Pre-disaster Investment Decisions for Strengthening a Highway Network

Srinivas Peeta [a], F. Sibel Salman[b], Dilek Gunnec[c], Kannan Viswanath[d]

[a]*School of Civil Engineering, Purdue University, W. Lafayette, IN 47907, USA*

[b]*Department of Industrial Engineering, Koc University, Istanbul 34450, Turkey*

[c]*Robert H. Smith School of Business, University of Maryland, MD 20742, USA*

[d]*Krannert School of Management, Purdue University, W. Lafayette, IN 47907, USA*

ABSTRACT

We address a pre-disaster planning problem that seeks to strengthen a highway network whose links are subject to random failures due to a disaster. Each link may be either operational or non-functional after the disaster. The link failure probabilities are assumed to be known a priori, and investment decreases the likelihood of failure. The planning problem seeks connectivity for first responders between various origin-destination (O-D) pairs and hence focuses on uncapacitated road conditions. The decision-maker's goal is to select the links to invest in under a limited budget with the objective of maximizing the post-disaster connectivity and minimizing traversal costs between the origin and destination nodes. The problem is modeled as a two-stage stochastic program in which the investment decisions in the first stage alter the survival probabilities of the corresponding links. We restructure the objective function into a monotonic non-increasing multilinear function and show that using the first order terms of this function leads to a knapsack problem whose solution is a local optimum to the original problem. Numerical experiments on real-world data related to strengthening Istanbul's urban highway system against earthquake risk illustrate the tractability of the method and provide practical insights for decision-makers.

Keywords: Networks, random link failures, retrofitting highways, earthquake mitigation, two-stage stochastic program, decision-dependent probability distribution.

Corresponding author: F. Sibel Salman; Tel.:+90 212 338 1707; Fax:+90 212 338 1548.

E-mail addresses: peeta@purdue.edu (S. Peeta), ssalman@ku.edu.tr (F. S. Salman), dgunnec@rhsmith.umd.edu (D. Gunnec), viswanak@purdue.edu (K.Viswanath)

1 INTRODUCTION

The functionality of civil infrastructure systems, such as transportation, water, energy, and communication, can be significantly affected by disasters, either natural or man-made. These infrastructure systems can be characterized as networks whose links are subject to random failure due to the disaster. For effective disaster response, it is vital that these networks remain operational, especially the transportation network.

Disaster management is a multi-stage process that starts with pre-disaster mitigation and preparedness that focus on long-term measures for reducing or eliminating risk, and extends to post-disaster response, recovery and re-construction. Investment in infrastructure systems plays an essential role in mitigation activities as it entails the need to strengthen the links of a network to enhance their survivability. The strengthening of all links to targeted safety levels may require unacceptable or even unaffordable expenditures; hence, a subset of the links should be selected. Most decision-makers approach this problem by prioritizing the links using a weighted combination of several factors without explicit linkage to the expected network performance. Instead, an optimization problem that captures how investments would alter the performance of the post-disaster network can be posed to provide a system level analysis. Then the objective is to maximize expected network performance along the needs of the response agencies, subject to a budget restriction.

In this study, we address the above investment problem for highway networks affected by earthquakes with the objective of maximizing the post-disaster connectivity and minimizing traversal costs between multiple origin and destination (O-D) nodes. The links represent the highways and the nodes represent junctions. The links are subject to random failures, which are assumed to be independent in the earthquake context due to the following reasons. In highway networks, investment decisions for earthquake related response planning imply the seismic retrofit of the bridges, which tend to be the weakest structural components and require several weeks to months or years to restore if they fail. Links without bridges can be made functional relatively quickly (in a few hours) even if they have some damage ([25]). Hence, link failure is typically associated with bridge failure under earthquakes. Bridges in the highway network tend to be structurally heterogeneous due to their type, design, age, load and maintenance levels. Further, geotechnical conditions can vary from one location to another, affecting a bridge's response based on soil characteristics and related elements. Hence bridges may have differential impacts under an earthquake; thereby, link failures need not be dependent.

In our study, the decision-maker has prior knowledge of link failure probabilities, which can be determined through structural analyses ([23]). The survival probability of a link is increased to a targeted level by investing in the seismic retrofit of bridges on that link (as discussed in [10]). A link will be either functional or non-functional after the earthquake, resulting in a network realization of surviving links. Associated with each link is also a traversal cost based on uncapacitated traffic conditions. Since the focus of the planning problem is ensuring connectivity for the first responders in the immediate aftermath of the earthquake, daily traffic patterns are no longer relevant. We seek to allocate a budget to a subset of links for investment so as to minimize the expected value of the weighted sum of traversal costs between the origin and destination nodes, across post-disaster surviving network realizations. The traversal cost under a realization is the least path cost among the surviving paths. If no path exists between an O-D pair in a network realization, the associated traversal cost is a fixed penalty cost. The weight of an O-D pair represents the importance of connecting it for post-disaster response. The problem is modeled as a two-stage stochastic program and an approximate solution approach is proposed.

The remainder of the paper is organized as follows. In Section 2, we discuss related literature and summarize our contributions. In Section 3, we present the two-stage stochastic program for a single O-D pair for ease of notation. Then, we extend the results to the multiple O-D case. Section 4 derives the structural results for the objective function. In Section 5, we use these results to obtain a local optimal solution. In Section 6, results from a computational study are presented for the multiple O-D Istanbul transportation network under earthquake risk. Here, sampling methods are used for tractability. We present some concluding comments in Section 7.

2 BACKGROUND

Our focus in this study is on allocating a budget among links of a road network in order to strengthen it and enhance its resilience against disasters. In this section we first review previous work on link improvement plans for disaster preparedness. Next, investment problems to increase capacity of networks with random demand are discussed. These problems differ from ours in two aspects: i) we increase the probability of having link functionality by investment rather than adding new capacity, ii) our goal is to provide access for first responders in the aftermath of a disaster and thus we do not take into consideration demand and capacity. There exists a considerable amount of work on assessing network reliability and functional performance for road networks. Such studies are useful in providing a basis to measure benefits of investment to links, but most of them focus on link capacity degradation in day-to-day incidents and demand variability. However, for the disaster situation, connectivity of short routes becomes the critical issue.

Most studies on developing link improvement plans for disaster response, both in practice and in the literature, focus primarily on factors specific to a link such as its physical condition, characteristics, and the cost to upgrade it to withstand disasters of specific severities. The links are then prioritized with respect to a score that incorporates these factors, as in Sohn et al. [31], Sohn [30], as well as Bana et al. [4] with application to Lisbon. Earlier, Basöz and Kiremidjian [5] reviewed approaches for prioritizing bridges for seismic retrofitting. Some of these methods were implemented in the USA, Japan, and New Zealand. Werner et al. [35], among others, recognize that strengthening a link has implications beyond its own survivability. Moghtaderi-Zadeh and Der Kiureghian [22] address the need for efficient upgrade of lifelines. For a given magnitude and location of an earthquake, they determine which links will fail and which ones will survive based on a distance threshold and calculate the probability that the network is functional, Then, they incrementally invest in critical components to increase the system reliability above a target value as a heuristic approach.

In recent years, formulations that use a systems perspective to address one or more objectives (such as response time, network accessibility, path redundancy) are being proposed. Viswanath and Peeta [33] formulate the multicommodity maximal covering network design problem to identify critical routes for earthquake response in a deterministic setting. Similar to our study, Sanchez-Silva et al. [26] optimize the allocation of resources among links in order to maximize the operational reliability of a transport network. They use a continuous time Markov chain with failure and repair rates for each link, such that the rates can be changed with investment. They also consider how network users react to failure of links along a route and the waiting time cost of the users until the link is repaired. However, they assume that only one link fails at a time, which makes their approach non-applicable to a disaster situation such as earthquakes, where many links may fail simultaneously.

Outside the disaster context, several studies exist on investing in the links of a stochastic network to improve its performance. Wollmer [36] studied the problem of transporting commodities from their source nodes to their demand nodes in a network with random link capacities. The objective is to identify an investment policy to add link capacities to minimize the expected total cost. The total cost is defined as the sum of the investment expenditure used to increase the capacities and the minimum transport cost after the realization of the updated capacities. A two-stage stochastic program is formulated and a solution procedure based on a cutting plane technique that exploits network structure is proposed. Wallace [34] also studied the problem of investing in the links of a network with random link capacities. A two-stage stochastic linear program is formulated to maximize the expected maximum flow from a source node to a sink node. Bounds for the second stage program are generated. Later, Wollmer [37] studied a variant of this problem with the objective of maximizing a linear combination of the expected maximum flow between two nodes and the negative of the investment cost. A two-stage linear program under uncertainty is formulated and solved using a cutting plane technique.

Studies on investing in a stochastic network, such as those discussed above, have primarily focused on the changes in the link capacities due to disruptions. The link capacities are considered to change with respect to a probability distribution [18,19] or use two possible states as either functional or nonfunctional (1 for functional and 0 otherwise) [26]. Chen et al. [8] claimed that when the disruption is a huge disaster rather than a simpler congestion, the 0-1 approach is most rational. Other than capacity reductions, the effect of disruptions in the network can be explained by increases in travel times [9,24]. By contrast, our problem focuses on ensuring connectivity and fast response time for first responders following a disaster. Hence, the purpose of investment in our study is to increase the link survival probabilities rather than capacity enhancements.

Network reliability and functional performance are inter-related objectives that characterize the vulnerability of networks subject to random link failures. Ball et al. [3] provide a comprehensive discussion on network reliability problems that are commonly characterized in terms of connectivity through non-failed links. Performability measures, such as overall or maximum delay, total travel time between O-D pairs, and average throughput, aim to measure the network performance when the network remains connected after component failures. They are typically tailored to the specific application domain. Sanso and Milot [27] define the performance of a transportation network as the network's ability to transport passengers from their origins to their destinations in a reasonable amount of time. For road networks, various connectivity, travel time and capacity reliability definitions have been proposed to assess network performance (see Chen et al. [8] and Clark and Watling [9] for the definitions). In connectivity reliability, under independent, probabilistic and binary mode of operation (functional or not), the probability of connectedness of an O-D pair is measured [9]. This also constitutes part of our system performance measure along with travel time in the surviving network.

A number of articles including [18],[19] have addressed the stochastic degradation of link capacity due to day-to-day traffic incidents rather than a disaster situation. These articles focus on travel time variability considering travelers' route choice behavior and traffic equilibrium. Along the similar lines, Clark and Watling [9] and Sumalee et al. [32] considered the effect of stochastic demand on network reliability. Lo et al. [18] considered stochastic demand and link capacity degradation simultaneously. Since work in this area remains out of the disaster context, the aim is not to reduce the vulnerability of a network to enhance its performance. Similarly, past studies in network reliability have not shown how to make link investment decisions to increase reliability. In this study we incorporate reliability and performance in a single framework and seek to enhance them through investment decisions.

In this paper, we: (i) propose a novel mathematical model for the strategic planning problem in disaster management to facilitate link investment decisions to enhance disaster response; (ii) derive structural results for the model that characterize system-level benefits of link investments; (iii) use the structural properties of the model to generate a locally optimal solution and illustrate the quality of the local optimum through computational tests; and (iv) demonstrate the computational tractability of the proposed method in a real-world application.

3 MATHEMATICAL MODEL

In this section we introduce the notation and formulate the investment problem as a two-stage stochastic program. We give the model for the single O-D case for ease of notation and understanding, without loss of generality.

We are given a directed network $G = (V, E)$ with node set V and arc set E, where the index i denotes a node and the index e denotes an arc in G. A pair of nodes is also specified: let O represent the origin node and D the destination node in G. From here on, we use the terms arc and link interchangeably without any distinction. We provide a list of the notation used throughout the article in the Appendix.

Let p_e $(0 < p_e < 1)$ denote the survival probability of link e. This probability can be increased to q_e by investing an amount equal to c_e (>0). We are given a budget B for investment in the links to increase their survival probabilities. The investment decision vector is denoted by $\boldsymbol{y} = (y_e)$, where y_e is binary-valued, taking the value 1 if there is an investment in link e, and 0 otherwise. After the occurrence of a disaster, each link would be either operational or non-operational. We use a binary-valued random variable ξ_e to denote the state of link e. That is, $\xi_e = 1$, if link e is operational after the disaster, and $\xi_e = 0$, otherwise. The vector of the random variables ξ_e for all links e in E is denoted by $\boldsymbol{\xi} = (\xi_e)$. It represents the network realization, and takes values in its support $\Xi \subseteq \{0, 1\}^{|E|}$ according to the decision-dependent probability distribution, $P(\cdot \mid \boldsymbol{y})$. A specific realization of $\boldsymbol{\xi}$ is denoted by $\tilde{\xi} = (\tilde{\xi}_e)$. A non-negative traversal cost t_e is specified, for all e in E. A unit flow is sent from O to D using the path with the least cost in the realized network. If there does not exist a path connecting O to D in this realization, a fixed penalty cost M $(< \infty)$ is incurred. This cost may represent the cost of an alternative mode of transportation such as the use of a helicopter, or may be used to represent the relative importance of connectivity. The flow vector from O to D is denoted by $\boldsymbol{x}(\tilde{\xi}) = (x_e(\tilde{\xi}))$, where $x_e(\tilde{\xi})$ is binary-valued, taking the value 1 if there is a unit flow through link e in the network realization $\tilde{\xi}$, and 0 otherwise.

We make several assumptions in formulating our problem.

Assumption 1: *Each link in G appears in at least one path from O to D.*

This is without loss of generality since any link which does not appear in any path from O to D can be dropped from G as it will not be invested in.

Assumption 2: *Link failures are independent.*

This assumption is often made in the network reliability literature [3]. In our context, as discussed in Section 1, due to the differences in the structural properties of the bridges, even links subject to the same earthquake magnitude may fail independently.

Assumption 3: *$M > T_{max}$, where we denote the maximum path cost from O to D in G by T_{max}.* The assumption implies that any O-D path in G is preferred over no such path being available.

The two-stage stochastic program **P** is given below.

Program **P**

First Stage:

$$Z = \min_{\boldsymbol{y}} F(\boldsymbol{y}) = \min_{\boldsymbol{y}} E_{\xi|\boldsymbol{y}}\left[f(\xi)\right] \tag{1}$$

subject to:

$$\sum_{e \in E} c_e y_e \leq B \tag{2}$$

$$y_e = 0 \text{ or } 1 \qquad \forall\, e \in E \tag{3}$$

Second Stage:

$$z(\tilde{\xi}) = \min \sum_{e \in E} t_e x_e(\tilde{\xi}) \tag{4}$$

subject to:

$$\sum_{e=(i,j)\in E} x_e(\tilde{\xi}) - \sum_{e=(j,i)\in E} x_e(\tilde{\xi}) = \begin{cases} 1, if\ i = O \\ -1, if\ i = D \\ 0, otherwise \end{cases} \qquad \forall i \in V \tag{5}$$

$$x_e(\tilde{\xi}) \leq \tilde{\xi}_e \qquad \forall\, e \in E \tag{6}$$

$$0 \leq x_e(\tilde{\xi}) \leq 1 \qquad \forall\, e \in E \tag{7}$$

where the first stage objective function, $F(\boldsymbol{y}) = E_{\xi|\boldsymbol{y}}[f(\xi)]$, is the expectation of $f(\xi)$ with respect to the random vector ξ for a given investment vector $\boldsymbol{y}$ and can be expanded as:

$F(\boldsymbol{y}) = \sum_{\tilde{\xi} \in \Xi} P(\xi = \tilde{\xi} | \boldsymbol{y}) f(\tilde{\xi})$. Here, $f(\tilde{\xi}) = z(\tilde{\xi})$, if the second stage problem for the realization $\tilde{\xi}$ is feasible, and $f(\tilde{\xi}) = M$, otherwise; thus, its value is equal to the least traversal cost from O to D in the network realization if a path exists from O to D, or the penalty cost M, if O-D is disconnected. Note that $P(\xi = \tilde{\xi} \,|\, \boldsymbol{y})$ is the probability that $\tilde{\xi}$ is realized given that the investment vector is $\boldsymbol{y}$. Due to independence assumption of the link failures, this value can be specified as:

$$P(\xi = \tilde{\xi} | \boldsymbol{y}) = \prod_{\forall e \in E} \{ \tilde{\xi}_e [(1 - y_e) p_e + y_e q_e] + (1 - \tilde{\xi}_e)[(1 - y_e)(1 - p_e) + y_e (1 - q_e)] \}. \quad (8)$$

This expression illustrates the decision-dependent nature of the probabilities in our model. Constraint (2) is the budget restriction on the total investment. Constraint set (3) is the integrality restrictions on the first stage decision variables. The second stage objective is given in (4). Constraint set (5) is the flow conservation constraint. Constraint set (6) precludes flow in links that are non-operational in $\tilde{\xi}$. Constraint set (7) defines the second stage flow variables which take either 0 or 1 value.

The complexity of **P** is indicated by the fact that even the computation of $F(\boldsymbol{y})$ for a given $\boldsymbol{y}$ vector is #P-Complete [3], the counting analogue of NP-Complete, due to $2^{|E|}$ possible network realizations. In addition, the need to evaluate the probability in (8) for each investment vector further increases complexity.

Stochastic Integer Programming (SIP) problems are inherently large scale, and decomposition methods such as Bender's decomposition are widely used to solve them [17]. When integer variables exist in the second stage, the non-convexity of the value function ([6]), creates further computational difficulty but combinations of decomposition and branch and cut methods have been used with increasing success [7], [28], [29] (for a survey on structural properties and algorithms for SIP models, see [15] and [28]).

The problem in this paper requires only the first stage variables to be integer but the probability distributions of the random parameters depend on these variables. A class of stochastic programs with decision-dependent uncertainty in which the optimization decisions influence the time of information discovery for a subset of the uncertain parameters was defined in Jonsbraten et al. [14] and later in Goel and Grossmann [11]. In [11], a Lagrangean duality based branch and bound algorithm is proposed, whereas in [14] an implicit enumeration algorithm is given. However, in our problem the decisions alter the probability distribution itself by making one network realization more likely than the others. Studies on such problems is very limited; Ahmed [1] presents some examples relating to network design, server selection and facility location. Models with decision-dependent probabilities are typically known to be quite difficult to solve and one way to solve them is by combining a search of the feasible space with sampling. The Sample Average Approximation (SAA) method ([2], [16]) approximates the expected objective function of the stochastic problem by means of a random sample from the set of scenarios. The resulting problem is then solved by deterministic optimization techniques. Here, we also use sampling to construct an approximate feasible solution but we do not utilize an iterative search method.

4 STRUCTURAL RESULTS FOR THE OBJECTIVE FUNCTION

In this section, we give an equivalent deterministic program of **P** by a path-based approach. Then we derive the multilinear functional form of $F(\boldsymbol{y})$ and characterize its coefficients. Consequently, we prove the monotone non-increasing property of $F(\boldsymbol{y})$. We also extend these results to the multiple O-D case.

4.1 Equivalent Deterministic Program of P

Let us first group the network realizations by defining S as the set of realizations that have O-D connectivity, and S^c as its complement. Expanding the expectation in $F(\boldsymbol{y})$ by conditioning it over network realizations in S and S^c, we have:

$$F(\boldsymbol{y}) = \sum_{\tilde{\xi}\in \boldsymbol{S}} P(\xi = \tilde{\xi}\,|\boldsymbol{y}) f(\tilde{\xi}) + \sum_{\tilde{\xi}\in \boldsymbol{S}^c} P(\xi = \tilde{\xi}\,|\boldsymbol{y}) f(\tilde{\xi}). \tag{9}$$

We next define $P(S|\boldsymbol{y}) = \sum_{\tilde{\xi}\in \boldsymbol{S}} P(\xi = \tilde{\xi}\,|\boldsymbol{y})$ as the probability that O is connected to D for a given $\boldsymbol{y}$, and $P(S^c|\boldsymbol{y}) = 1 - P(S|\boldsymbol{y})$ as its complement. With this notation,

$$F(\boldsymbol{y}) = \{\sum_{\tilde{\xi}\in S} \frac{P(\xi=\tilde{\xi}\,|\boldsymbol{y})}{P(S|\boldsymbol{y})} z(\tilde{\xi})\}\, P(S|\boldsymbol{y}) + MP(S^c|\boldsymbol{y}). \tag{10}$$

Let $F_S(\boldsymbol{y})$ denote the *expected least path cost* in case of O-D connectedness. Then, the objective function becomes:

$$F(\boldsymbol{y}) = F_S(\boldsymbol{y})P(S|\boldsymbol{y}) + MP(S^c|\boldsymbol{y}). \tag{11}$$

The objective function can now be viewed as combining two criteria: reliability and functional performance. Here, reliability refers to the probability that O is connected to D, and the functional performance is measured by the expected least path cost.

Let the set $K = \{ \pi_1, \pi_2, \ldots, \pi_{|K|} \}$ be the set of all paths from O to D in G, where $T(\pi_k)$ is the traversal cost of the path π_k. For practical purposes, we can identify a subset of paths from K using a k-shortest path algorithm to preclude the enumeration of all paths. Let the random variable π represent a path from O to D with the least traversal cost when O-D connectivity exist; here π takes values from the set K. Given a network realization $\tilde{\xi}$, let $I(k, \tilde{\xi})$ be an indicator variable for all $k \in K$. If any of the *O-D* paths in K has survived in this realization, we seek the one with the least cost path. Let k^* be the index of a least cost path in K. If several surviving paths have the least cost, then we pick one of them arbitrarily. We set $I(k^*, \tilde{\xi}) = 1$ and $I(k, \tilde{\xi}) = 0$ for all $k \in K - \{k^*\}$.

The expected least path cost $F_S(\boldsymbol{y})$ can now be expressed as

$$F_S(\boldsymbol{y}) = \sum_{k=1}^{|K|} \frac{P(\pi_k|\boldsymbol{y})}{P(S|\boldsymbol{y})} T(\pi_k),$$

where $P(\pi_k|\boldsymbol{y}) = P(\pi = \pi_k|\boldsymbol{y})$ is the probability that π_k is a least cost path given that the investment vector is $\boldsymbol{y}$. Thus, (11) can be rewritten as:

$$F(\boldsymbol{y}) = \sum_{k=1}^{|K|} P(\pi_k|\boldsymbol{y})\, T(\pi_k) + MP(S^c|\boldsymbol{y}). \quad (12)$$

Having this expression for the objective function $F(\boldsymbol{y})$ obviates the need to solve the second stage problem. We now have a single stage equivalent program of the two-stage program **P**.

$$Z = \min_{y \in Y}\{\sum_{k=1}^{|K|} P(\pi_k|\boldsymbol{y})\, T(\pi_k) + MP(S^c|\boldsymbol{y})\}$$

where $Y = \{\boldsymbol{y} | \sum_{e\in E} c_e y_e \leq B, \boldsymbol{y} \in \{0,1\}^{|E|}\}$.

This reformulated program is a 0-1 integer program with a single constraint but the objective function complicates its solution. An explicit expression of the objective function is elusive and depends on the network structure as both the $P(\pi_k|\boldsymbol{y})$ need to be derived and $P(S^c|\boldsymbol{y})$ has to be calculated (a computationally difficult task in itself, as mentioned earlier). However, we circumvent this difficulty by the following construction.

4.2 Multilinear Functional Form

The function $F(\boldsymbol{y})$ is defined only at the vertices of the unit hypercube, $H = \{ \boldsymbol{y} \mid 0 \leq y_e \leq 1, e \in E \}$. Given the discrete nature of the feasible set of **P**, we relax the integrality restrictions on the components of $\boldsymbol{y}$. This allows $F(\boldsymbol{y})$ to be continuously differentiable in the domain H, and hence enables the consideration of its Taylor series expansion in the neighborhood of some $\boldsymbol{y_0} \in H$:

$$F(\boldsymbol{y}) = F(\boldsymbol{y_0}) + \sum_{e\in E} g_e(\boldsymbol{y_0})(y_e - y_{0e}) + \frac{1}{2!}\sum_{e_1\in E}\sum_{e_2\in E} g_{e_1e_2}(\boldsymbol{y_0})(y_{e_1} - y_{0e_1})(y_{e_2} - y_{0e_2}) +$$
$$\cdots + \frac{1}{|E|!}\sum_{e_1\in E}\sum_{e_2\in E}\cdots\sum_{e_{|E|}\in E} g_{e_1e_2\cdots e_{|E|}}(\boldsymbol{y_0})(y_{e_1} - y_{0e_1})(y_{e_2} - y_{0e_2})\cdots\left(y_{e_{|E|}} - y_{0e_{|E|}}\right) + \cdots. \quad (13)$$

Here $g_e(\boldsymbol{y_0}) = \frac{\partial F(y)}{\partial y_e}|_{y=y_0}$ is the first order derivative with respect to investment in link e at $\boldsymbol{y_0}$, $g_{e_1e_2}(\boldsymbol{y_0}) = \frac{\partial^2 F(y)}{\partial y_{e_1}\partial y_{e_2}}|_{y=y_0}$ is the second order derivative with respect to investment in links e_1 and e_2 at $\boldsymbol{y_0}$, and so forth. Next, we have three Lemmas that characterize the derivatives. For convenience, let Δp_e denote $q_e - p_e$, and define $\boldsymbol{u_e}$ as the unit vector of dimension $|E|$ having 1 at component e and 0 at the remaining components. In Lemma 1, we characterize the first order derivative coefficient in (13).

Lemma 1: $g_e(\boldsymbol{y}) = F(\boldsymbol{y} + \boldsymbol{u_e}) - F(\boldsymbol{y})$.

Proof: Combining (9) and (12), in the path-based approach we have

$$F(\boldsymbol{y}) = \sum_{\tilde{\xi}\in \boldsymbol{S}} P(\xi = \tilde{\xi}|\boldsymbol{y})(\sum_{k=1}^{|K|} I(k, \tilde{\xi})\, T(\pi_k)) + \sum_{\tilde{\xi}\in \boldsymbol{S^c}} P(\xi = \tilde{\xi}\,|\boldsymbol{y})M. \quad (14)$$

Recall from equation (8), $P(\xi = \tilde{\xi} \mid \boldsymbol{y}) = \prod_{\forall e \in E} P(\tilde{\xi}_{\boldsymbol{e}} | y_e)$, where

$$P(\xi_{\boldsymbol{e}} | y_e) = \xi_e[(1 - y_e)p_e + y_e q_e] + (1 - \xi_e)[(1 - y_e)(1 - p_e) + y_e(1 - q_e)]. \quad (15)$$

If we rearrange and simplify the above, we obtain

$$P(\xi_{\boldsymbol{e}} | y_e) = \Delta p_e(2\xi_e - 1)y_e + \xi_e p_e + (1 - \xi_e)(1 - p_e). \quad (16)$$

Here we see clearly that $P(\xi_{\boldsymbol{e}} | y_e)$ is linear in y_e and thus $P(\xi = \tilde{\xi} \mid \boldsymbol{y})$ is multilinear. This in turn implies that $F(\boldsymbol{y})$ is multilinear since it consists of the sum of multilinear terms, as seen in (14).

Differentiating $P(\xi_{\boldsymbol{e}} | y_e)$ with respect to y_e gives $\Delta p_e(2\xi_e - 1)$ and furthermore,

$$\frac{\partial}{\partial y_e} P(\xi_{\boldsymbol{e}} | y_e) = \Delta p_e(2\xi_e - 1) = P(\xi_{\boldsymbol{e}} | y_e = 1) - P(\xi_{\boldsymbol{e}} | y_e = 0). \quad (17)$$

Since $F(\boldsymbol{y}) = \sum_{\tilde{\xi} \in \Xi} P(\xi = \tilde{\xi} | \boldsymbol{y}) f(\tilde{\xi}) = \sum_{\tilde{\xi} \in \Xi} [\prod_{\forall e \in E} P(\tilde{\xi}_{\boldsymbol{e}} | y_e)] f(\tilde{\xi})$, we obtain

$$g_e(\boldsymbol{y}) = \sum_{\xi} [P(\xi_{\boldsymbol{e}} | y_e = 1) - P(\xi_{\boldsymbol{e}} | y_e = 0)] \frac{P(\xi | \mathbf{y})}{P(\xi_{\boldsymbol{e}} | y_e)} f(\xi) \quad (18)$$

$$= \sum_{\xi} \frac{P(\xi_{\boldsymbol{e}} | y_e = 1)}{P(\xi_{\boldsymbol{e}} | y_e)} P(\xi | \boldsymbol{y}) f(\xi) - \sum_{\xi} \frac{P(\xi_{\boldsymbol{e}} | y_e = 0)}{P(\xi_{\boldsymbol{e}} | y_e)} P(\xi | \boldsymbol{y}) f(\xi)$$

$$= F(\boldsymbol{y} + \boldsymbol{u_e}) - F(\boldsymbol{y}). \quad \square$$

Next, Lemma 2 characterizes the second order derivative in (13).

Lemma 2: $g_{e_1 e_2}(\boldsymbol{y}) = F(\boldsymbol{y} + u_{e_1} + u_{e_2}) - F(\boldsymbol{y} + u_{e_2}) - F(\boldsymbol{y} + u_{e_1}) + F(\boldsymbol{y})$, *if* $e_1 \neq e_2$

= 0, if $e_1 = e_2$.

Proof: In the case when $e_1 \neq e_2$, differentiating $g_{e_1}(\boldsymbol{y})$, as given in Lemma 1, with respect to y_{e_2}, and performing algebraic manipulations yields the result. For the case when $e_1 = e_2$, $g_{e_1 e_2}(\boldsymbol{y}) = 0$ since $g_{e_1}(\boldsymbol{y})$ is independent of y_{e_1} as seen from (17). $\square$

We now introduce additional notation; let $\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}$ be vectors of dimension $|E|$, and $\|v_i\|$ be the length of the vector $\boldsymbol{v_i}$. This enables us to write

$g_{e_1 e_2}(\boldsymbol{y}) = \sum_{v_1 = 0, u_{e_1}} \sum_{v_2 = 0, u_{e_2}} (-1)^{2 - (\|v_1\| + \|v_2\|)} F(\boldsymbol{y} + v_1 + v_2)$, where each summation is defined over the two vectors specified. Next, the result of Lemma 2 is extended to the higher order derivatives.

Lemma 3: *For* $m \geq 3$,

$g_{e_1 \cdots e_m}(\boldsymbol{y}) = \sum_{v_1 = 0, u_{e_1}} \cdots \sum_{v_m = 0, u_{e_m}} (-1)^{m - (\|v_1\| + \cdots + \|v_m\|)} F(\boldsymbol{y} + v_1 + \cdots + v_m)$, *if* $e_1, \ldots, e_m$ *are distinct links, and* $g_{e_1 \cdots e_m}(\boldsymbol{y}) = 0$, *otherwise.*

Proof: The proof is by induction on m, and is omitted here.$\square$

Of particular interest is the expansion corresponding to $y_0 = \mathbf{0}$ (the vector of zeros) since it is a feasible solution to every instance of **P**. We denote the value of the expansion at $y_0 = \mathbf{0}$ as $\Phi(\boldsymbol{y})$.

We can interpret $g_e(\mathbf{0})$ as the marginal system-level benefit of investing in link *e* alone, and $g_{e_1 \cdots e_m}(\mathbf{0})$ as the marginal system-level benefit of investing simultaneously in links e_1 to e_m. Lemmas 1, 2 and 3 lead us to a key result of this paper, which enables the characterization of the structure of the objective function.

Theorem 1: *The multilinear function* $\Phi(\boldsymbol{y}) = F(\mathbf{0}) + \sum_{m=1}^{|E|}(\sum_{e_1,\cdots,e_m} g_{e_1 \cdots e_m}(\mathbf{0}) y_{e_1} \cdots y_{e_m})$ *coincides with* $F(\boldsymbol{y})$, for all $\boldsymbol{y}$ *in* $\{0, 1\}^{|E|}$.

Proof: Setting $\boldsymbol{y_0} = \mathbf{0}$ in the r.h.s. of (13) gives us the expression:
$\Phi(\boldsymbol{y}) = F(\mathbf{0}) + \sum_{m=1}^{|E|}(\frac{1}{m!}\sum_{e_1 \in E} \cdots \sum_{e_m \in E} g_{e_1 \cdots e_m}(\mathbf{0})\, y_{e_1} \cdots y_{e_m})$. There are *m*! number of $g_{e_1 \cdots e_m}(\mathbf{0})$ terms due to *m*! permutations for a given $e_1, \ldots, e_m$. Hence, the above becomes, $\Phi(\boldsymbol{y}) = F(\mathbf{0}) + \sum_{m=1}^{|E|}(\sum_{e_1,\cdots,e_m} g_{e_1 \cdots e_m}(\mathbf{0})\, y_{e_1} \cdots y_{e_m})$. Given $\boldsymbol{y}$, suppose we reindex the links such that the set of invested links is $W = \{1, \ldots, w\}$, i.e. $\boldsymbol{y_1} = \cdots = \boldsymbol{y_w} = 1$ and $\boldsymbol{y_{w+1}} = \cdots = \boldsymbol{y_{|E|}} = 0$. After applying Lemmas 1, 2 and 3, we get

$$\Phi(\boldsymbol{y}) = F(\mathbf{0}) + \sum_{m=1}^{w} \sum_{e_1,\cdots,e_m \in W}\{\sum_{v_1=0,u_{e_1}} \cdots \sum_{v_m=0,u_{e_m}} (-1)^{m-(\|v_1\|+\cdots+\|v_m\|)} F(v_1 + \cdots + v_m)\}.$$

Simplifying the above expression, all terms vanish except $F(\boldsymbol{u_1} + \cdots + \boldsymbol{u_w}) = F(\boldsymbol{y})$. □

Since the multilinear function in Theorem 1 yields the exact value of $F(\boldsymbol{y})$, for all feasible solutions, i.e. $\forall\, \boldsymbol{y} \in Y$, we have the following result.

Corollary 1: *Solving program* ***P*** *is equivalent to minimizing the multilinear function* $\Phi(\boldsymbol{y})$ *over the set Y*.

4.3 Extension to the multiple O-D case

We next consider the investment problem with the objective of minimizing the expected weighted sum of traversal costs of multiple O-D pairs. When the connectivity of some O-D pairs is considered to be more critical for post-disaster emergency response, different weights can be assigned to the pairs. Let *N* denote the total number of O-D pairs in *G* and *n* denote the index of an O-D pair. For an O-D pair *n*, the corresponding penalty cost is M_n and the weight is w_n. Let $f_n(\tilde{\xi})$ be the traversal cost for the n^{th} O-D pair under realization $\tilde{\xi}$.

Let $F_n(\boldsymbol{y}) = E_{\xi|\boldsymbol{y}}[f_n(\tilde{\xi})]$, and $\Phi_{\text{n}}(\boldsymbol{y})$ be the corresponding multilinear function specified as: $F_n(\mathbf{0}) + \sum_{m=1}^{|E|}(\sum_{e_1,\cdots,e_m} g_{e_1 \cdots e_m}(\mathbf{0})\, y_{e_1} \cdots y_{e_m})$. Then, the objective function is:
$F(\boldsymbol{y}) = E_{\xi|\boldsymbol{y}}[\sum_{n=1}^{N} w_n\, f_n(\xi)\,] = \sum_{n=1}^{N} E_{\xi|\boldsymbol{y}}[w_n\, f_n(\xi)] = \sum_{n=1}^{N} w_n F_n(\boldsymbol{y})$. As shown in Theorem 1, $F_n(\boldsymbol{y})$ coincides with $\Phi_{\text{n}}(\boldsymbol{y})$ for all $\boldsymbol{y}$ in $\{0,1\}^{|E|}$. Therefore, $\sum_{n=1}^{N} w_n F_n(\boldsymbol{y}) = \sum_{n=1}^{N} w_n \Phi_{\text{n}}(\boldsymbol{y})$ for all $\mathbf{y}$ in $\{0,1\}^{|E|}$ so that $\sum_{n=1}^{N} w_n F_n(\boldsymbol{y}) = F(\mathbf{0}) + \sum_{m=1}^{|E|}(\sum_{e_1,\cdots,e_m} g_{e_1 \cdots e_m}(\mathbf{0})\, y_{e_1} \cdots y_{e_m})$. Here, $F(\mathbf{0}) =$

$\sum_{n=1}^{N} w_n F_n(\mathbf{0})$ and $g_{e_1 \cdots e_m}(\mathbf{0}) = \sum_{n=1}^{N} w_n g_{e_1 \cdots e_m}(\mathbf{0})_n$. Thus, Theorem 1 and Corollary 1 follow for the multiple O-D case as well.

4.4 Monotonic Non-increasing Property of $F(y)$

Theorem 2: *The function $F(\boldsymbol{y})$ is monotonic non-increasing with $\boldsymbol{y}$.*

Proof : The result follows from the proposition proved next.

Proposition 1: $g_e(\boldsymbol{y}) \leq 0$, *for all* $e \in E$. *In addition, if* $q_e < 1, \forall e \in E$, *then* $g_e(\boldsymbol{y}) < 0$, $\forall e \in E$.

Proof : Since the events $\xi_e = 0$ and $\xi_e = 1$ are mutually exclusive, by conditioning we obtain:

$$g_e(\boldsymbol{y}) = \Delta p_e \{ \sum_{\tilde{\xi} \in \Xi} P(\xi = \tilde{\xi} | \boldsymbol{y}, \xi_e = 1) f(\tilde{\xi}) - \sum_{\tilde{\xi} \in \Xi} P(\xi = \tilde{\xi} | \boldsymbol{y}, \xi_e = 0) f(\tilde{\xi}) \}. \qquad (19)$$

We let $\tilde{w}_{-e} \cup \{e\}$ represent the realization $\tilde{w}_{-e}$ to which link e is added and made operational. We then have

$$g_e(\boldsymbol{y}) = \Delta p_e \{ \sum_{\tilde{w}_{-e} \in \Xi_{-e}} P(w_{-e} = \tilde{w}_{-e}, \xi_e = 1 | \boldsymbol{y}, \xi_e = 1) f(\tilde{w}_{-e} \cup \{e\}) - \sum_{\tilde{w}_{-e} \in \Xi_{-e}} P(w_{-e} = \tilde{w}_{-e}, \xi_e = 0 | \boldsymbol{y}, \xi_e = 0) | \boldsymbol{y}) f(\tilde{w}_{-e}) \}.$$

Due to the independence of link failures,

$$g_e(\boldsymbol{y}) = \Delta p_e \sum_{\tilde{w}_{-e} \in \Xi_{-e}} P(w_{-e} = \tilde{w}_{-e} | \boldsymbol{y}) \{ f(\tilde{w}_{-e} \cup \{e\}) - f(\tilde{w}_{-e}) \}.$$

Note that $f(w_{-e}): \Xi_{-e} \rightarrow R^+$ is a non-increasing set function, that is $f(\tilde{w}_{-e} \cup \{e\}) \leq f(\tilde{w}_{-e})$ $\forall \tilde{w}_{-e} \in \Xi_{-e}$, or equivalently, the addition of a link to $\tilde{w}_{-e}$ will never worsen the value of the traversal cost from O to D. This proves the non-positivity of $g_e(\boldsymbol{y})$. To prove the strict inequality, we consider two sets. The first set consists of network realizations from Ξ_{-e} in which $\tilde{w}_{-e}$ is disconnected for the O-D pair and $\tilde{w}_{-e} \cup \{e\}$ is connected. Due to Assumption 1, there exists at least one $\tilde{w}_{-e}$ belonging to the first set. In cases where survival probabilities after investment become 1, the marginal benefit of an additional investment may become 0. To understand this, let π_{+e} be a path that connects the O-D pair when link e is added to $\tilde{w}_{-e}$; if $P(\pi = \pi_{+e}) > 0$, then due to Assumption 2, $f(\tilde{w}_{-e} \cup \{e\}) - f(\tilde{w}_{-e}) < 0$ for this realization; thus $g_e(\mathbf{0}) < 0$. However, if $P(\pi = \pi_{+e}) = 0$, then $g_e(\mathbf{0}) \leq 0$. This is the case when some links have survival probability after investment equal to 1 so that investing in such inks may prevent other links from becoming part of a shortest path. Note that if $P(\pi = \pi_{+e}) = 0$ for all such paths, then $g_e(\mathbf{0}) = 0$. □

5 FIRST ORDER APPROXIMATE SOLUTION PROCEDURE

The optimization of a multilinear function over the set $\{0, 1\}^{|E|}$ is NP-hard [21]. Given this complexity, we develop an approximate solution procedure for **P**. We approximate the objective function using the first order terms of the function $\Phi(\boldsymbol{y})$ in Theorem 1. We define $\Phi^{(1)}(\boldsymbol{y}) = F(\mathbf{0}) + \sum_{e \in E} g_e(\mathbf{0}) y_e$ and approximate $F(\boldsymbol{y})$ with $\Phi^{(1)}(\boldsymbol{y})$. The drawback of this approach is that by disregarding the second and subsequent higher order terms, we cannot capture the effect of simultaneous investments in more than one link. However, with the first order approximation we gain tractability since $\Phi^{(1)}(\boldsymbol{y})$ is linear in y. Without the constant term $F(\mathbf{0})$, the approximate program is:

P-approx

$$\min_y \sum_{e\in E} g_e(0) y_e$$

subject to

$$\sum_{e\in E} c_e y_e \leq B$$

$$y_e = 0 \text{ or } 1 \qquad \forall\, e \in E$$

Here, $g_e(\mathbf{0})$ can be evaluated from $F(\boldsymbol{u_e}) - F(\mathbf{0})$ using Lemma 1 after k-shortest paths are found. However, while taking the expectation, the number of network realizations is $2^{|E|}$. Therefore, a Sample Average Approximation approach can be used to estimate this value when the number of links gets large. Once the $g_e(\mathbf{0})$ are calculated exactly or approximately, **P-approx** is a 0-1 knapsack problem that can be solved efficiently either in pseudo-polynomial time using dynamic programming (see [20]), or by branch-and-bound using a standard MIP solver. Let $\boldsymbol{y_a}$ be an optimal solution of **P-approx** when $g_e(\mathbf{0})$ are calculated exactly. Next, we show that this solution is a strict local optimum of the original problem **P**.

Theorem 3: *An optimal solution to **P-approx**, $\boldsymbol{y_a}$, is a local optimum of **P**.*

Proof: The solution $\boldsymbol{y_a}$ is an extreme point of the unit hypercube H. On the contrary, assume that this solution is not a local optimum of **P**. Then, there must be a feasible extreme point solution, $\tilde{\boldsymbol{y}}$, neighboring $\boldsymbol{y_a}$, such that $F(\tilde{y}) < F(\boldsymbol{y_a})$ such that the Hamming distance between $\boldsymbol{y_a}$ and any of its neighboring solutions is 1, that is, $\|\boldsymbol{y_a} - \tilde{y}\| = 1$. From Proposition 1, we have $\tilde{\boldsymbol{y}} > \boldsymbol{y_a}$, that is, for all $e \in E$, $\tilde{y}_e \geq y_{a_e}$ and at least for one component the inequality is strict. Since $g_e(\mathbf{0}) \leq 0\ \forall\ e \in E$, $\Phi^{(1)}(\tilde{\boldsymbol{y}}) \leq \Phi^{(1)}(\boldsymbol{y_a})$. Since **P-approx** has the same constraint set as **P**, $\tilde{\boldsymbol{y}}$ is both feasible and optimal to **P-approx** as well. □

6 COMPUTATIONAL STUDY AND INSIGHTS

In this section, we provide computational results using a case study to demonstrate the practical applicability of our approach. We also illustrate the quality of the local optimum solution compared to the global optimum for instances in which the global optimum could be obtained by enumeration.

The computational study is based on highway networks from Istanbul, Turkey. Istanbul has been affected by two major earthquakes in 1999 with epicenters about 250 km from it that caused \$10-\$25 billion in damage. Parsons et al. [24] report that the probability of a major earthquake epicentered in Istanbul itself in the next few decades is 62.6±15%; such an earthquake would cause substantially higher damage than the 1999 ones due to the high population density. The government plans to invest $400 million to strengthen critical public infrastructure for earthquake resistance. A key element of this plan is to retrofit the highway system to ensure maximum accessibility and functionality after an earthquake. This implies the seismic retrofit of bridges/viaducts which tend to be the weakest structural elements in the highway system. Selecting the highway elements to be strengthened provides a real setting to analyze the practical use of our model.

The relevant data is obtained from the 2003 Master Earthquake Plan (MEP) of the Istanbul municipality, and focuses on the two main highways TEM and E-5 in the city and the bridges/viaducts located on them. The associated map is depicted in Figure 1. It shows the southern part of the city which is the most densely populated and seismically risk-prone area. The city is separated into the European and Asian sides by the Strait of Bosphorus and two bridges connect the two sides. In our experiments, we consider two networks to analyze our model. The first one consists of 25 nodes and 30 links and includes both sides as shown in Figure 2. It is used to analyze the performance and the computational scalability of the model. The second network has 8 nodes and 9 links (Figure 3) and represents only the Asian side. It is considered because the global optimum can be obtained by enumeration, thereby providing a benchmark for the quality of the local optimum obtained through our approach. The link traversal costs are proportional to the distances between the nodes.

The initial link survival probabilities are typically determined by structural engineers using domain-specific information. In this study, we use data from the MEP (Figure 1) that classifies bridges/viaducts as “less risky” and “very risky”, to determine the probabilities. By identifying the numbers of less risky and very risky bridges/viaducts on each link, and determining a weighted score which is then translated into the survival probabilities, five initial link survival probability levels (ranging from 0.5 to 0.8) are assigned to each link. The link survival probabilities are assumed to be 1 after the investment, based on feedback from the structural engineers involved in the retrofitting plan. The investment cost for each link is calculated as a weighted score proportional to the link length and the number of bridges/viaducts located on it. The traversal costs (lengths), survival probabilities and the investment costs are given in Table 1.

Table 1. Link lengths, investment costs and survival probabilities.

Link	*length*	c_e	p_e	Link	*length*	c_e	p_e	Link	*length*	c_e	p_e
1	2.46	80	0.8	11	7.11	940	0.55	21	1.80	40	0.8
2	2.20	80	0.8	12	4.03	160	0.8	22	1.97	160	0.7
3	8.00	320	0.8	13	5.02	620	0.6	23	1.61	40	0.8

4	2.56	260	0.7	14	4.55	1180	0.5	24	8.09	620	0.6
5	4.57	160	0.8	15	1.36	40	0.8	25	2.87	260	0.7
6	3.44	420	0.6	16	4.26	940	0.55	26	6.35	780	0.6
7	4.19	160	0.8	17	3.64	300	0.7	27	2.27	800	0.55
8	5.60	620	0.6	18	4.19	520	0.6	28	3.91	120	0.8
9	3.71	120	0.8	19	1.98	40	0.8	29	4.11	220	0.7
10	4.44	340	0.7	20	2.45	800	0.55	30	2.27	500	0.6

The total budget needed to invest in all links is 11640 units. We consider three budget levels for the experiments: B1=1164, B2=2328, and B3=3492, corresponding to strengthening approximately 10%, 20% and 30% of the links, respectively.

The O-D pairs are chosen based on detailed analyses of four most likely earthquake scenarios for the region of interest given in the Japan International Cooperation Agency Report, 2002. In these scenarios, the expected number of collapsed buildings, and the number of fatalities and injuries in each district of the region are estimated. The most-damaging earthquake scenario provides the basis for the selection of the O-D pairs. The origins correspond to the districts with the highest expected number of injured people. The destination nodes are the districts which have a large medical support capacity. For the 30-link network, the following O-D pairs are chosen: (14,20), (14,7), (12,18), (9,7) and (4,8). For the 9-link network, (17,19) and (15,22) are selected. Each O-D pair has up to 6 different paths connecting them. Links belonging to each path are given in Table 2 with corresponding traversal costs.

Table 2. Paths for each O-D pair with corresponding traversal costs (lengths).

O-D pair 14-20	Links								Total length
Path 1	21	22	25						6.65
Path 2	21	22	26	29	30	28			20.41
Path 3	20	17	18	23	24	26	25		29.20
Path 4	20	17	18	23	24	29	30	28	30.27
									M=31
O-D pair 14-7	**Links**								**Total length**
Path 1	20	16	10						11.14
Path 2	20	17	14	13	10				20.09
Path 3	20	17	14	11	12	9			25.48
Path 4	20	16	13	11	12	9			26.58
Path 5	20	17	14	11	6	7	9		29.08
Path 6	20	16	13	11	6	7	9		30.17
									M=31
O-D pair 12-18	**Links**								**Total length**
Path 1	17	20	21	22					9.86
Path 2	14	13	16	20	21	22			20.05
Path 3	18	23	24	26					20.24
Path 4	18	23	24	29	30	28	25		27.06
									M=28

O-D pair 9-7	Links							Total length
Path 1	13	10						9.46
Path 2	11	12	9					14.85
Path 3	14	17	16	10				16.88
Path 4	11	6	7	9				18.45
								M=19

O-D pair 4-8	Links							Total length
Path 1	3	4	6					14.00
Path 2	5	8	9	12				17.91
Path 3	3	4	7	12				18.79
Path 4	5	8	9	7	6			21.51
Path 5	5	8	10	13	11			26.73
Path 6	5	8	10	16	17	14	11	34.15
								M=35

O-D pair 17-19	Links							Total length
Path 1	31	22	25	27				13.52
Path 2	24	26	25	27				19.59
Path 3	24	29	30	28	27			20.65
Path 4	31	22	26	29	30	28	27	27.28
								M=30

O-D pair 15-22	Links							Total length
Path 1	22	25	28					8.75
Path 2	22	26	29	30				14.70
Path 3	31	24	29	30				20.87
Path 4	31	24	26	25	28			27.62
								M=30

6.1 Monte Carlo Sampling-Based Implementation Procedure

This section describes a Monte Carlo sampling procedure to implement the proposed method in practice. It is important to note that while computation time is not a key factor in the deployment of this method due to its pre-disaster planning context, we nevertheless need a procedure that is efficient for tractability.

Section 4.3 showed that the objective coefficients $g_e(\mathbf{0})$ of the approximate integer program **P-approx** are computed as $F(\boldsymbol{u_e})$-$F(\mathbf{0})$ for a given O-D pair. However, calculating $F(\boldsymbol{y})$ for any $\boldsymbol{y}$ requires exploring an exponential number of possible network realizations. For example, for the 30-link network, $30\text{x}2^{30}$ ($\approx$ 32 billion) cases should be explored. As a practical approach to overcome this difficulty, Monte Carlo Sampling is used to estimate the $F(\boldsymbol{u_e})$ and $F(\mathbf{0})$ values. First, the k-shortest paths are determined in advance for the O-D pair under consideration (as in Table 2). To estimate $F(\boldsymbol{u_e})$ for each link e, one million random network realizations are generated such that the links are either operational or non-operational according to the probabilities determined by the investment vector $\boldsymbol{u_e}$. For each realization, the O-D connectivity of the predetermined k-shortest paths is checked in terms of the increasing order of traversal cost to find the minimum cost operational path. If the O-D pair is not connected in that realization, the traversal cost is taken as M. The average of these 1 million traversal costs is the estimated $F(\boldsymbol{u_e})$. The procedure is repeated for each link with a different set of realizations generated for the corresponding post-investment probabilities. Finally, one million random network realizations are generated to estimate $F(\mathbf{0})$, and the $g_e(\mathbf{0})$ estimates are obtained. This computation takes about 380 seconds of CPU time for one O-D pair in the 30-link network on a PC with 2x2.8 GHz Xeon processor and 5 GB RAM using the sampling algorithm implemented in Matlab 7.0.

Sample Size and Convergence. We investigate the convergence of the estimated $F(\boldsymbol{u_e})$ values with increasing sample size of the generated realizations for the 30-link network. Table 3 shows estimated $F(\boldsymbol{u_e})$ values, denoted by $\hat{F}(\boldsymbol{u_e})$, for a representative subset of links and O-D pair (14,7), for sample sizes varying from 10 to 1,000,000. Results indicate that the estimated values converge rapidly as the sample size approaches to 1,000,000.

Table 3. Convergence of the estimated $\hat{F}(\boldsymbol{u_e})$ values for different sample sizes.

Sample Size / $\hat{F}(\boldsymbol{u_e})$	10	100	1000	10,000	100,000	200,000	400,000	600,000	800,000	1,000,000
e=1	68.5476	99.7284	90.4219	89.2341	89.0514	88.8871	88.915	88.932	89.0732	89.0106
e=2	87.3419	87.2498	91.5962	89.1253	88.8526	89.0942	88.9732	88.9392	89.0564	89.0394
e=3	89.7802	87.2862	88.5659	88.8772	88.7807	88.9159	89.1911	89.0184	88.9748	88.9921
e=29	65.5699	83.5401	88.2562	88.7485	89.0428	89.1445	88.983	89.0426	89.0936	89.0118
e=30	76.4559	95.9406	88.9307	88.6466	89.0565	88.9696	89.0181	88.9746	88.9642	88.9543

Confidence Intervals. The estimations are subject to error and a confidence interval is computed for each estimated value by the standard formula $\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = \bar{x} \pm \Delta$, where $\bar{x}$ is taken as $\hat{F}(\boldsymbol{u_e})$, σ as the sample standard deviation, n as 1,000,000 and for a desired confidence of 90%, z is taken to be 1.645. With these parameters, the Δ values for the O-D pair (14,7) are found to be at most 0.0677 illustrating that sufficient accuracy is obtained with the selected sample size.

6.2 Insights on the Solution Method and Parameters

This section provides insights on the interpretation of the $g_e(\mathbf{0})$ values and the effect of parameter M on the solution using the 30-link network, the five selected O-D pairs and the paths given in Table 2.

Interpretation of the $g_e(\mathbf{0})$ Values. The $g_e(\mathbf{0})$ values for the five O-D pairs are summed up to form the objective coefficients of the knapsack problem **P-approx.** These calculated coefficients are given in Table 4 for two cases: high and low M. That is, first M is set to a high value, 120, for all O-D pairs to give more importance to connectivity and the first row results are obtained after sampling. The second row results are calculated using minimal M values for each O-D, as specified in Table 2. Table 5 shows the solutions to the knapsack problems using these objective coefficients for the two cases, each one under increasing budget levels, B1, B2 and B3. Solving a single knapsack problem takes less than one second on our computing platform.

Table 4. Objective coefficients of **P-approx** for the 30-link network cases.

Link	1	2	3	4	5	6	7	8	9	10
High M	0.4796	0.0909	-7.5522	-12.962	-4.6874	-10.1104	-3.5077	-12.6736	-9.62	-23.4098
Low M	0.2411	0.2723	0.2486	0.2546	0.2583	0.2665	0.2172	0.3085	0.0025	-3.0429

Link	11	12	13	14	15	16	17	18	19	20
High M	-22.5208	-7.959	-20.1996	-10.7251	-0.2549	-19.1451	-12.0672	-11.7344	0.299	-45.7208
Low M	-0.5788	0.061	-2.3111	-0.5424	0.2497	-3.0092	-2.8857	-1.1547	0.2678	-9.7512

Link	21	22	23	24	25	26	27	28	29	30
High M	-16.122	-27.9923	-4.6131	-12.0463	-16.4982	-9.5883	0.0067	-1.7972	-2.5347	-4.4025
Low M	-4.0536	-7.1832	-0.2923	-1.1597	-3.4785	-1.3181	0.2753	0.1773	0.1187	0.0333

Table 5. The knapsack problem solutions for the 30-link network cases.

(a) ***High M*.**

Link	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
B1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0
B2	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	1	1	1	1	0	1	0	0	0	0	0
B3	0	0	0	1	1	0	1	0	1	1	0	1	1	0	1	0	1	0	0	1	1	1	1	0	1	0	0	0	0	0

(b) ***Low M*.**

Link	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
B1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0
B2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	1	1	0	1	0	0	0	0	0
B3	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0

The objective coefficient for a link represents the marginal system-level benefit of investing in that link alone; hence links with the most negative objective coefficients are good candidates for investment. In Table 4, the objective coefficients with the most negative values appear in links 10, 11, 13, 16, 20, and 22, for M =120, and links 10, 16, 20, 21, 22, and 25, for M =31. Note that all of the coefficients should be non-negative due to Proposition 1, but Table 4 contains some positive values, resulting from the accumulation of the sampling errors; thus, these coefficients should be considered as zero. When the solutions are examined in Table 5, we note that links 10, 20, 21, 22, 23 and 25 have been invested in under most budget levels. Links 11, 14 and 16 have high investment costs, while 23 has very low cost. Correspondingly, although 16 has a high negative coefficient, it is not invested in; whereas link 23 is invested in despite its very low objective coefficient.

Effect of Parameter M. Table 5 illustrates the effect of M on the solution. As discussed earlier, larger M implies greater emphasis on connectivity. In Table 5, we note that link 9 is invested in under all budget levels when M is high, and under none when M is low. Similar results are seen for link 4. Links 4 and 9 provide key options for connectivity as seen in Figure 2.

6.3 Quality of the Local Optimum Solution

As stated earlier, the 9-link network was considered primarily to analyze the solution quality as it allows for the enumeration of the solution. In order to increase the number of possible paths between the O-D pairs, we added a new link, link 31 (Figure 3) that represents the coastal road. Tables 6 and 7 report investment vectors, $\boldsymbol{y_a}$, obtained by solving **P-approx** when its objective coefficients are calculated with respect to sample size 50,000, for O-D pairs 17-19 and 15-22, respectively. M is set to 30 for both O-D pairs. Results are given for 20 cases with randomly generated survival probabilities before and after investment, and costs of investment for each link. Cases i and $i+1$, i=1,3,5,…,19 correspond to the same set of survival probabilities and link traversal costs but Case i allows a budget of 20% and Case $i+1$, 50% of the total amount needed to invest in all the links. $F(\boldsymbol{y_a})$ is the objective function value of the problem **P** at the proposed approximate solution $\boldsymbol{y_a}$ which is obtained by enumerating all network realizations. $F(\boldsymbol{y_a})$ is compared to $F(\boldsymbol{y_{opt}})$, where the optimal solution $\boldsymbol{y_{opt}}$ is found by enumerating all possible investment vectors. We were able to find exact solutions in 11 cases. In the rest of the cases, average deviation from the optimal objective value is 0.5113 which is less than 2% of the optimal value on the average, whereas the maximum percentage error is 10%.

Table 6. Solution of **P-approx** and comparison with global optimum for the O-D pair 17-19.

					Link					
Case	22	24	25	26	27 28	29	30	31	$F(\mathbf{y_a})$	$F(\mathbf{y_a})$- $F(y_{opt})$

Case	22	24	25	26	27	28	29	30	31	$F(\mathbf{y_a})$	$F(\mathbf{y_a})$- $F(y_{opt})$
1	0	0	1	0	0	0	1	0	0	29.0764	0.5745
2	1	0	1	0	1	0	0	1	0	28.8807	1.6487
3	1	0	0	0	0	1	1	0	0	24.1741	0
4	1	0	0	1	0	0	1	0	1	21.7124	0.1309
5	0	1	0	0	1	0	0	0	1	19.8715	2.0411
6	1	1	0	1	1	0	0	1	1	17.2518	0
7	1	1	0	0	0	0	0	0	1	23.6556	0
8	1	1	1	0	0	0	1	0	1	22.6957	0.3263
9	0	1	0	1	1	0	0	0	0	25.7185	0
10	0	1	1	1	1	0	0	0	1	24.7722	0

Table 7. Solution of **P-approx** and comparison with global optimum for the O-D pair 15-22.

Case	22	24	25	26	27	28	29	30	31	$F(\mathbf{y_a})$	$F(\mathbf{y_a})$- $F(y_{opt})$
11	0	0	1	0	0	0	0	1	0	15.2981	0
12	0	0	1	1	0	1	1	1	0	11.6103	0
13	0	0	1	0	0	1	1	1	0	9.7503	0.264
14	1	0	1	0	0	1	1	1	0	9.1319	0
15	0	0	1	0	0	1	0	0	0	16.1811	0
16	1	1	1	0	0	1	0	1	0	15.814	0.0107
17	1	1	0	0	0	1	0	0	1	20.9825	0
18	1	1	1	0	0	0	1	0	1	19.3724	0.0596
19	0	0	1	0	0	1	1	1	0	26.1578	0
20	1	0	1	0	0	1	1	1	0	23.5729	0.0574

7 CONCLUSIONS

We addressed a pre-disaster planning problem for earthquake disaster response that aims to strengthen the links of a stochastic network through investment. The strength of each link is measured by the probability that all of its weak components such as bridges and viaducts, remain operational after a possible disaster. The objective is to minimize the expected traversal cost for multiple O-D pairs across the post-disaster network realizations. A two-stage stochastic program is presented where the first stage identifies which links to invest in and the second stage determines the minimum traversal costs between the O-D pairs. By differentiating between connected and disconnected realizations, reliability is explicitly considered as part of the objective function. A path-based approach was used to reformulate the program as an equivalent deterministic program.

By relaxing the integrality of investment variables and applying Taylor series expansion, a multilinear function was derived which coincides with the objective function at the feasible integer solutions. Coefficients of this function denote the marginal system-level benefits due to investment. Using only the first order terms of the multilinear function as an approximation yielded a knapsack problem whose optimal solution is shown to be a local optimum to the original problem. The proposed approach captures the system-level effects of investment, thereby addressing the trade-offs between system-level benefits of investment, the budget limitations, and the investment costs.

Numerical experiments are illustrated on a real-world case related to strengthening Istanbul's urban highway system against earthquake risk. The problem was solved on a 30-link in less than 7 minutes by utilizing Monte Carlo sampling of the network realizations. The experiments provided insights on the effects of problem parameters on the solutions. The quality of the solutions was investigated on a 9-link network by comparing the approximate solutions with those obtained through enumeration. The proposed method found the global optimum in more than half of all cases in negligible computation time, and yielded an average percentage error of 2% in the remaining cases.

The proposed model and the solution approach can be used by local and central government agencies to aid investment decisions to upgrade a highway network for disaster response.

8 ACKNOWLEDGEMENTS

We acknowledge support from the Indiana Department of Transportation for research in disaster management, which motivated this research topic. We also acknowledge financial support from the Turkish Science and Technology Foundation (TUBITAK) and data support from Dr. Metin Çancı at the Istanbul Metropolitan Planning Department of the Municipality.

9 APPENDIX: LIST OF NOTATION

$G(V,E)$	a directed network with node set *V* and arc set *E*.
p_e	survival probability of link *e* without investment.
q_e	survival probability of link *e* after being invested in, where $\Delta p_e = q_e - p_e$.
c_e	cost of investing in link *e*.
B	budget.
y_e	1, if there is an investment in link *e*; 0, otherwise.
$y = (y_e)$	the investment decision vector for all links in *E*.
ξ_e	1, if link *e* is operational after the disaster; 0, otherwise.
$\xi = (\xi_e)$	the vector of the random variables ξ_e for all links in *E* with a specific realization $\tilde{\xi} = (\tilde{\xi}_e)$
t_e	non-negative traversal cost for link *e*.
M	a fixed penalty cost.
$T_{\max}$	the maximum path cost from *O* to *D* in *G*.
$x_e(\tilde{\xi})$	1, if there is a unit flow through link *e* in the network realization $\tilde{\xi}$; 0, otherwise.
$x(\tilde{\xi}) = (x_e(\tilde{\xi}))$	the flow vector for realization $\tilde{\xi}$.
$f(\xi)$	least path cost in the network realization ξ if it exists, or the penalty cost *M* if O-D is not connected.
$F(y)$	expectation of $f(\xi)$ with respect to the random variable ξ for a given investment vector *y*.
S	the set of network realizations that have O-D connectivity.
S^c	the complement set of S.
$F_S(y)$	expected traversal cost over all realizations with O-D connectivity for the investment vector *y*.
π	random variable representing the path with the least cost from *O* to *D* when connectivity exists.
$I\{k,\tilde{\xi}\}$	1, if the least cost path in the network realization $\tilde{\xi}$ is π_k; 0, otherwise.
$T(\pi_k)$	traversal cost of path π_k.
$T_{\max}$	the maximum path cost from *O* to *D* in *G*.
u_e	the unit vector of dimension $\|E\|$ having 1 at component *e* and 0 at the remaining components.
Ξ_{-e}	the set of network realizations of *G-e*.
w_{-e}	a random variable for the network realization of *G-e* with a specific realization $\tilde{w}_{-e}$.

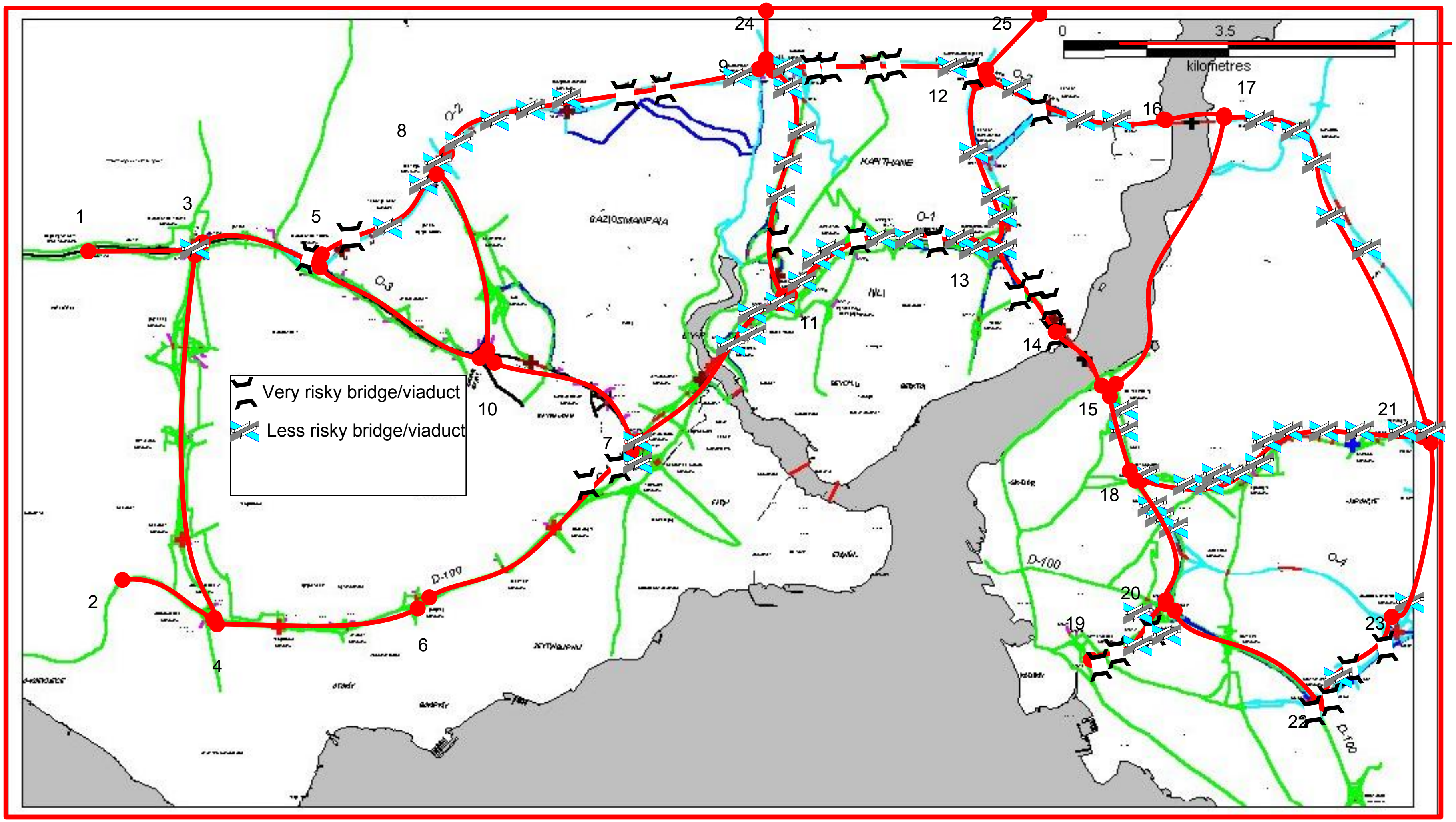

Figure 1. The network shown on the Istanbul city map with risky bridges and viaducts on major highways.

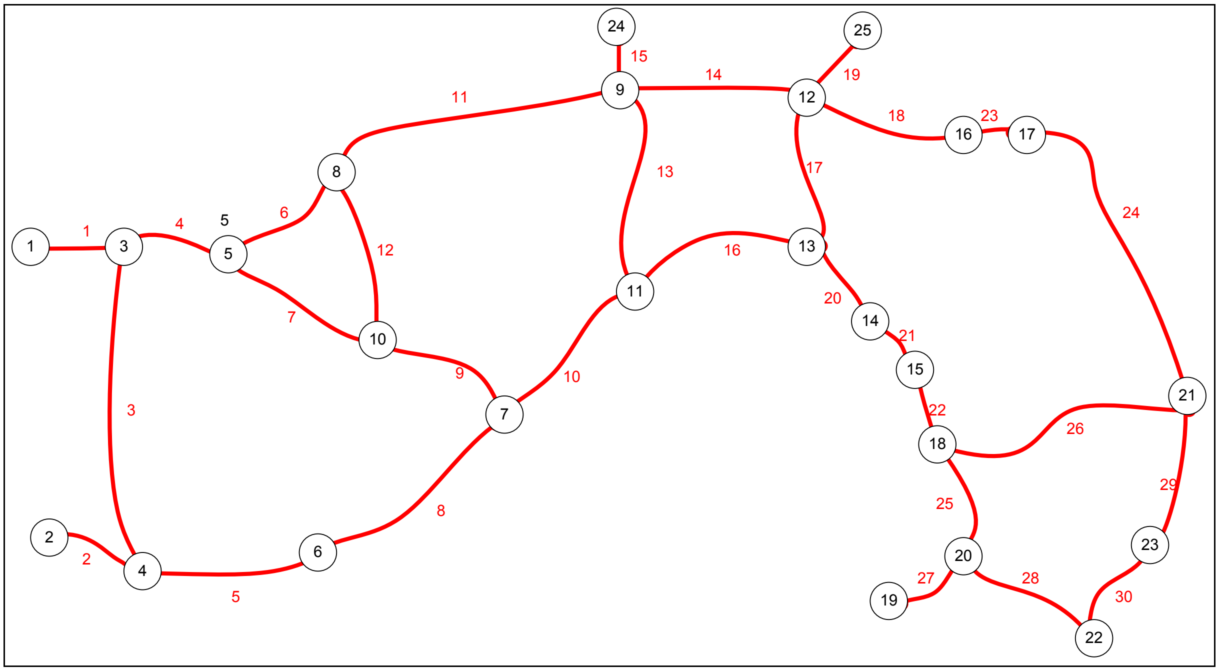

Figure 2. 30-link network of Istanbul.

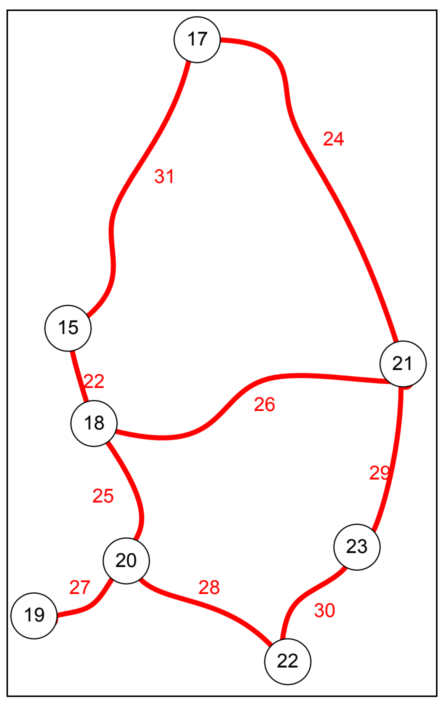

Figure 3. 9-link network of the Asian side of Istanbul.

10 REFERENCES

[1] Ahmed S. Strategic planning under uncertainty: Stochastic integer programming approaches. Ph.D. thesis, University of Illinois at Urbana-Champaign, 2000.

[2] Ahmed S, Shapiro A. The sample average approximation method for stochastic programs with integer recourse. SIAM Journal of Optimization 2002; 12: 479-502.

[3] Ball MO, Colbourn CJ, Colbourn JS. Network reliability. In: Ball MO et al. (Eds.). Handbooks of Operations Research and Management Science: Network Models. Amsterdam: Elsevier Science; 1995.

[4] Bana e Costa CA, Oliveira CS, Vieira V. Prioritization of bridges and tunnels in earthquake risk mitigation using multicriteria decision analysis: Application to Lisbon. Omega 2008; 36(3): 442-50.

[5] Basöz N, Kiremidjian AS. Prioritization of bridges for seismic retrofitting. Report 114, John Blume Earthquake Engineering Center, Department of Civil Engineering, Stanford University, Stanford, CA; 1995.

[6] Blair CE, Jeroslow RG. The value function of an integer program. Math. Program. 1982; 23: 237–273.

[7] Carøe CC, Tind J. L-shaped decomposition of two-stage stochastic programs with integer recourse. Math. Program. 1998; 83: 451–464.

[8] Chen A, Yang H, Lo HK, Tang WH. Capacity reliability of a road network: An assessment methodology and numerical results. Transportation Research Part B-Methodological 2002; 36(3): 225-52.

[9] Clark S, Watling D. Modeling network travel time reliability under stochastic demand. Transportation Research Part B-Methodological 2005; 39(2): 119-40.

[10] Cooper JD, Friedland IM, Buckle IG, Nimis RB, Bobb NM. The Northridge Earthquake: Progress Made, Lessons Learned in Seismic-resistant Bridge Design. Public Roads 1994; 58(1):

[11] Goel V, Grossmann IE. A class of stochastic programs with decision dependent uncertainty. Math. Programming, Ser. B 2006; 108: 355–394.

[12] Goel V, Grossmann IE. A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves. Computers and Chemical Engineering 2004; 28: 1409–1429.

[13] Goel V, Grossmann IE, El-Bakry AS, Mulkay EL. A novel branch and bound algorithm for optimal development of gas fields under uncertainty in reserves. Computers and Chemical Engineering 2006; 30: 1076–1092.

[14] Jonsbraten TW, Wets RJB, Woodruff DL. A class of stochastic programs with decision dependent random elements. Annals of Operations Research 1998; 82: 83–106.

[15] Klein Haneveld WK, van der Vlerk MH. Stochastic integer programming: General models and algorithms. Annals of Operations Research 1999; 85: 39–57.

[16] Kleywegt AJ, Shapiro A, Homem-de-Mello T. The sample average approximation method for stochastic discrete optimization. SIAM Journal on Optimization 2002; 12(2): 479–502.

[17] Laporte G, Louveaux FV. The integer L-shaped method for stochastic integer programs with complete recourse. Oper. Res. Lett. 1993; 13: 133–142.

[18] Lo HK, Luo XW, Siu BWY. Degradable transport network: Travel time budget of travelers with heterogeneous risk aversion. Transportation Research Part B-Methodological 2006; 40(9): 792-806.

[19] Lo HK, Tung YK. Network with degradable links: Capacity analysis and design. Transportation Research Part B-Methodological 2003; 37(4): 345-63.

[20] Martello S, Toth P. Knapsack Problems. In: Algorithms and Computer Implementations. Wiley; 1990.

[21] Masti CL, Vidyasagar M. A stochastic high-order connectionist network for solving inferencing problems. In: Proceedings of Int. Joint. Conf. Neural Networks 1991. Singapore: 911-6.

[22] Moghtaderi-Zadeh M, Der Kiureghian A. Reliability upgrading of lifeline networks for post-earthquake serviceability. Earthq. Engrg. Struct. Dyn. 1983;11(4):557-66.

[23] Padgett JE, DesRoches R. Methodology for the development of analytical fragility curves for retrofitted bridges. Earthq. Engrg. Struct. Dyn. 2008;37:1157-1174.

[24] Parsons T, Toda S, Stein RS, Barka A, Dieterich JH. Heightened odds of large earthquakes near Istanbul: An interaction-based probability calculation. Science 2000;288:661-5.

[25] Ramirez J, Peeta S, Sozen M, Garcia L, Viswanath K. Emergency earthquake routes for the state of Indiana; Part I: Criteria for selection of primary routes: transportation aspects; Part II: Route seismic vulnerability aspects. Report Number FHWA/IN/JTRP-2003/22 2005. Joint Transportation Research Program, Purdue University. 504 pages.

[26] Sanchez-Silva M, Daniels M, Lleras G, Patino D. A transport network reliability model for the efficient assignment of resources. Transportation Research Part B-Methodological 2005;39(1):47-63.
[27] Sanso B, Milot L. Performability of a congested urban transportation network when accident information is available. Transportation Science 1999; 33(1): 68-79.
[28] Schultz R. Stochastic programming with integer variables. Math. Program. Ser. B 2003; 97: 285–309.
[29] Sen S, Sherali HD. Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. Math. Program. Ser. A 2006; 106: 203–223.
[30] Sohn J. Evaluating the significance of highway network links under the flood damage: An accessibility approach. Transportation Research Part A: Policy and Practice 2006; 40(6): 491-506.
[31] Sohn J, Kim TJ, Hewings GJD, Lee JS, Jang S. Retrofit priority of transportation network links under an earthquake. Journal of Urban Planning and Development 2003; 129(4): 195–210.
[32] Sumalee A, Watling DP, Nakayama S. Reliable network design problem: Case with uncertain demand and total travel time reliability. Transportation Research Record: Journal of the Transportation Research Board 2006; 1964: 81–90.
[33] Viswanath K, Peeta S. The multicommodity maximal covering network design problem for planning critical routes for earthquake response. Trans. Res. Rec. 2003; 1857: 1-10.
[34] Wallace, SW. Investing in arcs in a network to maximize the expected max flow. Networks 1987; 17: 87-103.
[35] Werner SD, Taylor CE, Moore JE. Loss estimation due to seismic risks to highway systems. Earthquake Spectra 1997; 13(4): 585-604.
[36] Wollmer RD. Investment in stochastic minimum cost generalized multi-commodity networks with application to coal transport. Networks 1980; 10: 351-62.
[37] Wollmer RD. Investments in stochastic maximum flow networks. Ann. Oper. Res. 1991; 31: 459-468.