cost. To ensure that both commodity demands are satisfied in the expected value case (where when an arc fails, it fails totally), all unreliable arcs must have a backup arc. For this example, the backup arcs are perfectly reliable and have unlimited capacity, denoted as (*). All arcs, except the two backup arcs, have a unit flow cost of 1; the backup arcs have a cost of m, where m is strictly greater than 1 (the objective is to send as much flow as possible over the primary arcs). The overall objective is to determine if the above properties hold for all $m > 1$. This example network is first evaluated to determine the lower and upper bound costs; then, the expected cost is assessed.

The example network arc parameters, O-D demands and path sets are displayed

Figure A.1 Two-Commodity Example Network



in Tables A.1, 2, and 3 respectively.

Table A.1 Network Arc Parameters

i	r_i	u_i	c_i
1	1	4	1
2	.9	2	1
3	.9	5	1
4	1	1	1
5	1	3	1
6	1	*	m
7	1	*	m

Table A.2 O-D Demands

k	b^k	nodes
1	2	(1, 4)
2	3	(3, 2)

Table A.3 O-D Path Sets

k	p	arcs	R_p^k	c_p^k
1	1	1-2-3	.81	3
	2	1-2-7	.9	$2 + m$
	3	1-6-3	.9	$2 + m$
	4	1-6-7	1	$1 + 2m$
2	5	3-4-1	.9	3
	6	3-5	.9	2
	7	7-4-1	1	$2 + m$
	8	7-5	1	$1 + m$

The lower and upper bound network costs are obtained by solving the linear programs shown in Figures 3.1 and 3.2. The path flows for each bound are displayed below in Tables A.4 and A.5:

Table A.4 O-D Path Flows (Lower Bound)

k	p	f_p^k	cost
1	1	1.8	5.4
	2	0	0
	3	0	0
	4	.2	.2 + .4 m
2	5	0	0
	6	2.7	5.4
	7	0	0
	8	.3	.3 + .3 m
Z_{LB}^* =			11.3 + .7 m

Table A.5 O-D Path Flows (Upper Bound)

k	p	f_p^k	cost
1	1	2	6
	2	0	0
	3	0	0
	4	.32	.32 + .64 m
2	5	0	0
	6	3	6
	7	.33	.66 + .33 m
	8	0	0
Z_{UB}^* =			12.98 + .97 m

To obtain the expected cost we must evaluate the set of failure states S where $s = 1 \ldots 4$ represents the possible combinations of operating and failed arcs, and $P(s)$ the corresponding state probability. The expected value model is presented in Figure 2.2. The four possible failure states are displayed in Figure A.2 and the results summarized in Table A.6.

Table A.6 Expected Value Summary Results

s	$P(s)$	$Z(s)$
1	.81	12
2	.09	$10 + 2\,m$
3	.09	$7 + 5\,m$
4	.01	$5 + 7\,m$
Z_{EV} =		$11.3 + .7\,m$

The results for this example indicate that the lower bound cost and the expected cost are equivalent for all m, but the upper bound cost is strictly greater than both. To further investigate this relationship, the two-commodity network was modified to include three unreliable and three backup arcs, resulting in $2^3 = 8$ possible failure states. Specifically, arc $i = 5$ was assigned a reliability of 0.9 and arc $i = 8$ was included as the backup. The total number of paths was increased to 10 (4 for $k = 1$ and 6 for $k = 2$). The details of this extension are not included in this appendix; however, the overall results for both bounds and expected value are summarized in Table A.7.

Figure A.2 Network Failure State Configurations

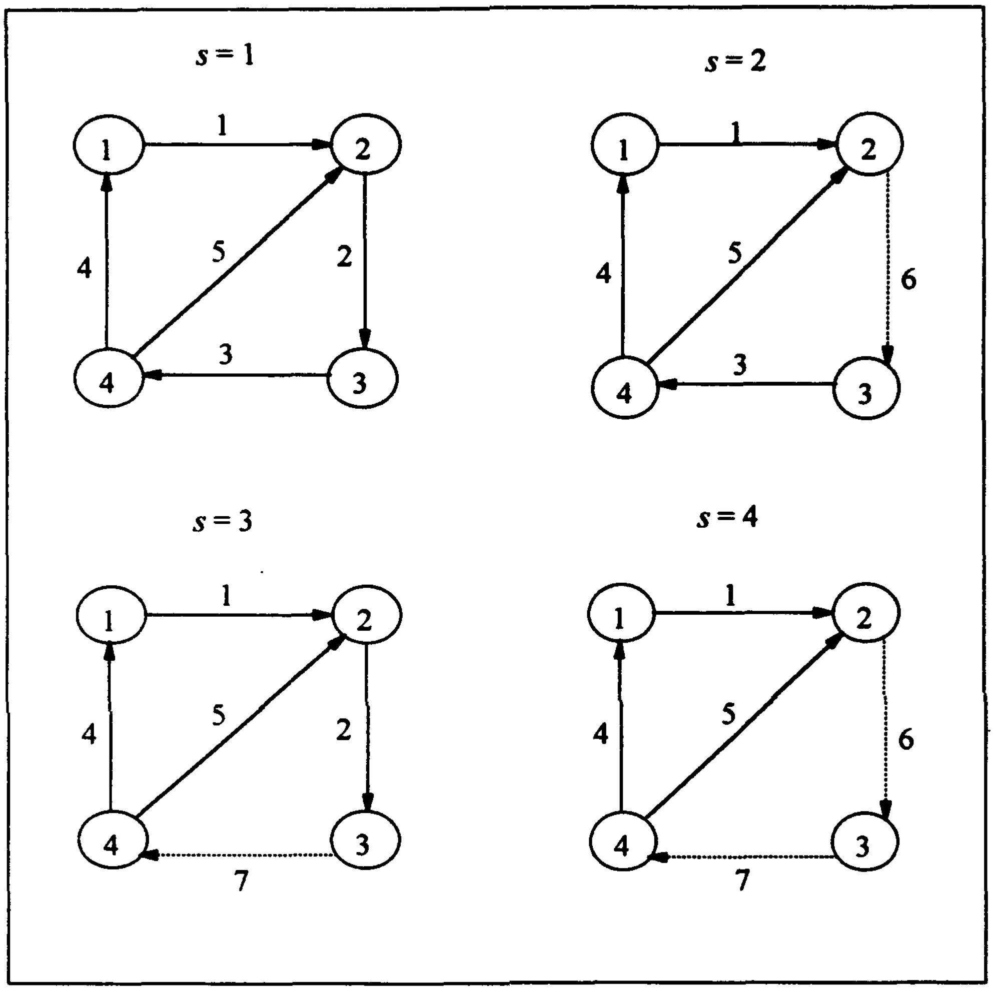


Table A.7 Extended Example Summary Results

$Z_{LB}^* =$	$11.6 + .7\ m$
$Z_{EV} =$	$11.19 + .9\ m$
$Z_{UB}^* =$	$12.42 + 1.88\ m$

The relationships for both examples are illustrated in Figures A.3 and A.4 where it appears that the above properties hold; in fact, it appears that the lower bound model is a tighter bound on the expected value.

Property 4: Tightness of Bounds.

For all arcs where reliability is strictly less than 1.0, as arc reliability increases toward 1.0 the bounds on the minimum cost flow converge on the totally reliable minimum cost. For communication networks, where arc reliabilities generally tend toward unity [28], this behavior is expected. Conversely, if the arc reliabilities decrease, then the bounds progressively widen. The obvious case, where $r_i = 1$ for all i, is that:

$$Z^* = Z_{LB}^* = Z_{EV} = Z_{UB}^*$$

Figure A.3 Linear Relationship of Objective Functions (2 Failing Arcs)

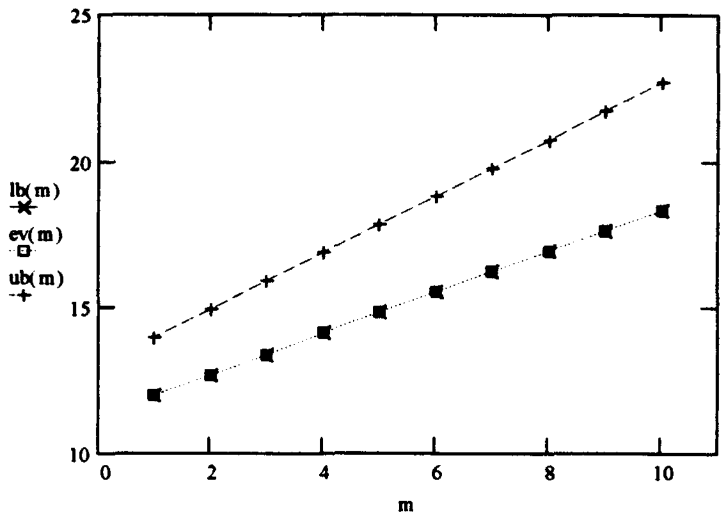
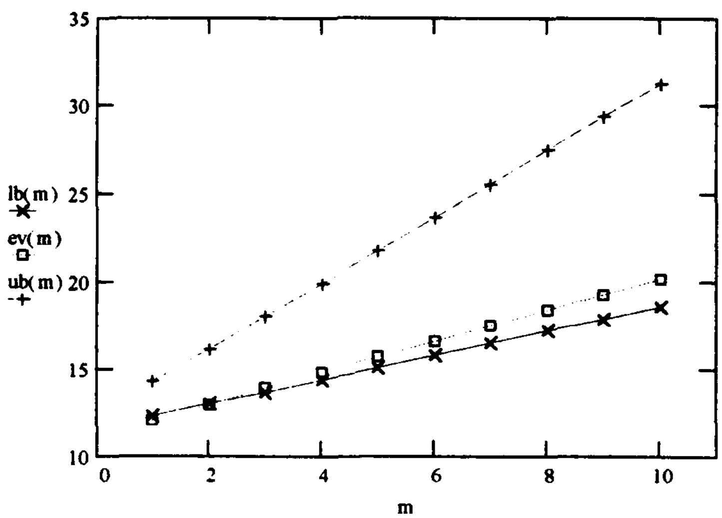


Figure A.4 Linear Relationship of Objective Functions (3 Failing Arcs)



Appendix B. *Network Bounds*

B.1 GAMS Model for Lower Bound Network Flow

```
GAMS 2.20 VAX VMS                                            3-JAN-1994
THREE-USER COMMUNICATION NETWORK (LOWER BOUND)

     SETS
          I        arcs               /1 * 16/
          K        O-D pairs          / USER1 * USER3/
          P        paths              / PATH1 * PATH10/;

     PARAMETERS
          B(K)     O-D pair volumes
                   / USER1  8
                     USER2  6
                     USER3  5 /

          C(I)     arc costs
                   / 1    2
                     2    1
                     3    2
                     4    3
                     5    1
                     6    3
                     7    3
                     8    2
                     9    2
                     10   1
                     11   1
                     12   3
                     13   3
                     14   2
                     15   2
                     16   2 /

          U(I)     arc capacities
                   / 1    7
                     2    5
                     3    5
                     4    10
                     5    4
                     6    8
                     7    4
                     8    10
                     9    5
                     10   3
                     11   5
                     12   9
                     13   10
                     14   9
                     15   6
                     16   4 /

```

```
        R(I)     arc reliabilities
                 / 1   1
                   2   .95
                   3   1
                   4   1
                   5   .98
                   6   .98
                   7   1
                   8   .7
                   9   .8
                   10  1
                   11  1
                   12  1
                   13  1
                   14  .98
                   15  .95
                   16  .95 / ;

TABLE D(P,K)           path-user assignment matrix
                  USER1  USER2  USER3
      PATH1         1
      PATH2         1
      PATH3         1
      PATH4         1
      PATH5                1
      PATH6                1
      PATH7                1
      PATH8                1
      PATH9                       1
      PATH10                      1  ;

TABLE A(P,K,I)         path-arc incidence matrix by user
                  1   2   3   4   5   6   7   8   9   10  11  12  13  14  15  16
PATH1.USER1       1           1       1                           1
PATH2.USER1       1           1                               1       1
PATH3.USER1           1               1       1                   1
PATH4.USER1           1                       1               1       1
PATH5.USER2               1   1                   1   1
PATH6.USER2               1   1                               1               1
PATH7.USER2                       1       1       1   1
PATH8.USER2                       1       1                   1               1
PATH9.USER3                           1       1           1       1
PATH10.USER3                                                          1   1 ;

PARAMETER EU(I)    expected arc capacities ;

          EU(I) = R(I) * U(I) ;

PARAMETER CP(P,K)  path costs;

          CP(P,K) = SUM(I, D(P,K) * A(P,K,I) * C(I)) ;

VARIABLES
     F(P,K)  amount of flow on path p from user k
     Z       total costs (in equivalent miles)  ;

POSITIVE VARIABLE F ;
```

```
EQUATIONS
     COST              define objective function
     VOLUME(K)         satisfy O-D pair volume demands
     CAPACITY(I)       satisfy arc capacity constraints ;

COST ..        Z =E= SUM(K, SUM(P, CP(P,K) * D(P,K) * F(P,K))) ;

VOLUME(K) ..         SUM(P, D(P,K) * F(P,K)) =E= B(K) ;

CAPACITY(I) ..       SUM((K,P), A(P,K,I) * F(P,K)) =L= EU(I) ;

MODEL NK3LB /ALL/ ;

SOLVE NK3LB USING LP MINIMIZING Z ;

DISPLAY F.L;

PARAMETER FARC(I)    arc flows;

     FARC(I) = SUM(P, SUM(K, A(P,K,I) * F.L(P,K)));

DISPLAY FARC;

PARAMETER ROUTE(I,K)    routing matrix;

     ROUTE(I,K) = (SUM(P, A(P,K,I) * F.L(P,K))) / B(K);

DISPLAY ROUTE;

PARAMETER BI(I,K)    arc-user assignment matrix;

     BI(I.K) = 1$(SUM(P, A(P,K,I) * F.L(P,K)) GT 0);

DISPLAY BI;

PARAMETER PROP(I)    proportion of flow on arc i for users k of i;

     PROP(I) = FARC(I) / (SUM(K$(BI(I,K) NE 0), BI(I,K) * B(K)));

DISPLAY PROP;
```

B.2 GAMS Model for Upper Bound Network Flow

```
GAMS 2.20 VAX VMS                                          3-JAN-1994
THREE-USER COMMUNICATION NETWORK (UPPER BOUND)


        SETS
             I        arcs                /1 * 16/
             K        O-D pairs           / USER1 * USER3/
             P        paths               / PATH1 * PATH10/;

        PARAMETERS
             B(K)     O-D pair volumes
                      / USER1  8
                        USER2  6
                        USER3  5 /

             C(I)     arc costs
                      / 1   2
                        2   1
                        3   2
                        4   3
                        5   1
                        6   3
                        7   3
                        8   2
                        9   2
                        10  1
                        11  1
                        12  3
                        13  3
                        14  2
                        15  2
                        16  2 /

             U(I)     arc capacities
                      / 1   7
                        2   5
                        3   5
                        4   10
                        5   4
                        6   8
                        7   4
                        8   10
                        9   5
                        10  3
                        11  5
                        12  9
                        13  10
                        14  9
                        15  6
                        16  4 /

             R(I)     arc reliabilities
                      / 1   1
                        2   .95
                        3   1
                        4   1
                        5   .98
```

```
                    6   .98
                    7   1
                    8   .7
                    9   .8
                    10  1
                    11  1
                    12  1
                    13  1
                    14  .98
                    15  .95
                    16  .95 /

        RP(P)     path reliabilities
                 / PATH1    .98
                   PATH2    .98
                   PATH3    .6517
                   PATH4    .6517
                   PATH5    .8
                   PATH6    .95
                   PATH7    .784
                   PATH8    .931
                   PATH9    .686
                   PATH10   .931 / ;

TABLE D(P,K)             path-user assignment matrix
                    USER1  USER2  USER3
      PATH1           1
      PATH2           1
      PATH3           1
      PATH4           1
      PATH5                  1
      PATH6                  1
      PATH7                  1
      PATH8                  1
      PATH9                         1
      PATH10                        1  ;

TABLE A(P,K,I)           path-arc incidence matrix by user
                   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16
PATH1.USER1        1           1       1                           1
PATH2.USER1        1           1                               1       1
PATH3.USER1            1               1       1                   1
PATH4.USER1            1                       1               1       1
PATH5.USER2                1   1                   1   1
PATH6.USER2                1   1                               1               1
PATH7.USER2                        1       1       1   1
PATH8.USER2                        1       1                   1               1
PATH9.USER3                            1       1           1       1
PATH10.USER3                                                           1   1 ;

PARAMETER CP(P,K)  path costs;

          CP(P,K) = SUM(I, D(P,K) * A(P,K,I) * C(I)) ;

VARIABLES
     F(P,K)  amount of flow on path p from user k
     Z       total costs (in equivalent miles)  ;

POSITIVE VARIABLE F ;
```

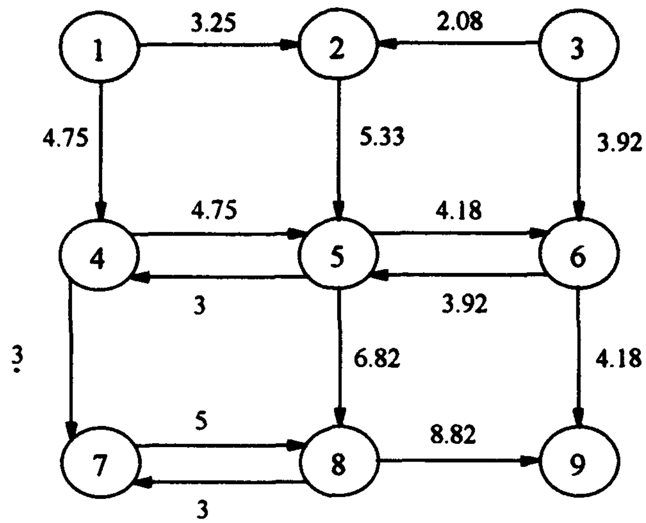
```
EQUATIONS
     COST            define objective function
     VOLUME(K)       satisfy O-D pair volume demands
     CAPACITY(I)     satisfy arc capacity constraints ;

COST ..       Z =E= SUM(K, SUM(P, CP(P,K) * D(P,K) * F(P,K))) ;

VOLUME(K) ..        SUM(P, RP(P) * D(P,K) * F(P,K)) =E= B(K) ;

CAPACITY(I) ..      SUM((K,P), A(P,K,I) * F(P,K)) =L= U(I) ;

MODEL NK3UB /ALL/ ;

SOLVE NK3UB USING LP MINIMIZING Z ;

DISPLAY F.L;

PARAMETER FARC(I)    arc flows;

     FARC(I) = SUM(P, SUM(K, A(P,K,I) * F.L(P,K)));

DISPLAY FARC;

PARAMETER BU(K)    revised O-D demands;

     BU(K) = SUM(P, F.L(P,K));

DISPLAY BU;

PARAMETER ROUTE(I,K)    routing matrix;

     ROUTE(I,K) = (SUM(P, A(P,K,I) * F.L(P,K))) / BU(K);

DISPLAY ROUTE;

PARAMETER BI(I,K)    arc-user assignment matrix;

     BI(I,K) = 1$(SUM(P, A(P,K,I) * F.L(P,K)) GT 0);

DISPLAY BI;

PARAMETER PROP(I)    proportion of flow on arc i for users k of i;

     PROP(I) = FARC(I) / (SUM(K$(BI(I,K) NE 0), BI(I,K) * BU(K)));

DISPLAY PROP;
```

B.3 Optimal Flow Patterns

The optimal arc flow values for both the lower and upper bound networks in both the "normal" and "compromise" scenarios are illustrated in Figures B.1 and B.2. Note that the thicker arcs in the "compromise" scenarios denote the "compromise" arcs.

Figure B.1 Lower Bound Network Flow Patterns

"Normal" Scenario



"Compromise" Scenario

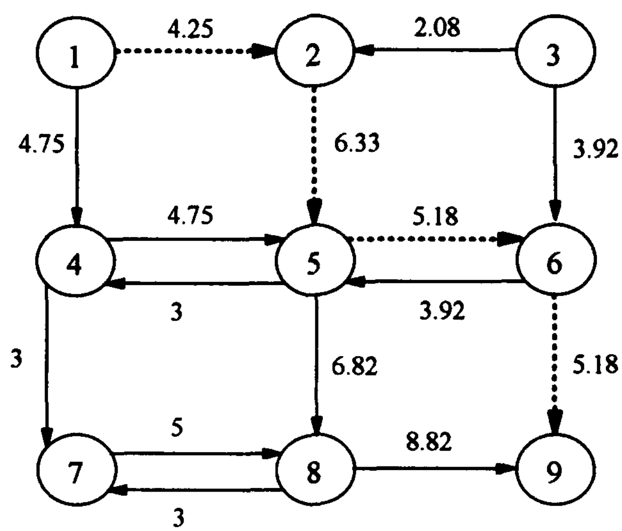
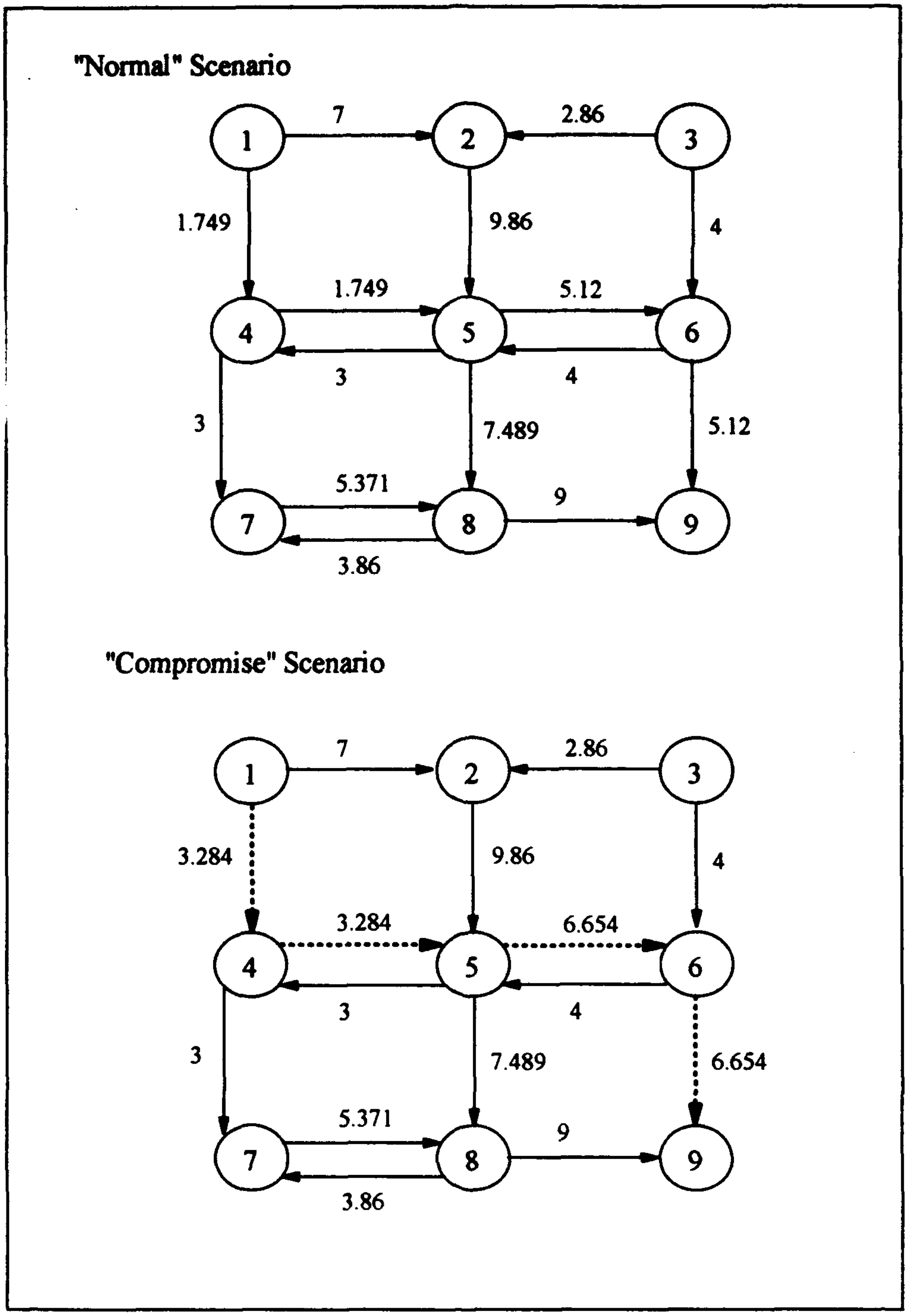


Figure B.2 Upper Bound Network Flow Patterns



Appendix C. *Monitor Location Models*

C.1 *GAMS Model for Lower Bound Monitor Placement*

```
GAMS 2.20 VAX VMS                                        24-JAN-1994
THREE-USER COMMUNICATION NETWORK - LOCATION MODEL, N-SET (LOWER BOUND)

         SETS
              I         arcs               /1 * 16/
              K         O-D pairs          / USER1 * USER3/
              KP        k' set             / 1, 2, 3, 4, 5 /
              PAIR(K,KP)   k in k'         / USER1.1, USER1.4, USER1.5,
                                             USER2.2, USER2.4,
                                             USER3.3, USER3.5 /
              J         weight index       /1 * 11/ ;

         PARAMETERS
              C(I)      arc costs
                        / 1   2
                          2   1
 .
 .
                          16  2 /

              U(I)      arc capacities
                        / 1   7
                          2   5
 .
 .
                          16  4 /

              R(I)      arc reliabilities
                        / 1   1
                          2   .95
 .
 .
                          16  .95 /

              FARC(I)    arc flow values
                        / 1   3.25
                          2   4.75
                          3   2.08
                          4   5.33
                          5   3.92
                          6   4.18
                          7   3.92
                          8   4.75
                          9   3
                          10  3
                          11  5
                          12  6.82
                          13  4.18
                          14  8.82
                          15  5
                          16  3     /
```

```
        PROP(I)  arc flow proportions
                / 1   .406
                  2   .594
                  3   .347
                  4   .381
                  5   .653
                  6   .522
                  7   .653
                  8   .594
                  9   .5
                  10  .5
                  11  0
                  12  .487
                  13  .522
                  14  .678
                  15  1
                  16  .5   /

        NSET(*)   effecient solutions for varying weights ;

   TABLE E(I,KP)  arc-user coverage assignment matrix
                    1  2  3  4  5
              1     1
              2     1
              3        1
              4              1
              5        1
              6     1
              7        1
              8     1
              9        1
              10       1
              11
              12             1
              13    1
              14                1
              15          1
              16       1           ;

   PARAMETER WA(I)    remaining arc capacities ;

           WA(I) = U(I) - FARC(I) ;

   PARAMETER W(I)     arc consideration;

           W(I) = 1$(WA(I) GT 0) ;

   PARAMETER V(I)     arc reliability to cost ratio;

           V(I) = R(I) / C(I) ;

   SCALAR M           number of monitors;

           M = 2 ;

   SCALAR WEIGHT      ;

           WEIGHT = 0.0 ;
```

```
VARIABLES
     X(I,KP)  binary monitor location decision variable
     Y1       criterion function 1
     Y2       criterion function 2
     Y        compromise function (Y1 and Y2) ;
BINARY VARIABLE X ;
EQUATIONS
     CF1               objective function (CF1 only)
     CF2               objective function (CF2 only)
     REDUND(KP)        redundancy avoidance constraints
     GOAL              number of available monitors
     COVER(K)          minimum user coverage constraints ;
CF1 ..  Y1 =E= SUM(I, SUM(KP, E(I,KP)*W(I)*V(I)*X(I,KP))) ;
CF2 ..  Y2 =E= SUM(I, SUM(KP, E(I,KP)*W(I)*PROP(I)*X(I,KP))) ;
REDUND(KP) ..  SUM(I, E(I,KP)*X(I,KP)) =L= 1 ;
GOAL ..        SUM(I, SUM(KP, E(I,KP)*X(I,KP))) =E= M ;
COVER(K) ..    SUM(PAIR(K,KP), SUM(I, E(I,KP)*X(I,KP))) =G= 1 ;
MODEL OPTCF1 /CF1, REDUND, GOAL, COVER/ ;
MODEL OPTCF2 /CF2, REDUND, GOAL, COVER/ ;
SOLVE OPTCF1 USING MIP MAXIMIZING Y1 ;
SOLVE OPTCF2 USING MIP MAXIMIZING Y2 ;
DISPLAY "CF1 AT OPTIMUM", Y1.L;
DISPLAY "CF2 AT OPTIMUM", Y2.L;
PARAMETER ALPHA(I)   compromise coefficient ;
   ALPHA(I) = W(I)*(((WEIGHT*V(I))/Y1.L) + (((1 -
         WEIGHT)*PROP(I))/Y2.L)) ;
EQUATION
     CS                compromise solution (CF1 and CF2) ;
CS ..   Y =E= SUM(I, SUM(KP, E(I,KP)*ALPHA(I)*X(I,KP)));
MODEL COMPSOL /CS, REDUND, GOAL, COVER/ ;
SOLVE COMPSOL USING MIP MAXIMIZING Y ;
NSET('1') = Y.L
SCALAR SHORT   percent shortfall from ideal solution ;
       SHORT = (1 - Y.L) * 100 ;
DISPLAY "PERCENT SHORTFALL FOR WEIGHT =", WEIGHT, SHORT ;
DISPLAY "MONITOR LOCATION STRATEGY (LOWER BOUND)", X.L ;
WEIGHT = 0.1 ;
```

```
PARAMETER ALPHA(I)   compromise coefficient ;

   ALPHA(I) = W(I)*(((WEIGHT*V(I))/Y1.L) + (((1 -
         WEIGHT)*PROP(I))/Y2.L)) ;

SOLVE COMPSOL USING MIP MAXIMIZING Y ;

NSET('2') = Y.L

SCALAR SHORT   percent shortfall from ideal solution ;

         SHORT = (1 - Y.L) * 100 ;

DISPLAY "PERCENT SHORTFALL FOR WEIGHT 1 =", WEIGHT, SHORT ;

DISPLAY "MONITOR LOCATION STRATEGY (LOWER BOUND)", X.L ;

WEIGHT = 0.2 ;

PARAMETER ALPHA(I)   compromise coefficient ;

   ALPHA(I) = W(I)*(((WEIGHT*V(I))/Y1.L) + (((1 -
         WEIGHT)*PROP(I))/Y2.L)) ;

SOLVE COMPSOL USING MIP MAXIMIZING Y ;

NSET('3') = Y.L

SCALAR SHORT   percent shortfall from ideal solution ;

         SHORT = (1 - Y.L) * 100 ;

DISPLAY "PERCENT SHORTFALL FOR WEIGHT 1 =", WEIGHT, SHORT ;

DISPLAY "MONITOR LOCATION STRATEGY (LOWER BOUND)", X.L ;
.
.
.
WEIGHT = 1.0 ;

PARAMETER ALPHA(I)   compromise coefficient ;

   ALPHA(I) = W(I)*(((WEIGHT*V(I))/Y1.L) + (((1 -
         WEIGHT)*PROP(I))/Y2.L)) ;

SOLVE COMPSOL USING MIP MAXIMIZING Y ;

NSET('11') = Y.L

SCALAR SHORT   percent shortfall from ideal solution ;

         SHORT = (1 - Y.L) * 100 ;

DISPLAY "PERCENT SHORTFALL FOR WEIGHT 1 =", WEIGHT, SHORT ;

DISPLAY "MONITOR LOCATION STRATEGY (LOWER BOUND)", X.L ;

DISPLAY NSET ;
```

C.2 GAMS Model for Upper Bound Monitor Placement

```
GAMS 2.20 VAX VMS                                              24-JAN-1994
THREE-USER COMMUNICATION NETWORK - LOCATION MODEL, N-SET (UPPER BOUND)

        SETS
             I          arcs              /1 * 16/
             K          O-D pairs         / USER1 * USER3/
             KP         k' set            / 1, 2, 3, 4, 5 /
             PAIR(K,KP)  k in k'          / USER1.1, USER1.4, USER1.5,
                                            USER2.2, USER2.4,
                                            USER3.3, USER3.5 /
             J          weight index      /1 * 11/ ;

        PARAMETERS
             C(I)       arc costs
                        / 1   2
                          2   1
.
.
                    .     16  2 /

             U(I)       arc capacities
                        / 1   7
                          2   5
.
.
                          16  4 /

             R(I)       arc reliabilities
                        / 1   1
                          2   .95
.
.
                          16  .95 /

             FARC(I)     arc flow values
                        / 1   7
                          2   1.749
                          3   2.86
                          4   9.86
                          5   4
                          6   5.12
                          7   4
                          8   1.749
                          9   3
                          10  3
                          11  5
                          12  7.489
                          13  5.12
                          14  9
                          15  5.371
                          16  3.86    /

             PROP(I)    arc flow proportions
                        / 1   .8
                          2   .2
                          3   .417
                          4   .632
```

```
               5   .583
               6   .585
               7   .583
               8   .2
               9   .437
               10  .437
               11  0
               12  .48
               13  .585
               14  .637
               15  1
               16  .563   /

     NSET(*)   effecient solutions for varying weights ;

TABLE E(I,KP)  arc-user coverage assignment matrix
                1  2  3  4  5
           1    1
           2    1
           3       1
           4             1
           5       1
           6    1
           7       1
           8    1
           9       1
           10      1
           11
           12            1
           13   1
           14               1
           15         1
           16      1           ;

PARAMETER WA(I)    remaining arc capacities ;

          WA(I) = U(I) - FARC(I) ;

PARAMETER W(I)     arc consideration;

          W(I) = 1$(WA(I) GT 0) ;

PARAMETER V(I)     arc reliability to cost ratio;

          V(I) = R(I) / C(I) ;

SCALAR M           number of monitors;

          M = 2 ;

SCALAR WEIGHT      ;

          WEIGHT = 0.0 ;

VARIABLES
     X(I,KP)  binary monitor location decision variable
     Y1       criterion function 1
     Y2       criterion function 2
     Y        compromise function (Y1 and Y2) ;
```

```
BINARY VARIABLE X ;

EQUATIONS
     CF1              objective function (CF1 only)
     CF2              objective function (CF2 only)
     REDUND(KP)       redundancy avoidance constraints
     GOAL             number of available monitors
     COVER(K)         minimum user coverage constraints ;

CF1 ..  Y1 =E= SUM(I, SUM(KP, E(I,KP)*W(I)*V(I)*X(I,KP))) ;

CF2 ..  Y2 =E= SUM(I, SUM(KP, E(I,KP)*W(I)*PROP(I)*X(I,KP))) ;

REDUND(KP) ..  SUM(I, E(I,KP)*X(I,KP)) =L= 1 ;

GOAL ..        SUM(I, SUM(KP, E(I,KP)*X(I,KP))) =E= M ;

COVER(K) ..    SUM(PAIR(K,KP), SUM(I, E(I,KP)*X(I,KP))) =G= 1 ;

MODEL OPTCF1 /CF1, REDUND, GOAL, COVER/ ;
MODEL OPTCF2 /CF2, REDUND, GOAL, COVER/ ;

SOLVE OPTCF1 USING MIP MAXIMIZING Y1 ;
SOLVE OPTCF2 USING MIP MAXIMIZING Y2 ;

DISPLAY "CF1 AT OPTIMUM", Y1.L;
DISPLAY "CF2 AT OPTIMUM", Y2.L;

PARAMETER ALPHA(I)   compromise coefficient ;

   ALPHA(I) = W(I)*(((WEIGHT*V(I))/Y1.L) + (((1 -
         WEIGHT)*PROP(I))/Y2.L)) ;

EQUATION
     CS               compromise solution (CF1 and CF2) ;

CS ..   Y =E= SUM(I, SUM(KP, E(I,KP)*ALPHA(I)*X(I,KP)));

MODEL COMPSOL /CS, REDUND, GOAL, COVER/ ;

SOLVE COMPSOL USING MIP MAXIMIZING Y ;

NSET('1') = Y.L

SCALAR SHORT   percent shortfall from ideal solution ;

       SHORT = (1 - Y.L) * 100 ;

DISPLAY "PERCENT SHORTFALL FOR WEIGHT =", WEIGHT, SHORT ;

DISPLAY "MONITOR LOCATION STRATEGY (UPPER BOUND)", X.L ;

WEIGHT = 0.1 ;

PARAMETER ALPHA(I)   compromise coefficient ;

   ALPHA(I) = W(I)*(((WEIGHT*V(I))/Y1.L) + (((1 -
         WEIGHT)*PROP(I))/Y2.L)) ;

SOLVE COMPSOL USING MIP MAXIMIZING Y ;
```

```
NSET('2') = Y.L

SCALAR SHORT   percent shortfall from ideal solution ;

       SHORT = (1 - Y.L) * 100 ;

DISPLAY "PERCENT SHORTFALL FOR WEIGHT 1 =", WEIGHT, SHORT ;

DISPLAY "MONITOR LOCATION STRATEGY (UPPER BOUND)", X.L ;
.
.
WEIGHT = 1.0 ;

PARAMETER ALPHA(I)   compromise coefficient ;

   ALPHA(I) = W(I)*(((WEIGHT*V(I))/Y1.L) + (((1 -
         WEIGHT)*PROP(I))/Y2.L)) ;

SOLVE COMPSOL USING MIP MAXIMIZING Y ;

NSET('11') = Y.L

SCALAR SHORT   percent shortfall from ideal solution ;

       SHORT = (1 - Y.L) * 100 ;

DISPLAY "PERCENT SHORTFALL FOR WEIGHT 1 =", WEIGHT, SHORT ;

DISPLAY "MONITOR LOCATION STRATEGY (UPPER BOUND)", X.L ;

DISPLAY NSET ;
```

Appendix D. *Algorithm Results for Strategy S1*

D.1 GAMS O-D Estimation Model (Both Bounds)

```
GAMS 2.20 VAX VMS                                        26-JAN-1994
THREE-USER COMMUNICATION NETWORK - O-D ESTIMATION MODEL (LB & UB)

        SETS
             S          selected arcs (monitors) /4, 12, 15/
             K          O-D pairs        / USER1 * USER3/
             B          network bounds   / L, U /
             P          norm exponents   / p1, p2, pn / ;

        PARAMETERS
             BL(K)      O-D pair external flow (lower bound)
                        / USER1   8
                          USER2   6
                          USER3   5 /

             BU(K)      O-D pair external flow (upper bound)
                        / USER1   8.749
                          USER2   6.86
                          USER3   5.371 /

             FARCL(S)      observed flow on selected arcs (LB)
                           / 4     5.33
                             12    6.82
                             15    5   /

             FARCU(S)      observed flow on selected arcs (UB)
                           / 4     9.86
                             12    7.489
                             15    5.371     /

             F(*,*,*)      O-D estimate report ;

        TABLE GL(S,K)   routing matrix for selected arcs (LB)
                          USER1   USER2   USER3
                     4     .41     .35
                     12    .48     .50
                     15                      1 ;

        TABLE GU(S,K)   routing matrix for selected arcs (UB)
                          USER1   USER2   USER3
                     4     .8      .42
                     12    .42     .56
                     15                      1 ;

        VARIABLES
             FL(K)        O-D demand estimates for pair k (LB)
             FU(K)        O-D demand estimates for pair k (UB)
             DM(K)        deviational measure (minus)
             DP(K)        deviational measure (plus)
             D(K)         sum of deviational measures
             V            variable used in p=infinity
```

```
          Y1       P=1 function
          Y2       p=2 function
          YN       p=infinity function ;

     POSITIVE VARIABLES FL, FU, DM, DP, D, V ;

     EQUATIONS
          R1                  P=1 norm
          R2                  p=2 norm
          RN                  p=infinity norm
          DEV(K)              sum of deviational variables
          GOAL1(K)            constraint for p (inf)
          GOAL2L(K)           constraint for p (1 2 inf) LB
          GOAL2U(K)           constraint for p (1 2 inf) UB
          ARCL(S)             arc flow counts (LB)
          ARCU(S)             arc flow counts (UB) ;

     R1 ..   Y1 =E= SUM(K,  (DM(K) + DP(K))) ;

     R2 ..   Y2 =E= SUM(K,  (DM(K)**2 + 2*DM(K)*DP(K) + DP(K)**2)) ;

     RN ..   YN =E= V ;

     GOAL1(K) ..   V - DM(K) - DP(K) =G= 0 ;

     GOAL2L(K) ..  FL(K) + DM(K) - DP(K) =E= BL(K) ;

     GOAL2U(K) ..  FU(K) + DM(K) - DP(K) =E= BU(K) ;

     ARCL(S) ..    SUM(K, GL(S,K) * FL(K)) =E= FARCL(S) ;

     ARCU(S) ..    SUM(K, GU(S,K) * FU(K)) =E= FARCU(S) ;

     MODEL P1LB /R1, GOAL2L, ARCL/ ;
     SOLVE P1LB USING LP MINIMIZING Y1 ;
       F(K,'L','p1') = FL.L(K) ;

     MODEL P2LB /R2, GOAL2L, ARCL/ ;
     SOLVE P2LB USING NLP MINIMIZING Y2 ;
       F(K,'L','p2') = FL.L(K) ;

     MODEL PNLB /RN, GOAL1, GOAL2L, ARCL/ ;
     SOLVE PNLB USING LP MINIMIZING YN ;
       F(K,'L','pn') = FL.L(K) ;

     MODEL P1UB /R1, GOAL2U, ARCU/ ;
     SOLVE P1UB USING LP MINIMIZING Y1 ;
       F(K,'U','p1') = FU.L(K) ;

     MODEL P2UB /R2, GOAL2U, ARCU/ ;
     SOLVE P2UB USING NLP MINIMIZING Y2 ;
       F(K,'U','p2') = FU.L(K) ;

     MODEL PNUB /RN, GOAL1, GOAL2U, ARCU/ ;
     SOLVE PNUB USING LP MINIMIZING YN ;
       F(K,'U','pn') = FU.L(K) ;

     DISPLAY 'O-D ESTIMATES', F;
```

D.2 Strategy S1

Experiment #1, Strategy S1 = (4, 12, 15)
- relaxed equality constraints for compromise estimates

ORIGIN = 1

Bounds	**# of O-D Pairs**	**O-D Index**
$i := 1..2$	$t := 3$	$k := 1..t$

O-D External Flows:

p = 1, 2 and inf

Target: "Normal" Estimates: "Compromise" Estimates:

$$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \quad F_N := \begin{pmatrix} 7.514 & 8.749 \\ 6.427 & 6.811 \\ 5 & 5.371 \end{pmatrix} \quad F_C := \begin{pmatrix} 10.317 & 8.749 \\ 6 & 6.811 \\ 5 & 5.371 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \quad D_{N_k} := F_{N_{k,2}} - F_{N_{k,1}} \quad D_{C_k} := F_{C_{k,2}} - F_{C_{k,1}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \quad D_N = \begin{pmatrix} 1.235 \\ 0.384 \\ 0.371 \end{pmatrix} \quad D_C = \begin{pmatrix} -1.568 \\ 0.811 \\ 0.371 \end{pmatrix}$$

Estimation Error, MSE_{est}:
$$MSE_{est} := \frac{\sum_k \left(D_{N_k} - Y_k\right)^2}{t} \qquad MSE_{est} = 0.154$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_k := Yhat_k - \left(2 \cdot \sqrt{s2\ Yhat_k + MSE_{est}}\right)$$

$$EUCL_k := Yhat_k + \left(2 \cdot \sqrt{s2\ Yhat_k + MSE_{est}}\right)$$

$$ELCL = \begin{pmatrix} -0.335 \\ -0.266 \\ -0.266 \end{pmatrix} \quad Yhat = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \quad EUCL = \begin{pmatrix} 1.833 \\ 1.497 \\ 1.497 \end{pmatrix}$$

Compromise Estimates:

$$D_C = \begin{pmatrix} -1.568 \\ 0.811 \\ 0.371 \end{pmatrix}$$

So, only O-D Pair k = 1 is outside its respective confidence interval. Conclude that User 1 is the source of compromise.

Joint Confidence Interval for E(Y):

$$JLCL = Ybar - \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}}\right)$$

$$JUCL = Ybar + \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}}\right)$$

$JLCL = -0.132$ $\quad$ $Ybar = 0.66$ $\quad$ $JUCL = 1.452$

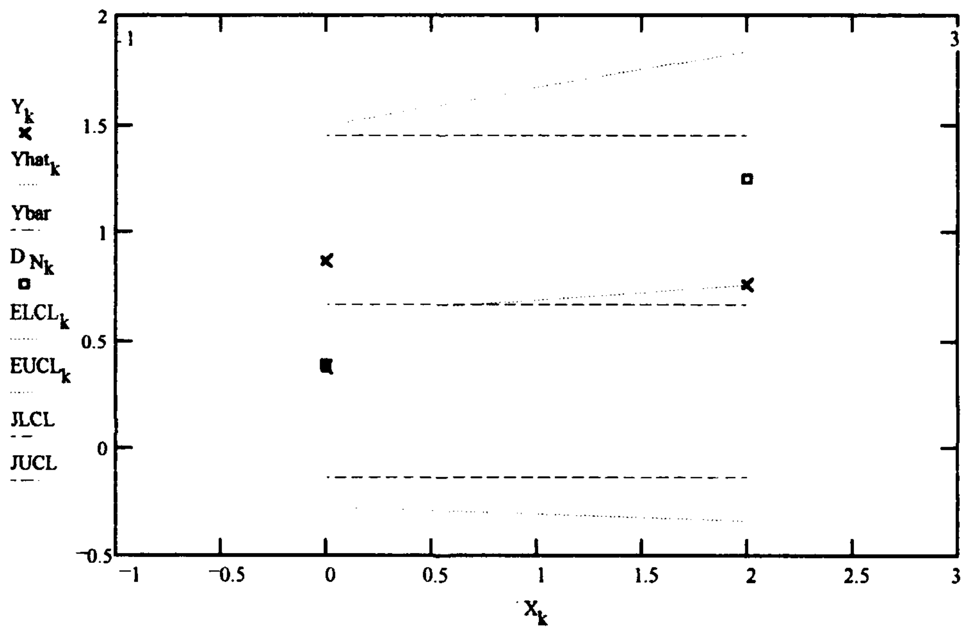
Compromise Mean Estimate:

$$Dbar_C = mean(D_C)$$

$Dbar_C = -0.129$

Here, however, can not conclude that a compromise has occured.

Graph:



D.3 *Initial Compromise Metric (Target Data Only)*

ORIGIN ≡ 1

Bounds (1 = LB; 2 = UB) # of O-D Pairs O-D Index

$i := 1..2$ $t := 3$ $k := 1..t$

Target O-D External Flows, b^k: $b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix}$

Deviations, Y_k:

$Y_k := b_{k,2} - b_{k,1}$ $Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix}$

Regression Model:

Independent Variable, X_k: $X := \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$

$Xbar := \text{mean}(X)$ $Xbar = 0.667$

$Ybar := \text{mean}(Y)$ $Ybar = 0.66$

Regression Parameters:

$$b_1 := \frac{\sum_k \left[(X_k - Xbar) \cdot (Y_k - Ybar) \right]}{\sum_k (X_k - Xbar)^2} \qquad b_1 = 0.067$$

$$b_0 := Ybar - b_1 \cdot Xbar \qquad b_0 = 0.616$$

Fitted Line:

$Yhat_k := b_0 + b_1 \cdot X_k$ $Yhat = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix}$

Analysis of Variance:

Sum of Squares:

Mean Squares:

$$SSTO := \sum_k (Y_k - Ybar)^2 \qquad SSTO = 0.131$$

$$SSR := \sum_k (Yhat_k - Ybar)^2 \qquad SSR = 0.012 \qquad MSR := \frac{SSR}{1} \qquad MSR = 0.012$$

$$SSE := \sum_k (Y_k - Yhat_k)^2 \qquad SSE = 0.12 \qquad MSE := \frac{SSE}{t-2} \qquad MSE = 0.12$$

Compromise Metrics (Target Data Only):

Individual Confidence Interval for E(Y_k), where we scale the intervals by a factor of :

$$s2\ Yhat_k := MSE \cdot \left[\frac{1}{t} + \frac{(X_k - Xbar)^2}{\sum_k (X_k - Xbar)^2} \right] \qquad s2\ Yhat = \begin{pmatrix} 0.12 \\ 0.06 \\ 0.06 \end{pmatrix}$$

$$LCL_k := Yhat_k - 2 \cdot \sqrt{s2\ Yhat_k} \qquad UCL_k := Yhat_k + 2 \cdot \sqrt{s2\ Yhat_k}$$

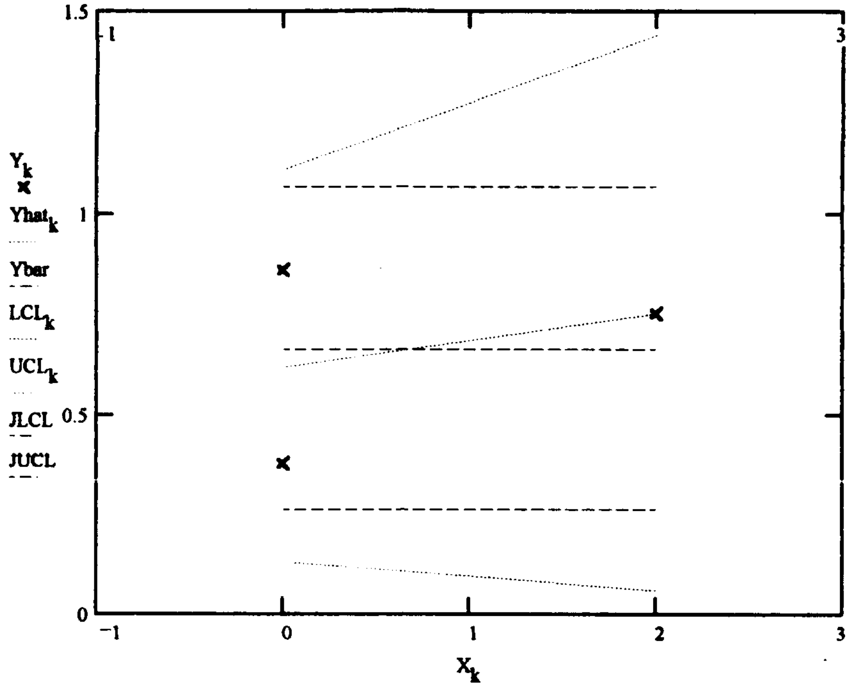
$$LCL = \begin{pmatrix} 0.057 \\ 0.127 \\ 0.127 \end{pmatrix} \quad Yhat = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \quad UCL = \begin{pmatrix} 1.441 \\ 1.105 \\ 1.105 \end{pmatrix}$$

Joint Confidence Interval for E(Y), where we scale the intervals by a factor of 2

$$JLCL := Ybar - 2 \cdot \sqrt{\frac{MSE}{t}} \qquad JUCL := Ybar + 2 \cdot \sqrt{\frac{MSE}{t}}$$

$$JLCL = 0.261 \quad Ybar = 0.66 \quad JUCL = 1.059$$

Graph:



Appendix E. *Experiment #1 Results*

E.1 Strategy S2

Experiment #1, Strategy S2 = (4, 5, 14)

ORIGIN ≡ 1

Bounds	# of O-D Pairs	O-D Index
$i := 1..2$	$t := 3$	$k := 1..t$

O-D External Flows:

p = 1, 2 and inf

Target: "Normal" Estimates: "Compromise" Estimates:

$$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \qquad F_N := \begin{pmatrix} 7.852 & 8.704 \\ 6.031 & 6.897 \\ 5.051 & 5.344 \end{pmatrix} \qquad F_C := \begin{pmatrix} 10.291 & 8.704 \\ 6.031 & 6.897 \\ 3.88 & 5.344 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \qquad D_{N_k} := F_{N_{k,2}} - F_{N_{k,1}} \qquad D_{C_k} := F_{C_{k,2}} - F_{C_{k,1}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \qquad D_N = \begin{pmatrix} 0.852 \\ 0.866 \\ 0.293 \end{pmatrix} \qquad D_C = \begin{pmatrix} -1.587 \\ 0.866 \\ 1.464 \end{pmatrix}$$

Estimation Error, MSE_{est}:

$$MSE_{est} := \frac{\sum_k \left(D_{N_k} - Y_k\right)^2}{t} \qquad MSE_{est} = 0.006$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_k := Yhat_k - \left(2 \cdot \sqrt{s2\ Yhat_k} + \sqrt{MSE_{est}}\right) \qquad EUCL_k := Yhat_k + \left(2 \cdot \sqrt{s2\ Yhat_k} + \sqrt{MSE_{est}}\right)$$

$$\text{ELCL} = \begin{pmatrix} -0.017 \\ 0.052 \\ 0.052 \end{pmatrix} \quad \text{Yhat} = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \quad \text{EUCL} = \begin{pmatrix} 1.515 \\ 1.179 \\ 1.179 \end{pmatrix}$$

Compromise Estimates:

$$D_C = \begin{pmatrix} -1.587 \\ 0.866 \\ 1.464 \end{pmatrix}$$

So, both O-D Pairs k = 1 and 3 are outside their respective confidence intervals. Conclude that either one or both are contributing sources of compromise.

Joint Confidence Interval for E(Y):

$$\text{JLCL} := \text{Ybar} - \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{\text{est}}}\right) \qquad \text{JUCL} := \text{Ybar} + \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{\text{est}}}\right)$$

$$\text{JLCL} = 0.186 \qquad \text{Ybar} = 0.66 \qquad \text{JUCL} = 1.134$$

Compromise Mean Estimate:

$$\text{Dbar}_C := \text{mean}(D_C)$$

$$\text{Dbar}_C = 0.248$$

Here, however, can not conclude that a compromise has occured.

E.2 Strategy S3

Experiment #1, Strategy S3 = (4, 5, 15)

ORIGIN ≡ 1

Bounds	# of O-D Pairs	O-D Index
$i := 1..2$	$t := 3$	$k := 1..t$

O-D External Flows:

p = 1, 2 and inf

Target:	"Normal" Estimates:	"Compromise" Estimates:
$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix}$	$F_N := \begin{pmatrix} 7.852 & 8.704 \\ 6.031 & 6.897 \\ 5 & 5.371 \end{pmatrix}$	$F_C := \begin{pmatrix} 10.291 & 8.704 \\ 6.031 & 6.897 \\ 5 & 5.371 \end{pmatrix}$

Deviations:

$Y_k := b_{k,2} - b_{k,1}$ $\quad D_{N_k} := F_{N_{k,2}} - F_{N_{k,1}}$ $\quad D_{C_k} := F_{C_{k,2}} - F_{C_{k,1}}$

$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix}$ $\quad D_N = \begin{pmatrix} 0.852 \\ 0.866 \\ 0.371 \end{pmatrix}$ $\quad D_C = \begin{pmatrix} -1.587 \\ 0.866 \\ 0.371 \end{pmatrix}$

Estimation Error, MSE_{est}: $\quad MSE_{est} := \dfrac{\sum_k \left(D_{N_k} - Y_k\right)^2}{t}$ $\quad MSE_{est} = 0.004$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_k := Yhat_k - \left(2 \cdot \sqrt{s2_{Yhat_k} + \sqrt{MSE_{est}}}\right)$$

$$EUCL_k := Yhat_k + \left(2 \cdot \sqrt{s2_{Yhat_k} + \sqrt{MSE_{est}}}\right)$$

$ELCL = \begin{pmatrix} -0.002 \\ 0.067 \\ 0.067 \end{pmatrix}$ $\quad Yhat = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix}$ $\quad EUCL = \begin{pmatrix} 1.5 \\ 1.164 \\ 1.164 \end{pmatrix}$

Compromise Estimates:

$$D_C = \begin{pmatrix} -1.587 \\ 0.866 \\ 0.371 \end{pmatrix}$$

So, only O-D Pair k = 1 is outside its respective confidence interval. Conclude that User 1 is the contributing source of compromise.

Joint Confidence Interval for E(Y):

$$JLCL := Ybar - \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}} \right)$$

$$JUCL := Ybar + \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}} \right)$$

$JLCL = 0.201$ $\quad$ $Ybar = 0.66$ $\quad$ $JUCL = 1.119$

Compromise Mean Estimate:

$$Dbar_C := mean(D_C)$$

$Dbar_C = -0.117$

Here, can also conclude that a compromise has occured.

E.3 Strategy S4

Experiment #1, Strategy S4 = (4, 14, 15)

ORIGIN ≡ 1

Bounds	**# of O-D Pairs**	**O-D Index**
$i := 1..2$	$t := 3$	$k := 1..t$

O-D External Flows:

p = 1, 2 and inf

Target:	"Normal" Estimates:	"Compromise" Estimates:

$$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \qquad F_N := \begin{pmatrix} 7.958 & 8.64 \\ 5.906 & 7.018 \\ 5 & 5.371 \end{pmatrix} \qquad F_C := \begin{pmatrix} 7.958 & 8.64 \\ 8.763 & 7.018 \\ 5 & 5.371 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \qquad D_{N_k} := F_{N_{k,2}} - F_{N_{k,1}} \qquad D_{C_k} := F_{C_{k,2}} - F_{C_{k,1}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \qquad D_N = \begin{pmatrix} 0.682 \\ 1.112 \\ 0.371 \end{pmatrix} \qquad D_C = \begin{pmatrix} 0.682 \\ -1.745 \\ 0.371 \end{pmatrix}$$

Estimation Error, MSE_{est}:

$$MSE_{est} := \frac{\sum_k \left(D_{N_k} - Y_k\right)^2}{t} \qquad MSE_{est} = 0.023$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_k := Yhat_k - \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right)$$

$$EUCL_k := Yhat_k - \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right)$$

$$ELCL = \begin{pmatrix} -0.093 \\ -0.024 \\ -0.024 \end{pmatrix} \qquad Yhat = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \qquad EUCL = \begin{pmatrix} 1.591 \\ 1.255 \\ 1.255 \end{pmatrix}$$

Compromise Estimates:

$$D_C = \begin{pmatrix} 0.682 \\ -1.745 \\ 0.371 \end{pmatrix}$$

So, only O-D Pair k = 2 is outside its respective confidence interval. Conclude that User 2 is the contributing source of compromise.

Joint Confidence Interval for E(Y):

$$JLCL := Ybar - \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}} \right)$$

$$JUCL := Ybar + \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}} \right)$$

$JLCL = 0.11$ $\quad$ $Ybar = 0.66$ $\quad$ $JUCL = 1.21$

Compromise Mean Estimate:

$$Dbar_C := mean(D_C)$$

$Dbar_C = -0.231$

Here, can also conclude that a compromise has occured.

E.4 Strategy S5

Experiment #1, Strategy S5 = (5, 14, 15)

ORIGIN ≡ 1

Bounds	# of O-D Pairs	O-D Index
$i := 1..2$	$t := 3$	$k := 1..t$

O-D External Flows:

p = 1, 2 and inf

Target: "Normal" Estimates: "Compromise" Estimates:

$$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \quad F_N := \begin{pmatrix} 7.958 & 8.64 \\ 6.031 & 6.897 \\ 5 & 5.371 \end{pmatrix} \quad F_C := \begin{pmatrix} 7.958 & 8.64 \\ 6.031 & 6.897 \\ 5 & 5.371 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \qquad D_{N_k} := F_{N_{k,2}} - F_{N_{k,1}} \qquad D_{C_k} := F_{C_{k,2}} - F_{C_{k,1}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \quad D_N = \begin{pmatrix} 0.682 \\ 0.866 \\ 0.371 \end{pmatrix} \quad D_C = \begin{pmatrix} 0.682 \\ 0.866 \\ 0.371 \end{pmatrix}$$

Estimation Error, MSE_{est}:

$$MSE_{est} := \frac{\sum_k \left(D_{N_k} - Y_k\right)^2}{t} \qquad MSE_{est} = 0.002$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_k := Yhat_k - \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right)$$

$$EUCL_k := Yhat_k + \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right)$$

$$ELCL = \begin{pmatrix} 0.019 \\ 0.088 \\ 0.088 \end{pmatrix} \quad Yhat = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \quad EUCL = \begin{pmatrix} 1.479 \\ 1.143 \\ 1.143 \end{pmatrix}$$

Compromise Estimates:

$$D_C = \begin{pmatrix} 0.682 \\ 0.866 \\ 0.371 \end{pmatrix}$$

So, no O-D Pairs are outside their respective confidence interval. Conclude that no compromise exists.

Joint Confidence Interval for E(Y):

$$JLCL := Ybar - \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}}\right)$$

$$JUCL := Ybar + \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}}\right)$$

$JLCL = 0.222$ $\quad$ $Ybar = 0.66$ $\quad$ $JUCL = 1.098$

Compromise Mean Estimate:

$$Dbar_C := mean(D_C)$$

$Dbar_C = 0.64$

Here, can not conclude that a compromise has occured.

Appendix F. *Experiment #2 Results*

F.1 Strategy S6

Experiment #2, Strategy S6 = (4, 12)
- relaxed equality constraints for compromise estimates (LB only)

ORIGIN ≡ 1

Bounds	# of O-D Pairs	O-D Index
$i := 1..2$	$t := 3$	$k := 1..t$

O-D External Flows:

p = 1, 2 and inf

Target: "Normal" Estimates: "Compromise" Estimates:

$$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \qquad F_N := \begin{pmatrix} 7.514 & 8.749 \\ 6.427 & 6.811 \\ 5 & 5.371 \end{pmatrix} \qquad F_C := \begin{pmatrix} 10.317 & 8.749 \\ 6 & 6.811 \\ 5 & 5.371 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \qquad D_{N_k} := F_{N_{k,2}} - F_{N_{k,1}} \qquad D_{C_k} := F_{C_{k,2}} - F_{C_{k,1}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \qquad D_N = \begin{pmatrix} 1.235 \\ 0.384 \\ 0.371 \end{pmatrix} \qquad D_C = \begin{pmatrix} -1.568 \\ 0.811 \\ 0.371 \end{pmatrix}$$

Estimation Error, MSE_{est}:
$$MSE_{est} := \frac{\sum_k \left(D_{N_k} - Y_k\right)^2}{t} \qquad MSE_{est} = 0.154$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_k := Yhat_k - \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right) \qquad EUCL_k := Yhat_k + \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right)$$

$$\text{ELCL} = \begin{pmatrix} -0.335 \\ -0.266 \\ -0.266 \end{pmatrix} \quad \text{Yhat} = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \quad \text{EUCL} = \begin{pmatrix} 1.833 \\ 1.497 \\ 1.497 \end{pmatrix}$$

Compromise Estimates:

$$D_C = \begin{pmatrix} -1.568 \\ 0.811 \\ 0.371 \end{pmatrix}$$

So, only O-D Pair k = 1 is outside its respective confidence interval. Conclude that User 1 is the source of compromise.

Joint Confidence Interval for E(Y):

$$\text{JLCL} := \text{Ybar} - \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{est}}\right)$$

$$\text{JUCL} := \text{Ybar} + \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{est}}\right)$$

$$\text{JLCL} = -0.132 \qquad \text{Ybar} = 0.66 \qquad \text{JUCL} = 1.452$$

Compromise Mean Estimate:

$$\text{Dbar}_C := \text{mean}(D_C)$$

$\text{Dbar}_C = -0.129$

Here, however, can not conclude that a compromise has occured.

F.2 Strategy S7

Experiment #2, Strategy S7 = (4, 15)

ORIGIN ≡ 1

$p := 1..3$ **lp-norm exponents (where 3 = infinity)**

$t := 3$ **# of O-D Pairs**

$k := 1..t$ **O-D Index**

O-D External Flows: "Normal" Estimates:

Target: Lower Bound Upper Bound

$$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \quad Flb_N := \begin{pmatrix} 7.878 & 7.929 & 7.934 \\ 6 & 5.94 & 5.934 \\ 5 & 5 & 5 \end{pmatrix} \quad Fub_N := \begin{pmatrix} 8.723 & 8.729 & 8.732 \\ 6.86 & 6.85 & 6.843 \\ 5.371 & 5.371 & 5.371 \end{pmatrix}$$

"Compromise" Estimates:

$$Flb_C := \begin{pmatrix} 10.317 & 9.34 & 9.25 \\ 6 & 7.144 & 7.25 \\ 5 & 5 & 5 \end{pmatrix} \quad Fub_C := \begin{pmatrix} 8.723 & 8.729 & 8.732 \\ 6.86 & 6.85 & 6.843 \\ 5.371 & 5.371 & 5.371 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \quad D_{N_{k,p}} := Fub_{N_{k,p}} - Flb_{N_{k,p}} \quad D_{C_{k,p}} := Fub_{C_{k,p}} - Flb_{C_{k,p}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \quad D_N = \begin{pmatrix} 0.845 & 0.8 & 0.798 \\ 0.86 & 0.91 & 0.909 \\ 0.371 & 0.371 & 0.371 \end{pmatrix} \quad D_C = \begin{pmatrix} -1.594 & -0.611 & -0.518 \\ 0.86 & -0.294 & -0.407 \\ 0.371 & 0.371 & 0.371 \end{pmatrix}$$

Estimation Error, MSE_{est}:

$$MSE_{est_p} := \frac{\sum_k \left(D_{N_{k,p}} - Y_k\right)^2}{t} \quad MSE_{est} = \begin{pmatrix} 0.0031 \\ 0.0017 \\ 0.0016 \end{pmatrix}$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_{k,p} := Yhat_k - \left(2\cdot\sqrt{s2\ Yhat_k} + \sqrt{MSE_{est_p}}\right) \quad EUCL_{k,p} := Yhat_k + \left(2\cdot\sqrt{s2\ Yhat_k} + \sqrt{MSE_{est_p}}\right)$$

$$\text{ELCL} = \begin{pmatrix} 0.002 & 0.016 & 0.017 \\ 0.071 & 0.085 & 0.086 \\ 0.071 & 0.085 & 0.086 \end{pmatrix} \quad \text{Yhat} = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \quad \text{EUCL} = \begin{pmatrix} 1.496 & 1.482 & 1.481 \\ 1.16 & 1.146 & 1.145 \\ 1.16 & 1.146 & 1.145 \end{pmatrix}$$

Compromise Estimates:

$p = 1$: $\quad \text{ELCL}_{k,1} = \boxed{\begin{matrix} 0.002 \\ 0.071 \\ 0.071 \end{matrix}} \quad {}^{D}C_{k,1} = \boxed{\begin{matrix} -1.594 \\ 0.86 \\ 0.371 \end{matrix}} \quad \text{EUCL}_{k,1} = \boxed{\begin{matrix} 1.496 \\ 1.16 \\ 1.16 \end{matrix}}$

User 1 is the source of compromise

$p = 2$: $\quad \text{ELCL}_{k,2} = \boxed{\begin{matrix} 0.016 \\ 0.085 \\ 0.085 \end{matrix}} \quad {}^{D}C_{k,2} = \boxed{\begin{matrix} -0.611 \\ -0.294 \\ 0.371 \end{matrix}} \quad \text{EUCL}_{k,2} = \boxed{\begin{matrix} 1.482 \\ 1.146 \\ 1.146 \end{matrix}}$

Users 1 and/or 2 are the sources of compromise

$p = \text{inf}$: $\quad \text{ELCL}_{k,3} = \boxed{\begin{matrix} 0.017 \\ 0.086 \\ 0.086 \end{matrix}} \quad {}^{D}C_{k,3} = \boxed{\begin{matrix} -0.518 \\ -0.407 \\ 0.371 \end{matrix}} \quad \text{EUCL}_{k,3} = \boxed{\begin{matrix} 1.481 \\ 1.145 \\ 1.145 \end{matrix}}$

Users 1 and/or 2 are the sources of compromise

Joint Confidence Interval for E(Y):

$$\text{JLCL}_p = \text{Ybar} - \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{\text{est}_p}}\right) \qquad \text{JUCL}_p = \text{Ybar} + \left(2 \cdot \sqrt{\frac{\text{M}}{t}} \quad \sqrt{\text{MSE}_{\text{est}_p}}\right)$$

$$\text{JLCL} = \begin{pmatrix} 0.205 \\ 0.219 \\ 0.221 \end{pmatrix} \quad \text{Ybar} = 0.66 \quad \text{JUCL} = \begin{pmatrix} 1.115 \\ 1.101 \\ 1.099 \end{pmatrix}$$

Compromise Mean Estimates:

$$\text{Dbar}_{C_p} = \frac{\sum_k {}^{D}C_{k,p}}{t} \qquad \text{Dbar}_C = \begin{pmatrix} -0.121 \\ -0.178 \\ -0.185 \end{pmatrix}$$

Here, can conclude for each level of p that a compromise has occured.

F.3 *Strategy S8*

Experiment #2, Strategy S8 = (12, 15)

ORIGIN ≡ 1

$p := 1..3$ **lp-norm exponents (where 3 = infinity)**

$t := 3$ **# of O-D Pairs**

$k := 1..t$ **O-D Index**

O-D External Flows:

"Normal" Estimates:

Target: Lower Bound Upper Bound

$$b = \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \quad Flb_N = \begin{pmatrix} 8 & 7.98 & 7.98 \\ 5.96 & 5.979 & 5.98 \\ 5 & 5 & 5 \end{pmatrix} \quad Fub_N = \begin{pmatrix} 8.749 & 8.726 & 8.721 \\ 6.811 & 6.829 & 6.832 \\ 5.371 & 5.371 & 5.371 \end{pmatrix}$$

"Compromise" Estimates:

$$Flb_C = \begin{pmatrix} 8 & 7.98 & 7.98 \\ 5.96 & 5.979 & 5.98 \\ 5 & 5 & 5 \end{pmatrix} \quad Fub_C = \begin{pmatrix} 8.749 & 8.726 & 8.721 \\ 6.811 & 6.829 & 6.832 \\ 5.371 & 5.371 & 5.371 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \quad D_{N_{k,p}} := Fub_{N_{k,p}} - Flb_{N_{k,p}} \quad D_{C_{k,p}} := Fub_{C_{k,p}} - Flb_{C_{k,p}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \quad D_N = \begin{pmatrix} 0.749 & 0.746 & 0.741 \\ 0.851 & 0.85 & 0.852 \\ 0.371 & 0.371 & 0.371 \end{pmatrix} \quad D_C = \begin{pmatrix} 0.749 & 0.746 & 0.741 \\ 0.851 & 0.85 & 0.852 \\ 0.371 & 0.371 & 0.371 \end{pmatrix}$$

Estimation Error, MSE_{est}:

$$MSE_{est_p} := \frac{\sum_k \left(D_{N_{k,p}} - Y_k\right)^2}{t} \quad MSE_{est} = \begin{bmatrix} 2.7 \cdot 10^{-5} \\ 3.6333 \cdot 10^{-5} \\ 4.2667 \cdot 10^{-5} \end{bmatrix}$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_{k,p} := Yhat_k - \left(2 \cdot \sqrt{s2\ Yhat_k} + \sqrt{MSE_{est_p}}\right) \quad EUCL_{k,p} := Yhat_k + \left(2 \cdot \sqrt{s2\ Yhat_k} + \sqrt{MSE_{est_p}}\right)$$

$$\text{ELCL} = \begin{pmatrix} 0.052 & 0.051 & 0.051 \\ 0.121 & 0.12 & 0.12 \\ 0.121 & 0.12 & 0.12 \end{pmatrix} \quad \text{Yhat} = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \quad \text{EUCL} = \begin{pmatrix} 1.446 & 1.447 & 1.447 \\ 1.11 & 1.111 & 1.111 \\ 1.11 & 1.111 & 1.111 \end{pmatrix}$$

Compromise Estimates:

p = 1:

$$\text{ELCL}_{k,1} = \boxed{\begin{matrix} 0.052 \\ 0.121 \\ 0.121 \end{matrix}} \quad {}^{D}C_{k,1} = \boxed{\begin{matrix} 0.749 \\ 0.851 \\ 0.371 \end{matrix}} \quad \text{EUCL}_{k,1} = \boxed{\begin{matrix} 1.446 \\ 1.11 \\ 1.11 \end{matrix}}$$

No compromise

p = 2:

$$\text{ELCL}_{k,2} = \boxed{\begin{matrix} 0.051 \\ 0.12 \\ 0.12 \end{matrix}} \quad {}^{D}C_{k,2} = \boxed{\begin{matrix} 0.746 \\ 0.85 \\ 0.371 \end{matrix}} \quad \text{EUCL}_{k,2} = \boxed{\begin{matrix} 1.447 \\ 1.111 \\ 1.111 \end{matrix}}$$

No compromise

p = inf:

$$\text{ELCL}_{k,3} = \boxed{\begin{matrix} 0.051 \\ 0.12 \\ 0.12 \end{matrix}} \quad {}^{D}C_{k,3} = \boxed{\begin{matrix} 0.741 \\ 0.852 \\ 0.371 \end{matrix}} \quad \text{EUCL}_{k,3} = \boxed{\begin{matrix} 1.447 \\ 1.111 \\ 1.111 \end{matrix}}$$

No compromise

Joint Confidence Interval for E(Y):

$$\text{JLCL}_p := \text{Ybar} - \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{\text{est}_p}}\right) \qquad \text{JUCL}_p := \text{Ybar} + \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{\text{est}_p}}\right)$$

$$\text{JLCL} = \begin{pmatrix} 0.256 \\ 0.255 \\ 0.254 \end{pmatrix} \quad \text{Ybar} = 0.66 \quad \text{JUCL} = \begin{pmatrix} 1.064 \\ 1.065 \\ 1.066 \end{pmatrix}$$

Compromise Mean Estimates:

$$\text{Dbar}_{C_p} := \frac{\sum_k {}^{D}C_{k,p}}{t} \qquad \text{Dbar}_C = \begin{pmatrix} 0.657 \\ 0.656 \\ 0.655 \end{pmatrix}$$

No compromise

F.4 Strategy S9

Experiment #2, Strategy S9 = (4, 5)

ORIGIN ≡ 1

Bounds	**# of O-D Pairs**	**O-D Index**
$i := 1..2$	$t := 3$	$k := 1..t$

O-D External Flows:

p = 1, 2 and inf

Target: "Normal" Estimates: "Compromise" Estimates:

$$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \qquad F_N := \begin{pmatrix} 7.852 & 8.704 \\ 6.031 & 6.897 \\ 5 & 5.371 \end{pmatrix} \qquad F_C := \begin{pmatrix} 10.291 & 8.704 \\ 6.031 & 6.897 \\ 5 & 5.371 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \qquad D_{N_k} := F_{N_{k,2}} - F_{N_{k,1}} \qquad D_{C_k} := F_{C_{k,2}} - F_{C_{k,1}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \qquad D_N = \begin{pmatrix} 0.852 \\ 0.866 \\ 0.371 \end{pmatrix} \qquad D_C = \begin{pmatrix} -1.587 \\ 0.866 \\ 0.371 \end{pmatrix}$$

Estimation Error, MSE_{est}:

$$MSE_{est} := \frac{\sum_k \left(D_{N_k} - Y_k\right)^2}{t} \qquad MSE_{est} = 0.004$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_k := Yhat_k - \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right) \qquad EUCL_k := Yhat_k + \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right)$$

$$ELCL = \begin{pmatrix} -0.002 \\ 0.067 \\ 0.067 \end{pmatrix} \qquad Yhat = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \qquad EUCL = \begin{pmatrix} 1.5 \\ 1.164 \\ 1.164 \end{pmatrix}$$

Compromise Estimates:

$$D_C = \begin{pmatrix} -1.587 \\ 0.866 \\ 0.371 \end{pmatrix}$$

So, only O-D Pair k = 1 is outside its respective confidence interval. Conclude that User 1 is the source of compromise.

Joint Confidence Interval for E(Y):

$$JLCL := Ybar - \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}}\right)$$

$$JUCL := Ybar + \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}}\right)$$

$JLCL = 0.201$ $\quad$ $Ybar = 0.66$ $\quad$ $JUCL = 1.119$

Compromise Mean Estimate:

$$Dbar_C := mean\left(D_C\right)$$

$Dbar_C = -0.117$

Here, can conclude that a compromise has occured.

F.5 Strategy S10

Experiment #2, Strategy S10 = (4, 14)

ORIGIN = 1

$p := 1..3$	lp-norm exponents (where 3 = infinity)
$t := 3$	# of O-D Pairs
$k := 1..t$	O-D Index

O-D External Flows: "Normal" Estimates:

Target: Lower Bound Upper Bound

$$b = \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \quad Flb_N = \begin{pmatrix} 7.958 & 7.932 & 7.934 \\ 5.906 & 5.937 & 5.934 \\ 5 & 5.013 & 5.012 \end{pmatrix} \quad Fub_N = \begin{pmatrix} 8.724 & 8.726 & 8.717 \\ 6.86 & 6.856 & 6.873 \\ 5.336 & 5.335 & 5.339 \end{pmatrix}$$

"Compromise" Estimates:

$$Flb_C = \begin{pmatrix} 8 & 9.218 & 9.25 \\ 8.714 & 7.287 & 7.25 \\ 4.98 & 4.395 & 4.38 \end{pmatrix} \quad Fub_C = \begin{pmatrix} 8.724 & 8.726 & 8.717 \\ 6.86 & 6.856 & 6.873 \\ 5.336 & 5.335 & 5.339 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \qquad D_{N_{k,p}} := Fub_{N_{k,p}} - Flb_{N_{k,p}} \qquad D_{C_{k,p}} := Fub_{C_{k,p}} - Flb_{C_{k,p}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \quad D_N = \begin{pmatrix} 0.766 & 0.794 & 0.783 \\ 0.954 & 0.919 & 0.939 \\ 0.336 & 0.322 & 0.327 \end{pmatrix} \quad D_C = \begin{pmatrix} 0.724 & -0.492 & -0.533 \\ -1.854 & -0.431 & -0.377 \\ 0.356 & 0.94 & 0.959 \end{pmatrix}$$

Estimation Error, MSE_{est}:

$$MSE_{est_p} := \frac{\sum_k \left(D_{N_{k,p}} - Y_k\right)^2}{t} \qquad MSE_{est} = \begin{pmatrix} 0.0035 \\ 0.0026 \\ 0.0031 \end{pmatrix}$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_{k,p} := Yhat_k - \left(2 \cdot \sqrt{s2_{Yhat_k} + MSE_{est_p}}\right) \qquad EUCL_{k,p} := Yhat_k + \left(2 \cdot \sqrt{s2_{Yhat_k} + MSE_{est_p}}\right)$$

$$\text{ELCL} = \begin{pmatrix} -0.001 & 0.006 & 0.002 \\ 0.068 & 0.075 & 0.071 \\ 0.068 & 0.075 & 0.071 \end{pmatrix} \quad \text{Yhat} = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \quad \text{EUCL} = \begin{pmatrix} 1.499 & 1.492 & 1.496 \\ 1.163 & 1.156 & 1.16 \\ 1.163 & 1.156 & 1.16 \end{pmatrix}$$

Compromise Estimates:

p = 1:

$\text{ELCL}_{k,1}$	${}^{D}C_{k,1}$	$\text{EUCL}_{k,1}$
-0.001	0.724	1.499
0.068	-1.854	1.163
0.068	0.356	1.163

User 2 is the source of compromise

p = 2:

$\text{ELCL}_{k,2}$	${}^{D}C_{k,2}$	$\text{EUCL}_{k,2}$
0.006	-0.492	1.492
0.075	-0.431	1.156
0.075	0.94	1.156

Users 1 and/or 2 are the sources of compromise

p = inf:

$\text{ELCL}_{k,3}$	${}^{D}C_{k,3}$	$\text{EUCL}_{k,3}$
0.002	-0.533	1.496
0.071	-0.377	1.16
0.071	0.959	1.16

Users 1 and/or 2 are the sources of compromise

Joint Confidence Interval for E(Y):

$$\text{JLCL}_p = \text{Ybar} - \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{\text{est}_p}}\right) \qquad \text{JUCL}_p = \text{Ybar} + \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{\text{est}_p}}\right)$$

$$\text{JLCL} = \begin{pmatrix} 0.202 \\ 0.209 \\ 0.205 \end{pmatrix} \quad \text{Ybar} = 0.66 \quad \text{JUCL} = \begin{pmatrix} 1.118 \\ 1.111 \\ 1.115 \end{pmatrix}$$

Compromise Mean Estimates:

$$\text{Dbar}_{C_p} = \frac{\sum_k {}^{D}C_{k,p}}{t} \qquad \text{Dbar}_C = \begin{pmatrix} -0.258 \\ 0.006 \\ 0.016 \end{pmatrix}$$

Only p = 2 and infinity detect potential compromise

F.6 Strategy S11

Experiment #2, Strategy S11 = (5, 14)

ORIGIN $\equiv 1$

$p := 1..3$ lp-norm exponents (where 3 = infinity)

$t := 3$ # of O-D Pairs

$k := 1..t$ O-D Index

O-D External Flows: "Normal" Estimates:

Target: Lower Bound Upper Bound

$$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \quad Flb_N := \begin{pmatrix} 8 & 7.992 & 8 \\ 6.031 & 6.031 & 6.031 \\ 4.98 & 4.984 & 4.98 \end{pmatrix} \quad Fub_N := \begin{pmatrix} 8.749 & 8.733 & 8.728 \\ 6.897 & 6.897 & 6.897 \\ 5.325 & 5.332 & 5.334 \end{pmatrix}$$

"Compromise" Estimates:

$$Flb_C := \begin{pmatrix} 8 & 7.992 & 8 \\ 6.031 & 6.031 & 6.031 \\ 4.98 & 4.984 & 4.98 \end{pmatrix} \quad Fub_C := \begin{pmatrix} 8.749 & 8.733 & 8.728 \\ 6.897 & 6.897 & 6.897 \\ 5.325 & 5.332 & 5.334 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \quad D_{N_{k,p}} := Fub_{N_{k,p}} - Flb_{N_{k,p}} \quad D_{C_{k,p}} := Fub_{C_{k,p}} - Flb_{C_{k,p}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \quad D_N = \begin{pmatrix} 0.749 & 0.741 & 0.728 \\ 0.866 & 0.866 & 0.866 \\ 0.345 & 0.348 & 0.354 \end{pmatrix} \quad D_C = \begin{pmatrix} 0.749 & 0.741 & 0.728 \\ 0.866 & 0.866 & 0.866 \\ 0.345 & 0.348 & 0.354 \end{pmatrix}$$

Estimation Error, MSE_{est}:

$$MSE_{est_p} := \frac{\sum_k \left(D_{N_{k,p}} - Y_k\right)^2}{t} \qquad MSE_{est} = \begin{bmatrix} 2.3733 \cdot 10^{-4} \\ 2.0967 \cdot 10^{-4} \\ 2.5533 \cdot 10^{-4} \end{bmatrix}$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_{k,p} := Yhat_k - \left(2 \cdot \sqrt{s2\ Yhat_k} + \sqrt{MSE_{est_p}}\right) \qquad EUCL_{k,p} := Yhat_k + \left(2 \cdot \sqrt{s2\ Yhat_k} + \sqrt{MSE_{est_p}}\right)$$

$$\text{ELCL} = \begin{pmatrix} 0.042 & 0.043 & 0.041 \\ 0.111 & 0.112 & 0.111 \\ 0.111 & 0.112 & 0.111 \end{pmatrix} \quad \text{Yhat} = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \quad \text{EUCL} = \begin{pmatrix} 1.456 & 1.455 & 1.457 \\ 1.12 & 1.119 & 1.12 \\ 1.12 & 1.119 & 1.12 \end{pmatrix}$$

Compromise Estimates:

p = 1:

$\text{ELCL}_{k,1}$	$^{D}C_{k,1}$	$\text{EUCL}_{k,1}$
0.042	0.749	1.456
0.111	0.866	1.12
0.111	0.345	1.12

No compromise

p = 2:

$\text{ELCL}_{k,2}$	$^{D}C_{k,2}$	$\text{EUCL}_{k,2}$
0.043	0.741	1.455
0.112	0.866	1.119
0.112	0.348	1.119

No compromise

p = inf:

$\text{ELCL}_{k,3}$	$^{D}C_{k,3}$	$\text{EUCL}_{k,3}$
0.041	0.728	1.457
0.111	0.866	1.12
0.111	0.354	1.12

No compromise

Joint Confidence Interval for E(*Y*):

$$\text{JLCL}_p = \text{Ybar} - \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{\text{est}_p}}\right) \qquad \text{JUCL}_p = \text{Ybar} + \left(2 \cdot \sqrt{\frac{\text{MSE}}{t}} + \sqrt{\text{MSE}_{\text{est}_p}}\right)$$

$$\text{JLCL} = \begin{pmatrix} 0.245 \\ 0.246 \\ 0.245 \end{pmatrix} \quad \text{Ybar} = 0.66 \quad \text{JUCL} = \begin{pmatrix} 1.075 \\ 1.074 \\ 1.075 \end{pmatrix}$$

Compromise Mean Estimates:

$$\text{Dbar}_{C_p} = \frac{\sum_k {}^{D}C_{k,p}}{t} \qquad \text{Dbar}_C = \begin{pmatrix} 0.653 \\ 0.652 \\ 0.649 \end{pmatrix}$$

No compromise

F.7 Strategy S12

Experiment #2, Strategy S12 = (5, 15)

ORIGIN ≡ 1

Bounds	**# of O-D Pairs**	**O-D Index**
$i := 1..2$	$t := 3$	$k := 1..t$

O-D External Flows:

p = 1, 2 and inf

Target:	"Normal" Estimates:	"Compromise" Estimates:
$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix}$	$F_N := \begin{pmatrix} 8 & 8.749 \\ 6.031 & 6.897 \\ 5 & 5.371 \end{pmatrix}$	$F_C := \begin{pmatrix} 8 & 8.749 \\ 6.031 & 6.897 \\ 5 & 5.371 \end{pmatrix}$

Deviations:

$Y_k := b_{k,2} - b_{k,1}$ $\quad D_{N_k} := F_{N_{k,2}} - F_{N_{k,1}}$ $\quad D_{C_k} := F_{C_{k,2}} - F_{C_{k,1}}$

$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix}$ $\quad D_N = \begin{pmatrix} 0.749 \\ 0.866 \\ 0.371 \end{pmatrix}$ $\quad D_C = \begin{pmatrix} 0.749 \\ 0.866 \\ 0.371 \end{pmatrix}$

Estimation Error, MSE_{est}:
$$MSE_{est} := \frac{\sum_k \left(D_{N_k} - Y_k\right)^2}{t} \qquad MSE_{est} = 1.2 \cdot 10^{-5}$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_k := Yhat_k - \left(2 \cdot \sqrt{s2_{Yhat_k}} - \sqrt{MSE_{est}}\right)$$

$$EUCL_k := Yhat_k + \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right)$$

$ELCL = \begin{pmatrix} 0.054 \\ 0.123 \\ 0.123 \end{pmatrix}$ $\quad Yhat = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix}$ $\quad EUCL = \begin{pmatrix} 1.444 \\ 1.108 \\ 1.108 \end{pmatrix}$

Compromise Estimates:

$$D_C = \begin{pmatrix} 0.749 \\ 0.866 \\ 0.371 \end{pmatrix}$$

No compromise.

Joint Confidence Interval for E(Y):

$$JLCL := Ybar - \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}}\right)$$

$$JUCL := Ybar + \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}}\right)$$

$JLCL = 0.257$ $\quad$ $Ybar = 0.66$ $\quad$ $JUCL = 1.063$

Compromise Mean Estimate:

$$Dbar_C := mean(D_C)$$

$Dbar_C = 0.662$

Here, can not conclude that a compromise has occured.

F.8 Strategy S13

Experiment #2, Strategy S13 = (14, 15)

ORIGIN ≡ 1

Bounds	**# of O-D Pairs**	**O-D Index**
$i := 1..2$	$t := 3$	$k := 1..t$

O-D External Flows:

p = 1, 2 and inf

Target: "Normal" Estimates: "Compromise" Estimates:

$$b := \begin{pmatrix} 8 & 8.749 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \qquad F_N := \begin{pmatrix} 7.958 & 8.64 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix} \qquad F_C := \begin{pmatrix} 7.958 & 8.64 \\ 6 & 6.86 \\ 5 & 5.371 \end{pmatrix}$$

Deviations:

$$Y_k := b_{k,2} - b_{k,1} \qquad D_{N_k} := F_{N_{k,2}} - F_{N_{k,1}} \qquad D_{C_k} := F_{C_{k,2}} - F_{C_{k,1}}$$

$$Y = \begin{pmatrix} 0.749 \\ 0.86 \\ 0.371 \end{pmatrix} \qquad D_N = \begin{pmatrix} 0.682 \\ 0.86 \\ 0.371 \end{pmatrix} \qquad D_C = \begin{pmatrix} 0.682 \\ 0.86 \\ 0.371 \end{pmatrix}$$

Estimation Error, MSE_{est}: $$MSE_{est} := \frac{\sum_k \left(D_{N_k} - Y_k\right)^2}{t} \qquad MSE_{est} = 0.001$$

Compromise Metrics:

Individual Confidence Interval for $E(Y_k)$:

$$ELCL_k := Yhat_k - \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right)$$

$$EUCL_k := Yhat_k + \left(2 \cdot \sqrt{s2_{Yhat_k}} + \sqrt{MSE_{est}}\right)$$

$$ELCL = \begin{pmatrix} 0.019 \\ 0.088 \\ 0.088 \end{pmatrix} \qquad Yhat = \begin{pmatrix} 0.749 \\ 0.616 \\ 0.616 \end{pmatrix} \qquad EUCL = \begin{pmatrix} 1.479 \\ 1.143 \\ 1.143 \end{pmatrix}$$

Compromise Estimates:

$$D_C = \begin{pmatrix} 0.682 \\ 0.86 \\ 0.371 \end{pmatrix}$$

No compromise.

Joint Confidence Interval for E(*Y*):

$$JLCL := Ybar - \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}}\right)$$

$$JUCL := Ybar + \left(2 \cdot \sqrt{\frac{MSE}{t}} + \sqrt{MSE_{est}}\right)$$

$JLCL = 0.222$ $\quad Ybar = 0.66$ $\quad JUCL = 1.098$

Compromise Mean Estimate:

$$Dbar_C := mean(D_C)$$

$Dbar_C = 0.638$

Here, can not conclude that a compromise has occured.

Vita

Captain Alan Robinson was born on 13 May, 1964 in St. Nazaire, France. He graduated from the S.H.A.P.E. International High School at S.H.A.P.E., Belgium. He attended Rutgers University at New Brunswick, New Jersey, from which he received the degree of Bachelor of Arts in Statistics in May 1987. He was commissioned in the United States Air Force in May 1987. His first assignment was as a Cost Analysis Officer for the Small ICBM program at Norton AFB, California. While at Norton, he received a Master of Science in Systems Management from the University of Southern California. He entered the School of Engineering, Air Force Institute of Technology, in August 1992.

Permanent Address: 6845-A Castlerock Trail
Centerville, Ohio 45459

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
March 1994

3. REPORT TYPE AND DATES COVERED
Master's Thesis

4. TITLE AND SUBTITLE
ASSESSING THE VULNERABILITY OF MULTI-COMMODITY NETWORKS WITH FAILING COMPONENTS

5. FUNDING NUMBERS

6. AUTHOR(S)
Alan R. Robinson, Capt, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology, WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GOR/ENS/94M-12

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Lt Col Larry Pulcher
R&E Bldg, R5
9800 Savage Rd
Ft meade MD 20755

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

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

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This research proposes an analytical approach for assessing flow disturbance, or "compromise," based on limited sampling of arc flow information in multi-commodity, or multiple origin-destination (O-D), networks with failing arcs. There were three objectives established for this research. The first objective was to bound the expected flow, given the arcs fail with certain probabilities, which was accomplished by reviewing current approaches for single-commodity networks and extending the results to the multi-commodity case. The second objective was to determine the best placement of flow monitors to obtain the most accurate estimates of O-D pair volumes. This was accomplished using a multi-criteria approach for defining and evaluating all possible monitor placement strategies satisfying monitor availability. The O-D pair volumes were estimated using the l_p-norm metric for varied levels of p. The final objective was to define a compromise metric providing confident assessments on the occurrence of "compromise." This was accomplished using simple regression techniques to generate confidence intervals around the expected flow for each O-D pair. The approach proposed in this research is provided as an initial look into "compromise" assessment based on limited network information.

14. SUBJECT TERMS
Networks, Reliability, Vulnerability, Origin-Destination Matrix, Multi-Commodity Networks

15. NUMBER OF PAGES
168

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL