1-1-2014

Reliable Design and Operations of Infrastructure Systems

Yu An
University of South Florida, yan2@mail.usf.edu

Reliable Design and Operations of Infrastructure Systems

by

Yu An

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Co-Major Professor: Bo Zeng, Ph.D.
Co-Major Professor: Yu Zhang, Ph.D.
Tapas Das, Ph.D.
Kingsley Reeves, Ph.D.
Shengyong Wang, Ph.D.

Date of Approval:
November 3, 2014

Keywords: Hub-and-spoke Network, Facility Location Problem, Unit Commitment,
Reliability, Robust Optimization

Acknowledgments

First and foremost, I want to express my sincere thanks to my advisors Dr. Bo Zeng and Dr. Yu Zhang who guided me to a research area I really like. The completion of this dissertation could not have been accomplished without their great support, guidance, and patience. I also wish to thank my committee members Dr. Tapas Das, Dr. Kingsley Reeves, and Dr. Shengyong Wang for their valuable suggestions.

Deep gratitude is also due to a fellow in my group: Dr. Long Zhao. I learned a lot from his attitudes towards work and life. Special thanks also to other friends in the IE department: Wei Yuan, Anna Danandeh, Chen Kan, Gang Liu, and Dr. Xi Zhang.

Finally, a honorable mention goes to my friends and supervisors in the Walt Disney Company: Dr. Ludwig Kuznia, Dr. Haining Yu, and Dr. Tianke Feng. The internship opportunity greatly enriches my knowledge and experience.

Table of Contents

List of Tables

List of Figures

Abstract

The reliability issue of the infrastructure systems has become one of the major concerns of the system operators. This dissertation is a collection of four published and working papers that address the specific reliable design and operations problems from three different application settings: transportation/telecommunications network, distribution network, and power plant. In these four projects, key random factors like site disruption and uncertain demand are explicitly considered and proper research tools including stochastic programming, robust optimization, and variants of robust optimization are applied to formulate the problems based on which the important and challenging modelling elements (nonlinear congestion, disruption caused demand variation, etc.) can be introduced and studied. Besides, for each of the optimization models, we also develop advanced solution algorithms that can solve large-scale instances within a short amount of time and devise comprehensive numerical experiments to derive insights. The modelling techniques and solution methods can be easily extended to study reliability issues in other applications.

1 Introduction

Infrastructure is the physical entity and the corresponding organizational structure that integrate interconnected elements to provide and maintain the fundamental function in the normal operations of a society or an enterprise (OnlineCompactOxfordEnglishDictionary, 2014). It includes a wide range of engineered assets. For a society, it could encompass transportation systems, power generation and transmissions systems, communication systems, so on and so forth. While for a company, it may refer to a supply chain of commodities or services. Infrastructure systems exist in almost all aspects of social activities and are critical to ensure the smooth functioning of a society or an organization (Murray and Grubesic, 2007a). The importance of infrastructures is best demonstrated by its magnitude of use. For example, over 645 millions of passengers were moved by the U.S. air transportation system in 2014 (BureauofTransportationStatistics, 2014). The U.S. internet backbone reached data volume of 48 million terabytes with advertisement revenue alone hits $31 billion in 2011 (MinnesotaInternetTrafficStudies (2014) and InteractiveAdvertisingBureau (2014)).

Given the universality and significance of infrastructure systems, the researches in this area are prosperous both on general development policies and specific applications (Ostrom et al. (1993), Martin and Rogers (1995), and Gil et al. (2012)). We mention that when it comes to the quantitative studies of the infrastructures, optimization tools are often utilized to obtain cost-effective designs and operations schedules (Frangopol and Liu (2007), Saranga and Kumar (2006), and Amin (2003)).

Societal functions are highly dependent on the normal operations of infrastructures. However, in the real world, the functioning of an infrastructure is often disrupted by various sorts of events including natural disasters, labor strikes, and equipment failures. Due to the interdependent property of different infrastructure networks (Murray and Grubesic (2007a), Pederson et al. (2006), and Dueñas-Osorio and Vemuru (2009)), a disruption can often negatively influence the service in a extensive geographical area resulting in enormous economic losses. Take air transportation system for example, domestic flight delays were found to cost the U.S. economy $31.2 billion in 2007, including direct costs to airlines and passengers, lost demand, and forgone GDP (AirlinesforAmerica, 2014). Another more recent example is the 2011 southwest blackout that left nearly 2.7 million customers in Arizona and California without power for up to 12 hours. It was caused by a 15 minute power system disturbance and lead to cascading outages that resulted in a estimated cost of $100 million (NorthAmericanElectricReliabilityCorporation (NERC) and Jergler (2014)).

Therefore, the reliability issue of infrastructure systems has become the major concern of homeland security (OfficialWebsiteofDepartmentofHomelandSecurity, 2014) and one of the most important research topics in the related literature, one may refer to Amin (2003), Rinaldi et al. (2001), Moslehi and Kumar (2010), Murray and Grubesic (2007b), and Conrad et al. (2006).

From the major findings in the field, we discovered three issues of the existing researches: (*i*) While lots of studies focused on the diagnosis and assessment of the reliability and vulnerability of infrastructures, the key factor of reliability is not considered in the design stage for many types of infrastructures. For example, in an air transportation network, hub airport disruptions can result in local or even complete malfunction of the whole network, the operators usually rely on strategies like delaying, cancelling, or cancelling to mitigate the negative effects of hub failures (Janic (2005) and Ball et al. (2006)). However the effect of those measures are largely restricted by the initial design which did not consider hub disruptions; (*ii*) For most of the proposed studies on reliable designs of infrastructure systems, some newly developed formulation tools and solution methods are not applied and hence some important insights have not been discovered by the current literature. In the reliable design of distribution networks, for instance, the demand variation of a site due to disruptions exists in the real applications (Ergun et al., 2010) and could drastically influence the network configuration. But most models are based on stochastic programming (Cui et al. (2010), Snyder and Daskin (2005), and Lim et al. (2009)) and the demand change factor is difficult to be included because of the computational challenges it presents; (*iii*) In some area like power systems, the conventional research tools to deal with the random factors like stochastic programming (Takriti et al. (1996), Papavasiliou and Oren (2013) and Constantinescu et al. (2011)) and robust optimization (Bertsimas et al. (2013b), Xiong and Jirutitijaroen (2012), and Zhao and Zeng (2012b)) need to be improved to better address the modelling needs. The stochastic programming enumeration of all possible scenarios may cause prohibitive computational burden while for the robust optimization, although only the worst scenario is considered, may lead to over conservative solutions.

Motivated by the gap in the research, in this dissertation, we present four studies arising from the key infrastructure systems in modern world (air transportation system, distribution network, and power system). We successfully addressed the reliable design or operation issues in these areas by utilizing novel modelling techniques and effective solution algorithms. Note that since overcoming the random risk factors requires the good knowledge and understanding of the infrastructure under study and the settings of these researches are significantly different from each other, we separated them in four chapters with comprehensive literature review and clear problem statement given to each of them respectively. Extensive numerical experiment are also conducted to derive meaningful insights in every chapter. The rest of the dissertation is organized as follows.

Chapter 2 solves a design problem of the hub-and-spoke air transportation network by explicitly considering the single random hub failures in a compact stochastic programming model. Different from the classical model, we introduce backup hub for each affected route in the network under each disruption scenario so that the random hub failure can be addressed in the design stage. With Lagrangian relaxation and branch-and-bound applied jointly, difficult cases are successfully solved to optimality. We also showcases the significant improvement of service quality brought by our model.

Based on the work in Chapter 2, Chapter 3 extends the reliable design problem of the hub-and-spoke network. The formulation is more close to the real world applications by integrating highly challenging factors like multiple hub disruptions and hub congestions in a two-stage robust optimization framework. The newly developed column-and-constraint generation algorithm (Zeng and Zhao, 2013) and linearization techniques are presented which demonstrated good computational performance in solving our cases.

The modelling ability of two-stage robust optimization is further explored in Chapter 4 where it is used to address a reliable facility location problem. Similar to the formulations in Chapter 1 and 2, in our reliable facility location network, the affected customers can be reassigned to other functional facilities when original ones are down due to random disruptions. Challenging factors like demand change due to site disruptions and facility capacity, which are widely neglected in the current literature, are introduced into our model. Through in-depth computational experiments, we investigate the influence of those factors and compare the solutions obtained by RO and SP to gain new insights on distribution network design.

In Chapter 5, we turn our attention to the area of power system operation reliability. A novel modelling method that combines the advantages of SP and RO is proposed and applied in solving a unit commitment problem subjects to uncertain demands. Multiple uncertainty sets are introduced into the classical robust optimization. Based on our computational study, we can see that it requires less information of random demand than stochastic programming and is able to achieve less conservative solution (on-off schedule of generation units) than RO.

Conclusion will be given in Chapter 6. We mention that the modelling techniques and solution methods we developed, together with the insights we obtained from numerical experiments, can be easily extended to solve other infrastructure design and operation problems. The works presented in this dissertation have the potential to fill the gap in the existing literature.

2 The Reliable Hub-and-spoke Design Problem: Models and Algorithms[1]

2.1 Background and Motivation

The hub-and-spoke system has been widely employed in various industrial applications, such as transportation and telecommunications system designs. It is a fully interconnected network with material/information flow between any two nodes being processed at a small number of critical nodes (i.e., *hubs*) so that the operators can benefit from the economies of scale by consolidating flows from and to spoke nodes and increasing the utilization of equipment and staff at those critical nodes. Clearly,a hub-and-spoke network heavily relies on hubs to make the whole system functional, and therefore it is vulnerable to any disruptions and degradations of hubs. Traditional hub-and-spoke network design solves the problem of hub location and allocations of spoke nodes to hubs, assuming network components work properly. In practice, nevertheless, operators have to face various disruptions and apply disruption management techniques to recover the system. Such an issue is most prominently demonstrated in air transportation where severe weather, labor strikes, terrorism threats, and runway incursions disrupt regular operations and make airports partially or completely unavailable (Palpant et al. (2009) and Løve and Sørensen (2001)).

To deal with the vulnerability issue of the hub-and-spoke system, several mitigation strategies have been proposed and implemented, such as delaying, canceling, and rerouting in air transportation (Janic (2005) and Ball et al. (2006)) and network peering in telecommunications systems (O'Kelly et al., 2006). However, most of mitigation strategies are reactive, which are often costly to implement and inefficient, given that the initial network is designed for perfect conditions. For example, it is observed in (Bratu and Barnhart, 2006) that, although the disrupted passengers were only three percent of the total passengers, they suffered 39 percent of the total passenger transportation delays with much lower customer satisfaction. Clearly, the initial network design affects the selections of backup hubs and al-

[1]This chapter is under review for publication in Transportation Research Part B: Methodological

Figure 2.1: Regular and Alternative Routes

ternative routes, which affects the cost of mitigation operations. Therefore, to achieve both economic advantage and system reliability, the network design problem should consider both the hub locations and regular route designs as well as the backup hubs and alternative route designs under disruptions in a holistic modeling framework. Therefore, in this chapter, we propose a reliable hub-and-spoke network design strategy by explicitly considering the hub unavailability, i.e., backup hub and alternative route decisions will be considered in the design stage and related cost will be included in the objective function of the design problem. With this strategy, we aim to develop a new type of optimization models to minimize the operating cost considering both the normal situation, which is disruption free, and disrupted situations where survived hubs serve as backup hubs for rerouting disrupted flights due to unavailable hubs. As illustrated in Figure 2.1, where the solid line denotes a regular route for the flight from Tampa to San Francisco and the dotted line denotes an alternative route using Dallas as a backup hub if the Miami hub is unavailable. This strategy will not only benefit airlines but also other industries who adopted hub-and-spoke distribution paradigm with which they can build and operate their networks with both reliability and economic advantages.

Compared to classical models, the introduction of backup hubs and alternative routes drastically increases the complexity of the network design problem. As the choice of backup hubs and alternative routes depends on the hubs in regular routes, a large number of nonlinear terms are introduced to capture the dependency. As a result, nonlinear mixed integer formulations are constructed. Their structures are further investigated and solution methods developed. To the best of our knowledge, our study is the first analytical work on the reliable hub-and-spoke design with consideration of backup hubs and alternative routes. The developed algorithm is easy to implement and can solve practical instances in a reasonable amount of time. Numerical study demonstrates that our reliable models can serve more passengers under the disruption situations and sensitivity analysis shows that the resulting designs are robust to hub unavailability.

The proposed reliable hub-and-spoke network design also yields a set of useful tools for practitioners, such as airlines, to re-structure their networks or to identify strategic partners to hedge against various disruptions and achieve better performance.

The rest of the chapter is organized as follows. In Section 2.2, literature review on hub-and-spoke design is presented as well as recent research on reliable facility location models. In Section 2.3, the reliable single allocation hub-and-spoke model is formulated and the solution methods are elaborated. In Section 2.4, the study is extended to the reliable multiple allocation model. Section 2.5 demonstrates computational performance of the developed algorithms using the CAB data set from airline operations as the case study and provides comparisons between our reliable hub-and-spoke design models and classical models. In addition, system design and performance with proposed model are analyzed and discussed, including sensitivity analysis and the demonstration of applying proposed model to a recent airlines merger. Section 2.6 concludes this chapter with some discussions on future research directions.

2.2 Literature Review

The hub-and-spoke design problem is conventionally called *hub location problem* (HLP), which is concerned with locating hub facilities and allocating spoke nodes to hubs. There are generally two basic structures: single allocation (SA) and multiple allocation (MA). In SA hub-and-spoke model, all outbound/inbound flows of any node must travel directly from/to a specific hub. In MA model, flows of a given node can go directly from/to different hubs. When the number of hubs, denoted by p, is given, the problem is called the p-hub median problem (HMP). In the remainder of this chapter, we use SA-HMP or MA-HMP to denote the corresponding design problem. O'Kelly (1987) proposes the first mathematical formulation for HMP and presented the first quantitative analysis on this type of network structure using the Civil Aeronautics Board (CAB) data set. Since then, as hub-and-spoke structures are of significant theoretical and practical values, a large number of studies have been conducted on developing models with more practical features and on designing efficient algorithms.

We first briefly describe a few important results on formulation and algorithm design. Ernst and Krishnamoorthy (1996) and Ernst and Krishnamoorthy (1998a) formulate SA-HMP and MA-HMP, respectively, based on the idea of "multicommodity flow". Skorin-Kapov et al. (1996) propose mixed integer formulations for both SA-HMP and MA-HMP that yield tight linear relaxations. As for the customized algorithm development, Branch-and-Bound process and Lagrangian relaxation have been widely used to obtain exact solutions (Ernst and Krishnamoorthy (1998b) and Pirkul and Schilling (1998)). Different from the p-hub median problem, the *hub location problem with fixed costs* treats the number of hubs as a decision variable and seeks to minimize the transportation cost and the construction cost where a fixed construction cost is associated with a decision of hub location. O'Kelly (1992) and Campell (1994) study a few formulations of HLP with fixed costs. There are also extensive literature in search of effective solution algorithms for these problems, see Cunha and Silva (2007), Chen (2007), Cánovas et al. (2007), and Contreras et al. (2011a) for examples. One may refer to Alumur and Kara (2008) and Campbell and O'Kelly (2012) for a comprehensive review of modeling techniques and solution methods of HLP. In the remainder of this chapter, unless we explicitly mention, the hub-and-spoke network design problem indicates p-hub median problem.

Recent studies focused on extending classical SA and MA models by incorporating practical factors, such as hub congestion (Elhedhli and Wu, 2010; Grove and O'Kelly, 1986), hub capacity (Contreras et al., 2012), nonlinear economies of scale (de Camargo et al., 2009a), and dynamic/stochastic nature of demand and cost (Contreras et al. (2011b) and Contreras et al. (2011c)).

Nearly all studies on HLP assumed that the chosen hubs would always operate functionally as planned. Nevertheless, in practice, hubs could fail due to different reasons. As the typical cases in air transportation industry, adverse weather often significantly deteriorates the availability of a hub airport and results in huge disruption costs. Similar situations have been observed in facility-and-client based supply chain and logistics systems, where facilities, same as hubs, play the central role and their locations are derived using facility location models. Note that, different from hub-and-spoke design, there is no inter-facility transportation in those systems. To deal with facility disruptions, a facility location model with backup strategy, referred to as the *reliable facility location model*, was introduced by Snyder and Daskin (2005). Since then, this type of research has received significant attention (Cui et al. (2010), Li and Ouyang (2010); Lim et al. (2009), Li (2011), and An et al. (2014)). It is commonly observed that the resulting nonlinear optimization formulations are computationally challenging. Hence, customized algorithms are needed for solving real-size problems, among which Lagrangian relaxation methods and their Branch-and-Bound extensions are the major solution strategies (Snyder and Daskin (2005), Li and Ouyang (2010), Cui et al. (2010), Lim et al. (2009), and Li (2011)).

In contrast to reliable facility location problems that have attracted the attention of many researchers, up to now, only several recent studies considered reliable hub-and-spoke networks. In Kim and O'Kelly (2009), given that each arc or hub has a reliability (same as availability in this chapter), they build SA and MA models to derive an optimal network structure that maximizes the expected network flow, without considering backup hubs and alternative routes. Kim (2008) proposes a p-hub protection model based on single allocation structure with primary and secondary routes presented. The authors then utilized a heuristic method (tabu search) to solve the real instances with up to 20 nodes. In Zeng et al. (2010), reliable SA and MA models with consideration of hub unavailability and alternative routes have been developed and a heuristic algorithm has been implemented. The authors observe that, different from the reliable facility location model, reliable hub-and-spoke models are much more complicated. Indeed, this type of problems have not been analytically investigated with advanced solution methods to deal with real-size problems. Given that many infrastructure systems, e.g., air transportation and telecommunications systems, adopt hub-and-spoke structures where system reliability is of extremely high importance, we perform an analytical study on reliable hub-and-spoke models and develop efficient algorithms to solve practical instances. A framework of model evaluation under correlated hub disruptions will also be proposed. The study fills in the gap in existing literature and advances the research in reliable hub-and-spoke network design.

2.3 Reliable Single Allocation Hub-and-spoke Model

We are aware that multiple allocation hub-and-spoke model is widely applied instead of single allocation model in air transportation industry. Although we apply aviation case studies later in our study, for the sake of completeness, we study the formulation and solution algorithms for both reliable single and multiple allocation models in Section 3 and 4. There are two major assumptions in our study. First, we focus on solving the problem with single disruptions. We are aware that under some circumstances, multiple disruptions and even massive disruptions could occur. For example, the volcano ash crisis in Europe in 2010 and 2011 caused the closure of many major airports in that region and Sandy hurricane in 2013 made all three commercial airports in New York area malfunctioned for days. Nevertheless, in airline industry, one carrier often strategically locates its hubs far from each other in its hub-and-spoke network. On one hand, it helps the carrier to fully take advantage of the discounted inter-hub transportation. On the other hand, it can prevent the carrier from being affected by multiple simultaneous hub failures due to the same cause. As an example, during the Grisvötn volcano eruption in Iceland in 2011, two neighboring airports, i.e., Airport Hamburg (HAM) and Airport Bremen (BRE) in Germany (BBC, 2011a,b), had to be closed. Although HAM and BRE serve as hubs for 19 legacy airlines and low cost carriers, the majority of airlines (17 out of 19) operate just one of the two airports as their hubs (Flylowcostairlines.com (2012) and Mygermancity.com (2012)). Under single disruption and normal condition, our model can provide an optimal routing strategy while when multiple disruptions occur, airline companies can take "principle of proximity" and assign disrupted routes to closest functional hubs. We demonstrate, in later section, that the optimized solutions from single disruption model provides better network set-up under multiple disruption scenarios compared to the outcomes from classical model. We will continue tackle the multiple disruption problem in out future study. Second, we assume that for routes going through two hubs, the alternative route is still required to go through the unaffected hub airport. The main reason for doing so is to alleviate the possible impacts of rerouting at tactical operational level to the maintenance scheduling. Certain maintenance requirements have to be followed in airline industry. Type A maintenance check is required every 500-800 flight hours and needs 20-50 man-hours. It can be done overnight at an airport gate or hangar. For other types of checks (B, C, and D), the man-hours needed are much longer and many of them have to be performed at hubs, which usually act the role of maintenance bases. Furthermore, MA structure is adopted in designing alternative routes, regardless of the SA or MA structure used for determining regular routes.

2.3.1 Reliable SA Model: Definition and Formulation

In a single allocation problem, every node is assigned to a single hub and all the inbound and outbound flows of this node are routed through that hub. Let $\mathbf{N} = \{0, 1, ..., |\mathbf{N} - 1|\}$ be the set of nodes and $\mathbf{H} \subseteq \mathbf{N}$ be the set of candidate hub locations for this reliable hub-and-spoke design model, i.e., *R-SAHMP*. We assume that $\mathbf{H} = \mathbf{N}$ throughout this chapter. Then, a node $i \in N$, is with $q_i \in [0, 1]$ to represent its disruption probability. We denote a flow by its source (i) and destination (j) nodes, i.e., an $i - j$ flow. A route of $i - j$ flow can be represented by a 4-tuple (i, k, m, j), where k and m represent the first and the second hubs on the route. Unit transportation cost between a pair of nodes i and j is c_{ij} and the traffic volume between them is w_{ij}. A discount factor of economies of scale, $0 < \alpha < 1$, is applied to the inter-hub links. So, for $i - j$ flow taking the route (i, k, m, j), the cost of transporting one unit flow is $F_{ikmj} = c_{ik} + \alpha c_{km} + c_{mj}$. Decision variables in R-SAHMP include hub location and allocation variable $\mathbf{Y}$, route variable $\mathbf{X}$ and backup hub variables $\mathbf{U}$ and $\mathbf{V}$.

$$Y_{ik} = \begin{cases} 1, & i \text{ is assigned to hub } k, \\ 0, & \text{otherwise;} \end{cases}$$

$$X_{ikmj} = \begin{cases} 1, & i-j \text{ flow is routed through hubs } k \text{ and } m, \\ 0, & \text{otherwise;} \end{cases}$$

$$U_{ijn} = \begin{cases} 1, & \text{hub } n \text{ is the backup hub for the first hub in the route of} \\ & i-j \text{ flow,} \\ 0, & \text{otherwise;} \end{cases}$$

$$V_{ijn} = \begin{cases} 1, & \text{hub } n \text{ is the backup hub for the second hub in the route} \\ & \text{of } i-j \text{ flow,} \\ 0, & \text{otherwise.} \end{cases}$$

In our formulation, w_{ij} is used to denote the traffic volume between nodes i and j, which is the *transportable* flow with both i and j functional. In other words, if one of these two nodes fails, there will be no traffic between them. We recognize that for air transportation, in practice w_{ij} might not be constant given that passengers might migrate from a disrupted airport to another airport nearby to complete their critical travel plans. Nevertheless, modeling such migration requires additional information that varies from airport to airport and causes our models intractable in the designing stage. More important, to keep models general for different hub-and-spoke networks where the migration phenomenon may not occur, we assume that w_{ij} is a constant. We also adopt a convention in many literature (e.g., O'Kelly et al. (1996), Pirkul and Schilling (1998), and Sohn and Park (1998)) that $w_{ij} = w_{ji}$. Given this symmetric structure, in this study we design the network only considering flow from i to j with $j > i$. Note that this assumption also indicates that the first backup hub for route (i,k,m,j) will be the second one for route (j,m,k,i). Next, we present R-SAHMP that generalizes and extends the classical SA hub-and-spoke model developed by Skorin-Kapov et al. (1996).

$$\min \sum_{i\in\mathbf{N}} \sum_{k\in\mathbf{H}/\{i\}} \sum_{m\in\mathbf{H}} \sum_{\substack{j\in\mathbf{N}/\{m\} \\ j>i}} F_{ikmj} w_{ij} (1 - q_k - q_m^k) X_{ikmj}$$

$$+ \sum_{i\in\mathbf{N}} \sum_{\substack{j\in\mathbf{N} \\ j>i}} \Bigg(\sum_{m\in\mathbf{H}/\{j\}} F_{iimj} w_{ij} (1 - q_m^i) X_{iimj}$$

$$+\sum_{k\in\mathbf{H}/\{i\}} F_{ikjj} w_{ij} (1-q_k^j) X_{ikjj} + F_{iijj} w_{ij} X_{iijj} \Bigg)$$

$$+\rho\Bigg(\sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{H}}\sum_{m\in\mathbf{H}/\{k\}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{n\in\mathbf{H}} (F_{inmj} w_{ij} q_k X_{ikmj} U_{ijn} + F_{iknj} w_{ij} q_m X_{ikmj} V_{ijn})$$

$$+\sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{H}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{n\in\mathbf{H}} F_{innj} w_{ij} q_k X_{ikkj} U_{ijn}\Bigg) \tag{2.1}$$

s.t.

$$\sum_{m\in\mathbf{H}} X_{ikmj} = Y_{ik} \qquad \forall i, j>i, k \tag{2.2}$$

$$\sum_{k\in\mathbf{H}} X_{ikmj} = Y_{jm} \qquad \forall i, j>i, m \tag{2.3}$$

$$\sum_{k\in\mathbf{H}} Y_{ik} = 1 \qquad \forall i \tag{2.4}$$

$$\sum_{k\in\mathbf{H}} Y_{kk} = p \tag{2.5}$$

$$U_{ijk} + \sum_{m\in\mathbf{H}} X_{ikmj} \le Y_{kk} \qquad \forall i, j>i, k \tag{2.6}$$

$$\sum_{k\in\mathbf{H}} U_{ijk} = 1 - \sum_{m\in\mathbf{H}} X_{iimj} - \sum_{m\in\mathbf{H}} X_{ijmj} \qquad \forall i, j>i \tag{2.7}$$

$$V_{ijm} + \sum_{k\in\mathbf{H}} X_{ikmj} \le Y_{mm} \qquad \forall i, j>i, m \tag{2.8}$$

$$\sum_{m\in\mathbf{H}} V_{ijm} = 1 - \sum_{k\in\mathbf{H}} X_{ikjj} - \sum_{k\in\mathbf{H}} X_{ikij} \qquad \forall i, j>i \tag{2.9}$$

$$X_{ikmj} \in \{0,1\}\ \forall i, j>i, k, m;\ Y_{ik} \in \{0,1\}\ \forall i, k;\ U_{ijk}, V_{ijk} \in \{0,1\}\ \forall i, j>i, k. \tag{2.10}$$

In the R-SAHMP, the objective function is the expected transportation cost considering both the regular and the disrupted situations. Specifically, the first term represents the regular transportation cost for traffic flows with both source and destination at spoke nodes. The probability of regular transportation, given the assumption that in a hub-and-spoke network at most one hub will fail under disruption situation, is computed as $1 - q_k - q_m^k$, where q_m^k is 0 if $k = m$ and q_m otherwise. By introducing q_m^k in this way, one formula can capture both cases, i.e., the route is operable with the probability $1 - q_k - q_m$ if two hubs are different and reduces to $1 - q_k$ when hubs k and m are identical. The second term represents the regular transportation cost for traffic flows with source or destination at a hub node. The third term in the objective function is the cost of disruption mitigation by diverting flows to alternative routes, which is penalized by a coefficient ρ ($\rho > 1$) to represent the impact of disruption to the overall cost in this transportation network (Welman et al. (2010)).

Constraints (2.2)-(2.5) are classical constraints for the SA p-median problem (Skorin-Kapov et al., 1996). Constraints (2.6) and (2.8) ensure that regular hubs and backup hubs can only be the nodes chosen to be hubs and the regular hubs and the backup hubs must be different. Constraints (2.7) and (2.9) are used to ensure that backup routes are selected for all disrupted flows, except the cases where either the source or the destination node of a flow is a hub.

Existing studies have approved that traditional SA hub-and-spoke model is NP-hard. The proposed R-SAHMP problem can be reduced to the traditional one if all nodes are always reliable, so it is also an NP-hard problem. Not only the entire problem is NP-hard, even when all hubs are fixed, the allocation and routing problem in R-SAHMP is still NP-hard (Sohn and Park (2000)). Nevertheless, once all hubs and spoke node allocations are fixed, the design for regular and alternative routes is polynomially solvable. Note that R-SAHMP is an integer quadratic program as its objective function has multiple terms that involve products of two binary variables. Thus, we used the standard linearization method to convert it into a linear model. We also adopted a recent linearization strategy (Chaovalitwongse et al. (2004), Sherali and Smith (2007), and He et al. (2012)) to derive a more compact linear reformulation of R-SAHMP. The linearized formulations of the above two methods and computational results are presented in the Appendix A and Section 2.5, respectively.

2.3.2 Lagrangian Relaxation and Branch-and-Bound

Existing professional mixed integer programming solvers can be applied to seek solutions of the linearized formulas of R-SAHMP. However, due to the large number of variables and constraints in the model, it takes excessive running times (see computational results presented in Section 2.5). Hence, a Lagrangian relaxation (LR) algorithm is developed after exploring the structure of R-SAHMP. Compared with other algorithms or commercial solvers, the Lagrangian relaxation algorithm often yields high-quality approximate or optimal solutions with much less computational time (Contreras et al., 2011b; Pirkul and Schilling, 1998). Actually, the proposed Lagrangian relaxation technique is able to directly deal with the nonlinear R-SAHMP without linearizing the formulation. Furthermore, variable fixing and Branch-and-Bound methods are implemented to identify an optimal solution if the Lagrangian relaxation algorithm fails to obtain it.

2.3.2.1 Lagrangian Lower Bound

For *R-SAHMP*, we dualize the constraints (2.2), (2.3), (2.4), (2.6), and (2.8) with $\delta_{ijk,1}$, $\delta_{ijm,2}$, β_i, $\gamma_{ijk,1} \geq 0$, and $\gamma_{ijm,2} \geq 0$ as their Lagrangian multipliers, respectively. As a result, we obtain the following relaxation:

$$
\begin{aligned}
& f(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \\
& \min \sum_{i \in \mathbf{N}} \sum_{k \in \mathbf{H}} \bar{C}_{ik} Y_{ik} - \sum_{i \in \mathbf{N}} \beta_i \\
& + \sum_{i \in \mathbf{N}} \sum_{k \in \mathbf{H}/\{i\}} \sum_{m \in \mathbf{H}} \sum_{\substack{j \in \mathbf{N}/\{m\} \\ j>i}} (F_{ikmj} w_{ij} (1 - q_k - q_m^k) + \delta_{ijk,1} + \gamma_{ijk,1} + \delta_{ijm,2} + \gamma_{ijm,2}) X_{ikmj} \\
& + \sum_{i \in \mathbf{N}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \sum_{m \in \mathbf{H}/\{j\}} (F_{iimj} w_{ij} (1 - q_m^i) + \delta_{iji,1} + \gamma_{iji,1} + \delta_{ijm,2} + \gamma_{ijm,2}) X_{iimj} \\
& + \sum_{i \in \mathbf{N}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \sum_{k \in \mathbf{H}/\{i\}} (F_{ikjj} w_{ij} (1 - q_k^j) + \delta_{ijk,1} + \gamma_{ijk,1} + \delta_{ijj,2} + \gamma_{ijj,2}) X_{ikjj} \\
& + \sum_{i \in \mathbf{N}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} (F_{iijj} w_{ij} + \delta_{iji,1} + \gamma_{iji,1} + \delta_{ijj,2} + \gamma_{ijj,2}) X_{iijj} \\
& + \sum_{i \in \mathbf{N}} \sum_{k \in \mathbf{H}} \sum_{m \in \mathbf{H}/\{k\}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \sum_{n \in \mathbf{H}} \rho F_{inmj} w_{ij} q_k X_{ikmj} U_{ijn} + \sum_{i \in \mathbf{N}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \sum_{k \in \mathbf{H}} \gamma_{ijk,1} U_{ijk} \\
& + \sum_{i \in \mathbf{N}} \sum_{k \in \mathbf{H}} \sum_{m \in \mathbf{H}/\{k\}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \sum_{n \in \mathbf{H}} \rho F_{iknj} w_{ij} q_m X_{ikmj} V_{ijn} + \sum_{i \in \mathbf{N}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \sum_{m \in \mathbf{H}} \gamma_{ijm,2} V_{ijm} \\
& + \sum_{i \in \mathbf{N}} \sum_{k \in \mathbf{H}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \sum_{n \in \mathbf{H}} \rho F_{innj} w_{ij} q_k X_{ikkj} U_{ijn}
\end{aligned} \tag{2.11}
$$

s.t.

$$\text{Contraints } (2.5), (2.7), (2.9), (2.10)$$

$$Y_{ik} \le Y_{kk} \qquad \forall i, k \tag{2.12}$$

where

$$\bar{C}_{ik} = \begin{cases} \beta_i - \sum\limits_{\substack{j \in \mathbf{N} \\ j > i}} \delta_{ijk,1} - \sum\limits_{\substack{j \in \mathbf{N} \\ j > i}} \delta_{jik,2}, & \text{if } i \neq k; \\ \beta_k - \sum\limits_{\substack{i \in \mathbf{N} \\ i > k}} \delta_{kik,1} - \sum\limits_{\substack{i \in \mathbf{N} \\ i < k}} \delta_{ikk,2} - \sum\limits_{i \in \mathbf{N}} \sum\limits_{\substack{j \in \mathbf{N} \\ j > i}} (\gamma_{ijk,1} + \gamma_{ijk,2}), & \text{otherwise.} \end{cases}$$

Note that (2.12) is implied in R-SAHMP and can be derived from (2.2) and (2.6).

Since $\mathbf{X}$ and $\mathbf{Y}$ variables are not linked any more in the relaxed formulation, the problem can be decomposed into two independent subproblems (SAsub-1 and SAsub-2). An optimal solution to the relaxed problem can be obtained by solving the two subproblems and combining their optimal solutions. The form of SAsub-1 is given below.

$$\min \{ \sum_{i \in \mathbf{N}} \sum_{k \in \mathbf{H}} \bar{C}_{ik} Y_{ik} - \sum_{i \in \mathbf{N}} \beta_i : \sum_{k \in \mathbf{H}} Y_{kk} = p, \quad Y_{ik} \le Y_{kk} \ \forall i, k, \ Y_{ik} \in \{0, 1\} \ \forall i, k. \}$$

SAsub-1 contains only the allocation variable $\mathbf{Y}$ and is solved by a procedure as follows. Note that it can be completed within $O(|\mathbf{N}|^2)$.

(i) For i, k $(i \neq k)$, set $Y_{ik} = 1$ if $\bar{C}_{ik} < 0$ and $Y_{ik} = 0$ otherwise. Compute $S_k = \sum_{i \in \mathbf{N}} \bar{C}_{ik} Y_{ik}$, for each k.

(ii) Sort S_k's in ascending order, choose p of the nodes with smaller S_k, and set the corresponding $Y_{kk} = 1$ and set the remaining Y_{kk}'s to 0. Calculate the optimal value of SAsub-1 by $\sum_{k \in \mathbf{H}} S_k Y_{kk} - \sum_{i \in \mathbf{N}} \beta_i$.

(iii) For i, k $(i \neq k)$, set Y_{ik} to 0 if $Y_{kk} = 0$.

The SAsub-2 is provided below.

$$
\begin{aligned}
\min \sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{H}/\{i\}}\sum_{m\in\mathbf{H}}\sum_{\substack{j\in\mathbf{N}/\{m\}\\ j>i}} &(F_{ikmj}w_{ij}(1-q_k-q_m^k)+\delta_{ijk,1}+\gamma_{ijk,1}+\delta_{ijm,2}+\gamma_{ijm,2})X_{ikmj}\\
&+\sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{m\in\mathbf{H}/\{j\}}(F_{iimj}w_{ij}(1-q_m^i)+\delta_{iji,1}+\gamma_{iji,1}+\delta_{ijm,2}+\gamma_{ijm,2})X_{iimj}\\
&+\sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{k\in\mathbf{H}/\{i\}}(F_{ikjj}w_{ij}(1-q_k^j)+\delta_{ijk,1}+\gamma_{ijk,1}+\delta_{ijj,2}+\gamma_{ijj,2})X_{ikjj}\\
&+\sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}(F_{iijj}w_{ij}+\delta_{iji,1}+\gamma_{iji,1}+\delta_{ijj,2}+\gamma_{ijj,2})X_{iijj}\\
&+\sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{H}}\sum_{m\in\mathbf{H}/\{k\}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{n\in\mathbf{H}}\rho F_{inmj}w_{ij}q_kX_{ikmj}U_{ijn}+\sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{k\in\mathbf{H}}\gamma_{ijk,1}U_{ijk}\\
&+\sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{H}}\sum_{m\in\mathbf{H}/\{k\}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{n\in\mathbf{H}}\rho F_{iknj}w_{ij}q_mX_{ikmj}V_{ijn}+\sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{m\in\mathbf{H}}\gamma_{ijm,2}V_{ijm}\\
&+\sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{H}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{n\in\mathbf{H}}\rho F_{innj}w_{ij}q_kX_{ikkj}U_{ijn}
\end{aligned}
\tag{2.13}
$$

s.t.

$$\text{Constraints } (2.7), (2.9)$$

$$\sum_{k\in\mathbf{H}}\sum_{m\in\mathbf{H}}X_{ikmj}=1 \qquad \forall i, j>i \tag{2.14}$$

$$U_{ijk}+\sum_{m\in\mathbf{H}}X_{ikmj}\le 1 \qquad \forall i, j>i, k \tag{2.15}$$

$$V_{ijm}+\sum_{k\in\mathbf{H}}X_{ikmj}\le 1 \qquad \forall i, j>i, m \tag{2.16}$$

$$X_{ikmj}\in\{0,1\}\ \forall i, j>i, k, m; U_{ijk}, V_{ijk}\in\{0,1\}\ \forall i, j>i, k, \tag{2.17}$$

SAsub-2 includes the regular route variable $\mathbf{X}$ and the backup hub variables $\mathbf{U}$ and $\mathbf{V}$. Constraints (2.14), (2.15) and (2.16) are redundant in the original model. Nevertheless, including them in SAsub-2 yields solutions that are more likely to be feasible to the original problem and therefore strengthens the lower bound obtained from Lagrangian relaxation. Note that the constraints in (2.14) require that each $i-j$ flow has to go through one or two nodes to reach its destination; constraints in (2.15) and (2.16) ensure that for each $i-j$ flow, the first/second node in its backup route must be different from the first/second node of its regular route. Note that, if a regular route is a single-hub route, so is its alternative route. Furthermore, in SAsub-2, an optimal solution for one $i-j$ flow, i.e., a set of X^*_{ikmj}, U^*_{ijn}, and V^*_{ijn}, is independent of those of others. So, it is sufficient to consider each individual $i-j$ flow with the corresponding cost function from (2.13) and constraints from SAsub-2.

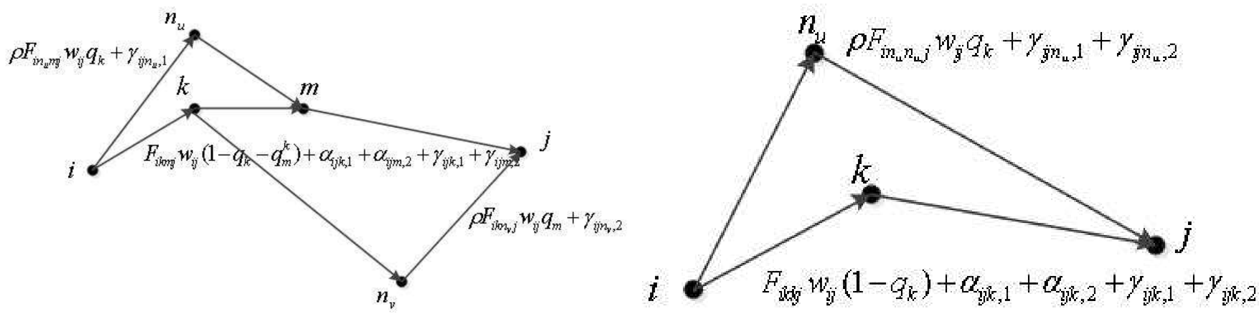
Although the cost function is nonlinear, every feasible solution has a clear combinatorial structure. As shown in Figure 2.2(a), if $i-j$ flow takes (i,k,m,j) as its regular route satisfying $i \neq k$ and $j \neq m$, a cost of $F_{ikmj}w_{ij}(1-q_k-q^k_m)+\delta_{ijk,1}+\gamma_{ijk,1}+\delta_{ijm,2}+\gamma_{ijm,2}$ will be incurred; if this flow takes n_u (n_v, respectively) as the backup hub for k (m, respectively), a cost of $\rho F_{inmj}w_{ij}q_k+\gamma_{ijn_u,1}$ ($\rho F_{iknj}w_{ij}V_{ijn}q_m+\gamma_{ijn_v,2}$, respectively) will be incurred additionally. A similar situation on transportation cost can be observed in Figure 2.2(b) when the $i-j$ flow selects a single-hub regular route. Such observations motivate us to develop the following enumeration procedure to identify an optimal solution to the $i-j$ flow.

(i) For one pair of (k,m), i.e., a given regular route, obtain its best alternative route (or best backup hubs) by computing all possible backup hubs that are different from k and m and selecting the alternative routes (or a single alternative route if $k=m$) with the least transportation cost. Compute the total transportation costs from both the regular route and the alternative routes.

(ii) Repeat Step 1 for all (k,m) pairs and identify the pair that provides the least total transportation cost. Denote that pair by (k^*,m^*) and its corresponding best backup hubs by n^*_u and n^*_v.

(iii) Obtain an optimal solution to $i-j$ flow by setting $X_{ikmj}=1$ if $k=k^*$ and $m=m^*$, and otherwise to zero; setting $U_{ijn}=1$ if $n=n^*_u$, and otherwise to zero; setting $V_{ijn}=1$ if $n=n^*_v$, and otherwise to zero.

The computational complexity of this procedure for one $i-j$ flow is $O(|\mathbf{N}|^4)$ and therefore SAsub-2 can be solved within $O(|\mathbf{N}|^6)$.



(a) A Two-hub Route ($i \neq k$, $j \neq m$) and Its Alternative Routes

(b) A Single-hub Route ($k \neq i, j$) and Its Alternative Route

Figure 2.2: Transportation Cost of Solutions to SAsub-2

2.3.2.2 Upper Bound and Multiplier Updating

To obtain a feasible solution as well as an upper bound, we apply a procedure similar to the one in Pirkul and Schilling (1998) that exploits the solution of SAsub-1. Specifically, for each node $i \in \mathbf{N}$, its allocation will be retained if (2.4) is not violated. For the node with allocation infeasible to (2.4), given that hubs are already fixed after solving SAsub-1, we select the lowest cost allocation. After determining $\mathbf{Y}$ variables, the regular route for each $i - j$ flow is determined and its alternative routes can also be obtained by evaluating hubs not in the regular route and selecting the best ones.

We apply the classical subgradient algorithm described in Fisher (2004) to iteratively update the Lagrangian multipliers and to search for the best lower bound. Parameters such as step-size multiplier and maximum number of iterations are usually set up while applying the algorithm. The values of such parameters for the experimental study are described in Section 2.5.1.

2.3.2.3 Variable Fixing

Variable fixing is an approach that uses both primal information from a feasible solution and dual information from Lagrangian multipliers to fix some variables in Lagrangian solution procedure. It has been proven to be effective in reducing search space and computation time(Snyder and Daskin (2005) and Contreras et al. (2011b)).

Assume that we have the current best upper bound UB, $(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ are the current Lagrangian multipliers, and $(\mathbf{Y}^*, \mathbf{X}^*)$ is the corresponding optimal solution to the Lagrangian relaxed problem. Let $f(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 | C)$ be the optimal objective function value for $(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ under some condition C. Then, we have the following results.

Proposition 1 *When UB is strictly greater than LB,*
(i) if $Y^*_{kk} = 1$ and $f(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 | Y_{kk} = 0) > UB$ for some k, $Y_{kk} = 1$ in any optimal solution; (ii) if $Y^*_{kk} = 0$ and $f(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 | Y_{kk} = 1) > UB$ for some k, we have $Y_{kk} = 0$ in any optimal solution.

Proof. We provide the proof for (i). Results in (ii) can be proven using similar arguments.

Note that $f(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 | Y_{kk} = 0)$ is a lower bound to R-SAHMP with a spoke node located in k for the given Lagrangian multipliers $(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$. So, if

$$f(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 | Y_{kk} = 0) > UB,$$

any solution to R-SAHMP with a spoke node in k will generate more cost than the current best feasible solution. Therefore, we have $Y_{kk} = 1$ in any optimal solution to R-SAHMP. ■

We mention that although more variable fixing rules can be developed, such as rules for the case of $Y^*_{ik} = 0$, they will either be time-consuming to implement or have less impact on the Lagrangian relaxation. So, we only perform variable fixing procedure on Y^*_{kk} variable for each k with the best multipliers ever found once the Lagrangian procedure is terminated.

2.3.2.4 Branch-and-Bound Strategies

If the subgradient method reaches the maximum number of iterations and the gap is still larger than the tolerance, the Lagrangian relaxation algorithm discussed in the previous section will be embedded in a Branch-and-Bound framework to further reduce the gap.

The Branch-and-Bound technique with Lagrangian relaxation has been implemented in reliable facility location models (Cui et al. (2010), Snyder and Daskin (2005), and Li and Ouyang (2011)). The results imply that branching on facility location variables is sufficient for determining an optimal network structure (Cui et al. (2010), Snyder and Daskin (2005), and Li and Ouyang (2011)). However, this is not the case for R-SAHMP. Note that for a classical single allocation hub-and-spoke model, given fixed hubs, the remaining problem on spoke node allocation has been proven to be NP-hard. Thus, a more sophisticated two-stage Branch-and-Bound framework is developed and implemented in a width-first manner.

The first stage Branch-and-Bound is similar to that used for reliable facility location models in Snyder and Daskin (2005), where branching is made for Y_{kk} (hub location) variables. At each Branch-and-Bound node, the hub location variable $Y_{k^*k^*}$ selected for branching is the unfixed open hub with the greatest assigned flow (without considering alternative routes), i.e.,

$$k^* = \arg\max_{k\in\mathbf{N}}\{\sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{m\in\mathbf{H}} w_{ij}X_{ikmj} + \sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{m\in\mathbf{H}/\{k\}} w_{ij}X_{imkj}\}.$$

$Y_{k^*k^*}$ is forced to be 0 and then 1. The first stage Branch-and-Bound process will be terminated either with an optimal (including ϵ-optimal) solution or with p hubs forced to open (or equivalently, $|\mathbf{N}| - p$ hubs forced to close).

In the latter case, the second stage Branch-and-Bound method is applied to close the gap. Branching is made for Y_{ik} (allocation) variables for spoke node i. First, the *level of violation*, v^i, for spoke node i is computed. Given the current solution to the relaxed problem, the total number of violations to constraints in (2.2), (2.3), and (2.4) for each i are then calculated. The spoke node with the largest violation level v^i, say i^*, is selected for branching. Then, we partition the hub set $\mathbf{H}$ (note that hub locations are already determined) into two sets $\mathbf{H}_1$ and $\mathbf{H}_2$ and create two nodes. Correspondingly, constraint $\sum_{k\in\mathbf{H}_1} Y_{i^*k} = 1$ is added to the left-hand node and constraint $\sum_{k\in\mathbf{H}_2} Y_{i^*k} = 1$ to the right-hand node. Once hub and spoke node allocation decisions are made, the remaining problem, including regular route and alternative route decisions, is polynomially solvable, which implies that no further branching is necessary.

During the whole Branch-and-Bound procedure, the set of Lagrangian multipliers that yields the smallest gap at a given node is passed to its child nodes as initial multipliers.

2.4 Reliable Multiple Allocation Hub-and-spoke Model

In this section, we consider the reliable MA-HLP model (R-MAHMP). Compared with the single allocation model, the multiple allocation model does not restrict flows from one source (or to one destination) to route through the same hub. As a result, we do not need to introduce spoke-hub allocation variables but simply introduce binary variables to define hubs.

2.4.1 Reliable MA Model: Definition and Formulation

The formulation for R-MAHMP is given below, most constraints reflect the requirements similar to those in R-SAHMP.

$$
\begin{aligned}
\min \; & \sum_{i\in\mathbf{N}} \sum_{k\in\mathbf{H}/\{i\}} \sum_{m\in\mathbf{H}} \sum_{\substack{j\in\mathbf{N}/\{m\}\\ j>i}} F_{ikmj} w_{ij} (1 - q_k - q_m^k) X_{ikmj} \\
& + \sum_{i\in\mathbf{N}} \sum_{\substack{j\in\mathbf{N}\\ j>i}} \Bigg(\sum_{m\in\mathbf{H}/\{j\}} F_{iimj} w_{ij} (1 - q_m^i) X_{iimj} \\
& + \sum_{k\in\mathbf{H}/\{i\}} F_{ikjj} w_{ij} (1 - q_k^j) X_{ikjj} + F_{iijj} w_{ij} X_{iijj} \Bigg) \\
& + \rho \Bigg(\sum_{i\in\mathbf{N}} \sum_{k\in\mathbf{H}} \sum_{m\in\mathbf{H}/\{k\}} \sum_{\substack{j\in\mathbf{N}\\ j>i}} \sum_{n\in\mathbf{H}} (F_{inmj} w_{ij} q_k X_{ikmj} U_{ijn} + F_{iknj} w_{ij} q_m X_{ikmj} V_{ijn}) \\
& + \sum_{i\in\mathbf{N}} \sum_{k\in\mathbf{H}} \sum_{\substack{j\in\mathbf{N}\\ j>i}} \sum_{n\in\mathbf{H}} F_{innj} w_{ij} q_k X_{ikkj} U_{ijn} \Bigg) \qquad (2.18)
\end{aligned}
$$

s.t.

$$\sum_{k\in\mathbf{H}} X_{ikjj} = Y_j \qquad \forall i, j > i \qquad (2.19)$$

$$\sum_{m\in\mathbf{H}} X_{iimj} = Y_i \qquad \forall i, j > i \qquad (2.20)$$

$$\sum_{i\in\mathbf{N}} Y_i = p \qquad (2.21)$$

$$\sum_{k\in\mathbf{H}} \sum_{m\in\mathbf{H}} X_{ikmj} = 1 \qquad \forall i, j > i \qquad (2.22)$$

$$U_{ijk} + \sum_{m\in\mathbf{H}} X_{ikmj} \le Y_k \qquad \forall i, j > i, k \qquad (2.23)$$

$$\sum_{k\in\mathbf{H}} U_{ijk} = 1 - \sum_{m\in\mathbf{H}} X_{iimj} - \sum_{m\in\mathbf{H}} X_{ijmj} \qquad \forall i, j > i \qquad (2.24)$$

$$V_{ijm} + \sum_{k \in \mathbf{H}} X_{ikmj} \leq Y_{m \in \mathbf{H}} \qquad \forall i, j > i, m \tag{2.25}$$

$$\sum_{m \in \mathbf{H}} V_{ijm} = 1 - \sum_{k \in \mathbf{H}} X_{ikjj} - \sum_{k \in \mathbf{H}} X_{ikij} \qquad \forall i, j > i \tag{2.26}$$

$$X_{ikmj} \in \{0,1\} \ \forall i, j > i, k, m; \ \ Y_k \in \{0,1\} \ \forall k; \ \ U_{ijk}, V_{ijk} \in \{0,1\} \ \forall i, j > i, k \tag{2.27}$$

we use a binary variable Y_k to indicate whether k is a hub. Constraints (2.19)-(2.20) imply that if i (or j) is a hub, it must be the first (or the second) hub in the routes of all flows from i (or to j). Constraints (2.22) require that each $i - j$ flow must have a route through hub(s).

Compared to the R-SAHMP, R-MAHMP is much simpler. First, Campell (1994) states that, for the classical MA-HLP, since there is no capacity restriction on links, each $i - j$ flow should be routed through the least-cost hub pair. So one optimal solution would always force the X variables to be 1 or 0 and therefore there is no need to restrict X variables to be binary. Second, MA-HLP is polynomial solvable if p is fixed. In fact, these two observations still hold in R-MAHMP. For a given p, the R-MAHMP problem is polynomially solvable, and there exists one optimal solution such that all the flow variables X_{ikmj} take either 0 or 1 for all $i, j > i, k$, and m.

2.4.2 Solution Methods for R-MAHMP

Note that the two linearization approaches described in Appendix A could be applied to R-MAHMP with little modification. So, we only describe the development of a Lagrangian relaxation algorithm for R-MAHMP. We dualize constraints linking the route variables $\mathbf{X}$ and the hub variables $\mathbf{Y}$ and solve two resulting subproblems separately. By dualizing constraints in (2.19), (2.20), (2.23), and (2.25) with Lagrangian multipliers $\delta_{ij,1}$, $\delta_{ij,2}$, $\gamma_{ijk,1} \geq 0$, and $\gamma_{ijm,2} \geq 0$, we obtain subproblems MAsub-1 and MAsub-2 as follows.

$$\min\{\sum_{k \in \mathbf{H}} \bar{C}_k Y_k : \sum_{k \in \mathbf{H}} Y_k = p, \quad Y_k \in \{0,1\} \ \forall k\}$$

where $\bar{C}_k = -\sum_{i\in\mathbf{N},i<k}\delta_{ik,1} - \sum_{i\in\mathbf{N},i>k}\delta_{ki,2} - \sum_{i\in\mathbf{N}}\sum_{j\in\mathbf{N},j>i}(\gamma_{ijk,1}+\gamma_{ijk,2})$. Clearly, MAsub-1 can be solved by sorting variables' coefficients and selecting smallest p of them.

$$
\begin{aligned}
\min \ & \sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{H}/\{i\}}\sum_{m\in\mathbf{H}}\sum_{\substack{j\in\mathbf{N}/\{m\}\\ j>i}}(F_{ikmj}w_{ij}(1-q_k-q_m^k)+\gamma_{ijk,1}+\gamma_{ijm,2})X_{ikmj} \\
& + \sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{m\in\mathbf{H}/\{j\}}(F_{iimj}w_{ij}(1-q_m^i)+\delta_{ij,2}+\gamma_{iji,1}+\gamma_{ijm,2})X_{iimj} \\
& + \sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{k\in\mathbf{H}/\{i\}}(F_{ikjj}w_{ij}(1-q_k^j)+\delta_{ij,1}+\gamma_{ijk,1}+\gamma_{ijj,2})X_{ikjj} \\
& + \sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}(F_{iijj}w_{ij}+\delta_{iji,1}+\gamma_{iji,1}+\delta_{ijj,2}+\gamma_{ijj,2})X_{iijj} \\
& + \sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{H}}\sum_{m\in\mathbf{H}/\{k\}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{n\in\mathbf{H}}\rho F_{inmj}w_{ij}q_kX_{ikmj}U_{ijn} + \sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{k\in\mathbf{H}}\gamma_{ijk,1}U_{ijk} \\
& + \sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{H}}\sum_{m\in\mathbf{H}/\{k\}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{n\in\mathbf{H}}\rho F_{iknj}w_{ij}q_mX_{ikmj}V_{ijn} + \sum_{i\in\mathbf{N}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{m\in\mathbf{H}}\gamma_{ijm,2}V_{ijm} \\
& + \sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{H}}\sum_{\substack{j\in\mathbf{N}\\ j>i}}\sum_{n\in\mathbf{H}}\rho F_{innj}w_{ij}q_kX_{ikkj}U_{ijn}
\end{aligned}
\tag{2.28}
$$

s.t.

$$\text{Constraints } (2.22), (2.24), (2.26), (2.27)$$

$$U_{ijk} + \sum_{m\in\mathbf{H}} X_{ikmj} \le 1 \qquad \forall i, j > i, k \tag{2.29}$$

$$V_{ijm} + \sum_{k\in\mathbf{H}} X_{ikmj} \le 1 \qquad \forall i, j > i, m \tag{2.30}$$

similar to SAsub-2, constraints (2.29) and (2.30) are supplied to get a tighter lower bound. Again, MA-sub-2 can be solved by using the combinatorial structure of each single $i-j$ flow. To obtain a feasible solution, as well as an upper bound, we take advantage of the result from MAsub-1 to fix hubs. Then, an optimal solution for those given hubs can be determined by deriving an optimal route for each individual $i-j$ flow.

Lagrangian multipliers are updated iteratively by applying the classical subgradient algorithm. Also, a variable fixing strategy and a Branch-and-Bound technique which consider only hub location variables, are developed and implemented. Given that optimal routing decisions can be obtained in polynomial time if all hubs are fixed, this Branch-and-Bound procedure is guaranteed to be completed by branching on hub location variables only, which has a similar complexity to that of reliable facility location models in Snyder and Daskin (2005), Li and Ouyang (2011), and Cui et al. (2010)

2.5 Computational Experiments

2.5.1 Data and Design of Experiments

We test our algorithms on the widely-used CAB data set (O'Kelly, 1987), which contains the distance between two nodes (interpreted as the transportation cost $\mathbf{c}$) and origin-destination traffic flow $\mathbf{w}$.We set the disruption rate q_i to a random number within $[0.01, 0.05]$ for $i \in \mathbf{N}$.

We consider 36 combinations structured from setting the number of nodes $|\mathbf{N}| = 10, 15, 20,$ 25, the number of hubs $p = 3, 5, 7$, and inter-hub transportation cost discount factor $\alpha =$ $0.3, 0.5, 0.7$. Because rerouting flows will cause more operations and much longer waiting times, we set ρ to 2 to represent this effect (Welman et al., 2010).

The aforementioned instances provide a test bed for both R-SAHMP and R-MAHMP models. We set the optimality tolerance, ϵ, to 0.1% for all solution methods, including the off-the-shelf MIP solver CPLEX 12.1 that is adopted for benchmark. For the Lagrangian relaxation/Branch-and-Bound algorithm, the initial values of all multipliers are set to zero. The step-size multiplier, Δ, is set to 6; the maximum number of iterations allowed to obtain an improvement of the lower bound is set to 50, i.e., when 50 consecutive iterations fail to improve the lower bound, Δ will be halved and the Lagrangian multipliers will be reset to the values used to get the best lower bound. The maximum number of iterations at the root node in the Branch-and-Bound tree is set to 3000 and at a child node it is set to 200. In the implementation of subgradient method, we terminate the Lagrangian procedure if one of the following conditions is met: (i) all Lagrangian multipliers are zero, which implies the current solution is proven to be optimal; (ii) the difference between the upper and lower bounds is below a threshold value ϵ, i.e., an ϵ-optimal solution is found; (iii) the maximum number of iterations, 3000, is reached. If (iii) happens, the variable fixing procedure starts, then if applying variable fixing fails to reduce the gap to less than ϵ, Branch-and-Bound is embedded into the Lagrangian relaxation algorithm. The maximum computation time is set as 3600 seconds. The problem is reported as unsolvable if no optimal solution is obtained within 3600 seconds.

All algorithms are implemented in C++, and all instances are tested on a Dell Optiplex 760 desktop computer (Intel Core 2 Duo CPU, 3.0GHz, 3.25GB of RAM) in Windows XP environment.

2.5.2 Performance of Lagrangian Relaxation and Branch-and-Bound

Table 2.1 summarizes the computational results of our Lagrangian relaxation and Branch-and-Bound methods for instances of R-SAHMP and R-MAHMP. The column marked *Iter.* indicates the total number of Lagrangian iterations in all Branch-and-Bound nodes; the column marked *Gap(%)* provides the smallest relative gap we have achieved within the time limit. The column *BB_Nodes* shows the total number of nodes evaluated in the procedure of Branch-and-Bound (excluding the root node); the column marked *Time(s)* presents the total computational time in seconds for obtaining optimal solution, if some instances cannot be solved due to time limit or memory issue, we use T or M, respectively, to represent the reason.

Similarly, Table 2.2 presents computational results of CPLEX 12.1 used to solve two types of linearized formulations, i.e., those obtained by the standard and a compact linearization methods, for R-SAHMP and R-MAHMP. Detailed derivations and concrete linear formulations are presented in the appendix. Results of instances with $|\mathbf{N}| > 15$ are omitted because CPLEX fails to deal with larger instances within 3600 seconds.

The outcomes of the computational experiments show that: (i) The commercial solver CPLEX is of a very limited capability to solve practical instances with more than 10 nodes. With compact linearization formulation, the solver can provide feasible solutions; (ii) the Lagrangian relaxation algorithm with variable fixing and Branch-and-Bound is efficient in solving reliable models. All 72 instances can be solved to optimality within 1000s; (iii) the Branch-and-Bound technique is necessary to derive optimal solutions for quite a few instances. This observation clearly shows that reliable models are more challenging than the

Table 2.1: Computation of R-SAHMP and R-MAHMP

$\lvert\mathbf{N}\rvert$	p	α	R-SAHMP				R-MAHMP			
			Iter.	*BB_Nodes*	*Gap(%)*	*Time(s)*	*Iter.*	*BB_Nodes*	*Gap(%)*	*Time(s)*
10	3	0.3	250	0	0.099	1.3	878	2	0.100	2.6
	5	0.3	565	0	0.100	3.7	1057	0	0.090	3.5
	7	0.3	184	0	0.098	1.8	604	0	0.100	1.4
	3	0.5	257	0	0.098	4.6	830	2	0.000	2.6
	5	0.5	1902	6	0.095	17.6	866	0	0.097	2.3
	7	0.5	184	0	0.099	1.7	587	0	0.099	1.5
	3	0.7	182	2	0.099	1.5	607	2	0.070	2.7
	5	0.7	1515	4	0.096	5.8	731	0	0.098	1.8
	7	0.7	323	0	0.098	2.7	561	0	0.095	2.8
15	3	0.3	1015	2	0.016	15.0	1455	4	0.000	20.0
	5	0.3	1353	4	0.099	27.3	596	0	0.097	4.3
	7	0.3	1722	6	0.099	30.4	716	0	0.100	8.4
	3	0.5	1362	4	0.095	24.4	910	4	0.100	18.4
	5	0.5	1701	6	0.080	31.3	563	0	0.096	2.3
	7	0.5	1313	6	0.090	21.7	635	2	0.100	15.7
	3	0.7	980	2	0.099	21.6	1958	8	0.098	27.6
	5	0.7	1540	4	0.099	31.4	573	0	0.092	3.4
	7	0.7	512	0	0.099	13.7	564	2	0.099	12.7
20	3	0.3	482	0	0.098	32.3	1979	6	0.000	92.0
	5	0.3	553	0	0.099	37.6	608	2	0.000	36.1
	7	0.3	118	0	0.100	8.1	581	0	0.100	33.2
	3	0.5	1762	6	0.098	116.1	1441	6	0.000	68.1
	5	0.5	1584	4	0.099	107.8	605	0	0.100	26.6
	7	0.5	589	0	0.099	63.7	971	4	0.100	77.8
	3	0.7	3925	16	0.097	177.6	1660	8	0.044	82.8
	5	0.7	3871	14	0.099	188.2	722	0	0.100	38.4
	7	0.7	2095	8	0.097	138.8	561	0	0.100	28.9
25	3	0.3	2020	6	0.098	365.3	2845	8	0.000	338.0
	5	0.3	965	2	0.100	221.3	1709	6	0.100	268.8
	7	0.3	812	2	0.100	239.2	1601	6	0.000	257.5
	3	0.5	1745	6	0.097	361.2	2914	10	0.092	375.2
	5	0.5	2774	10	0.099	435.5	727	0	0.100	97.7
	7	0.5	201	0	0.082	33.0	1587	8	0.100	323.4
	3	0.7	765	4	0.076	121.7	3313	12	0.000	416.1
	5	0.7	7318	34	0.096	953.8	3126	12	0.099	457.8
	7	0.7	879	2	0.100	249.2	613	0	0.100	31.2

Table 2.2: Solver Performance for R-SAHMP and R-MAHMP

$\|\mathbf{N}\|$	p	α	R-SAHMP				R-MAHMP			
			StdLinear		*CptLinear*		*StdLinear*		*CptLinear*	
			Time(s)	*Gap(%)*	*Time(s)*	*Gap(%)*	*Time(s)*	*Gap(%)*	*Time(s)*	*Gap(%)*
10	3	0.3	33.7	0.032	T	0.514	641.1	0.100	1456.7	0.100
	5	0.3	24.5	0.047	T	1.827	3516.3	0.100	T	1.896
	7	0.3	5.2	0.000	2.2	0.000	138.4	0.100	4.4	0.100
	3	0.5	40.4	0.000	T	2.069	343.5	0.099	T	0.321
	5	0.5	35.3	0.000	T	3.414	T	0.164	2041.3	0.000
	7	0.5	7.1	0.006	4.5	0.094	520.9	0.100	76.5	0.100
	3	0.7	50.1	0.000	T	2.007	407.6	0.100	1335.4	0.100
	5	0.7	39.2	0.010	T	1.660	M	0.760	T	1.951
	7	0.7	7.6	0.000	19.7	0.099	M	0.330	M	0.740
15	3	0.3	M	NA	T	4.030	M	16.360	T	4.440
	5	0.3	M	NA	M	5.070	M	14.480	T	5.441
	7	0.3	M	NA	T	4.789	M	18.660	T	3.669
	3	0.5	M	NA	T	3.729	M	11.650	T	4.531
	5	0.5	M	NA	M	5.340	M	14.960	T	4.620
	7	0.5	M	NA	T	4.020	M	13.560	T	3.117
	3	0.7	M	NA	T	4.907	M	10.110	T	3.949
	5	0.7	M	NA	M	4.560	M	9.770	T	3.723
	7	0.7	M	NA	T	3.480	M	9.600	T	2.662

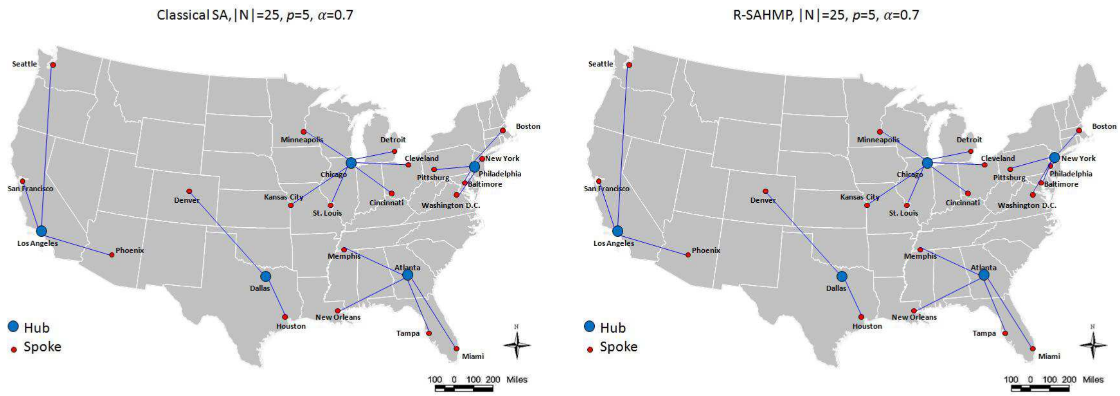
classical ones for which study presented in Pirkul and Schilling (1998) shows that Lagrangian relaxation method itself is sufficient to solve CAB instances; (iv) comparing reliable SA and MA models, the former often involves more Branch-and-bound nodes and longer computation times, which also confirms that the former one is of a higher complexity level than the latter one.

2.5.3 Analysis and Discussion on System Design and Performance

In this section, we discuss the impact of reliable design paradigm on the system configurations and performance. The network configurations are compared with those determined by the classical hub-and-spoke models, which actually are special cases of the proposed R-SAHMP and R-MAHMP with the disruption probability $\mathbf{q}$=0.

2.5.3.1 Impact of Hub Unavailability on System Design

Hub locations and spoke node allocations of reliable models could be different from those of classical models. Figure 2.3 demonstrates a case with $|\mathbf{N}| = 25$ (their associated disruption probabilities are presented in Table B.1 in the appendix), $p = 5$, and $\alpha = 0.7$. Note that with classical hub-and-spoke network design, Philadelphia is selected as one of the hubs. Nevertheless, when the reliability issue is considered in the design, this hub is replaced by New York, and the spoke nodes in the service region of Philadelphia are re-allocated to New York as well.



(a) Configuration from Classical Model (b) Configuration from Reliable Model

Figure 2.3: Optimal System Configurations in Different SA Models

Expected numbers of served passengers are calculated as the performance metrics and compared for different network configurations. It is a better measurement of airlines service quality for this study because the objective functions of reliable hub-and-spoke network models include the costs under both normal and disruption conditions which make them incomparable with the objective functions from classical models that only count the costs under normal condition. The following formulas are used to calculate the expected numbers of served passengers of classical (Psg_c) and reliable hub-and-spoke networks (Psg_r) respectively.

Given the disruption probabilities presented in Table B.1, for the particular case discussed in this subsection, the classical network configuration is expected to transport $4,126,900$ passengers and the reliable one $4,270,000$ passengers (by both regular and alternative routes) with a 3.47% improvement. In fact, we want to highlight that, even without considering backup hubs and alternative routes, the derived reliable network system can transport more passengers ($4,127,250$) just by its regular routes than the classical network configuration. Such an observation indicates that it is necessary to consider the availability issue of network components when we design the network system for better performance.

Table 2.3: Comparison of Served Passengers

$\lvert\mathbf{N}\rvert$	p	SA model Classical Psg_c	SA model Reliable Psg_r	SA model Reliable $Improvement(\%)$	MA model Classical Psg_c	MA model Reliable Psg_r	MA model Reliable $Improvement(\%)$
10	3	484653	499513	3.066	490297	499513	1.845
	5	487181	499513	2.531	494180	499513	1.068
	7	494730	499513	0.967	495343	499513	0.835
15	3	1155060	1182470	2.373	1162180	1182470	1.716
	5	1149840	1182470	2.838	1164140	1182470	1.550
	7	1154940	1182470	2.384	1169760	1182470	1.075
20	3	2781810	2877300	3.433	2820550	2877300	1.972
	5	2801900	2877300	2.691	2832790	2877300	1.547
	7	2803800	2877300	2.621	2845150	2877300	1.117
25	3	4135680	4270000	3.248	4163530	4270000	2.493
	5	4126900	4270000	3.467	4166670	4270000	2.420
	7	4133240	4270000	3.309	4210840	4270000	1.385

$$Psg_c = \sum_{i\in\mathbf{N}} \sum_{k\in\mathbf{H}/\{i\}} \sum_{m\in\mathbf{H}} \sum_{\substack{j\in\mathbf{N}/\{m\} \\ j>i}} w_{ij}(1 - q_k - q_m^k)X_{ikmj}$$

$$\sum_{i\in\mathbf{N}} \sum_{\substack{j\in\mathbf{N} \\ j>i}} \left(\sum_{m\in\mathbf{H}/\{j\}} w_{ij}(1 - q_m^i)X_{iimj} + \sum_{k\in\mathbf{H}/\{i\}} w_{ij}(1 - q_k^j)X_{ikjj} + w_{ij}X_{iijj} \right) \tag{2.31}$$

$$\begin{aligned} Psg_r = \ & Psg_c + \sum_{i\in\mathbf{N}} \sum_{\substack{j\in\mathbf{N} \\ j>i}} \sum_{k\in\mathbf{H}} \sum_{m\in\mathbf{H}/\{k\}} \sum_{n\in\mathbf{H}} w_{ij}q_kX_{ikmj}U_{ijn} \\ & + \sum_{i\in\mathbf{N}} \sum_{\substack{j\in\mathbf{N} \\ j>i}} \sum_{k\in\mathbf{H}} \sum_{m\in\mathbf{H}/\{k\}} \sum_{n\in\mathbf{H}} w_{ij}q_mX_{ikmj}V_{ijn} \\ & + \sum_{i\in\mathbf{N}} \sum_{\substack{j\in\mathbf{N} \\ j>i}} \sum_{k\in\mathbf{H}} \sum_{n\in\mathbf{H}} w_{ij}q_kX_{ikkj}U_{ijn} \end{aligned} \tag{2.32}$$

2.5.3.2 Performance of Reliable Hub-and-spoke Networks

The expected numbers of served passengers of reliable models and those of the classical models are further compared for more scenarios. Results are listed in Table 2.3. In the table, the performance measures (Psg_c and Psg_r) for the classical and reliable model are presented with numerical values and the relative improvements (denoted by $Improvement$) achieved by the reliable model are shown in percentages. In all experiments, the inter-hub transportation cost discount factor α is set to 0.7.

Note that since our model can handle any single hub disruption, the number of served passengers is exactly the total transportable flow $\sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{N}, j>i} w_{ij}$, which is constant for each fixed $|\mathbf{N}|$. It is observed that the reliable network always transports more passengers compared to classical model, with the magnitude increasing with the growth of the network scale $|\mathbf{N}|$. Therefore, in terms of the expected number of served passengers, the reliable models clearly outperform the classical ones.

2.5.3.3 Verification with Correlated Multiple Disruptions

One assumption we made in developing reliable models is that no more than one hub will fail at any time. In some extreme cases, such an assumption may not valid and multiple failures could occur simultaneously. So, in this section, we perform numerical experiments to evaluate the influence of the single disruption (SD) assumption. We study the optimal network configurations obtained from our models in an environment that correlated multiple disruption (MD) may occur. Letting the random variable D_k be the status for any hub k, i.e., $D_k = 1$ when hub k is down and 0 otherwise, we use the following equations to recalculate the expected number of passengers to be served with possible multiple hub disruptions in the real situation.

$$
\begin{gathered}
Psg'_c = \sum_{i \in \mathbf{N}} \sum_{k \in \mathbf{H}/\{i\}} \sum_{m \in \mathbf{H}} \sum_{\substack{j \in \mathbf{N}/\{m\} \\ j>i}} w_{ij} P(D_k = 0, D_m = 0) X_{ikmj} \\
\sum_{i \in \mathbf{N}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \left(\sum_{m \in \mathbf{H}/\{j\}} w_{ij}(1 - q^i_m) X_{iimj} + \sum_{k \in \mathbf{H}/\{i\}} w_{ij}(1 - q^j_k) X_{ikjj} + w_{ij} X_{iijj} \right),
\end{gathered} \tag{2.33}
$$

$$
\begin{gathered}
Psg'_r = \; Psg'_c \\
+ \sum_{i \in \mathbf{N}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \sum_{k \in \mathbf{H}} \sum_{m \in \mathbf{H}/\{k\}} \sum_{n \in \mathbf{H}} w_{ij} P(D_k = 1, D_m = 0, D_n = 0) X_{ikmj} U_{ijn} \\
+ \sum_{i \in \mathbf{N}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \sum_{k \in \mathbf{H}} \sum_{m \in \mathbf{H}/\{k\}} \sum_{n \in \mathbf{H}} w_{ij} P(D_k = 0, D_m = 1, D_n = 0) X_{ikmj} V_{ijn} \\
+ \sum_{i \in \mathbf{N}} \sum_{\substack{j \in \mathbf{N} \\ j>i}} \sum_{k \in \mathbf{H}} \sum_{n \in \mathbf{H}} w_{ij} P(D_k = 1, D_n = 0) X_{ikkj} U_{ijn}.
\end{gathered} \tag{2.34}
$$

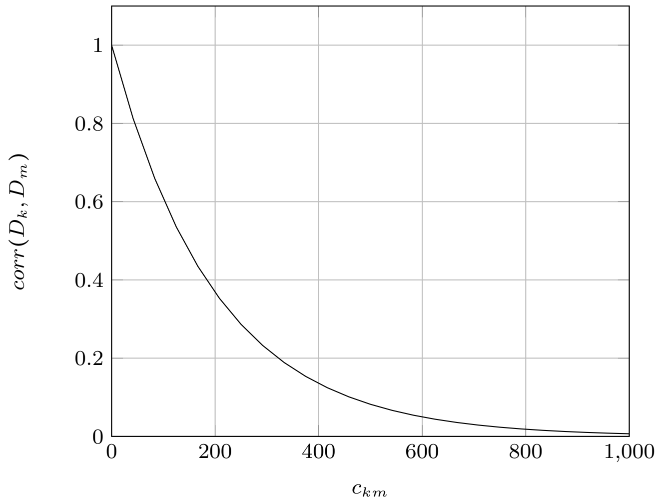


Figure 2.4: Curve of Correlation between D_k and D_m

Given that $P(D_k = 1) = q_k$ for any hub k, by setting a correlation $corr(D_k, D_m)$ of any pair of random variables (D_k, D_m) and assuming a relationship between $P(D_m = 0, D_n = 0|D_k = 1)$ and $P(D_m = 0, D_n = 0)$ we can obtain the probabilities in (2.33) and (2.34). Specifically, we want the correlation between given nodes k and m decreases as the distance c_{km} grows, so we choose $corr(D_k, D_m) = e^{-\Gamma_1 x}$ where Γ_1 is a positive constant. In order to avoid the situation in which the correlation decreases too fast, Γ_1 is set to $\frac{1}{200}$ (see Figure 2.4). Note that under this correlation assumption, the geographically close nodes can have high correlations. For instance, $corr(D_5, D_8) = 0.624$ (Cleveland and Detroit). Based on the correlation function, we can derive the required $P(D_k = 0, D_m = 0)$ and $P(D_k = 1, D_n = 0)$. For the probabilities involving three nodes like $P(D_k = 1, D_m = 0, D_n = 0)$ we further assume that $P(D_m = 0, D_n = 0|D_k = 1) = P(D_m = 0, D_n = 0)(1 - \frac{e^{-\frac{c_{km}+c_{kn}}{2}}}{\Gamma_2})$, i.e., $P(D_m = 0, D_n = 0|D_k = 1)$ is related to but smaller than $P(D_m = 0, D_n = 0)$ and also determined by the average distance $\frac{c_{km}+c_{kn}}{2}$, then $P(D_k = 1, D_m = 0, D_n = 0)$ can be easily calculated. See Appendix C for details. We mention that by changing the form of the correlation function, we can even model negative correlation. Therefore, (2.33) and (2.34) provide us a useful tool to evaluate a hub-and-spoke network in the real practice in which correlated multiple node failures may occur.

First, the relative decrease of expected served passengers with respect to that under the single disruption assumption is listed in the column "*Change*(%)" of Table 2.4. It is easy to observe that, in terms of expected served passengers, the influence of the multiple hub disruptions to the system performance is small (all less than 0.5%). Next, expected served passengers of the reliable model and the classical model under multiple disruptions are computed and listed in Table 2.5. According to the results, proposed reliable models outperform the classical ones under correlated multiple disruptions as well.

Table 2.4: Relative Change of Passengers with Different Assumptions

		SA model			MA model		
$\|\mathbf{N}\|$	p	SD	MD	$Change(\%)$	SD	MD	$Change(\%)$
10	3	499513	497023	-0.498	499513	497804	-0.342
	5	499513	497444	-0.414	499513	498744	-0.154
	7	499513	498563	-0.190	499513	498853	-0.132
15	3	1182470	1181630	-0.071	1182470	1181060	-0.119
	5	1182470	1179170	-0.279	1182470	1179760	-0.229
	7	1182470	1179450	-0.255	1182470	1180440	-0.172
20	3	2877300	2871970	-0.185	2877300	2874430	-0.100
	5	2877300	2873540	-0.131	2877300	2874820	-0.086
	7	2877300	2870280	-0.244	2877300	2873950	-0.116
25	3	4270010	4262770	-0.170	4270000	4263220	-0.159
	5	4270000	4263040	-0.163	4270000	4265340	-0.109
	7	4270000	4262290	-0.181	4270000	4264660	-0.125

Table 2.5: Performance of Reliable Models with Multiple Disruption Assumption

		SA model			MA model		
		Classical	Reliable		Classical	Reliable	
$\|\mathbf{N}\|$	p	Psg'_c	Psg'_r	$Improvement(\%)$	Psg'_c	Psg'_r	$Improvement(\%)$
10	3	488612	497023	1.721	490613	497804	1.466
	5	491128	497444	1.286	494187	498744	0.922
	7	494733	498563	0.774	496130	498853	0.549
15	3	1155140	1181630	2.293	1163250	1181060	1.531
	5	1157080	1179170	1.909	1164630	1179760	1.299
	7	1162000	1179450	1.502	1171440	1180440	0.768
20	3	2783080	2871970	3.194	2828600	2874430	1.620
	5	2802660	2873540	2.529	2833540	2874820	1.457
	7	2805300	2870280	2.316	2846600	2873950	0.961
25	3	4137710	4262770	3.022	4164820	4263220	2.363
	5	4129250	4263040	3.240	4182190	4265340	1.988
	7	4183440	4262290	1.885	4217820	4264660	1.111

Finally, a sensitivity analysis of failure rates on system configurations is conducted both for classical and reliable models. Assuming that all nodes have the same hub disruption probability, we investigate the impact of small variation in failure rate $\mathbf{q}$ on the aforementioned performance measures, Psg'_c and Psg'_r. Both low ($\mathbf{q}$ = 0.009) and high probability scenarios ($\mathbf{q}$ = 0.04) are considered for multiple disruption scenarios. In Table 2.6, numerical results for $|\mathbf{N}| = 25$, $p = 3, 5, 7$, and $\alpha = 0.7$ are presented, with the columns Psg'_c and Psg'_r

representing the expected number of passengers of the corresponding network configuration with initial hub failure rates and the column $Change(\%)$ representing the percentage change from Psg_c' to Psg_r' when $\mathbf{q}$ is increased by 0.001 while keeping the network configuration fixed.

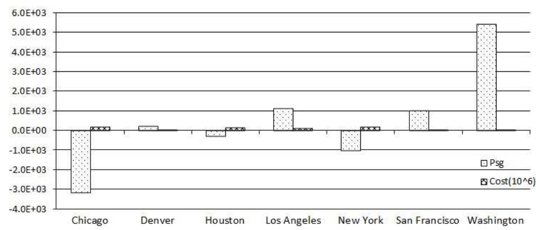
A clear observation is that the reliable model is much less sensitive than the classical model to the variations of hub availability. The reliable networks have a higher survivability and are more robust to disruptions. Such observations again demonstrate the importance of taking into account hub unavailabilities in designing robust hub-and-spoke networks.

Table 2.6: Sensitivity Analysis of Failure Rates Under Multiple Disruptions

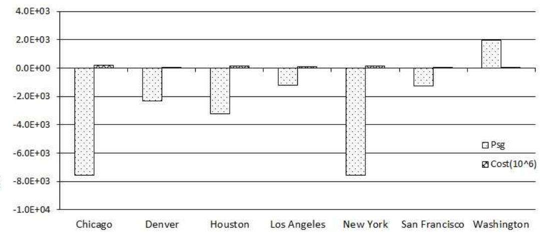
Model	p	$\mathbf{q}$	Classical		Reliable	
			Psg_c'	$Change(\%)$	Psg_r'	$Change(\%)$
SA	3	0.009	4226770	-0.114	4268530	-0.005
		0.04	4079250	-0.116	4258290	-0.011
	5	0.009	4236470	-0.088	4269210	-0.003
		0.04	4122120	-0.089	4262140	-0.008
	7	0.009	4239740	-0.079	4267870	-0.006
		0.04	4136310	-0.080	4256210	-0.011
MA	3	0.009	4248990	-0.055	4269440	-0.002
		0.04	4176820	-0.056	4264540	-0.005
	5	0.009	4250050	-0.052	4269560	-0.002
		0.04	4181600	-0.053	4265030	-0.005
	7	0.009	4250620	-0.051	4268440	-0.005
		0.04	4184070	-0.051	4259930	-0.008

2.5.3.4 Application of Proposed Reliable Models

The recent merger between United and Continental Airlines brings the new United Airlines (UA) eight domestic hubs. The hub at Cleveland Hopkins Airport shares a great functional similarity with the hub at Chicago O'Hare and is expected to be closed to save cost by industrial experts (Grossman (2010)). In this section, we apply the proposed reliable models to UA network and evaluate different network configurations in a quantitative way. Our analysis uses the proposed reliable MA network with CAB data set under the correlated multiple disruption assumption with current eight hubs in UA. Parameter $\mathbf{q}$ is shown in Table B.1, $|\mathbf{N}|$ is set as 25 and $\alpha = 0.7$. We evaluate two performance measurements, i.e., the expected number of served passengers and the expected transportation cost, under different single hub closing options. We point out that our study is simply for demonstration as UA's coverage and traffic flows may be very different from those from CAB data set.



(a) Results for $\mathbf{q}$ in Table B.1



(b) Results for $\mathbf{q}$ in Table B.1 with q_5 = 0.025

Figure 2.5: Relative Changes of Passengers and Transportation Cost

We first compute the impact of closing Cleveland and obtain corresponding results: the expected number of served passengers is 4.25603×10^6 and the expected transportation cost is 3.54629×10^9. Then, we compute results of closing any of other hubs and calculate the differences compared with the result of closing Cleveland. The outcomes are presented in Figure 2.5. For example, closing the hub in New York will result in 4.255×10^6 served passengers and a transportation cost of 3.71088×10^9, which are 1030 less passengers and 1.6459×10^8 more cost compared to the performance of closing the Cleveland hub (as shown in Figure 2.5(a)).

It is observed that the disruption probability of Cleveland (q_5) is relatively high in Table B.1 (0.047 in the range of 0.012 to 0.050 for all 25 nodes). A different scenario with q_5 equal to 0.025 is evaluated and the corresponding results are presented in Figure 2.5(b). We observe that, from the perspective of transportation cost, the hub in Cleveland is always the optimal choice to be closed. This quantitative analysis endorses the opinion from the industrial expert. Nevertheless, if the number of served passengers is of a higher priority, closing the hub at Washington DC becomes a better option. Although no current information of UA but the CAB data set is used, this quantitative analysis demonstrates that the proposed reliable models and algorithms can be used to provide decision support to the management of airlines to re-structure their networks. Similarly, they can be used by airlines for identifying strategic partners/alliance to hedge against disruptions and achieve their desired operational goals.

2.6 Conclusion

In this study, we construct reliable single and multiple allocation hub-and-spoke models that generalize their classical counterparts. Our models seek to build hub-and-spoke systems with backup hubs and alternative routes to better hedge against various disrupted situations in practice. Due to the complexity of the reliable models, we develop a set of easy-to-implement Lagrangian relaxation/Branch-and-Bound algorithms that can compute optimal solutions efficiently. Computational study demonstrates the effectiveness of these algorithms, as well as the superiority of the proposed models to classical models in terms of serving passengers and being robust subject to the variations of hub failure rates.

To the best of our knowledge, our work is the first analytical study on reliable hub-and-spoke network design problem. It theoretically extends the existing literature on reliable network design and also has a clear practical impact on transportation and telecommunications systems. The proposed models can be slightly modified to deal with different situations, such as just allowing a subset of nodes being chosen to be hubs and allowing a subset of flows to be rerouted. Therefore, they are powerful decision support tools for system designers to derive optimal system configuration with a desired trade-off among performance measures.

Nevertheless, the proposed models have significant caveats that need to be addressed in future research. Although it is demonstrated that the resulted network settings from proposed models outperform those from classical models under correlated multiple disruption scenarios, explicitly including multiple disruption into mathematical modeling is a desire and should be considered in future research. Furthermore, more complicated issues in practice, such as congestion effect, should also be taken into account.

3 Extended Reliable Hub-and-spoke System Design

3.1 Introduction and Previous Works

In this chapter, we conduct a deeper research of the reliable hub-and-spoke network on the basis of the work done in Chapter 2. Specifically, multiple node failures and hub congestions will be considered.

Chapter 2 embraced the single random hub failure in the hub-and-spoke system design. Although multiple hub failures rarely happen in the real world, the single disruption assumption is restrictive. Neglecting the multiple simultaneous hub failures may lead to a suboptimal network configuration under adverse weather or other extreme conditions. Moreover, in order to deal with the real applications of the hub-and-spoke system, more realistic features should be considered. For instance, a large volume of traffic is often required by an interhub link to maintain the economies of scale, while consolidating flows at hubs will lead to congestions. Take the air transportation industry for example, it is estimated that airport congestion costs US economy $32.9 billion in 2007 (Pita et al., 2012). Federal Aviation Administration (FAA) forecasts that the total number of US airline passengers will reach 1 billion in 2024 (Price, 2014), so the congestion effect will become more and more significant in the following years. Unfortunately, this main side effect is seldom studied in the hub-and-spoke system literature primarily because the congestion is normally modeled as a nonlinear function of the traffic flow and the introduction of nonlinear terms will cause the formulation highly difficult to solve. de Camargo et al. (2009b) designs the multiple allocation hub-and-spoke system with hub congestion, generalized Benders decomposition is applied and the nonlinear congestion is handled directly by the flow deviation algorithm. However, in many other applications, the nonlinear formulation has prohibitively large integrality gap which renders solving nonlinear models impossible. Elhedhli and Hu (2005) and Elhedhli and Wu (2010) consider hub-and-spoke networks with congestions applying tangent lines to form lower envelopes and approximate the nonlinear terms. Only single allocation is studied in both works and no hub failures are included in the model. We mention that the two emerging factors are actually closely related to each other. Hub congestion is a major reason for hub failures while diverting flows from the failed hub in turn could overload other existing hubs. Ideally, the hub congestion and hub disruptions should be considered together in the design stage of a hub-and-spoke network.

In our new formulation, we propose to use two-stage robust optimization to address an extended reliable hub-and-spoke design problem which considers multiple node failures and hub congestions. All possible disruptive scenarios (single and multiple hub failures) are represented in an uncertainty set and for each of them, the affected passengers are rerouted with backup routes. Hub congestions are included in the objective function and modeled as convex functions of traffic flows which are then linearized to avoid computational challenges brought by nonlinearity. In addition to that, we will also study a phenomenon in which the flow originated from a disrupted node, regardless of whether the origin is a hub or not, will not be served. This demand loss assumption is generally accepted in air transportation systems where the passengers departing from a disrupted airport can not be assigned to another airport due to long distance between two cities. The large scale model is successfully solved by customized column-and-constraint generation algorithm. The solution of our robust model is able to mitigate the recourse cost of failure scenarios comprehensively considering hub disruptions and congestions as well as their coupling effects mentioned above.

The rest of this chapter is organized as follows. Section 3.2 presents our route-based and flow-based formulations of reliable hub-and-spoke system design with congestions along with their variants with demand loss assumption. Section 3.3 describes the linearization method and the solution method applied to solve the cases. Section 3.4 gives the computational results using CAB data set with corresponding analysis and Section 3.5 provides the conclusion.

3.2 Formulation

3.2.1 Route-based Model

We adopt the multiple allocation which is widely applied in the real practice nowadays and assume that the number of hubs is fixed as p. The formulation has a structure of the two-stage robust optimization model. In the first stage, the model determines the hub location and the primary routing strategy under the normal situation. In the second stage, a worst case scenario is identified given the network configuration of the first stage and the recourse strategy is then found to minimize the recourse cost. Hub congestion costs are taken into account for both stages. Let the set of nodes be $\mathbf{N}$ and ρ be the weight of recourse cost. Following the notation in Chapter 2, we define the unit transportation cost $\mathbf{c}$ and $\mathbf{F}$, hub location variable $\mathbf{Y}$, and primary routing variable $\mathbf{X}$. Besides, the continuous decision variable W_{ikmj} is used to indicate the portion of traffic flow between i and j that uses backup hub k and m in a disruptive scenario $z \in A$, where A is the uncertainty set expressed as

$$\{\mathbf{z} \in \{0,1\}^{|\mathbf{N}|} : \sum_{k \in \mathbf{N}} z_k \leq \Gamma\}$$

λ_{ij} is used to represent the flow between i and j. P_k^0 and P_k are the traffic flows served by hub k under normal and disruptive situations, respectively. For the traffic flow volume P_k of hub k, the associated congestion cost is defined to be a convex function $C(P_k) = a_k(P_k)^{b_k}$, where a_k and b_k are constants controlling the curvature of the function.

For now, we assume that a disrupted node will cease to serve other flows if it is a hub but still generate flows originating from and receive flows going to itself that have to be routed through hubs. The formulation RoHMPC based on route variables (see Chapter 2 or Skorin-Kapov et al. (1996)) is given as follows.

$$\begin{aligned} \min_{\mathbf{X},\mathbf{Y},\mathbf{P}^0} (1-\rho)(\sum_i \sum_k \sum_m \sum_{j>i} F_{ikmj}\lambda_{ij}X_{ikmj} + \sum_k a_k(P_k^0)^{b_k}) + \\ \rho \max_{\mathbf{z}\in A} \min_{\mathbf{W},\mathbf{P}} (\sum_i \sum_k \sum_m \sum_{j>i} F_{ikmj}\lambda_{ij}W_{ikmj} + \sum_k a_k(P_k)^{b_k}) \end{aligned} \tag{3.1}$$

s.t.

$$\sum_k Y_k = p \tag{3.2}$$

$$\sum_k \sum_m X_{ikmj} = 1 \qquad \forall i, j > i \tag{3.3}$$

$$X_{iijj} \geq Y_i + Y_j - 1 \qquad \forall i, j > i \tag{3.4}$$

$$\sum_{m \neq j} X_{iimj} \geq Y_i - Y_j \qquad \forall i, j > i \tag{3.5}$$

$$\sum_{k \neq i} x_{ikjj} \geq Y_j - Y_i \qquad \forall i, j > i \tag{3.6}$$

$$\sum_i \sum_{j>i} \sum_m \lambda_{ij} X_{ikmj} + \sum_i \sum_{j>i} \sum_{m \neq k} \lambda_{ij} X_{imkj} = P_k^0 \qquad \forall k \tag{3.7}$$

$$\sum_m W_{ikmj} + \sum_{m \neq k} W_{imkj} \leq 1 - z_k \qquad \forall i, j > i, k \tag{3.8}$$

$$\sum_m W_{ikmj} + \sum_{m \neq k} W_{imkj} \leq Y_k \qquad \forall i, j > i, k \tag{3.9}$$

$$\sum_k \sum_m W_{ikmj} = 1 \qquad \forall i, j > i \tag{3.10}$$

$$w_{iijj} \geq (Y_i - z_i) + (Y_j - z_j) - 1 \qquad \forall i, j > i \tag{3.11}$$

$$\sum_m W_{iimj} \geq (Y_i - z_i) - (Y_j - z_j) \qquad \forall i, j > i \tag{3.12}$$

$$\sum_k W_{ikjj} \geq (Y_j - z_j) - (Y_i - z_i) \qquad \forall i, j > i \tag{3.13}$$

$$\sum_i \sum_{j>i} \sum_m \lambda_{ij} W_{ikmj} + \sum_i \sum_{j>i} \sum_{m \neq k} \lambda_{ij} W_{imkj} = P_k \qquad \forall k \tag{3.14}$$

$$X_{ikmj} \geq 0 \ \forall i, k, m, j > i; \ Y_k \in \{0, 1\} \ \forall k; \ P_k^0 \geq 0 \ \forall k;$$

$$W_{ikmj} \geq 0 \ \forall i, k, m, j > i; \ P_k \geq 0 \ \forall k \tag{3.15}$$

Note that in both stages, the transportation cost and congestion cost are simultaneously considered. Most constraints are directly borrowed from the model R-MAHMP in Chapter 2. Constraints (3.4) ensure that the flow λ_{ij} goes through hubs i and j when both of them are hubs. Constraints (3.5) and (3.6) restrict the number of hubs to two when i or j are hubs, respectively. Constraints (3.7) simply calculate the flow of hub k and build the relationship between $\mathbf{P}^0$ and $\mathbf{X}$. Constraints (3.11)-(3.14) in the recourse problem are similar to their counterparts in the first stage taking into account the hub availabilities.

The model RoHMPC has an important property that given fixed hub locations $\mathbf{y}^*$, the corresponding worst case scenario $\mathbf{z}^*$ will have $z_k^* \leq y_k^*$ for all k. In other words, the disrupted nodes will always be hubs in a worst case scenario. The proof of this property is straightforward under the assumption that the disrupted node still generate and receive flows that have to be routed by hubs. Since the worst scenario is the one with the largest recourse cost, if we have a worst scenario that has ${z_{k0}}^* = 1$ for a nonhub k_0, one can find a scenario with larger recourse cost by letting ${z_{k0}}^* = 0$ and ${z_{k1}}^* = 1$ where ${y_{k_1}}^* = 1$ and ${z_{k_1}}^* = 0$. The contradiction is obtained.

Next, we adopt the demand loss assumption and suppose that a disrupted node will not generate and receive traffic flows and stop serving other flows if it is a hub. The robust hub-and-spoke model with this assumption d-RoHMPC can be derived with little modification from RoHMPC. Only the second-stage constraints (3.8)-(3.14) need to be changed as given below.

$$\min_{\mathbf{X},\mathbf{Y},\mathbf{P}^0} (1-\rho)(\sum_i \sum_k \sum_m \sum_{j>i} F_{ikmj}\lambda_{ij}X_{ikmj} + \sum_k a_k(P_k^0)^{b_k}) +$$

$$\rho \max_{\mathbf{z}\in A} \min_{\mathbf{W},\mathbf{P}} (\sum_i \sum_k \sum_m \sum_{j>i} F_{ikmj}\lambda_{ij}W_{ikmj} + \sum_k a_k(P_k)^{b_k}) \tag{3.16}$$

s.t.

$$\text{Constraints (3.2)} - \text{(3.7)}$$

$$\sum_m W_{ikmj} + \sum_{m\neq k} W_{imkj} \leq 1 - z_k \qquad \forall i, j > i, k \tag{3.17}$$

$$\sum_m W_{ikmj} + \sum_{m\neq k} W_{imkj} \leq Y_k \qquad \forall i, j > i, k \tag{3.18}$$

$$\sum_k \sum_m W_{ikmj} \geq 1 - z_i - z_j \qquad \forall i, j > i \tag{3.19}$$

$$\sum_m W_{iimj} \geq Y_i - z_i - z_j \qquad \forall i, j > i \tag{3.20}$$

$$\sum_k W_{ikjj} \geq Y_j - z_j - z_i \qquad \forall i, j > i \tag{3.21}$$

$$\sum_i \sum_{j>i} \sum_m \lambda_{ij}W_{ikmj} + \sum_i \sum_{j>i} \sum_{m\neq k} \lambda_{ij}W_{imkj} = P_k \qquad \forall k \tag{3.22}$$

$$X_{ikmj} \geq 0 \ \forall i,k,m,j>i; \ Y_k \in \{0,1\} \ \forall k; \ P_k^0 \geq 0 \ \forall k;$$

$$W_{ikmj} \geq 0 \ \forall i,k,m,j>i; \ P_k \geq 0 \ \forall k \tag{3.23}$$

by adding "$-z_i - z_j$" to constraints in (3.10) we guarantee that the flow with disrupted origin or destination will not be served. Constraints in (3.12) and (3.13) are modified similarly.

3.2.2 Flow-based Model

The route-based model mentioned in the previous section is a tight formulation and has the advantage of small integrality gap. However, the four-index variable $\mathbf{X}$ may drastically increase the problem size. In order to control the problem scale and avoid the potential memory issues, we also adopt the flow-based model introduced in Ernst and Krishnamoorthy (1998b), whose variables have at most three indices, to build a robust model that is equivalent to RoHMPC.

The basic idea of the flow-based formulation is to use continuous flow variables $(\mathbf{S}, \mathbf{I}, \mathbf{Q})$ to indirectly describe the configuration of the network. Let S_{ik} denote the flow going from node i to hub k, I_{ikm} be the amount of flow originating at node i going through hubs k and m, and Q_{imj} be the flow originating at node i that is routed to node j using m as the second hub. In our two-stage robust model, we also need $(\mathbf{U}, \mathbf{V}, \mathbf{T})$ as the counterpart of $(\mathbf{S}, \mathbf{I}, \mathbf{Q})$ in the second stage. The flow-based RoHMPCf is given as follows.

$$\min_{\mathbf{Y},\mathbf{S},\mathbf{Q},\mathbf{I},\mathbf{P^0}} (1-\rho)\sum_i [\sum_k c_{ik} S_{ik} + \sum_k \sum_m \gamma c_{km} I_{ikm} + \sum_m \sum_{j>i} c_{mj} Q_{imj}] + (1-\rho)\sum_k a_k (P_k^0)^{b_k} +$$
$$\rho \max_{\mathbf{z}\in A} \min_{\mathbf{U},\mathbf{V},\mathbf{T},\mathbf{P}} \sum_i [\sum_k c_{ik} U_{ik} + \sum_k \sum_m \gamma c_{km} V_{ikm} + \sum_m \sum_{j>i} c_{mj} T_{imj}] + \sum_k a_k (P_k)^{b_k} \tag{3.24}$$

s.t.

$$\sum_k Y_k = p \tag{3.25}$$

$$\sum_k S_{ik} = \sum_{j>i} \lambda_{ij} \qquad \forall i \tag{3.26}$$

$$\sum_m Q_{imj} = \lambda_{ij} \qquad \forall i, j > i \tag{3.27}$$

$$\sum_m I_{ikm} + \sum_j Q_{ikj} - \sum_m I_{imk} - S_{ik} = 0 \qquad \forall i, k \tag{3.28}$$

$$S_{ii} = (\sum_{j>i} \lambda_{ij}) y_i \qquad \forall i \tag{3.29}$$

$$Q_{ijj} = \lambda_{ij} Y_j \qquad \forall i, j > i \tag{3.30}$$

$$\sum_i (\sum_m I_{ikm} + \sum_j Q_{ikj}) = P_k^0 \qquad \forall k \tag{3.31}$$

$$\sum_k U_{ik} = \sum_{j>i} \lambda_{ij} \qquad \forall i \tag{3.32}$$

$$\sum_m T_{imj} = \lambda_{ij} \qquad \forall i, j > i \tag{3.33}$$

$$\sum_m V_{ikm} + \sum_j T_{ikj} - \sum_m V_{imk} - U_{ik} = 0 \qquad \forall i, k \tag{3.34}$$

$$U_{ii} = (\sum_{j>i} \lambda_{ij})(y_i - z_i) \qquad \forall i \tag{3.35}$$

$$T_{ijj} = \lambda_{ij}(y_j - z_j) \qquad \forall i, j > i \tag{3.36}$$

$$\sum_i (\sum_m V_{ikm} + \sum_j T_{ikj}) = P_k \qquad \forall k \tag{3.37}$$

$$Y_k \in \{0,1\} \ \forall k; \ S_{ik} \geq 0 \ \forall i,k; \ I_{ikm} \geq 0 \ \forall i,k,m; \ Q_{imj} \geq 0 \ \forall i,m,j > i; \ P_k^0 \geq 0 \ \forall k$$
$$U_{ik} \geq 0 \ \forall i,k; \ V_{ikm} \geq 0 \ \forall i,k,m; \ T_{imj} \geq 0 \ \forall i,m,j > i; \ P_k \geq 0 \ \forall k \tag{3.38}$$

Given the notation above, the constraints (3.25)-(3.27) are straightforward. Constraints (3.28) are the divergence equations. Constraints in (3.29)-(3.30) limit the number of hubs to two when i or j are hubs, respectively. Constraints in (3.31) serve the same function as (3.7) in the route-based model RoHMPC. Finally, we give the formulation of the flow-based model with the demand loss assumption (d-RoHMPCf) which is different from RoHMPCf only in second stage constraints (3.32)-(3.37).

$$\min_{\mathbf{Y},\mathbf{S},\mathbf{Q},\mathbf{I},\mathbf{P^0}} (1-\rho) \sum_i [\sum_k c_{ik} S_{ik} + \sum_k \sum_m \gamma c_{km} I_{ikm} + \sum_m \sum_{j>i} c_{mj} Q_{imj}] + (1-\rho) \sum_k a_k (P_k^0)^{b_k} +$$
$$\rho \max_{\mathbf{z} \in A} \min_{\mathbf{U},\mathbf{V},\mathbf{T},\mathbf{P}} \sum_i [\sum_k c_{ik} U_{ik} + \sum_k \sum_m \gamma c_{km} V_{ikm} + \sum_m \sum_{j>i} c_{mj} T_{imj}] + \sum_k a_k (P_k)^{b_k} \tag{3.39}$$

s.t.

$$\text{Constraints (3.25)} - (3.31)$$

$$\sum_k U_{ik} = \sum_{j>i} \lambda_{ij}(1 - z_i)(1 - z_j) \qquad \forall i \tag{3.40}$$

$$\sum_m T_{imj} = \lambda_{ij}(1 - z_i)(1 - z_j) \qquad \forall i, j > i \tag{3.41}$$

$$\sum_m V_{ikm} + \sum_j T_{ikj} - \sum_m V_{imk} - U_{ik} = 0 \qquad \forall i, k \tag{3.42}$$

$$U_{ii} = (\sum_{j>i} \lambda_{ij}) y_i (1 - z_i)(1 - z_j) \qquad \forall i \tag{3.43}$$

$$T_{ijj} = \lambda_{ij} y_j (1 - z_i)(1 - z_j) \qquad \forall i, j > i \tag{3.44}$$

$$\sum_i (\sum_m V_{ikm} + \sum_j T_{ikj}) = P_k \qquad \forall k \tag{3.45}$$

3.3 Solution Methods

3.3.1 Linearization of Congestion Cost

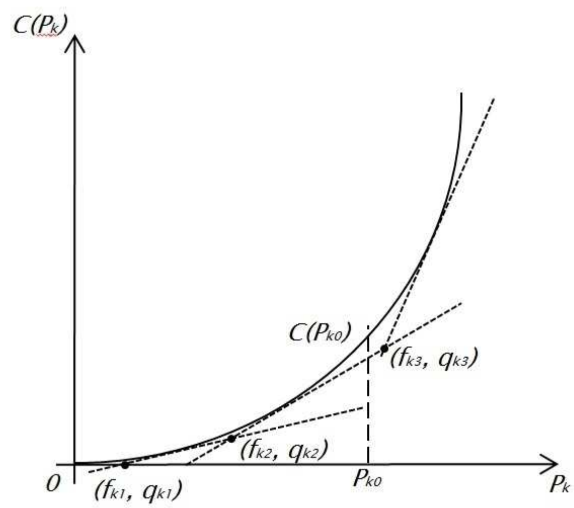


Figure 3.1: Linearization of $C(P_k)$

For illustration, we pick the model RoHMPC and take the second stage hub congestion of k for example. Nonlinear terms in other models and stages can be dealt with in a similar way. We first take nL points on the curve of $C(P_k)$, starting from the origin (0,0) and ending in $(TD, a_k(TD)^{b_k})$, where TD is the total demand flow of the network. At each point, a corresponding tangent line is drawn as shown in Figure 3.1. We denote the intersections of each pair of neighbouring tangent lines as (f_{k1}, q_{k1}), ... $(f_{k,nL-1}, q_{k,nL-1})$ and $(TD, a_k(TD)^{b_k})$ as $(f_{k,nL}, q_{k,nL})$. $C(P_k)$ is linearized by these points together with the two endpoints $(0, 0)$ and $(f_{k,nL}, q_{k,nL})$. By introducing a continuous variable between $\Lambda_{kl} \in [0, 1]$ to each hub

k and point l, the traffic flow P_{k0} is equal to $\sum_{l=0}^{nL} f_{kl}\Lambda_{kl}$ and $C(P_{k0})$ can be approximated by the convex combination $\sum_{l=0}^{nL} q_{kl}\Lambda_{kl}$. Observing that the tangent lines of sample points constitute a lower envelope of the original curve, we have $\sum_{l=0}^{nL} q_{kl}\Lambda_{kl} \leq C(P_{k0})$ but when nL is large enough $\sum_{l=0}^{nL} q_{kl}\Lambda_{kl}$ will be a good approximation of $C(P_{k0})$. In our model, we choose nL to be 30 according to our computational experience.

The linearized model for RoHMPC is obtained by replacing the objective function (3.1) with

$$\begin{aligned}&\min_{\mathbf{X},\mathbf{Y},\mathbf{P}^0}(1-\rho)(\sum_i\sum_k\sum_m\sum_{j>i}F_{ikmj}\lambda_{ij}X_{ikmj}+\sum_k\sum_l q^0_{kl}\Lambda^0_{kl}+\\&\rho\max_{\mathbf{z}\in A}\min_{\mathbf{W},\mathbf{P}}(\sum_i\sum_k\sum_m\sum_{j>i}F_{ikmj}\lambda_{ij}W_{ikmj}+\sum_k\sum_l q_{kl}\Lambda_{kl}),\end{aligned}\tag{3.46}$$

the constraints of the first stage (3.7) with

$$\sum_{l=0}^{nL}\Lambda^0_{kl}=Y_k \qquad \forall k \tag{3.47}$$

$$\sum_i\sum_{j>i}\sum_m\lambda_{ij}X_{ikmj}+\sum_i\sum_{j>i}\sum_{m\neq k}\lambda_{ij}X_{imkj}=\sum_{l=0}^{nL}f^0_{kl}\Lambda^0_{kl} \qquad \forall k, \tag{3.48}$$

and the constraints of the second stage (3.14) with

$$\sum_{l=0}^{nL}\Lambda_{kl}\geq Y_k-z_k \qquad \forall k \tag{3.49}$$

$$\sum_{l=0}^{nL}\Lambda_{kl}\leq 1 \qquad \forall k \tag{3.50}$$

$$\sum_i\sum_{j>i}\sum_m\lambda_{ij}W_{ikmj}+\sum_i\sum_{j>i}\sum_{m\neq k}\lambda_{ij}W_{imkj}=\sum_{l=0}^{nL}f_{kl}\Lambda_{kl} \qquad \forall k, \tag{3.51}$$

3.3.2 Customized Column-and-constraint Generation Algorithm

General robust models are highly difficult to solve. In Zeng and Zhao (2013), the authors propose a column-and-constraint generation algorithm which decompose the original problem into a master problem and a subproblem and fully utilize the information provided by the subproblem in each iteration. The method is shown to be effective (Hervet et al., 2013;

Zugno and Conejo, 2013). We will apply the column-and-constraint generation algorithm to seek solutions of our two-stage robust formulations. For demonstration, we will only take the linearized route-based model RoHMP for example and the flow-based models can be handled in a similar way.

An important issue of designing the customized column-and-constraint generation algorithm is to solve the subproblem in each iteration which typically involves dealing with nonlinearities. We first give the formulation of the subproblem obtained by taking the dual of the inner most "min" problem and merge it with the "max" problem in the second layer supposing hub locations are already fixed in the master problem, $\mathbf{Y} = \mathbf{Y}^*$. Note that we denote the dual variables of the constraints (3.8)-(3.13) and (3.49)-(3.51) as $\mathbf{d}^1$, $\mathbf{d}^2$, ..., $\mathbf{d}^9$, respectively. For notational convenience, we let $\mathcal{I}(\mathcal{C})$ denote a constant indicating whether the condition $\mathcal{C}$ is satisfied or not, taking value of 1 or 0, respectively. Following is the full formulation of the subproblem SubP.

$$\max_{\mathbf{z},\mathbf{d}^1,\ldots,\mathbf{d}^9} \sum_i \sum_{j>i} \sum_k (1 - z_k) d^1_{ijk} + \sum_i \sum_{j>i} \sum_k Y^*_k d^2_{ijk} + \sum_i \sum_{j>i} \lambda_{ij} d^2_{ij} + \sum_i \sum_{j>i} d^3_{ij}$$
$$+ \sum_i \sum_{j>i} ((Y^*_i - z_i) - (Y^*_j - z_j) - 1) d^4_{ij} + \sum_i \sum_{j>i} ((Y^*_i - z_i) - (Y^*_j - z_j)) d^5_{ij}$$
$$+ \sum_i \sum_{j>i} ((Y^*_j - z_j) - (Y^*_i - z_i)) d^6_{ij} + \sum_k (Y^*_k - z_k) d^7_k + \sum_k d^8_k \tag{3.52}$$

s.t.

$$d^1_{ijk} + d^2_{ijk} + d^3_{ij} + \lambda_{ij} d^9_k + \mathcal{I}(i = k, j = m) d^4_{ij} + \mathcal{I}(i = k, j \neq m) d^5_{ij} + \mathcal{I}(i \neq k, j = m) d^6_{ij} \leq F_{ikmj} \lambda_{ij}$$
$$\forall i, k, m \neq k, j > i \tag{3.53}$$

$$d^1_{ijk} + d^1_{ijm} + d^2_{ijk} + d^2_{ijm} + d^3_{ij} + \lambda_{ij} d^9_k + \lambda_{ij} d^9_m + \mathcal{I}(i = k, j = m) d^4_{ij} + \mathcal{I}(i = k, j \neq m) d^5_{ij}$$
$$+ \mathcal{I}(i \neq k, j = m) d^6_{ij} \leq F_{ikmj} \lambda_{ij} \qquad \forall i, k, m \neq k, j > i \tag{3.54}$$

$$d^7_k + d^8_k - f_{kl} d^9_k \leq q_{kl} \qquad \forall k, l \tag{3.55}$$

$$d^1_{ijk} \leq 0 \ \forall i, j > i, k; d^2_{ijk} \leq 0 \ \forall i, j > i, k; d^3_{ij} \ free \ \forall i, j > i; d^4_{ij} \geq 0 \ \forall i, j > i;$$
$$d^5_{ij} \geq 0 \ \forall i, j > i; d^6_{ij} \geq 0 \ \forall i, j > i; d^7_k \geq 0 \ \forall k; d^8_k \leq 0 \ \forall k; d^9_k \ free \ \forall k. \tag{3.56}$$

The complete procedure of column-and-constraint generation algorithm is given below.

(i) Set the lower bound $LB = -\infty$, upper bound $UB = \infty$, and iteration number $\tau = 0$.

(ii) Solve the following *master problem* (MP) and obtain an optimal solution $(\mathbf{X}^\tau, \mathbf{Y}^\tau, \mathbf{\Lambda}^{0\tau}, \eta^\tau)$ and set LB to the optimal value of the MP, which is shown below.

$$\min\ (1-\rho)\Big(\sum_i \sum_k \sum_m \sum_{j>i} F_{ikmj}\lambda_{ij}X_{ikmj} + \sum_k \sum_l q^0_{kl}\Lambda^0_{kl}\Big) + \rho\eta \tag{3.57}$$

s.t.

$$\text{Constraints } (3.2) - (3.6), (3.47) - (3.48)$$

$$\eta \geq \Big(\sum_i \sum_k \sum_m \sum_{j>i} F_{ikmj}\lambda_{ij}\hat{W}^n_{ikmj} + \sum_k \sum_l q_{kl}\hat{\Lambda}^n_{kl}\Big), \qquad \forall n = 1, 2, ..., \tau \tag{3.58}$$

$$\sum_m \hat{W}^n_{ikmj} + \sum_{m\neq k} \hat{W}^n_{imkj} \leq 1 - z^n_k \qquad \forall i, j > i, k, n = 1, 2, ..., \tau \tag{3.59}$$

$$\sum_m \hat{W}^n_{ikmj} + \sum_{m\neq k} \hat{W}^n_{imkj} \leq Y_k \qquad \forall i, j > i, k, n = 1, 2, ..., \tau \tag{3.60}$$

$$\sum_k \sum_m \hat{W}^n_{ikmj} = 1 \qquad \forall i, j > i, n = 1, 2, ..., \tau \tag{3.61}$$

$$\hat{W}^n_{iijj} \geq (Y_i - z^n_i) + (Y_j - z^n_j) - 1 \qquad \forall i, j > i, n = 1, 2, ..., \tau \tag{3.62}$$

$$\sum_{m\neq j} \hat{W}^n_{iimj} \geq (Y_i - z^n_i) - (Y_j - z^n_j) \qquad \forall i, j > i, n = 1, 2, ..., \tau \tag{3.63}$$

$$\sum_{k\neq i} \hat{W}^n_{ikjj} \geq (Y_j - z^n_j) - (Y_i - z^n_i) \qquad \forall i, j > i, n = 1, 2, ..., \tau \tag{3.64}$$

$$\sum_{l=0}^{nL} \hat{\Lambda}^n_{kl} \geq Y_k - z^n_k \qquad \forall k, n = 1, 2, ..., \tau \tag{3.65}$$

$$\sum_{l=0}^{nL} \hat{\Lambda}^n_{kl} \leq 1 \qquad \forall k, n = 1, 2, ..., \tau \tag{3.66}$$

$$\sum_i \sum_{j>i} \sum_m \lambda_{ij}\hat{W}^n_{ikmj} + \sum_i \sum_{j>i} \sum_{m\neq k} \lambda_{ij}\hat{W}^n_{imkj} = \sum_{l=0}^{nL} f_{kl}\hat{\Lambda}^n_{kl} \qquad \forall k, n = 1, 2, ..., \tau \tag{3.67}$$

$$X_{ikmj} \geq 0,\ \forall i, k, m, j > i;\ Y_j \in \{0, 1\},\ \forall j; \eta \geq 0;$$

$$\hat{W}^n_{ikmj} \geq 0,\ \forall i, k, m, j > i, n = 1, 2, ..., \tau; \hat{\Lambda}^n_{kl} \geq 0,\ \forall k, l, n = 1, 2, ..., \tau \tag{3.68}$$

(iii) Solve SubP with respect to $(\mathbf{X}^\tau, \mathbf{Y}^\tau, \mathbf{P}^{0\tau})$ and derive an optimal solution $(\mathbf{z}^\tau, \mathbf{W}^\tau, \mathbf{\Lambda}^\tau)$ and its optimal value $\mathcal{Q}^\tau$. Update

$$UB = min\{UB, (1-\rho)(\sum_i \sum_k \sum_m \sum_{j>i} F_{ikmj}\lambda_{ij}X^\tau_{ikmj} + \sum_k \sum_l q^0_{kl}\Lambda^{0\tau}_{kl}) + \rho\mathcal{Q}^n\}.$$

(iv) If $Gap = \frac{UB-LB}{LB} \leq \epsilon$, an ϵ-optimal solution is found, terminate. Otherwise, create recourse variables $(\hat{\mathbf{W}}^\tau, \hat{\mathbf{\Lambda}}^\tau)$ and corresponding constraints associated with the identified $\mathbf{z}^n$ (constraints (3.62) - (3.67)) and add them to MP. Update $\tau = \tau + 1$. Go to Step 2.

It has been theoretically proven in Zeng and Zhao (2013) that C&CG algorithm can converges in a finite number of iterations.

3.4 Computational Experiments and Analysis

As in Chapter 2, we test our formulations and algorithms on the CAB data set (O'Kelly, 1987). Because the robust models do not require probability information, we only need to determine the number of disrupted nodes Γ in set A. We consider 32 combinations of parameter settings: $|\mathbf{N}| = 10, 15, 20, 25$, $p = 3, 5, 7$, $\gamma = 0.7$, and Γ = 1,2,3. For the cases where $p = 3$, Γ will only take 1 and 2. a_k and b_k are set as 10 and 1.5, respectively, for all node k. The weight of the recourse cost ρ is set to 0.2 to represent our emphasis on the worst case scenario.

We choose 0.1% to be the optimality tolerance (ϵ). The off-the-shelf MIP solver CPLEX 12.5 is adopted to solve the MP and SubP in each iteration. Solution process will terminate after 3600 seconds or 1000 iterations.

Algorithms are implemented in C++ and all instances are tested on a Gateway laptop (Intel Dual Core , 3.2GHz, 4GB of RAM) in Windows Vista environment.

From Table 3.1, we can see that the column-and-constraint generation algorithm is very efficient in solving cases in RoHMPC and RoHMPCf. For the route-based model, the average computation time is 331.09 seconds and one case has memory issue. For the flow-based model, our method can find an optimal solution for all the cases with average computation time of 401.84 seconds. This observation confirms the effectiveness of RoHMPCf in reducing

Table 3.1: Computation of RoHMPC and RoHMPCf

$\|\mathbf{N}\|$	p	Γ	RoHMPC *Iter.*	RoHMPC *Time(s)*	RoHMPCf *Iter.*	RoHMPCf *Time(s)*	*OL*	*Oz*
10	3	1	2	5.1	2	3.2	1 4 6	6
	3	2	2	3.8	2	3.6	1 4 6	1 4
	5	1	2	4.3	2	3.4	1 4 5 7 9	9
	5	2	2	5.7	2	3.6	1 4 5 7 9	4 5
	5	3	3	7.1	3	5.2	1 4 5 8 9	4 8 9
	7	1	2	3.5	2	3.7	0 1 4 5 7 8 9	9
	7	2	2	6.2	2	4.5	0 1 4 5 7 8 9	7 9
	7	3	2	9.8	2	5.6	0 1 4 5 7 8 9	4 7 9
15	3	1	2	43.1	2	28.6	5 7 12	5
	3	2	3	21.3	3	15.8	5 10 12	10 12
	5	1	2	37.6	2	27.5	1 4 7 12 14	12
	5	2	2	33.4	2	24.7	1 4 7 12 14	4 12
	5	3	3	60.8	3	54.4	1 4 7 12 14	1 4 12
	7	1	2	28.3	2	23.2	1 4 5 7 9 12 14	12
	7	2	2	32.6	2	24.4	1 4 5 7 9 12 14	9 12
	7	3	3	103.9	3	60.1	1 4 5 7 9 12 14	7 9 12
20	3	1	3	679.3	3	641.1	1 4 10	1
	3	2	2	99.8	2	89.3	1 4 10	1 4
	5	1	2	240.2	2	213.9	1 4 12 18 19	1
	5	2	2	218.9	2	195.5	1 4 12 18 19	1 19
	5	3	3	208.7	3	187.6	1 4 5 10 19	4 5 10
	7	1	2	200.3	2	161.4	1 4 5 10 12 18 19	18
	7	2	2	160.5	2	132.3	1 4 5 10 12 18 19	1 19
	7	3	2	181.4	2	149.9	1 4 5 10 12 18 19	1 5 19
25	3	1	2	1376.3	2	1350.8	1 4 7	1
	3	2	2	409.4	2	439	1 4 10	1 4
	5	1	2	1390.8	2	1443.1	1 4 12 18 19	18
	5	2	3	923.4(M)	3	2710.7	1 4 12 18 19	12 18
	5	3	2	1135.6	2	1014.6	1 4 12 18 19	1 4 19
	7	1	2	1082.5	2	1004.4	1 4 12 14 18 19 23	18
	7	2	2	944.3	2	855.5	1 4 12 14 18 19 23	14 18
	7	3	2	952.1	3	1993.9	1 4 12 14 18 19 23	1 4 19

Table 3.2: Percentage of Single-hub Routes

γ	(a, b)	$percentage(\%)$
0.7	(10, 1.5)	80.91
	(1, 1.5)	79.05
	(1, 1.3)	47.62
0.5	(10, 1.5)	80.91
	(1, 1.5)	76.42
	(1, 1.3)	40

the model sizes. Note that we use OL and Oz to indicate the optimal locations of hubs and disrupted nodes in the worst case scenarios, respectively.

To justify our RoHMPC model and its flow-based variant, we take the case with $|\mathbf{N}| = 15$, $p = 3$, and $\Gamma = 2$ and compare the optimal network configurations obtained by RoHMPC formulations with $\rho = 0$ and $\rho = 0.2$. The former case has the optimal hubs Cleveland, Denver, and Memphis and the latter Cleveland, Kansas City, and Memphis. We observe that in the worst case scenario, the former case which does not consider hub failures will incur a 4.24% increase in the recourse cost than the latter case. If we further neglect the congestion effect, i.e., set $a_k = 0$ for all k, the optimal configuration (Atlanta, Houston, and Miami) will increase the recourse cost by 7.25% which demonstrates the necessity of considering hub failures and hub congestions.

In order to explore the contradictory effect between economies of scale and hub congestion, we choose a case with $|\mathbf{N}| = 15$, $p = 7$, and $\Gamma = 2$ and define single-hub route as the one with $X_{ikmj} > 0$ and $k = m$. Suppose that (a_k, b_k) take the value (a, b) for all k, then we calculate the percentage of single-hub routes from the total number of routes to estimate the usage of hub-and-spoke structure. Table 3.2 shows the results under different parameter settings. It is obvious that under conservative cases ($a = 10$ and $b = 1.5$) where congestion cost is high, the congestion effect dominate the economies of scale and the system tends to avoid two-hub routes since it will cause more congestion cost. This effect could be mitigated by either decreasing parameters a and b or increasing the benefit of using inter-hub links, i.e., decreasing γ.

Table 3.3: Computation of d-RoHMPC and d-RoHMPCf

$\|\mathbf{N}\|$	p	Γ	RoHMPC *Iter.*	RoHMPC *Time(s)*	RoHMPC *Gap(%)*	RoHMPCf *Iter.*	RoHMPCf *Time(s)*	RoHMPCf *Gap(%)*	*OL*	*Oz*
10	3	1	5	43.2		5	20.7		5 6 8	5
	3	2	10	252.4		10	109.8		5 6 8	5 8
	5	1	2	3.8		2	3.6		1 4 5 7 9	4
	5	2	3	19.6		3	14.4		1 4 5 7 9	1 4
	5	3	6	253.3		6	113.4		1 4 7 8 9	1 4 7
	7	1	2	3.4		2	12.6		0 1 4 5 7 8 9	4
	7	2	2	7.1		2	7.2		0 1 4 5 7 8 9	1 4
	7	3	2	22		2	17.1		0 1 4 5 7 8 9	1 4 7
15	3	1	7	T(M)	0.89	9	1408.9		0 5 7	7
	3	2	4	864.7(M)	8.19	15	3928.6(T)	3.71	6 8 14	6 14
	5	1	4	349.7		4	120.1		1 4 7 12 14	12
	5	2	7	865.2(M)	0.51	8	777.6		0 5 7 9 14	0 5
	5	3	2	1658.9(M)	5.92	9	4374.7(T)	1.62	0 5 7 9 14	0 5 14
	7	1	2	20.1		2	13.2		1 4 5 7 9 12 14	12
	7	2	2	78.2		2	34.7		1 4 5 7 9 12 14	4 12
	7	3	2	1643.4(M)	0.53	4	808.1		0 4 5 7 9 10 14	4 10 12

Finally, we investigate the computational performance under the demand loss assumption. The computational results are shown in Table 3.3. Gap information is also provided if the case can not be solved due to time or memory issues. Obviously,

3.5 Conclusion

In this chapter, we solve hub-and-spoke design problem with multiple hub failures and hub congestions. Linearization technique is first introduced to avoid directly solving the nonlinear models. Complete details of customized column-and-constraint generation method is then provided to solve the large scale cases. Flow-based models and demand loss factor are also considered in our work in this chapter. In the computational experiments, we demonstrated the effectiveness of the proposed algorithm, explore the influence of congestions on the usage of the hub-and-spoke structure, and show the computational performance of the solution method on the formulations with demand loss factors. One possible research direction in the further is to develop enhancement techniques to reduce the difficulty of considering demand loss factor in our model.

4 Reliable p-median Facility Location Problem: Two-stage Robust Models

4.1 Note to Reader

This chapter has been previously published ©2014 Elsevier. Reprinted, with permission, from Yu An, Bo Zeng, and Yu Zhang, Reliable p-median Facility Location Problem: Two-stage Robust Models and Algorithms, Transportation Research Part B: Methodological, June. 2014 [64]. The second and third author, Dr. Bo Zeng and Dr. Yu Zhang, contributed for part of literature review and the design of numerical experiments.

4.2 Introduction

The determination of facility locations and client assignments are among the most crucial issues in designing an efficient distribution network. To address these issues, various facility location models have been formulated and studied for decades, including those based on p-median and fixed-charge facility location formulations and their extensions (Daskin (1995), Drezner (1995), Revelle et al. (2008), and Melo et al. (2009)). The applications of those facility location models can be found in various industries, including manufacturing, retail, and healthcare (Barahona and Jensen (1998), Teo and Shu (2004), and Jia et al. (2007)). Although it is expected by designers that the distribution network works reliably, the system itself and/or its working environment could be seriously affected by various disruptions. For example, some facilities may be disabled by natural disasters, labor strikes, or terrorism threats. Since the material or information flows are generated, processed, and distributed by facilities, facility disruptions could significantly deteriorate the performance of the whole network and result in enormous economic losses (see the descriptions in Snyder et al. (2012) and references therein). In addition, in a disruptive situation, the whole system may need to deal with a demand pattern which is totally different from that in the normal disruption-free situation (Ergun et al. (2010)). Ignoring those issues may lead to a less reliable configuration of the distribution network that is not efficient in mitigating disruptions.

To consider disruptions in system design for better reliability, several recent studies, including Snyder and Daskin (2005), Berman et al. (2007), Chen et al. (2011), Cui et al. (2010), Li and Ouyang (2010), Lim et al. (2009), Li (2011), and Peng et al. (2011), propose to proactively consider disruptions and the incurred cost of countermeasures in the system design stage. The countermeasures, i.e., mitigation or recourse operations, are to reassign clients to survived facilities such that they can be served and the impact of disruptions can be minimized. Hence, the objective of system design is to minimize the (weighted) overall cost, including the operation cost in the normal situation when all facilities function properly, and the cost of mitigation in disruptive situations. To analytically represent this new design scheme, based on the explicit probabilistic information, several compact (nonlinear) mixed integer programs or scenario-based two-stage stochastic programming formulations are developed and customized exact or approximation algorithms are designed to solve real instances (Snyder and Daskin (2005), Chen et al. (2011), Cui et al. (2010), Li and Ouyang (2010), Lim et al. (2009), Shen et al. (2011), and Peng et al. (2011)).

Nevertheless, in many situations, either accurate method does not exist, or data are not sufficient to exactly characterize probability distributions, or data are contaminated to provide precise information. Under such situations, probabilistic models, e.g., the aforementioned two types of models, could be inappropriate or lead to infeasible solutions. To address this challenge, robust optimization (RO) method, which simply assumes an uncertainty set to capture random data, is developed to provide solutions that are robust to any perturbations within the uncertainty set. To model the situation where some decisions can be made and implemented after the uncertainty is revealed, robust optimization is extended to include the second stage recourse decisions so that the available information can be fully utilized to produce a less conservative solution. After their introduction, original robust optimization method and its two-stage extension have been applied in many operational and engineering areas (Ben-Tal et al. (2009) and Bertsimas et al. (2011)), such as facility location problems with random demands (Atamturk and Zhang (2007), Baron et al. (2011), and Gabrel et al. (2014)). In fact, comparing with demand uncertainty, disruptions are often less likely to be described by accurate probabilistic information. For example, earthquakes in California or hurricanes in Florida could cause facilities or client sites in those regions to be disrupted. However, it is very difficult to estimate the number of earthquakes or hurricanes in next 10 years based on historical/statistical data. Hence, in this chapter, we apply the concept of uncertainty set to capture the random disruptions and employ robust optimization method to study reliable facility location problems.

Specifically, we adopt two-stage robust optimization approach to investigate the reliable p-median problem, where location decisions are made before (*here-and-now*) and recourse (mitigation) decisions are made after disruptions being revealed (*wait-and-see*). We mention that such a modeling framework exactly captures the decision making sequence in real operations. In particular, due to its strong modeling capability, we are able to extend our study to consider facility capacities and demand changes due to disruptions. The former situation is very challenging for probabilistic models while the latter has not been analytically investigated in existing literature. We further implement two solution algorithms, i.e., *Benders decomposition* and *column-and-constraint generation* methods. The latter one is enhanced by a few improvement strategies based on structural properties. A set of numerical experiments are performed to generate insights on the algorithm performance and the network design.

The rest of the chapter is organized as follows. Section 4.3 reviews relevant literature on probabilistic models and two-stage robust optimization models. Section 4.4 introduces two-stage robust optimization reliable p-median models and analyzes their properties. Section 4.5 describes our solution algorithms. Numerical results and insights on system design are presented in Section 4.6, followed by Section 4.7, where the chapter is concluded and future research directions are discussed.

4.3 Literature Review

In this section, we briefly review two types of relevant studies on the facility location problem: probability based reliable facility location models and (two-stage) robust optimization formulations. Results on classical and deterministic facility location problems can be found in Daskin (1995) and Drezner (1995). For problems with uncertain demands and costs, readers are referred to a comprehensive review in Snyder (2006).

The research by Drezner (1987) is probably the first one studying facility location problem with unreliable facilities while Snyder and Daskin (2005) present the first reliable facility location models with inclusion of mitigation/recourse operations and costs. They implement Lagrangian relaxation algorithms within a branch-and-bound scheme to solve the resulting linear mixed integer programs for real instances. Chen et al. (2011) consider a combined facility location and inventory management system subject to facility failures. The authors develop an exact polynomial-time algorithm to handle the nonlinearity introduced by inventory costs and apply Lagrangian relaxation as the solution method. By relaxing the assumption that all sites share the same failure rate, Cui et al. (2010) build a nonlinear mixed integer program and develop both Lagrangian relaxation and continuum approximation (CA) methods for this challenging problem. To reduce the complexity of the nonlinear form, Lim et al. (2009) study a simplified model where clients are assigned to (unreliable) facilities and reliable backup facilities if needed. Shen et al. (2011) present both scenario-based stochastic programming and a nonlinear mixed integer programming model and show that they are generally equivalent. Also, a constant-ratio approximation algorithm for the case where all failure rates are identical is proposed. Li and Ouyang (2010) study a problem with correlated probabilistic disruptions and solve their model by CA method. Indeed, because CA technique could be useful to derive analytical insights, this approximation approach has been adopted to study the reliable facility location model in a competitive market environment by Wang and Ouyang (2013) and to investigate the effect of misestimation of (a single) failure rate on the network configuration by Lim et al. (2013). In addition to considering facility failures, Li (2011) study the problem with a fortification strategy where the unreliable facilities can be fortified by hardening operations under a budget. Recently, this line of research is extended to investigate more general reliable network design problems. Peng et al. (2011) consider a reliable multiple-echelon logistics network design problem where disruptions can happen in multiple echelons. An et al. (2011) study reliable hub-and-spoke network design problems in which hubs could be disrupted and affected flows will be rerouted through survived operational hubs. From those aforementioned studies, we observe that (i) either complicated nonlinear mixed integer programs or large-scale scenario-based stochastic programs are necessary to build the model. When professional solvers are not efficient to deal with those models, customized algorithms, either analytical or heuristic ones, are developed; (*ii*) some practical situations are not sufficiently investigated. For example, very limited research is done for capacitated models except Peng et al. (2011) and Lim et al. (2013), and no exact algorithm has been developed. Although it is noted that disruptions could cause different demand patterns (Ergun et al. (2010)), the impact of such type of demand changes has not been captured or included in the study of facility location problem.

Different from nonlinear mixed integer programs or scenario-based stochastic programs that are developed based on precise probabilistic information, robust optimization based location models, including those developed with two-stage robust optimization method, assume a probability-free uncertainty set and seek to determine locations that are robust to any perturbations in that uncertainty set. Baron et al. (2011) build a multi-period capacitated fixed charge (single-stage) robust location model and investigate the impact of different uncertainty sets on facility locations. Gülpınar et al. (2012) propose to use tractable (single-stage) robust optimization method to approximately solve stochastic facility location problem with a chance constraint. Atamturk and Zhang (2007), Gabrel et al. (2014), and Zeng and Zhao (2013) develop two-stage (tri-level) robust optimization formulations for location-transportation problems where locations and capacities are determined in the first stage and transportation decisions are adjusted after demand is realized. Different solution algorithms are proposed by them respectively, including an approximation algorithm (Atamturk and Zhang, 2007), Benders cutting plane algorithm (Gabrel et al., 2014), and the *column-and-constraint generation* (C&CG) algorithm (Zeng and Zhao, 2013). The *column-and-constraint generation* algorithm demonstrates a superior computational performance over Benders cutting plane method in the two-stage facility location problem subject to random demands.

We also note a survivable network design problem presented by Smith et al. (2007) which is formulated as a tri-level model where enemy's attack on arcs plays a role similar to site disruptions in the presented two-stage robust facility location models. It is different from our research by considering a commodity flow based network and employing Benders cutting plane as the solution method.

To make this chapter focused, we restrict this study to p-median problem and leave the study of two-stage RO formulations for another classical model, i.e., the fixed-charge facility location problem, as a future research direction. Research presented in this chapter makes the following contributions to the literature.

(i) To the best of our knowledge, no research has been done to apply two-stage RO to formulate reliable facility location problems with consideration of disruptions. Hence, this chapter presents the first set of reliable facility location formulations using two-stage robust optimization tools; (ii) because of the modeling advantages of two-stage RO, we consider real features that have received very limited or no attention. They are finite capacities of facilities and demand changes due to disruptions; (iii) in addition to some analytical study on these models, we customize and implement solution algorithms to perform numerical experiments. We also present management insights based on the numerical results from instances with real data.

4.4 Two-stage Robust p-median Reliable Models

In this section, we present our formulations on two-stage RO reliable p-median facility location problem. We first consider uncapacitated robust models and then extend our work to consider capacitated cases. Existing research generally ignores the demand changes due to disruptions. We show that, by using the two-stage robust optimization framework, demand changes can be easily incorporated. We also derive structural properties of these models.

4.4.1 Robust Uncapacitated p-median Facility Location Models

Different from stochastic programming models that explicitly consider all possible uncertain scenarios, (two-stage) robust optimization models use an uncertainty set to describe the concerned possible scenarios without depending on probability information. In the context of disruption description, we employ a cardinality constrained uncertainty set, which is probably the most used set to describe discrete uncertainties (see Atamturk and Zhang (2007) and Bertsimas and Sim (2004) for examples). Specifically, assuming that all sites in set J are homogeneous and considering all possible scenarios with up to k simultaneous disruptions, the uncertainty set, i.e., the disruption set in this chapter, can be represented as

$$A = \{\mathbf{z} \in \{0,1\}^{|J|} : \sum_{j \in J} z_j \leq k\}, \tag{4.1}$$

where z_j is the indicator variable for site j, i.e., $z_j = 1$ if site j is disrupted and $z_j = 0$ otherwise. Note that, although there may exist an exponential number of disruptive scenarios, this formulation provides an implicit but compact algebraic format to capture all of them. In the remainder of this chapter, unless explicitly mentioned, we employ this disruption set to perform our study.

Next, we develop our two-stage RO reliable p-median facility location models. Let I be the set of client sites (clients for short) and $J \subseteq I$ be the set of potential facility sites. Following the convention of previous research (Snyder and Daskin (2005), Chen et al. (2011), and Cui et al. (2010)), we assume that $I = J$. Each client $i \in I$ has a demand d_i and the unit cost of serving i by the facility at $j \in J$ is $c_{ij} \geq 0$ with $c_{ii} = 0$. We use $\mathbf{y}$ and $\mathbf{x}$ to denote the first stage (the normal situation without disruptions) decision variables: $y_j = 1$ means that a facility is located at j, $y_j = 0$ otherwise; $x_{ij} \in [0,1]$ represents the portion of i's demand served by j in the normal situation. Note that the first stage decision variables are to be fixed before any disruptive scenario $\mathbf{z}$ in set A is realized. In a disruptive scenario, as in Snyder and Daskin (2005) and Cui et al. (2010), a disrupted facility can not serve any client. However, system reliability can be achieved by implementing recourse or mitigation operations such as re-assigning clients to survived facilities. So, we introduce $\mathbf{w}$ and $\mathbf{q}$ to represent the second stage recourse operation decisions in a disruptive scenario, where $w_{ij} \in [0,1]$ represents the portion of demand d_i served by the survived facility at j and $q_i \in [0,1]$ represents the unsatisfied portion. Each unit of unsatisfied demand of d_i will incur a penalty M.

To the best of our knowledge, all existing formulations on facility location problem assume that all sites keep generating regular demands in spite of disruptions. However, under some situations, disruptions will introduce new demand patterns. On the one hand, demands of non-essential or luxury products often vanish in natural disaster-caused disruptions. On the other hand, some daily necessities and protection-based items, such as medicines and batteries, will significantly increase (Ergun et al. (2010)). To capture such phenomenon, we introduce a parameter θ to reflect the client demand change due to a disruption. Then, in a disruptive scenario, the demand of client i is set to $(1-\theta z_i)d_i$, which depends on the site disruption status z_i and θ. Clearly, by setting θ to a positive value (subject to ≤ 1) or to a negative value, we can model the disruption-caused demand reduction or increase, respectively. Hence, we consider $\theta \in (-\infty, 1]$ in the remainder of this chapter. Next, we present the two-stage RO reliable p-median facility location model (RO-PMP$_\theta$) with up to k simultaneous disruptions.

$$V_\theta(p,k,\rho) = \min_{\mathbf{x},\mathbf{y}} \ (1-\rho)\sum_i\sum_j c_{ij}d_ix_{ij} +$$

$$\rho \max_{\mathbf{z}\in A} \min_{(\mathbf{w},\mathbf{q})\in S(\mathbf{y},\mathbf{z})} \Big(\sum_i\sum_j c_{ij}(1-\theta z_i)d_iw_{ij} + \sum_i M(1-\theta z_i)d_iq_i\Big) \tag{4.2}$$

s.t.

$$x_{ij} \leq y_j, \qquad \forall i,j \tag{4.3}$$

$$\sum_j x_{ij} = 1, \qquad \forall i \tag{4.4}$$

$$\sum_j y_j = p, \tag{4.5}$$

$$x_{ij} \geq 0,\ \ \forall i,j;\ \ y_j \in \{0,1\},\ \ \forall j \tag{4.6}$$

where

$$S(\mathbf{y},\mathbf{z}) = \{w_{ij} \leq 1 - z_j, \qquad \forall i,j \tag{4.7}$$

$$w_{ij} \leq y_j, \qquad \forall i,j \tag{4.8}$$

$$\sum_j w_{ij} + q_i = 1, \qquad \forall i \tag{4.9}$$

$$w_{ij} \geq 0,\ \ \forall i,j; q_i \geq 0,\ \ \forall i.\ \} \tag{4.10}$$

The objective function in (4.2) seeks to minimize the weighted sum of the operation costs in the normal disruption-free situation and in the worst disruptive scenarios in A. The weight $\rho \in [0, 1]$ is a parameter reflecting the system designer's attitude towards the disruption cost. Clearly, a larger ρ indicates that the designer is more conservative and willing to configure the system in a way such that less recourse/mitigation operation costs will incur in disruptive situations. Constraints in (4.3)-(4.5) are from the classical p-median model and simply mean that a client can be assigned to a facility only if the facility is built, the entire demand of a client has to be served, and the total number of facilities is p, respectively.

The *max* operator identifies the disruptive scenario(s) in A yielding the largest operation cost, given the location $\mathbf{y}$. The second *min* seeks the least costly mitigation solution while the set $S(\mathbf{y}, \mathbf{z})$ defines the possible recourse operations. That is, given the definition of y_j and z_j, constraints (4.7) and (4.8) ensure that in any disruptive scenario, client i's demand can only be assigned to established *and* survived facilities. Then, constraints in (4.9) represent that the portion $(1-q_i)$ of i's demand has to be served and the rest will be lost and penalized.

In this chapter, our research focuses on the nontrivial cases where $k \leq p - 1$. Otherwise, there will be no mitigation operations in any worst disruption scenario and the problem reduces to the p-median formulation.

Note from (4.2)-(4.10) that the two-stage RO is a very adaptable modeling framework. By setting $\theta = 0$, we can compactly formulate the regular robust facility location problem without demand change, similar to existing studies on reliable facility location models (Cui et al., 2010; Peng et al., 2011; Snyder and Daskin, 2005). We can also consider the more involved situations by simply letting $\theta \neq 0$. As we mention, no existing work on the reliable facility location problem studies the impact of demand changes due to disruptions on the system design. One possible reason is that, if the demand change factor is considered, classical probabilistic models have to evaluate all possible scenarios while different scenarios will have different coefficients in their objective functions, which makes it very challenging to have a compact and tractable formulation. Nevertheless, the two-stage robust optimization scheme provides us a convenient modeling framework to address this issue. Indeed, even for the most sophisticated situation where demand changes are site dependent, our model can easily capture it by introducing site specific θ_i into its objective function (4.2). We leave research on this line as a future direction to keep our paper focused.

Although this robust formulation is a complicated tri-level optimization problem, we can analyze the impact of θ by deriving some structural properties.

Remark 4.4.1 *(i) The "closest" principle always holds when assigning demands to facilities. Specifically, in the normal disruption-free situation, demand d_i from client i is served by a facility that is closest to client i. In a disruptive situation, d_i is either served by a survived facility that is closest to i, or abandoned if the unit service cost from that facility is more than M. (ii) Let $v(\mathbf{y},\mathbf{z})$ be the optimal value of RO-PMP$_\theta$ for a given $\mathbf{y}$ and $\mathbf{z}$, including costs from both normal and disruptive situation $\mathbf{z}$. Because of (i), we have $v(\mathbf{y},\mathbf{z})$ is a linear non-increasing function with respect to $\theta \in (-\infty, 1]$. Furthermore, $v(\mathbf{y}) = \max_{\mathbf{z}} v(\mathbf{y},\mathbf{z})$ is a piecewise linear non-increasing function with respect to $\theta \in (-\infty, 1]$. Different pieces correspond to different $\mathbf{z}$. (iii) $V_\theta(p,k,\rho) = min_{\mathbf{y}} v(\mathbf{y})$ is a non-increasing quasi-convex function with respect to $\theta \in (-\infty, 1]$.*

Based on Remark 4.4.1, when $\theta \in [0,1]$, i.e., a disruption either does not affect demands or causes demand reduction, the worst case disruptions can be further characterized.

Lemma 4.4.2 *When M is sufficiently large, i.e., $M \geq \max_{i,j} c_{i,j}$, consider given facility location $\mathbf{y}^*$ and the disruption set A. If $\theta \in [0,1]$, the worst case disruptions and therefore demand reductions happen only at facility sites, i.e., those with $y_j^* = 1$.*

Proof. We prove it by contradiction. Consider a worst case disruptive situation $\mathbf{z}^1$ where a disruption happens at site j_0, on which there is no facility, i.e., $z_{j_0}^1 = 1$ and $y_{j_0}^* = 0$. Let C^1 be the operation cost under this disruptive situation.

As $p-k \geq 1$, there exists a facility, say j_1 with $y_{j_1}^* = 1$, survived in the disruptive situation $\mathbf{z}^1$. Consider two disruptive situations: $\mathbf{z}'$ where $z'_{j_0} = 0$, and $z'_j = z_j^1$ for $j \neq j_0$, and $\mathbf{z}^2$ where $z_{j_0}^2 = 0$, $z_{j_1}^2 = 1$ and $z_j^2 = z_j^1$ for $j \neq j_0$ and $j \neq j_1$. Denote the operation cost under $\mathbf{z}'$ by C', and that under $\mathbf{z}^2$ by C^2.

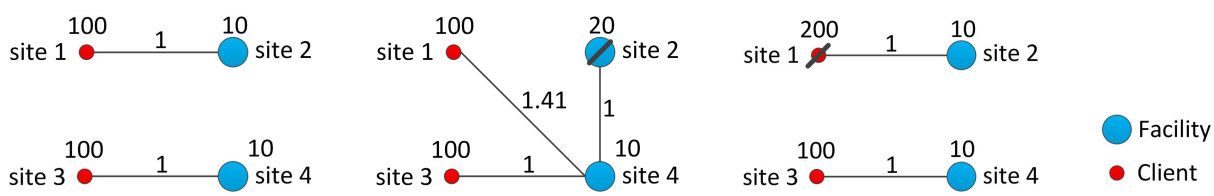
First, it is clear that $C' \geq C^1$, because demand from client j_0 must be served by some facility in $\mathbf{z}'$ while this demand is decreased in $\mathbf{z}^1$ and incurs less cost. Note that the equality could be achieved only if $\theta = 0$.

Second, under the disruptive situation $\mathbf{z}^2$, because the facility at j_1 is not available, all its served clients' demands, including the demand from j_1, will be served by other survived facilities, which will be further and more costly. Also, because $c_{j_1 j_1} = 0$, the demand from site j_1 will not incur any service cost in $\mathbf{z}'$. So, we have $C^2 \geq C'$. Note that the equality could be achieved only if $\theta = 1$.

Because $C^2 \geq C' \geq C^1$ and equalities cannot be achieved simultaneously (given θ only takes a single value), we have the desired contradiction.

■

However, the case with $\theta < 0$, i.e., a disruption will cause demand increase, is more complicated. Indeed, as demonstrated in the following example, for a given facility location solution $\mathbf{y}^*$, the worst case disruption may happen at a non-facility site.



(a) Normal Condition (b) $z_2 = 1$ (c) $z_1 = 1$

Figure 4.1: A 4-site Network with $\theta = -1$

Example 4.4.3 In the uncapacitated 4-site network ($\theta = -1$) in Figure 4.1.(a), two facilities are built on site 2 and 4. The unit service costs are symmetric and they are $c_{12} = c_{34} = c_{13} = c_{24} = 1$, and $c_{14} = c_{23} = 1.41$. The demands are $d_1 = d_3 = 100$, $d_2 = d_4 = 10$, and the penalty M is set to 15. Consider $k = 1$. We observe in Figure 4.1.(b) that a disruption at site 2 (or site 4, respectively) will disable its facility, cause a demand increase, and incur a recourse cost of 261. Nevertheless, any disruption at a client site, such as site 1 in Figure 4.1.(c), will cause a larger demand increase and incur a higher recourse cost of 300. Therefore, the worst case disruption could happen at a non-facility site.

Note from this example that we cannot properly evaluate the reliability of a distribution network by just considering facility disruptions, if demands increase along with disruptions. It further highlights the importance to adopt the presented robust p-median facility location model to analytically design the distribution network to hedge against disruptions.

Because the "closest" principle holds in both normal and disruptive situations, the following result is valid when demands must be met by existing facilities. Let $V_\theta(p,k,\rho|\mathbb{C})$ denote its optimal value under some condition $\mathbb{C}$. Then, we can evaluate the function $V_\theta(p,k,\rho)$ by its input parameters.

Lemma 4.4.4 *For a given facility location $\mathbf{y}_0$, let $C_r(\mathbf{y}_0)$ and $C_{\mathbf{z}^*}(\mathbf{y}_0)$ be the operating costs in the normal situation and a worst disruptive situation $\mathbf{z}^*$ in A, respectively. When M is sufficiently large, we have $C_{\mathbf{z}^*}(\mathbf{y}_0) \geq C_r(\mathbf{y}_0)$.*

Proposition 2 *When M is sufficiently large, the function $V_\theta(p,k,\rho)$ is (i) non-increasing with respect to p; (ii) non-decreasing with respect to k; and (iii) non-decreasing with respect to ρ.*

Proof. Statements in (i) and (ii) are straightforward. We give the proof for the statement (iii). Consider $\rho_1 \leq \rho_2$ and their corresponding optimal facility locations $\mathbf{y}_1$ and $\mathbf{y}_2$.

Clearly, as $\mathbf{y}_2$ may not be optimal when $\rho = \rho_1$, we have

$$V_\theta(p,k,\rho_1) = V_\theta(p,k,\rho_1|\mathbf{y}=\mathbf{y}_1) \leq V_\theta(p,k,\rho_1|\mathbf{y}=\mathbf{y}_2).$$

Given that $\rho_1 \leq \rho_2$, it follows from Lemma 4.4.4 that

$$V_\theta(p,k,\rho_1|\mathbf{y}=\mathbf{y}_2) \leq V_\theta(p,k,\rho_2|\mathbf{y}=\mathbf{y}_2).$$

Therefore, we have

$$V_\theta(p,k,\rho_1) = V_\theta(p,k,\rho_1|\mathbf{y}=\mathbf{y}_1) \leq V_\theta(p,k,\rho_2|\mathbf{y}=\mathbf{y}_2) = V_\theta(p,k,\rho_2).$$

■

We mention that results in Proposition 2 could be useful in algorithm design and implementation validation. For example, to deal with complicated instances with large p,k and

ρ, strong lower or upper bounds can be obtained from instances with small p, k and ρ, which are likely to be computationally simpler.

Next, we extend our study to the capacitated facility location problem, whose reliable models have received little research attention.

4.4.2 The Robust Capacitated Facility Location Model

The capacitated p-median facility location (CPMP) problem is an extension of the classical facility location model. Besides the same objective function and decision variables as in the classical uncapacitated facility location problem, it assumes that each potential facility has a capacity, i.e., an upper bound on the amount of demand that it can serve (Sridharan, 1995). Let K_j denote the capacity of site j. The two-stage robust capacitated p-median facility location problem (RO-CPMP$_\theta$) is shown as follows:

$$V_\theta^C(p,k,\rho) = \min_{\mathbf{x},\mathbf{y}} \ (1-\rho)\sum_i \sum_j c_{ij} d_i x_{ij} +$$

$$\rho \max_{\mathbf{z}\in A} \min_{(\mathbf{w},\mathbf{q})\in S^C(\mathbf{y},\mathbf{z})} (\sum_i \sum_j c_{ij}(1-\theta z_i) d_i w_{ij} + \sum_i M(1-\theta z_i) d_i q_i) \tag{4.11}$$

s.t.

$$\text{Constraints (4.3)} - \text{(4.6)}$$

$$\sum_i d_i x_{ij} \le K_j y_j, \qquad \forall j \tag{4.12}$$

with

$$S^C(\mathbf{y},\mathbf{z}) = \{(4.7) - (4.10)$$

$$\sum_i (1-\theta z_i) d_i w_{ij} \le K_j y_j, \qquad \forall j, \} \tag{4.13}$$

constraints (4.12) ensure that the total demand served by facility j does not exceed its capacity K_j. Constraints (4.13) impose the similar requirement on the survived facility j.

Because of the facility capacity constraints, RO-CPMP$_\theta$ is less trackable than the uncapacitated one. Nevertheless, under several mild assumptions, some properties can be derived.

We assume that (i) $K_j \geq d_j$ for all j, and (ii) the service costs satisfy the triangular inequality. The first assumption indicates that in both normal and disruptive situations, it is feasible to serve the whole demand of a (survived) facility site by the facility on it. The second assumption shows that it always leads to less cost to serve the demand of a (survived) facility site by the facility on it. Note that when $\theta < 0$, because demands increase along with disruptions, the total capacity from existing facilities may not be sufficient and therefore the penalty due to unmet demands will be incurred. When $\theta \in [0,1]$, a result similar to Lemma 4.4.2 can be easily derived.

Lemma 4.4.5 *Under assumptions* (i) *and* (ii)*, when* M *is sufficiently large, consider a given facility solution* $\mathbf{y}^*$ *and a disruption set* A*. We have that if* $\theta \in [0,1]$*, the worst case disruptions and therefore demand reductions happen only at facility sites, i.e., those with* $y_j^* = 1$*.*

Also, similar to Lemma 4.4.4 and Proposition 2, we derive the following results to evaluate $V_\theta^C(p,k,\rho)$ by its input parameters.

Lemma 4.4.6 *When* M *is sufficiently large, for a given facility solution* $\mathbf{y}_0$ *and a worst disruptive situation* $\mathbf{z}^*$*, we have* $C_{\mathbf{z}^*}(\mathbf{y}_0) \geq C_r(\mathbf{y}_0)$*.*

Proposition 3 *When* M *is sufficiently large, the function* $V_\theta^C(p,k,\rho)$ *is* (i) *non-increasing with respect to* p*;* (ii) *non-decreasing with respect to* k*; and* (iii) *non-decreasing with respect to* ρ*.*

4.5 Solution Algorithms

Two-stage RO models in general are very difficult to solve (Ben-Tal et al., 2004). When the second stage mitigation problem is a linear program (LP), as in each of the models we introduce so far, Benders decomposition method can be employed to seek optimal solutions (Bertsimas et al. (2013b) and Jiang et al. (2011)). However, Benders method is not efficient in dealing with real size instances. A different solution method, the *column-and-constraint generation* algorithm, denoted by C&CG algorithm, was developed in Zeng and Zhao (2013) recently, which shows a superior performance over Benders method in solving practical problems. In this chapter, we adopt C&CG method as the primary solution method to solve the proposed RO models. We first provide details of a customized C&CG method for our robust models and then present a set of enhancement strategies. We also briefly discuss the implementation of Benders decomposition method.

4.5.1 Implementation of C&CG Algorithm

We select RO-PMP$_\theta$ to describe the development of the customized C&CG algorithm. Because the capacitated robust model is of a similar structure, C&CG can be implemented with minor modifications.

C&CG algorithm is implemented within a two level master-sub problem framework. In the subproblem, for a given solution $(\mathbf{x}^*, \mathbf{y}^*)$ to the first stage decision problem, we solve the remaining *max-min* problem to identify the worst scenario. As the unsatisfied demand will be penalized in any disruptive situation, the second stage mitigation problem is always feasible. Hence, we can take the dual and obtain a *max-max* problem, which is actually a maximization problem. Specifically, let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{s}$ be the dual variables of the constraints (4.7), (4.8), and (4.9) respectively. The resulting nonlinear maximization formula of subproblem (NL-SubP) is as follows:

$$\max_{\mathbf{z},\mathbf{u},\mathbf{v},\mathbf{s}} \sum_i \sum_j (1 - z_j) u_{ij} + \sum_i \sum_j y_j^* v_{ij} + \sum_i s_i \tag{4.14}$$

s.t.

$$u_{ij} + v_{ij} + s_i \le c_{ij} d_i (1 - \theta z_i), \qquad \forall i, j \tag{4.15}$$

$$s_i \le M d_i (1 - \theta z_i), \qquad \forall i \tag{4.16}$$

$$\sum_{j \in J} z_j \le k, \tag{4.17}$$

$$u_{ij} \le 0,\ \forall i, j;\ v_{ij} \le 0,\ \forall i, j; s_i\ free,\ \forall i; z_j \in \{0, 1\},\ \forall j, \tag{4.18}$$

as the nonlinear terms are the products of a continuous variable and a binary variable, we can linearize this formulation by replacing them with a set of new variables, i.e., $U_{ij} = u_{ij} z_j$, and using *big-M* method. We denote the *big-M* as $\mathbb{M}$ to differentiate it from the penalty coefficient M.

As a result, the linearized subproblem SubP is:

$$\mathcal{Q} = \max_{\mathbf{z},\mathbf{u},\mathbf{v},\mathbf{s},\mathbf{U}} \sum_i \sum_j (u_{ij} - U_{ij} + y_j^* v_{ij}) + \sum_i s_i \tag{4.19}$$

s.t.

$$u_{ij} + v_{ij} + s_i \le c_{ij} d_i (1 - \theta z_i), \qquad \forall i, j \tag{4.20}$$

$$s_i \le M d_i (1 - \theta z_i), \qquad \forall i \tag{4.21}$$

$$U_{ij} \ge u_{ij}, \qquad \forall i, j \tag{4.22}$$

$$U_{ij} \ge -\mathbb{M} z_j, \qquad \forall i, j \tag{4.23}$$

$$U_{ij} \le u_{ij} + \mathbb{M}(1 - z_j), \qquad \forall i, j \tag{4.24}$$

$$\sum_{j \in J} z_j \le k, \tag{4.25}$$

$$u_{ij} \le 0,\ \forall i, j; U_{ij} \le 0,\ \forall i, j;\ v_{ij} \le 0,\ \forall i, j;$$

$$s_i\ free,\ \forall i;\ z_j \in \{0, 1\},\ \forall j. \tag{4.26}$$

Note that the linearized subproblem (SubP), which is a MIP problem, can be solved by a professional MIP solver. Next, we describe the details of the *column-and-constraint generation algorithm* along with the formulation of the *master problem*, which will be solved iteratively. In each iteration n, a significant scenario $\mathbf{z}^n$ will be identified through solving SubP. Then, a set of recourse variables $(\mathbf{w}^n, \mathbf{q}^n)$ and corresponding constraints in the forms of (4.28)-(4.31) associated with this particular scenario will be created and added to the *master problem*, whose complete set of variables are listed in (4.32). Let UB and LB be the upper and lower bounds respectively, Gap be the relative gap between UB and LB, n be the iteration index and ϵ be. Then, the procedures of column-and-constraint generation algorithm are given as follows.

(i) Set $LB = -\infty$, $UB = \infty$, and $n = 0$.

(ii) Solve the following *master problem* (MP) and obtain an optimal solution $(\mathbf{x}^n, \mathbf{y}^n, \eta^n)$ and set LB to the optimal value of the MP as below.

$$\min\ (1-\rho)\sum_i \sum_j c_{ij} d_i x_{ij} + \rho\eta \tag{4.27}$$

s.t.

$$\text{Constraints (4.3)} - \text{(4.6)}$$

$$\eta \geq \sum_i \sum_j c_{ij}(1-\theta z_i^l) d_i w_{ij}^l + \sum_i M(1-\theta z_i^l) d_i q_i^l, \qquad \forall l = 1, 2, ..., n \tag{4.28}$$

$$\sum_j w_{ij}^l + q_i^l = 1, \qquad \forall i, l = 1, 2, ..., n \tag{4.29}$$

$$w_{ij}^l \leq 1 - z_j^l, \qquad \forall i, j, l = 1, 2, ..., n \tag{4.30}$$

$$w_{ij}^l \leq y_j, \qquad \forall i, j, l = 1, 2, ..., n \tag{4.31}$$

$$x_{ij} \geq 0,\ \forall i, j;\ y_j \in \{0, 1\},\ \forall j; \eta\ free;$$

$$w_{ij}^l \geq 0,\ \forall i, j, l = 1, 2, ..., n; q_i^l \geq 0,\ \forall i, l = 1, 2, ..., n \tag{4.32}$$

(iii) Solve SubP with respect to $(\mathbf{x}^n, \mathbf{y}^n)$ and derive an optimal solution $(\mathbf{z}^n, \mathbf{u}^n, \mathbf{v}^n, \mathbf{s}^n)$ and its optimal value $\mathcal{Q}^n$. Update

$$UB = min\{UB, (1-\rho)\sum_i \sum_j c_{ij} d_i x_{ij}^n + \rho\mathcal{Q}^n\}.$$

(iv) If $Gap = \frac{UB-LB}{LB} \leq \epsilon$, an ϵ-optimal solution is found, terminate. Otherwise, create recourse variables $(\mathbf{w}^n, \mathbf{q}^n)$ and corresponding constraints associated with the identified $\mathbf{z}^n$ and add them to MP. Update $n = n + 1$. Go to Step 2. □

It has been proven in Zeng and Zhao (2013) that C&CG algorithm converges to an optimal solution in finite iterations. Different from C&CG method, after solving SubP, Benders decomposition method will iteratively supply a single cutting plane in the following form to its master problem that only carries the first stage decision variables $(\mathbf{x}, \mathbf{y})$

$$\eta \geq \sum_i \sum_j (1 - z_j^n) u_{ij}^n + \sum_i \sum_j v_{ij}^n y_j + \sum_i s_i^n.$$

Comparing these two types of algorithms, Zeng and Zhao (2013) theoretically show that (i) C&CG method is of a much less computational complexity. In our study, it depends on the cardinality of the disruption set. However, for Benders decomposition method, its computational complexity depends on the product of cardinality of the disruption set and the number of extreme points of the dual for the recourse problem; (ii) for C&CG method, its generated constraints are always stronger than those generated by Benders decomposition method. Because of its proven computational advantage and solution capability, C&CG method is recently employed to solve robust optimization problems in different applications (Zugno and Conejo (2013) and Hervet et al. (2013)).

4.5.2 Algorithm Enhancement

In this section, we study how to improve the computational performance of the basic C&CG method on solving reliable p-median facility location problems. In particular, note that the numbers of variables and constraints in MP will quickly increase over iterations, which may demand excessive amount of computational time for large instances. Hence, we develop a few enhancement strategies to reduce the computational expenses.

(I) Variable fixing: Variable fixing technique has been widely used within Lagrangian relaxation algorithms. It has been proven to be effective in reducing computation time for complicated network design problems (see Snyder and Daskin (2005) and Contreras et al. (2011b) for examples). Now, we extend this idea to improve C&CG method.

For any $i \in I$, we have the following results:

Proposition 4 *Assume that UB is derived from a feasible solution to RO-PMP$_\theta$.*
(i) if $V_\theta(p, k, \rho | y_i = 0) > UB$ for some i, we have $y_i = 1$ in any optimal solution; (ii) if $V_\theta(p, k, \rho | y_i = 1) > UB$ for some i, we have $y_i = 0$ in any optimal solution.

Consequently, we can implement the following *variable fixing* steps within C&CG method.

Corollary 4.5.1 *Assume that $\mathbf{y}^{\circ}$ is the best facility location solution obtained and the corresponding objective function value is UB°. The following operations can be implemented without losing any optimal solution:*
(i) if $y_i^{\circ} = 1$ and the optimal value of MP with an additional constraint $y_i = 0$ is strictly greater than UB°, fix $y_i = 1$ in MP; (ii) if $y_i^{\circ} = 0$ and the optimal value of MP with an additional constraint $y_i = 1$ is strictly greater than UB°, fix $y_i = 0$ in MP.

Note that once y_i is fixed, MP's feasible space will be reduced, which may lead to better solutions with less computational time. Although computing MP to optimality maybe difficult in practice, we can derive its lower bound within a time limit and use that lower bound to perform the aforementioned variable fixing operations; (II) multiple scenario generation: note that C&CG method generates, through solving SubP, a single scenario (i.e., its corresponding variables and constraints as in (4.28)-(4.32)) in each iteration. Because every scenario yields a valid lower bound, we actually can identify multiple significant scenarios, instead of a single optimal one. By supplying those scenarios to MP, we can further speed up the increase of the LB. Specifically, at each iteration n, given an optimal $\mathbf{z}^n$ that solves SubP, we create another disruptive scenario by modifying disrupted sites with the least demands to non-disrupted ones and changing non-disrupted sites with the largest demands to disrupted ones. Scenarios that replicate existing ones are eliminated in our implementation, although they will not affect the final solution. We observe that variable fixing and multiple scenario generation typically work better for large-scale instances. For small-scale instances, because they will either incur extra computational time on probing variables or lead to larger MP with more scenarios, they may not show as good performance as the basic method does; (III) good solutions of MP before convergence: note that MP gradually evolves into a large-scale MIP that is computationally intensive. Indeed, when *Gap* of C&CG method is large, it is not necessary to derive an optimal solution of MP and a good feasible solution could be sufficient to generate significant disruptive scenarios. So, in the beginning iterations where *Gap* is large, we can set a relatively larger optimality tolerance for a good solution to MP. As *Gap* becomes smaller, a smaller optimality tolerance will be adopted for a better solution and a more precise lower bound. Also, according to Proposition 2 and Proposition

Table 4.1: Cities in 25-site Data Set

No.	City	No.	City	No.	City	No.	City
0	Austin(TX)	7	St. Paul(MD)	14	Topeka(KS)	21	Pierre(SD)
1	Tallahassee(FL)	8	Baton Rouge(LA)	15	Charleston(WV)	22	Dover(DE)
2	Harrisburg(PA)	9	Frankfort(KY)	16	Salt Lake City(UT)	23	Washington(DC)
3	Columbus(OH)	10	Columbia(SC)	17	Lincoln(NE)	24	Montpelier(VT)
4	Richmond(VA)	11	Denver(CO)	18	Augusta(ME)		
5	Boston(MA)	12	Hartford(CT)	19	Boise City(ID)		
6	Annapolis(MD)	13	Des Moines(IA)	20	Helena(MT)		

3, we can impose bound constraint on the objective function, which could also reduce the computational time of the branch-and-bound process in solving MP.

4.6 Numerical Study and Analysis

In this section, we first describe data and experimental setup. Then, we demonstrate the results of a set of numerical experiments and present our insights on various reliable p-median models.

All of our experiments are performed on the 49-site data set described in Snyder and Daskin (2005), which includes information of demands and site coordinates. We also consider a data set of 25 sites that are randomly selected from the 49-site data set as shown in Table 4.1.

In the study, c_{ij} is the Euclidean distance between site i and j obtained from site coordinates. For capacitated models, the capacity of each site is randomly generated between $[D/10, 3D/10]$ where D is the total demand of all sites. If the capacity is smaller than its demand, we set the value of capacity equal to the demand. For all problems with 25 or 49 sites, we test them with different parameter values, i.e., θ = -1, 0, and 1, $\rho = 0.2, 0.4$, $p = 8, 10$, and $k = 1, 2$, and 3, totally 72 instances. For each instances, we also consider two cases where $M = 15$ and $M = \max_{i,j}\{c_{ij}\}$. The first value resembles a situation where an affected demand will be served by competitors if the service cost of using survived facilities is more than 15. The second value represents a situation where all demands must be served (if capacity is sufficient) in any disruptive scenario.

C&CG algorithm is our primary solution method and we apply it to solve all cases. For the comparison purpose, we also implement Benders decomposition method (BD) and benchmark it with C&CG on the RO-PMP$_1$ model with $|I| = 25$ and $M = 15$ to confirm the

Table 4.2: Computation of Benders Decomposition on The RO-PMP$_1$ (M = 15)

$\|I\|$	p	k	$\rho = 0.2$				$\rho = 0.4$			
			Time(s)	*Iter*	*Obj*	*Gap(%)*	*Time(s)*	*Iter*	*Obj*	*Gap(%)*
25	8	1	T	2321	1426.76	1.11	6994.2(m)	2308	1539.77	31.30
		2	T	2317	1525.92	7.54	T	2320	1721.44	38.53
		3	T	2350	1649.93	14.34	T	2351	1821.58	41.81
	10	1	6666.7(m)	2337	1024.11	6.61	5443.7(m)	2103	1067.20	33.19
		2	5274.8(m)	2101	1066.29	10.86	6887.9(m)	2377	1151.57	37.66
		3	6714.5(m)	2361	1119.92	14.56	6717.1(m)	2365	1292.44	44.47

efficiency of C&CG algorithm over BD method. The optimality tolerance ϵ is set as 0.1% and time limit 7200s. The master problems and the subproblems are solved by a mixed integer programming solver, CPLEX 12.5, at its default settings. All algorithms are implemented in C++ and tested on a Dell Optiplex 760 desktop computer (Intel Core 2 Duo CPU, 3.0GHz, 3.25GB of RAM) in Windows XP environment.

4.6.1 Algorithm Performance

Table 4.2 presents the performance of BD methods on instances of RO-PMP$_1$. Table 4.3-4.6 summarize the computational results of C&CG algorithm on the uncapacitated and capacitated reliable models with different M values. In those tables, the column *Time(s)* presents the computational time in seconds; the column *Iter* indicates the number of iterations; the column *Obj* shows the best objective value ever found; the column *Gap*(%) provides the relative gap in percentage if it is larger than ϵ. We use letter T in *Time(s)* column to indicate an instance which can not be solved within the time limit. If the algorithm terminates because the memory is not sufficient for the solver to compute MP or SubP, we add "(m)" after the computation time in column *Time(s)*.

Based on these tables, we observe that (i) C&CG algorithm performs hundreds of times faster and takes much fewer iterations than the classical Benders decomposition method. This result confirms the observations made in Zeng and Zhao (2013) for the location-transportation network design problem. Actually, compared to results in Zeng and Zhao (2013), a more significant superiority of the enhanced C&CG algorithm is observed in solving reliable p-median problems. To further demonstrate the computational advantage of C&CG algorithm, Figure 4.2 shows the convergence plots of two algorithms under the time limit of 15 seconds for a case ($\theta = 1$, M=15, $|I| = 25, \rho = 0.2$, $p = 8,$ and $k = 3$). Obvi-

Table 4.3: Computation of Uncapacitated Instances with $M = 15$

$\|I\|$	ρ	p	k	RO-PMP$_{-1}$				RO-PMP$_0$				RO-PMP$_1$			
				Time(s)	*Iter*	*Obj*	*Gap(%)*	*Time(s)*	*Iter*	*Obj*	*Gap(%)*	*Time(s)*	*Iter*	*Obj*	*Gap(%)*
25	0.2	8	1	0.8	2	1763.95		1.9	5	1558.09		1.9	5	1426.76	
			2	0.8	3	2233.97		2.8	6	1855.51		3.2	7	1525.92	
			3	40.2	19	2846.98		818.8	59	2292.82		11.3	12	1649.93	
		10	1	0.7	2	1364.19		0.7	2	1139.08		3.4	7	1024.11	
			2	2.1	6	1759.77		2.0	6	1374.09		5.9	9	1066.29	
			3	3.5	6	2088.41		6.4	10	1601.93		4.9	8	1119.92	
	0.4	8	1	0.7	2	2214.17		2.9	5	1802.45		2.0	7	1539.77	
			2	1.4	4	3086.90		16.3	14	2335.00		18.0	14	1721.44	
			3	184.4	32	4126.34		6627.6	97	3136.73		40.1	18	1821.58	
		10	1	0.7	2	1814.40		1.8	2	1364.19		3.2	7	1067.20	
			2	2.6	5	2503.45		3.9	7	1785.45		9.4	11	1151.57	
			3	4.3	8	3139.28		51.2	23	2235.93		25.8	15	1292.44	
49	0.2	8	1	4.8	4	6563.12		3.1	4	6145.54		2.2	3	5684.45	
			2	732.7	21	7502.22		48.2	10	6808.88		96.9	12	6020.47	
			3	T	32	8457.28	6.68	T	35	7426.71	5.64	6058.7	38	6364.45	
		10	1	2.1	3	5339.88		0.7	2	4950.32		9.1	6	4687.15	
			2	164.4	13	6090.26		126.3	12	5534.76		160.5	14	4982.95	
			3	T	30	7148.76	5.89	T	31	6111.83	4.27	737.5	20	5131.20	
	0.4	8	1	44.5	9	7405.65		4.5	4	6625.50		14.3	7	5898.97	
			2	T	37	8892.50	6.35	4208.4	30	8027.37		T	36	6528.61	0.23
			3	T	32	10625.10	15.87	T	28	9198.43	19.56	T	31	7215.81	8.11
		10	1	7.2	5	5988.67		3.7	4	5351.47		12.4	7	4796.70	
			2	T	27	7402.96	1.16	7165.2	31	6365.19		2153.1	24	5264.33	
			3	T	29	9337.32	17.71	T	28	7467.20	12.46	T	31	5697.75	4.52

Table 4.4: Computation of Uncapacitated Instances with $M = \max_{i,j}\{c_{ij}\}$

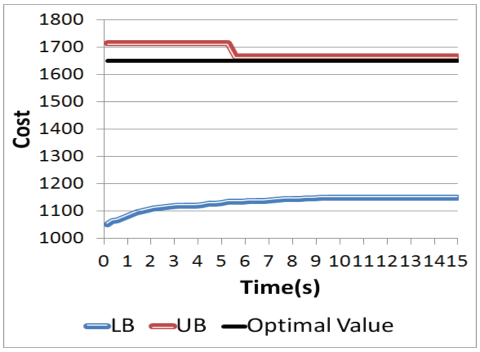
$\|I\|$	ρ	p	k	RO-PMP$_{-1}$				RO-PMP$_0$				RO-PMP$_1$			
				Time(s)	*Iter*	*Obj*	*Gap(%)*	*Time(s)*	*Iter*	*Obj*	*Gap(%)*	*Time(s)*	*Iter*	*Obj*	*Gap(%)*
25	0.2	8	1	1.9	2	1763.95		1.8	5	1558.09		1.3	5	1426.76	
			2	0.6	3	2233.97		2.3	6	1855.51		3.7	7	1525.92	
			3	177.6	32	2914.07		2287.6	71	2396.15		39.4	19	1734.64	
		10	1	1.8	2	1364.19		0.8	2	1139.08		6.6	9	1026.99	
			2	3.5	6	1759.77		2.7	6	1374.09		7.2	10	1083.22	
			3	3.8	6	2088.41		14.5	12	1616.81		5.6	8	1119.92	
	0.4	8	1	0.6	2	2214.17		2.8	5	1802.45		3.8	7	1539.77	
			2	1.4	4	3086.90		18.8	14	2335.00		25.1	15	1738.11	
			3	545.6	46	4197.12		T	88	3228.20	3.03	258.4	33	2010.03	
		10	1	1.7	2	1814.40		0.7	2	1364.19		7.7	10	1094.38	
			2	2.9	5	2503.45		4.6	7	1785.45		16.7	13	1183.81	
			3	6.0	8	3139.28		79.8	26	2252.60		35.1	17	1292.44	
49	0.2	8	1	5.5	4	6567.18		4.7	4	6149.60		1.8	3	5686.38	
			2	T	30	8132.07	3.18	T	35	7599.45	6.43	77.7	10	6026.61	
			3	T	29	9561.88	13.09	T	32	8329.70	11.02	5518.2	38	6432.67	
		10	1	2.5	3	5339.88		1.9	2	4950.32		10.3	6	4687.15	
			2	192.5	13	6090.26		138.1	12	5534.76		261.2	16	4992.81	
			3	T	32	7288.48	8.28	T	36	6520.93	11.16	1117.1	22	5131.20	
	0.4	8	1	55.1	9	7405.65		5.7	4	6633.61		5.8	4	5902.83	
			2	T	29	9974.33	10.97	T	34	9294.07	15.56	5316.4	30	6583.29	
			3	T	25	12920.90	25.70	T	27	10829.30	23.77	T	31	7384.12	9.61
		10	1	10.5	5	5988.67		4.7	4	5351.47		14.6	7	4796.70	
			2	T	26	7330.81	0.91	7112.8	31	6365.19		T	35	5316.31	0.13
			3	T	29	9673.43	19.77	T	32	8384.91	23.48	T	30	5792.46	7.16

Table 4.5: Computation of Capacitated Instances with $M = 15$

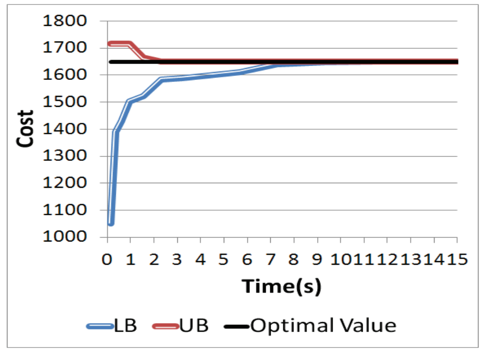
$\lvert I\rvert$	ρ	p	k	RO-CPMP$_{-1}$ $Time(s)$	$Iter$	Obj	$Gap(\%)$	RO-CPMP$_0$ $Time(s)$	$Iter$	Obj	$Gap(\%)$	RO-CPMP$_1$ $Time(s)$	$Iter$	Obj	$Gap(\%)$
25	0.2	8	1	0.9	2	2069.34		1.7	3	1783.39		2.8	4	1550.58	
			2	2.2	3	2686.70		4.8	6	2126.51		7.5	7	1638.56	
			3	4.6	5	3370.00		10.4	9	2466.17		6.7	7	1725.95	
		10	1	0.5	2	1474.23		0.6	2	1148.48		1.4	4	1036.93	
			2	1.2	2	1923.53		1.2	3	1383.48		4.6	6	1094.58	
			3	2.5	3	2405.02		2.8	3	1642.20		7.7	8	1180.54	
	0.4	8	1	1.3	2	2620.20		1.1	3	2024.04		3.8	5	1623.29	
			2	5.3	5	3809.90		25.9	11	2791.94		11.5	9	1799.26	
			3	40.6	13	5021.64		543.0	33	3495.95		79.9	18	2015.52	
		10	1	1.4	2	1981.33		0.7	2	1373.58		4.3	6	1083.45	
			2	1.8	2	2869.74		1.7	3	1843.59		12.1	10	1224.17	
			3	3.9	4	3832.70		14.8	10	2307.08		181.7	25	1403.52	
49	0.2	8	1	20.3	5	6946.01		13.2	4	6432.77		24.8	5	6180.25	
			2	7150.5	24	7970.07		T	26	7290.78	6.30	1165.5	19	6525.89	
			3	T	21	8977.90	8.49	T	21	8095.26	12.93	7173.6	33	6937.76	
		10	1	11.4	4	5616.50		7.6	4	5045.10		19.4	5	4778.10	
			2	3968.1	22	6493.59		3335.1	26	5752.22		442.7	15	5097.14	
			3	T	20	7361.09	8.55	T	24	6187.69	3.65	3476.3	23	5250.64	
	0.4	8	1	164.6	9	7730.58		28.3	6	6939.44		87.2	9	6426.95	
			2	T	22	9591.76	12.70	T	22	8582.31	15.45	T	25	7113.72	2.20
			3	T	16	11616.60	26.69	T	19	10144.40	25.84	T	22	7937.46	11.82
		10	1	111.7	8	6248.44		23.8	6	5434.73		39.1	7	4881.02	
			2	T	21	7838.19	6.14	T	28	6888.70	8.38	T	27	5552.00	1.79
			3	T	17	9496.51	18.43	T	22	7753.12	12.64	T	23	5885.55	6.35

Table 4.6: Computation of Capacitated Instances with $M = \max_{i,j}\{c_{ij}\}$

$\lvert I\rvert$	ρ	p	k	RO-CPMP$_{-1}$ $Time(s)$	$Iter$	Obj	$Gap(\%)$	RO-CPMP$_0$ $Time(s)$	$Iter$	Obj	$Gap(\%)$	RO-CPMP$_1$ $Time(s)$	$Iter$	Obj	$Gap(\%)$
25	0.2	8	1	2.1	2	2069.34		2.5	3	1783.39		3.2	5	1550.58	
			2	9.8	7	2729.23		5.7	6	2126.51		23.7	12	1752.15	
			3	27.4	9	3944.68		243.4	20	2772.17		83.6	17	1880.17	
		10	1	1.1	2	1474.23		0.8	2	1148.48		4.3	6	1037.56	
			2	2.6	2	1923.53		2.4	3	1383.48		5.8	6	1094.58	
			3	8.5	5	2486.07		11.2	8	1726.22		24.7	12	1226.52	
	0.4	8	1	1.7	2	2620.20		2.7	3	2024.04		5.1	6	1656.87	
			2	22.2	9	3884.64		38.1	12	2816.64		243.3	27	2024.80	
			3	440.8	21	5948.72		2387.6	37	3863.25		981.5	38	2263.44	
		10	1	1.6	2	1981.33		0.6	2	1373.58		9.1	8	1103.78	
			2	2.4	2	2869.74		2.2	3	1843.59		14.5	10	1224.17	
			3	12.7	7	3920.41		91.3	17	2425.86		1248.9	46	1520.56	
49	0.2	8	1	28.8	5	7008.87		82.7	8	6517.80		57.7	7	6197.10	
			2	T	19	8953.19	12.73	T	24	8175.68	6.71	5238.2	24	6585.34	
			3	T	17	11586.60	32.24	T	21	10005.10	30.47	T	22	7963.78	15.91
		10	1	16.5	4	5616.50		8.3	4	5045.10		33.4	6	4794.66	
			2	T	21	6671.10	3.45	T	29	5949.65	3.65	473.9	15	5097.14	
			3	T	20	8396.78	22.37	T	21	7215.95	18.27	T	26	5292.58	1.01
	0.4	8	1	414.1	10	7731.58		133.6	9	7086.20		276.4	11	6475.30	
			2	T	18	11470.30	24.19	T	21	10205.30	19.71	T	23	7232.63	5.55
			3	T	17	17546.90	49.89	T	19	13976.10	47.42	T	20	10561.40	34.83
		10	1	117.6	8	6248.44		26.4	6	5434.73		59.6	8	4914.15	
			2	T	20	8153.59	9.54	T	28	7209.75	14.43	T	27	5552.00	1.79
			3	T	19	10833.00	30.05	T	23	9732.88	34.73	T	22	5969.43	9.35



(a) Benders Decomposition Method



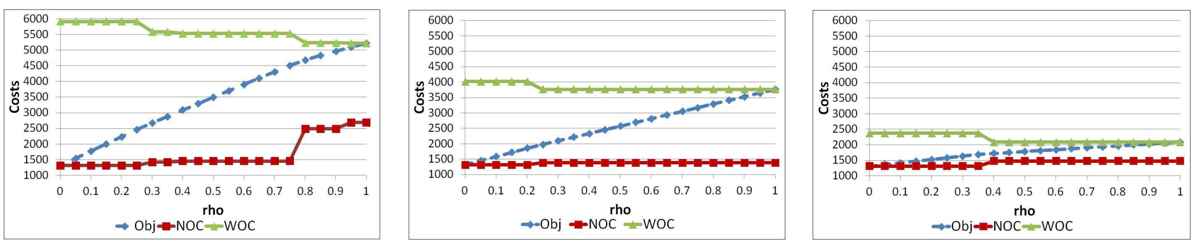
(b) C&CG Method

Figure 4.2: Convergence Plots within 15 Seconds

ously, these two methods present distinct patterns in Figure 4.2. In Figure 4.2(a) Benders decomposition method can not reduce the gap between upper bound and lower bound. In particular, it fails to improve lower bound. In Figure 4.2(b), however, upper and lower bounds of C&CG method quickly converge to the optimal value in a short time; (ii) the computation complexity of C&CG algorithm increases with the problem size $|I|$ and k, as well as the weight coefficient ρ. In all four tables, the most challenging instances are those with largest $|I|$, k, and ρ. Note that all instances with $k = 1$ are easy to compute. Most small size instances with $|I| = 25$ can be solved to optimality or with a small optimality gap while some instances with $|I| = 49$ are difficult. A closer analysis shows that SubP is easier to compute and the actual bottleneck is to solve MP, which will grow into a large MIP problem over iterations. As CPLEX, a general-purpose MIP solver, is currently called to solve MP, one possible direction of future research is to develop a specialized algorithm that takes advantage of the structure of MP for a faster computation; (iii) Including additional features does not incur significant computational expense. Compared with models without capacity restrictions or demand changes, capacitated ones are slightly harder while models with demand changes could be easier. Hence, our two-stage RO formulations of reliable p-median problems are computationally robust to additional features or restrictions; (iv) although for many instances the optimal objective values are the same for the different M values, the large penalty coefficient M generally negatively impacts the computational performance, which is more significant for instances in capacitated models. One explanation is that large penalty coefficient M forces demand that was served by a disrupted facility to be served by survived ones, instead of being simply treated as unmet demand. As a result, the optimization complexity increases.

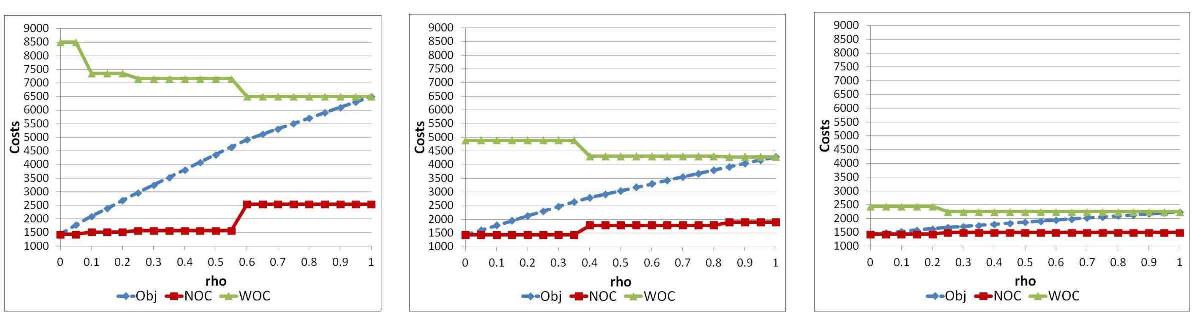
4.6.2 Impact of the Reliability

In this section, we investigate the effect of including the worst disruptive scenarios on the system configuration and operations. Specifically, for different ρ values, after deriving an optimal solution, we compute the corresponding normal operation cost (NOC) and worst case operation cost (WOC). Then, we plot those operation costs with respect to ρ. It is obvious that the classical p-median model can be obtained by setting ρ to 0 and the formulation to minimize only the worst case cost can be obtained by setting ρ to 1. Figure 4.3 and Figure 4.4 present our results for 25-site models with M = 15. Note that the weighted objective function values are also included.



(a) Cost Curves of RO-PMP$_{-1}$ (b) Cost Curves of RO-PMP$_0$ (c) Cost Curves of RO-PMP$_1$

Figure 4.3: Effect of ρ on The Robust Uncapacitated Facility Location Models



(a) Cost Curves of RO-CPMP$_{-1}$ (b) Cost Curves of RO-CPMP$_0$ (c) Cost Curves of RO-CPMP$_1$

Figure 4.4: Effect of ρ on The Robust Capacitated Facility Location Models

Clearly, the two cost functions demonstrate a monotone property, or a "staircase pattern", over ρ. Within a single stair the optimal system configurations are the same for different ρ values. Actually, the small number of stairs implies that the optimal system configuration based on two-stage RO is not very sensitive to different ρ. However, when ρ keeps increasing, WOC would decrease while NOC would increase. Sometimes a slight increase in NOC will lead to a significant decrease in WOC. Such a phenomenon is also observed in the stochastic programming based reliable facility location models (Snyder and Daskin, 2005). Overall, a desired trade-off between NOC and WOC can be achieved by selecting a configuration of an appropriate stair. Another observation is that the WOCs and hence the objective values of the models with demand changes are quite different from those of the classical models ($\theta = 0$). It is reasonable since, for example, in RO-PMP$_1$/RO-CPMP$_1$, under a disruptive scenario, the demands of disrupted clients will disappear and will not incur any cost, which counteracts the cost increase due to failed facilities.

4.6.3 Comparison of Stochastic Programming and RO Models

As mentioned in Section 4.3, stochastic programming (SP) has been used as the primary tool to develop models and corresponding algorithms to study the reliable facility location problems. To evaluate the solution quality of two-stage RO models, we compare the solutions from our model RO-PMP$_0$ and the SP model presented in Snyder and Daskin (2005) using the same data set and penalty coefficient (i.e., the uncapacitated instance with $|I| = 25$, $M = 15$ and $\theta = 0$). For the RO model, we first set the weight coefficient $\rho = 0.2$ and $1 - \rho = 0.8$ for the worst case and normal situation operation costs, respectively. We set the failure probability of each site to 0.01 in the SP model. It renders the probability of the normal disruption-free situation to 0.78, which roughly matches the weight of normal situation in the RO model. We compare two measures, i.e., NOC and WOC, for solutions of both models to evaluate their performances. For the SP model, WOC^{SP} is obtained by inserting its optimal facility locations to the RO model. Numerical results are provided in Table 4.7. We also consider a more conservative situation in which ρ is set to 0.4 in RO.

We note in Table 4.7 that, when $\rho = 0.2$, the qualities of SP and RO solutions are basically close in normal situations. An SP solution might have an equal or a little bit less NOC, while

Table 4.7: Comparison of SP and RO Models

ρ	p	k	SP		RO			
			NOC^{SP}	WOC^{SP}	NOC^{RO}	WOC^{RO}	$\Delta_{NOC}(\%)=\frac{NOC^{SP}-NOC^{RO}}{NOC^{RO}}$	$\Delta_{WOC}(\%)=\frac{WOC^{SP}-WOC^{RO}}{WOC^{RO}}$
0.2	8	2	1313.74	4022.6	1313.74	4022.6	0.00	0.00
		3	1313.74	6732.09	1417.77	5793.01	-7.34	16.21
	10	2	913.976	3214.53	913.976	3214.53	0.00	0.00
		3	913.976	4651.45	967.934	4137.91	-5.57	12.41
0.4	8	2	1313.74	4022.6	1380.83	3766.26	-4.86	6.81
		3	1313.74	6732.09	1546.41	5522.21	-15.05	21.91
	10	2	913.976	3214.53	981.016	2992.1	-6.83	7.43
		3	913.976	4651.45	967.934	4137.91	-5.57	12.41

its WOC could be significantly more, compared to an RO solution. When $\rho = 0.4$, because more consideration is placed on the worst case performance in RO, RO solutions might not be in favor of NOC while they have less WOC. Consequently, we observe more clearer differences in NOC and WOC between SP and RO solutions among all cases. Nevertheless, the difference in NOC for these two models is not drastic. Those results suggest that, SP and two-stage RO models are comparable, even when a relatively large weight is assigned to WOC in RO. Hence, our presented two-stage RO models are not overly conservative. Those results also indicate that, instead of relying on accurate probabilistic information to build an SP model, two-stage RO provides a dependable modeling alternative for practical usage that requires much less information support.

4.6.4 Effect of Demand Changes

As we mentioned earlier, demand changes due to disruptions have not been included or investigated in any existing reliable facility location models. So, it remains unknown that how demand changes will affect system design and operations, or how approximate the results we have if we ignore the demand change factor when it does exist. To explore the impact of demand changes, Table 4.8-4.9 present optimal configurations of the models with $\theta = -1, 1$, and 0 (M=15). The column OL represents the optimal locations of facilities. We then insert the optimal locations derived from $RO\text{-}PMP_0/RO\text{-}CPMP_0$ into $RO\text{-}PMP_{-1}$ and $RO\text{-}PMP_1/RO\text{-}CPMP_{-1}$ and $RO\text{-}CPMP_1$, and compute the associated operation costs in normal and the worst disruptive situations. The relative changes of NOCs and WOCs for RO-PMP$_{-1}$ model with respect to its own optimal NOCs and WOCs are presented in the column $\Delta^{-1}_{NOC}(\%)$ and $\Delta^{-1}_{WOC}(\%)$, respectively. Similar results for RO-PMP$_1$ are in columns

Table 4.8: Comparison between RO-PMP$_{-1}$, RO-PMP$_1$, and RO-PMP$_0$

ρ	p	k	RO-PMP$_{-1}$ OL	RO-PMP$_1$ OL	RO-PMP$_0$ OL	$\Delta^{-1}_{NOC}(\%)$	$\Delta^{-1}_{WOC}(\%)$	$\Delta^{1}_{NOC}(\%)$	$\Delta^{1}_{WOC}(\%)$
0.2	8	1	0 1 2 3 5 8 11 13	0 1 2 3 5 8 11 13	0 1 2 3 5 8 11 13	0.00	0.00	0.00	0.00
		2	0 1 2 3 5 8 11 13	0 1 2 3 5 8 11 13	0 1 2 3 5 8 11 13	0.00	0.00	0.00	0.00
		3	0 1 3 5 8 11 13 23	0 1 2 3 4 5 7 11	0 1 2 3 4 5 11 13	5.56	9.54	-4.09	13.82
	10	1	0 1 2 3 4 5 8 11 13 16	0 1 2 3 4 5 7 8 11 14	0 1 2 3 4 5 8 11 13 16	0.00	0.00	-6.83	23.62
		2	0 1 3 5 6 7 8 10 11 14	0 1 2 3 4 5 7 8 11 14	0 1 2 3 4 5 8 11 13 16	-9.01	15.36	-6.83	25.35
		3	0 1 3 5 7 8 10 11 14 23	0 1 2 3 4 5 8 11 13 16	0 1 2 3 4 5 8 10 11 13	-4.53	12.49	5.90	9.69
0.4	8	1	0 1 2 3 5 8 11 13	0 1 2 3 5 8 11 13	0 1 2 3 5 8 11 13	0.00	0.00	0.00	0.00
		2	0 1 3 5 6 8 11 14	0 1 2 3 4 5 7 11	0 1 2 3 8 11 12 13	-5.35	8.19	-6.59	29.26
		3	0 1 8 9 11 12 14 23	0 1 2 3 4 5 7 11	0 1 3 5 6 10 11 17	-12.20	10.87	4.61	27.68
	10	1	0 1 2 3 4 5 8 11 13 16	0 1 2 3 4 5 7 8 11 14	0 1 2 3 4 5 8 11 13 16	0.00	0.00	-6.83	23.62
		2	0 1 3 5 7 8 10 11 14 23	0 1 2 3 4 5 7 8 11 14	0 1 2 3 4 5 7 8 11 14	-3.24	5.60	0.00	0.00
		3	0 1 3 7 8 10 11 12 14 23	0 1 2 3 4 5 7 8 11 16	0 1 2 3 4 5 8 10 11 13	-10.45	15.38	-0.67	20.50

Table 4.9: Comparison between RO-CPMP$_{-1}$, RO-CPMP$_1$, and RO-CPMP$_0$

ρ	p	k	RO-CPMP$_{-1}$ OL	RO-CPMP$_1$ OL	RO-CPMP$_0$ OL	$\Delta^{-1}_{NOC}(\%)$	$\Delta^{-1}_{WOC}(\%)$	$\Delta^{1}_{NOC}(\%)$	$\Delta^{1}_{WOC}(\%)$
0.2	8	1	0 1 3 4 5 8 11 13	0 1 2 3 4 5 11 13	0 1 3 4 5 8 11 13	0.00	0.00	5.72	7.73
		2	0 1 3 4 5 8 11 13	0 1 2 3 4 5 11 13	0 1 2 3 4 5 11 13	-5.41	15.53	0.00	0.00
		3	0 1 3 4 8 11 12 13	0 1 2 3 4 5 11 13	0 1 2 3 4 5 11 13	-8.73	15.37	0.00	0.00
	10	1	0 1 2 3 4 5 8 11 13 16	0 1 2 3 4 5 7 8 11 14	0 1 2 3 4 5 8 11 13 16	0.00	0.00	-6.77	22.18
		2	0 1 2 3 4 5 8 10 11 13	0 1 2 3 4 5 8 11 13 16	0 1 2 3 4 5 8 11 13 16	-5.52	9.77	0.00	0.00
		3	0 1 2 3 4 5 8 10 11 13	0 1 2 3 4 5 8 11 13 16	0 1 2 3 4 5 8 10 11 13	0.00	0.00	5.84	-1.14
0.4	8	1	0 1 3 4 5 8 11 13	0 1 2 3 4 5 7 11	0 1 3 4 8 11 12 13	3.64	1.30	5.14	28.70
		2	0 1 3 4 8 11 12 13	0 1 2 3 4 5 7 11	0 1 2 3 4 5 8 17	13.15	1.62	18.97	34.97
		3	0 1 4 8 9 11 12 13	0 1 2 3 4 5 11 13	0 1 2 3 4 5 11 13	-19.30	23.20	0.00	0.00
	10	1	0 1 2 3 4 5 7 8 11 14	0 1 2 3 4 5 7 8 11 14	0 1 2 3 4 5 8 11 13 16	-6.77	6.05	-6.77	22.19
		2	0 1 2 3 4 5 8 10 11 13	0 1 2 3 4 5 7 8 11 16	0 1 2 3 4 5 8 11 13 16	-5.52	9.77	-6.15	12.30
		3	0 1 2 3 4 5 8 10 11 13	0 1 2 3 4 5 7 8 10 11	0 1 2 3 4 5 8 10 11 13	0.00	0.00	-5.83	11.88

$\Delta^1_{NOC}(\%)$ and $\Delta^1_{WOC}(\%)$. For a case, if the optimal facility locations are different among the three models, we will highlight the locations by underlines.

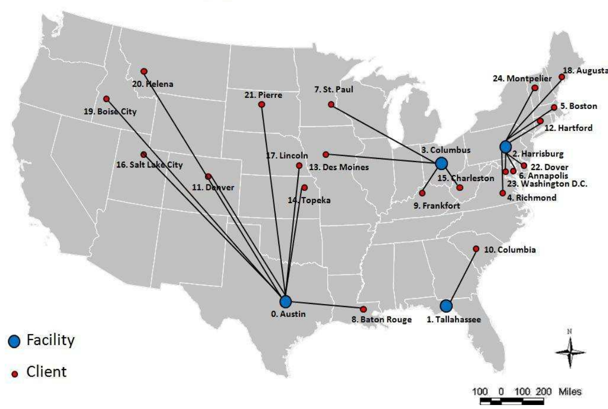
We observe that the demand change plays a significant role in determining system configuration. In all 24 instances, including uncapacitated and capacitated ones, there are 21 instances whose optimal facility locations are different for different θ values. In fact, when we put more weight on the worst disruptive situations or consider facility capacities, the impact of demand changes becomes more significant. For example, when $\rho = 0.4$, there are 11 instances (out of 12) on which those models yield different solutions. Furthermore, the different facility locations present different performances in both normal and the worst disruptive situations. From the columns $\Delta^{-1}_{NOC}(\%)$ and $\Delta^1_{NOC}(\%)$, we note that the system configuration derived without demand changes could incur more or less cost in the normal situation, which can hardly be predicted beforehand. In fact, the difference can be as significant as -19.30% or 18.97%, definitely a non-trivial value. Nevertheless, in the worst disruptive situation, it is generally observed that the system configuration derived without demand changes will incur much more operation cost, which can be up to 34.97% in a capacitated instance. Therefore, we can conclude that the demand change factor, if it exists in the practice, should not be ignored in system design, especially when the weight coefficient ρ is large and capacity needs to be considered.

To develop insights on system configuration, we plot optimal facility locations and normal situation assignments of a small-scale instance in Figure 4.5, where $p = 4$, $k = 1$ and $|I| = 25$ with $\rho = 0.2$ and $M = max_{i,j}\{c_{ij}\}$. We also include a relatively extreme case where $\theta = -5$, which indicates a four-time demand increase of a disrupted site.

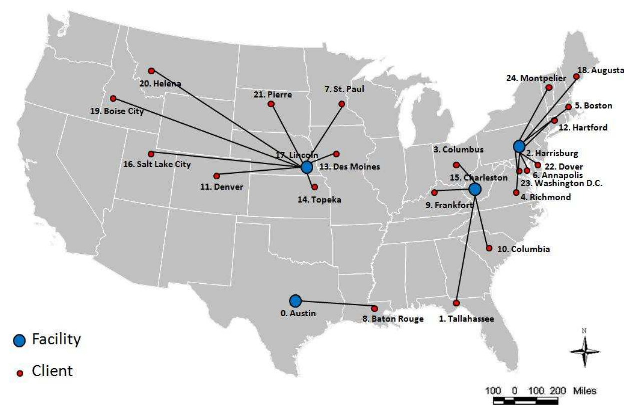
We mention that sites are numbered according to the descending order of demand quantities. In Figure 4.5, we observe a clear trend on selecting facility locations. When $\theta = 1$, four facilities are constructed on the sites of the smallest indices, i.e., those with largest demands. With θ becoming smaller, facilities are generally built on sites with smaller demands. One explanation is that: on the one hand, with θ becomes smaller, more demands should be served in disruptive situations; on the other hand, if sites with large demands are selected for facilities, although it saves cost in the normal situation, disruptions on those facility sites will incur very high service costs as demands will be served by facilities far away. Therefore, if demands increase under disruptions, to balance the cost and the risk, facilities should be built on sites with smaller demands to avoid the large cost in the worst scenarios.

4.6.5 A Correlated Disruption Set

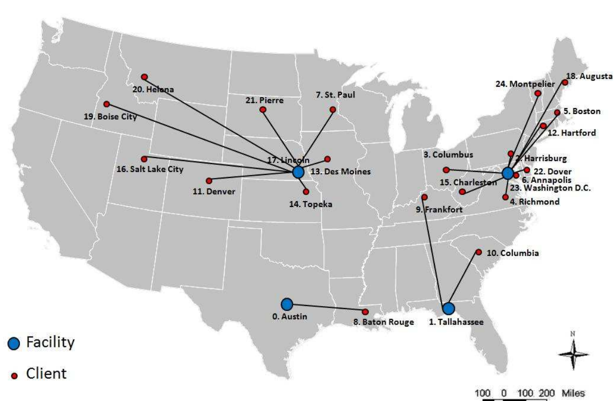
The disruption set with a simple cardinality restriction in Equation (4.1) indicates that all sites are of the identical failure possibility and there is little correlation among them. In this section, given the adaptability of our modeling framework, we investigate a different disruption set that carries some correlations. Specifically, we partition J into a few subsets and assume that sites in each subset are temporally or spatially correlated. Hence, the number of disruptions in each subset can be better estimated. Also, we have an overall budget constraint to restrict the total number of disruptions. The disruption set takes the



(a) System Configuration of RO-PMP$_1$



(b) System Configuration of RO-PMP$_0$



(c) System Configuration of RO-PMP$_{-1}$

(d) System Configuration of RO-PMP$_{-5}$

Figure 4.5: Optimal Configurations of Uncapacitated Models with Different θ

Table 4.10: Computation for Uncapacitated Instances with A_1

$\lvert I\rvert$	ρ	p	b	RO-PMP$_{-1}$				RO-PMP$_0$				RO-PMP$_1$			
				Time(s)	*Iter*	*Obj*	*Gap(%)*	*Time(s)*	*Iter*	*Obj*	*Gap(%)*	*Time(s)*	*Iter*	*Obj*	*Gap(%)*
25	0.2	8	15	1.1	2	1763.95		1.5	5	1558.09		2.4	5	1426.76	
			30	0.9	3	2233.97		3.6	7	1820.23		4.5	7	1525.92	
			40	9.8	11	2455.70		19.5	15	1989.57		6.4	10	1592.05	
		10	15	1.6	2	1364.19		0.4	2	1139.08		2.7	7	1024.11	
			30	2.7	6	1759.77		2.9	6	1374.09		4.8	8	1060.02	
			40	17.1	14	2003.68		11.8	12	1490.87		5.5	9	1088.26	
	0.4	8	15	0.8	2	2214.17		1.7	5	1802.45		3.4	7	1539.77	
			30	1.5	4	3086.90		12.2	12	2306.74		15.6	13	1674.14	
			40	22.6	16	3352.64		27.1	16	2458.94		18.9	14	1724.86	
		10	15	0.4	2	1814.40		0.6	2	1364.19		3.2	7	1067.20	
			30	2.5	5	2503.45		4.8	7	1785.45		6.7	9	1139.02	
			40	11.6	12	2883.19		43.4	19	2055.38		27.3	16	1221.07	
49	0.2	8	15	4.7	4	6563.12		3.1	4	6145.54		1.3	3	5684.45	
			30	73.5	11	7325.44		30.6	8	6631.47		203.4	16	6007.48	
			40	19.8	7	7447.91		10.6	5	6643.61		165.4	14	6046.56	
		10	15	2.1	3	5339.88		0.6	2	4950.32		9.6	6	4687.15	
			30	38.5	9	5865.65		8.4	5	5307.93		46.7	10	4887.14	
			40	266.4	15	6251.45		152.8	13	5559.63		995.9	23	5033.25	
	0.4	8	15	44.9	9	7405.65		4.9	4	6625.50		14.2	7	5898.97	
			30	570.3	17	8649.65		125.4	13	7393.31		972.1	21	6284.88	
			40	177.4	13	8954.28		116.1	11	7393.31		1064.6	20	6284.88	
		10	15	7.9	5	5988.67		3.2	4	5351.47		11.7	7	4796.70	
			30	1620.0	20	7158.15		68.8	11	5947.01		1302.4	22	5200.39	
			40	94.1	12	7575.79		81.2	11	6192.15		T	34	5357.96	1.38

following form:

$$A_1 = \{\mathbf{z} \in \{0,1\}^{|J|} : \sum_{j \in J_1} z_j \le k_1, \sum_{j \in J_2} z_j \le k_2, ..., \sum_{j \in J_L} z_j \le k_L, \sum_{l=1}^{L} \sum_{j \in J_l} a_j^l z_j \le b\}. \tag{4.33}$$

In our numerical study, $L = 2$, sites are randomly assigned to J_1 and J_2, $k_1 = 2$ and $k_2 = 1$, a_j^1 takes the value of 10 for all $j \in J_1$ and $a_j^2 = 15$ for all $j \in J_2$. Experiments are performed with $n = 25, 49$, $\rho = 0.2, 0.4$, $p = 8, 10$, $b = 15, 30, 40$, and $M = 15$. We mention that it is rather a simple set just for the demonstration of the impact of correlation. The computational performance of C&CG algorithm is presented in Table 4.10 and Table 4.11.

Note that with $b = 15, 30$, and 40 in A_1, the number of disruptions over the entire site set can be $1, 2$ and 3 respectively, which resembles the set we study in (4.1) with $k = 1, 2$ and 3. Nevertheless, comparing Table 4.10 with Table 4.3, and Table 4.11 with Table 4.5, we observe that: (i) the algorithm performance is generally better for the correlated set with less iterations; (ii) the objective function value, i.e., the weighted operation cost, is often smaller. Both points can be explained by the fact that the disruption set in (4.33) is a tighter and more precise description of all kinds of disruptive scenarios, if correlations exist; and A_1 is a

Table 4.11: Computation for Capacitated Instances with A_1

$\|I\|$	ρ	p	b	RO-CPMP$_{-1}$ *Time(s)*	*Iter*	*Obj*	*Gap(%)*	RO-CPMP$_0$ *Time(s)*	*Iter*	*Obj*	*Gap(%)*	RO-CPMP$_1$ *Time(s)*	*Iter*	*Obj*	*Gap(%)*
25	0.2	8	15	1.3	2	2069.34		2.6	3	1783.39		2.3	4	1550.58	
			30	1.2	3	2686.70		3.7	5	2126.51		4.6	6	1597.10	
			40	9.8	8	3253.49		11.9	8	2331.96		6.9	7	1626.47	
		10	15	0.7	2	1474.23		1.9	2	1148.48		2.3	4	1036.93	
			30	2.5	2	1923.53		1.7	3	1383.48		4.7	6	1084.08	
			40	13.1	9	2259.29		15.3	10	1612.96		5.6	6	1119.55	
	0.4	8	15	1.5	2	2620.20		1.4	3	2024.04		3.8	5	1623.29	
			30	5.7	5	3809.90		32.6	12	2791.94		15.2	11	1737.18	
			40	54.6	14	4670.89		11.6	7	2895.55		14.1	10	1791.85	
		10	15	1.5	2	1981.33		0.8	2	1373.58		4.4	6	1083.45	
			30	1.5	2	2869.74		1.8	3	1843.59		20.7	12	1224.17	
			40	18.4	11	3398.16		58.4	17	2211.98		23.1	12	1291.03	
49	0.2	8	15	21.6	5	6946.01		7.6	4	6432.77		14.7	5	6180.25	
			30	7126.0	23	7919.65		T	27	7062.09	1.40	723.5	15	6470.81	
			40	T	20	8214.50	0.16	452.5	11	7062.09		572.3	15	6470.81	
		10	15	11.8	4	5616.50		7.4	4	5045.10		10.4	5	4778.10	
			30	757.7	14	6290.50		224.6	12	5466.82		295.4	14	5064.34	
			40	T	21	6775.00	2.08	T	21	5796.18	1.22	5917.2	26	5215.26	
	0.4	8	15	172.6	9	7730.58		27.3	6	6939.44		87.6	9	6426.95	
			30	T	20	9585.18	12.30	T	25	8255.52	10.02	1373.8	16	6791.59	
			40	T	17	10139.70	7.01	7037.1	22	7948.09		1961.9	17	6791.59	
		10	15	115.4	8	6248.44		21.6	6	5434.73		21.0	7	4881.02	
			30	T	18	7881.48	5.07	7141.2	23	6276.43		T	24	5466.90	1.41
			40	T	19	8813.58	9.55	T	20	6696.92	5.93	T	23	5519.86	1.84

subset of that defined in (4.1) with an appropriate k. So, with more structural information available, both the master and subproblems are easier or smaller to compute than those with the cardinality set (4.1). Clearly, the cardinality set (4.1) is an overestimation of disruptions if disruptions actually happen according to the pattern defined in (4.33). Hence, in terms of system design and operations, the cardinality set (4.1) could lead to a more conservative system configuration with a higher objective function value as it is overprotective towards some *unrealistic* disruptive situations. So, with a proper description of the correlation, a less conservative system design can be achieved with the desired reliability level.

4.7 Conclusion

In this chapter, we propose two-stage robust optimization based models for the reliable p-median facility location problem. In particular, we demonstrate the strong modeling capability of two-stage robust optimization framework by considering two practical features, i.e., facility capacity and demand change due to site disruption, in a compact fashion, which otherwise would demand for large-scale scenario-based stochastic programming formulations and have received little attention. We study those models, and customize and implement exact computing algorithms, i.e., the column-and-constraint generation and Bender decomposition methods, to solve them. From a set of computational experiments, we note that (*i*) the column-and-constraint generation algorithm drastically outperforms the other method. Instances with up to 49 sites, including those with demand changes and capacities, can be solved exactly or with a reasonable gap; (*ii*) by assigning an appropriate weight to the operation cost from the worst disruptive situations in the objective function, a considerable decrease of that cost could be achieved by a small increase in the operation cost in the normal situation; (*iii*) demand changes due to disruptions should not be ignored in system design. Otherwise, a different network configuration could result in a huge increase of the operation cost in the worst disruptive situation; (*iv*) a description of disruption correlations, even in a simple form, makes those models less challenging and leads to less conservative system designs with the desired reliability.

A clear direction of the future research is to explore the problem structure to enhance the column-and-constraint generation algorithm for larger scale instances. So, a customized procedure can be developed to replace the professional solver for a better performance. Moreover, in real practice, one disruption can cause a complicated demand pattern. It could vary at different phases of the disruption and is also highly dependent on population and economic conditions of different locations (Ergun et al. (2010)). Hence, another interesting direction is to extend the recourse problem to a multi-period formulation so that it can capture dynamic and site-specific demand variations.

5 Exploring the Modeling Capacity of Two-Stage Robust Optimization

5.1 Note to Reader

This chapter has been previously published ©2014 IEEE. Reprinted, with permission, from Yu An and Bo Zeng, Exploring the Modeling Capacity of Two-Stage Robust Optimization: Variants of Robust Unit Commitment Model, IEEE Transactions on Power Systems, May. 2014 [PP]. The second author, Dr. Bo Zeng, contributed for part of literature review and numerical experiment.

5.2 Introduction

Recently, robust optimization (RO) techniques (Ben-Tal and Nemirovski (1998), Ben-Tal and Nemirovski (1999), Ben-Tal and Nemirovski (2000), Bertsimas and Sim (2003), and Bertsimas and Sim (2004)), especially two-stage robust optimization method (Ben-Tal et al., 2004), have attracted many researchers' attentions and been utilized to solve practical system design and operation problems. Different from classical stochastic programming models, an RO formulation has two features: (i) instead of assuming any probabilistic information on random factors, it assumes uncertain sets to capture randomness; (ii) instead of seeking for solutions with the optimal *expected* value, it derives a solution of the best performance with respect to worst cases in the uncertainty set. As a result, an RO model is less demanding on data to capture randomness and its solution is more reliable towards uncertainty. Therefore, robust optimization approaches are often adopted to address real problems, e.g., the operational problems in power industry (Hajimiragha et al. (2011), Zheng et al. (2012), Bertsimas et al. (2013b), Zhao and Zeng (2012b), Jiang et al. (2012), Xiong and Jirutitijaroen (2012), and Wang et al. (2012)), when it is challenging to construct a stochastic model to capture randomness and the system reliability is a critical concern.

Nevertheless, because a solution of the regular (single-stage) RO must hedge against any possible realization within the uncertainty set, it tends to be overly conservative and may not be cost-effective. To address the issue, two-stage (and multi-stage) robust optimization model (Ben-Tal et al., 2004) has been introduced to support decision making where decisions are partitioned into two stages, i.e., before and after uncertainty is disclosed. The first stage decisions still need to be made with respect to any realization in the uncertainty set while the second stage decisions can be made after the first stage decisions are determined and the uncertainty is revealed, which essentially enables the decision maker a recourse opportunity. Hence, comparing with that of the regular (single-stage) RO, a solution to two-stage RO is less conservative and more cost-effective. Note that this decision making structure nicely matches that of the day-ahead unit commitment (UC) problem, which makes use of dispatch as the recourse tool but must handle significant variability in loads, renewable energy generation, as well as various contingencies. Consequently, over the last few years, several two-stage robust unit commitment formulations have been developed and implemented to ensure reliable power generation and dispatch, see (Bertsimas et al. (2013b), Zhao and Zeng (2012b), Jiang et al. (2011), Jiang et al. (2012), Xiong and Jirutitijaroen (2012), Zhao et al. (2010), and Wang et al. (2012)).

As an optimization scheme at its early stage, we note that two key concepts of two-stage RO, i.e., the uncertainty set and the consideration of the worst case performance with recourse opportunities, jointly provide a very flexible mechanism that can actually be used to satisfy more complicated modeling needs. For example, on the one hand, when a single uncertainty set maybe too rough to describe the random factor, we can utilize different uncertainty sets to jointly define it. On the other hand, bounds on the worst case performances can be included as constraints to control the overall risks. As a result, the standard two-stage RO can be extended to capture different random situations, diverse data availabilities and qualities, and to meet various requirements. We mention that, although new models may be more involved than existing robust UC models in the literature, they generally can be solved rather efficiently by the recent *column-and-constraint generation* method with minor customizations (Zeng and Zhao, 2013).

Under this direction, we present two robust unit commitment variants in this chapter to demonstrate the advanced modeling capabilities of two-stage RO in solving practical problems. The first one is the *expanded robust unit commitment model* that considers the weighted summation of performances over multiple uncertainty sets. It actually yields solutions that are less conservative than those from the classical robust UC. The second one is the *risk constrained robust unit commitment model* that derives solutions subject to different bounds on the worst case performances in different uncertainty sets. Therefore, any feasible solution, if exists, will have guaranteed performances under those uncertainty sets.

Stochastic UC model is a popular model which utilizes probabilistic scenarios to capture random factors. Through those scenarios, critical properties in random factors, such as temporal and spatial correlations, can be explicitly included due to the narrative capacity of scenarios (Takriti et al. (1996), Papavasiliou and Oren (2013), and Constantinescu et al. (2011)). Furthermore, decision maker's risk consideration on specific scenarios, e.g., those which may fail to meet a predefined target, can be explicitly included by adding appropriate constraints on them (Li et al. (2007), Wu et al. (2008), and Abreu et al. (2012)). As shown in this paper, within robust optimization scheme, the usage of multiple uncertainty sets gives us a capable tool, comparable to the set of scenarios, to provide those modeling features. Indeed, stochastic UC and its risk constrained variant can be treated as special cases of our proposed robust UC models, when uncertainty sets reduce to scenarios. We also note that the risk constrained robust UC model is similar to the *globalized affinely adjustable robust counterpart* (Ben-Tal and Nemirovski, 2008) under two stage decision making framework, where violations of constraints due to randomness are bounded to a tolerable level. Indeed, because of the fully adjustable second stage decisions and the customized column-and-constraint generation method, our work carries stronger modeling capability and is able to produce exact solutions. Overall, we believe that, the presented models and solution method provide us useful tools to integrate practical data, decision maker's understanding on the underlying randomness, and system requirements to make unit commitment decisions.

The chapter is organized as follows. In Section 5.3, we first provide the *expanded robust UC model* and discuss its properties and connection with the scenario based stochastic UC model. Then, we introduce the *risk constrained robust UC model*, show a linkage to the risk constrained stochastic UC model, and present a customized column-and-constraint generation solution method. In Section 5.4, we perform a set of computational study of these two models and report numerical results. Section 5.5 concludes the chapter and discusses future research directions.

5.3 Two Robust Unit Commitment Variants

5.3.1 The Expanded Robust Unit Commitment Model

The classical robust unit commitment model is formulated as follows. We provide a compact matrix format for easy exposition while detailed formulations can be found in Bertsimas et al. (2013b), Zhao and Zeng (2012b), and Jiang et al. (2011).

$$\min_{\mathbf{x}} \mathbf{ax} + \max_{\mathbf{u}\in\mathbb{U}} \min_{\mathbf{y}\in\Omega(\mathbf{x},\mathbf{u})} \mathbf{gy}$$

s.t.

$$\mathbf{Dx} \geq \mathbf{f},\ \mathbf{x} \text{ binary}, \tag{5.1}$$

$$\Omega(\mathbf{x},\mathbf{u}) = \{\mathbf{y} : \mathbf{Ey} \leq \mathbf{e},$$

$$\mathbf{Ay} \leq \mathbf{L} - \mathbf{Gx} - \mathbf{Ru}, \mathbf{Fy} = \mathbf{d} - \mathbf{Tu}\ \},$$

where $\mathbf{x}$ represents the first stage decision variables, including binary start-up/shut-down and on-off decisions, $\mathbf{y}$ represents the second stage (i.e., recourse) decision variables, including economic dispatch and market buy/sell decisions that are continuous, $\mathbf{u}$ represents some uncertain factor, e.g., the renewable energy generation, load, or contingencies, whose randomness is captured by the uncertainty set $\mathbb{U}$. Note that, because of the two-stage decision making nature, the essential solution to the above robust model is the first stage decisions $\mathbf{x}$ while the second stage decisions $\mathbf{y}$ are made with perfect information on $\mathbf{u}$. Hence the set $\mathbb{U}$ plays a critical role in determining the quality of the first stage decisions.

Lemma 5.3.1 *Consider two uncertainty sets $\mathbb{U}_1$ and $\mathbb{U}_2$ such that $\mathbb{U}_1 \subseteq \mathbb{U}_2$, i.e., $\mathbb{U}_1$ is included in $\mathbb{U}_2$, and denote their corresponding optimal values of (5.1) by $\theta(\mathbb{U}_1)$ and $\theta(\mathbb{U}_2)$. We have $\theta(\mathbb{U}_1) \leq \theta(\mathbb{U}_2)$, i.e., θ is non-decreasing in $\mathbb{U}$ (in terms of set inclusion).*

Proof. Let $\theta(\mathbb{U}|\mathcal{C})$ represent the optimal value of (5.1) under a condition $\mathcal{C}$. Note that, for a given first-stage solution $\mathbf{x}^0$, because $\mathbb{U}_1 \subseteq \mathbb{U}_2$, we have

$$
\begin{aligned}
&\max_{\mathbf{u}\in\mathbb{U}_2}\min_{\mathbf{y}\in\Omega(\mathbf{x}^0,\mathbf{u})}\mathbf{g}\mathbf{y}\\
=\ &\max\{\max_{\mathbf{u}\in\mathbb{U}_1}\min_{\mathbf{y}\in\Omega(\mathbf{x}^0,\mathbf{u})}\mathbf{g}\mathbf{y},\ \max_{\mathbf{u}\in\mathbb{U}_2\setminus\mathbb{U}_1}\min_{\mathbf{y}\in\Omega(\mathbf{x}^0,\mathbf{u})}\mathbf{g}\mathbf{y}\}\\
\geq\ &\max_{\mathbf{u}\in\mathbb{U}_1}\min_{\mathbf{y}\in\Omega(\mathbf{x}^0,\mathbf{u})}\mathbf{g}\mathbf{y}.
\end{aligned}
$$

Hence, it follows that $\theta(\mathbb{U}_2|\mathbf{x}=\mathbf{x}^0)\geq\theta(\mathbb{U}_1|\mathbf{x}=\mathbf{x}^0)$. Let $\mathbf{x}^2$ denote an optimal solution of $\theta(\mathbb{U}_2)$. Consequently, we have $\theta(\mathbb{U}_2)=\theta(\mathbb{U}_2|\mathbf{x}=\mathbf{x}^2)\geq\theta(\mathbb{U}_1|\mathbf{x}=\mathbf{x}^2)\geq\theta(\mathbb{U}_1)$.

■

Hence, adopting a tight $\mathbb{U}$ leads to a less costly solution $\mathbf{x}$. However, if the scope of $\mathbb{U}$ is small, it cannot sufficiently depict the uncertain factor. For example, we construct $\mathbb{U}$ to capture random loads of a city in Florida using 7 consecutive days' historical data ($24\times$ 7 loads), which are presented in Figure 5.1. If the set is defined as a hypercubic set $\mathbb{U}=\{u_t:u_t\in[\overline{u}_t-\sigma_t,\overline{u}_t+\sigma_t]\}$, where $\overline{u}_t$ is the average load and σ_t is the standard deviation (STD) in hour t, it clearly cannot ensure a coverage on all load possibilities, which may cause us to take a risky first-stage UC solution infeasible to meet load or with a high recourse cost. Nevertheless, if the scope of $\mathbb{U}$ is large, we may take a solution that is costly and overly protective. As in Fig. 5.2, the set defined as $\mathbb{U}=\{u_t:u_t\in[\overline{u}_t-3\sigma_t,\overline{u}_t+3\sigma_t]\}$ might overstate that randomness and result in running units more than necessary.

To balance the risk and cost, one way is to construct and make use of a sophisticated uncertainty set to describe the randomness, which could be technically challenging. First, because of the nature of RO, extreme scenarios play a critical role in defining the uncertainty set. However, as they typically happen much less frequently, available data might not be sufficient to support a well-developed uncertainty set that not only precisely captures the critical dynamics in/among random factors but also provides an adequate coverage for those factors. Second, using a complicated uncertainty set, such as a general polyhedron, makes the derivation of the worst case situations computationally demanding (Zeng and Zhao, 2013) and therefore restricts the applications of robust UC model in real systems. We note that a novel data-driven method is developed recently to construct uncertainty sets for (single-stage) RO (Bertsimas et al., 2013a). Nevertheless, that method is less applicable in two-stage RO scheme given the fact that it focuses on the feasibility issue (of a single constraint) and the resulting uncertainty sets are described by computationally challenging nonlinear formulations.

We believe that one improved strategy is to expand the uncertainty description by using multiple sets in robust UC model, along with their respective recourse problems. Then, we can integrate the impacts from multiple sets under the same umbrella by assigning different weights to the worst case performances of those sets.

Specifically, let $\mathbb{U}_k,\ k = 1, \ldots, K$ denote K uncertainty sets and ρ_k be their weight coefficients normalized for the totality being one. The *expanded robust unit commitment model* is formulated as

$$\min_{\mathbf{x}} \mathbf{a}\mathbf{x} + \sum_k \rho_k \Big(\max_{\mathbf{u}\in\mathbb{U}_k} \min_{\mathbf{y}\in\Omega_k(\mathbf{x},\mathbf{u})} \mathbf{g}_k\mathbf{y} \Big)$$

s.t.

$$\begin{aligned} \mathbf{D}\mathbf{x} \geq \mathbf{f},\ \mathbf{x} \text{ binary}, \\ \Omega_k(\mathbf{x},\mathbf{u}) = \{\mathbf{y} : \mathbf{E}_k\mathbf{y} \leq \mathbf{e}_k, \\ \mathbf{A}_k\mathbf{y} \leq \mathbf{L}_k - \mathbf{G}_k\mathbf{x} - \mathbf{R}_k\mathbf{u}, \mathbf{F}_k\mathbf{y} = \mathbf{d}_k - \mathbf{T}_k\mathbf{u}\}. \end{aligned} \tag{5.2}$$

Note that Ω_k and its associated parameters are indexed with k. So we can model the situation in which different uncertainty sets or recourse problems are associated with different environments. For example, components of a real system may be operated with different ratings (e.g., different capacities) under the normal status and contingencies.

If Ω_k and $\mathbf{g}_k$ are identical for all k, the next result follows easily from Lemma 5.3.1.

Proposition 5 *Consider two uncertainty sets $\mathbb{U}_1$ and $\mathbb{U}_2$ such that $\mathbb{U}_1 \subseteq \mathbb{U}_2$, and two coefficients $\rho_1 \geq 0$ and $\rho_2 \geq 0$ such that $\rho_1 + \rho_2 = 1$. Denote the corresponding optimal value of (5.2) by $\Theta((\mathbb{U}_1, \mathbb{U}_2), (\rho_1, \rho_2))$. We have $\theta(\mathbb{U}_1) \leq \Theta((\mathbb{U}_1, \mathbb{U}_2), (\rho_1, \rho_2)) \leq \theta(\mathbb{U}_2)$. The equalities are achieved by setting $\rho_1 = 1$ and 0 respectively.*

Clearly, by jointly considering $\mathbb{U}_1$ and $\mathbb{U}_2$ through non-trivial ρ_1 and ρ_2, the expanded robust UC model will yield solutions that are less conservative than those derived exclusively

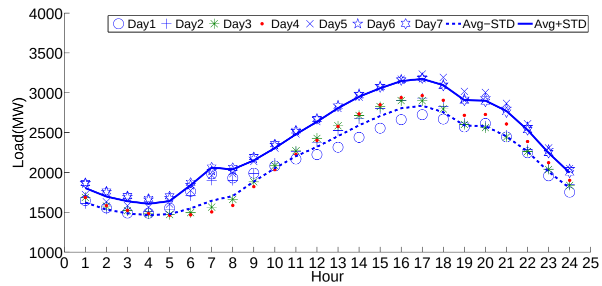


Figure 5.1: 24-hour Load Distribution and Single STD Description

under $\mathbb{U}_2$ while are more reliable than those derived exclusively under $\mathbb{U}_1$. Result in Proposition 5 can be generalized for the more general expanded robust UC models when Ω_k's and $\mathbf{g}_k$'s are identical.

Corollary 5.3.2 *Consider a collection of uncertainty sets* $\vec{\mathbb{U}} = \{\mathbb{U}_1, \ldots, \mathbb{U}_K\}$ *and a set of coefficients* $\vec{\rho} = \{\rho_1, \ldots, \rho_K\}$ *such that* $\sum_k \rho_k = 1$. *Denote the corresponding optimal value of (5.2) by* $\Theta(\vec{\mathbb{U}}, \vec{\rho})$. *We have*

$$\min_k \theta(\mathbb{U}_k) \leq \Theta(\vec{\mathbb{U}}, \vec{\rho}) \leq \max_k \theta(\mathbb{U}_k).$$

We mention that weight coefficients reflect decision maker's conservative/protective level and his understanding of the likelihoods of those sets. For example, $\mathbb{U}_1$ and $\mathbb{U}_2$ are defined as $\{u_t \in [\overline{u}_t - \sigma_t, \overline{u}_t + \sigma_t]\}$ and $\{u_t \in [\overline{u}_t - 3\sigma_t, \overline{u}_t + 3\sigma_t]\}$ respectively, to describe the overall load uncertainty. Based on data in Figure 5.1-5.2, although rigorous statistical analysis might not be obtainable, we are confident to conclude that the worst case situations of $\mathbb{U}_1$ are much more likely than those of $\mathbb{U}_2$. So, we can set ρ_1 to a value larger than ρ_2 to show our confidence. Actually, practitioners often perform relative evaluation on different uncertain situations, such as ranking them according to their likelihoods. Although no absolute quantities are derived to exactly describe the overall randomness, the qualitative information can be reliable and useful to practitioners. Hence, assigning weights to uncertainty sets provides us a flexible function to take advantage of such dependable information.

We further mention that employing multiple sets helps us to reduce the impact of unrealistic worst case situations. First, note that a single uncertainty set should be of a large

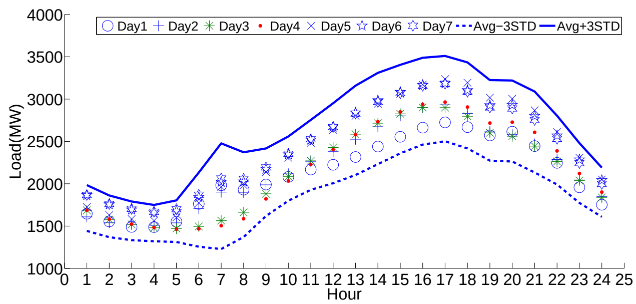


Figure 5.2: 24-hour Load Distribution and Three STD Description

scope to cover all concerned possibilities. Then, some of its worst case situations could be unrealistic, which however cause a big impact on the final UC solution. Nevertheless, in the expanded robust UC model, even that large-scope uncertainty set is adopted as one of multiple uncertainty sets, the impact of its worst case situations will be modulated by its weight coefficient in the objective function, which could be of much less importance in determining UC solutions.

Second, introducing multiple sets allows us to better explore available data to identify simple but critical patterns of random factors. Using our load example, the arguably most popular way to build an uncertainty set is to impose a *budget of uncertainty constraint* (Bertsimas et al. (2013b), Bertsimas et al. (2011), Zhao and Zeng (2012b), and Jiang et al. (2011)) on top of the hypercubic set:

$$\mathbb{U} = \{u_t : u_t \in [\overline{u}_t - \hat{u}_t, \overline{u}_t + \hat{u}_t], \sum_t \frac{|u_t - \overline{u}_t|}{\hat{u}_t} \le \Gamma\} \tag{5.3}$$

where the parameter Γ is to bound the aggregated deviations from nominal values, which can also be interpreted as the overall likelihood of u_t's taking their upper/lower bounds. This set has a nice property that when Γ is integral, its worst case situations will only be those with u_t at its bound or its nominal value for all t. Hence, $\mathbb{U}$ can be reduced to an equivalent set with only discrete points of that property, which renders itself computationally friendly (Zhao and Zeng (2012b) and Jiang et al. (2011)). Nevertheless, when $\hat{u}_t$ is large to cover all possibilities, because only a few observations can be made on u_t reaching its lower or upper bound, very limited analysis can be done to help us eliminate unrealistic ones from those satisfying that property. This issue can actually be mitigated by considering two or more uncertainty sets. For example, for a set defined with a smaller $\hat{u}_t$, as more observations can be made on u_t taking value equal to or beyond $\overline{u}_t \pm \hat{u}_t$, we can compute the correlation among those events. If a perfect correlation is derived for u_{t_1} and u_{t_2} reaching their bounds, we can introduce the following equality

$$\frac{u_{t_1} - \overline{u}_{t_1}}{\hat{u}_{t_1}} = \frac{u_{t_2} - \overline{u}_{t_2}}{\hat{u}_{t_2}}. \tag{5.4}$$

Note that including (5.4) into (5.3)'s discrete equivalence does not change its structure but simply eliminates some unrealistic worst case scenarios that do not carry the aforementioned correlation. If we consider the original continuous set of (5.3), when the equality may not strictly hold for u_ts within their bounds, (5.4) could be a reasonable and computationally friendly approximation of the underlying correlation, given that many observations show that loads reach bounds simultaneously if they do. Certainly, general inequality representations of correlations should be more accurate, which are our next step research. In Appendix F, a demonstration of this idea is presented.

As another advantage of considering multiple uncertainty sets, Γ for the set with a small $\hat{u}_t$ can be obtained in a more objective fashion with relatively more observations on u_t reaching its bounds. For a set defined with a larger $\hat{u}_t$, we can set its Γ to a smaller value given it is less likely that u_t reaches its bounds. Clearly, such a strategy can also avoid overly estimating Γ and reduce the unrealistic scenarios.

We recognize that the aforementioned strategies are basic steps to build meaningful uncertainty sets that capture the essence of random factors. When UC model is considered within a power grid system, uncertainties in loads, renewable energy injection, as well as system contingencies (Papavasiliou and Oren (2013), Constantinescu et al. (2011), and Street et al. (2011)), require more accurate uncertainty descriptions to reflect system nature and real environment. For instance, the likelihoods of $N-1$, $N-2$, or more general $N-k$ contingencies are definitely different. It indicates that multiple uncertainty sets with different weight coefficients will be more appropriate. It further demands for advanced methods to analyze the physical grid to define correlations among contingencies and to assign weight coefficient $\rho_1, \rho_2, \ldots, \rho_k$ to those sets. Also, geographically distributed renewable energy generation sources often demonstrate very strong spatial correlations, in addition to randomness presented within individual sources. For those situations, one possible strategy is to first construct uncertainty sets individually and then include correlation descriptions for neighboring generation sources. Overall, it is worth studying and developing more sophisticated methods to construct uncertainty sets of simple structures that can capture the inherent nature of random factors.

Next we show a connection between our expanded robust UC and the classical scenario based stochastic UC model.

Proposition 6 *Suppose the randomness of the uncertainty factor is completely captured by scenarios* $\mathbf{u}_k, k = 1, \ldots, K$*. Let* $\mathbb{U}_k$ *be a singleton* $\{\mathbf{u}_k\}$ *and set* ρ_k *equal to the corresponding realization probability, then the expanded robust UC is equivalent to the stochastic programming UC model.*

Proof. Note that when $\mathbb{U}_k = \{\mathbf{u}_k\}$, the max operator can be eliminated from the formulation. Hence, we have

$$\min_{\mathbf{x}} \mathbf{a}\mathbf{x} + \sum_k \rho_k \Big(\min_{\mathbf{y} \in \Omega_k(\mathbf{x}, \mathbf{u}_k)} \mathbf{g}_k \mathbf{y} \Big)$$

s.t.

$$\text{Constraints in (5.2).}$$

Because economic dispatch and buy/sell decisions are made specific to individual scenarios, we can replace $\mathbf{y}$ by introducing recourse variables $\mathbf{y}_k$ for the resource problem associated with $\mathbf{u}_k/\Omega_k$. Also, the second min operator can be removed given that it aligns with the first min operator. Hence, the overall min-max-min formulation can be simplified as

$$\min_{\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_K} \mathbf{a}\mathbf{x} + \sum_k \rho_k \mathbf{g}_k \mathbf{y}_k$$

s.t.

$$\mathbf{D}\mathbf{x} \geq \mathbf{f}, \ \mathbf{x} \text{ binary},$$

$$\mathbf{E}_k\mathbf{y}_k \leq \mathbf{e}_k, \ \forall k$$

$$\mathbf{A}_k\mathbf{y}_k \leq \mathbf{L}_k - \mathbf{G}_k\mathbf{x} - \mathbf{R}_k\mathbf{u}_k, \ \forall k$$

$$\mathbf{F}_k\mathbf{y}_k = \mathbf{d}_k - \mathbf{T}_k\mathbf{u}_k, \ \forall k$$

which is exactly the scenario based stochastic programming UC model.

■

According to Proposition 6, we can conclude that the expanded robust UC model is a complete and flexible modeling framework to handle various randomness. Decision makers can customize their uncertainty sets and adjust weight coefficients for their conveniences, according to data availability, data quality, system requirements, their conservative/protective level, and computational capability.

For the situation where the number of scenarios is not large and we are confident that they precisely capture the randomness, we believe that the classical stochastic program should be a good choice. If data size is big, the expanded robust model can be a reasonable alternative. Under such a situation, those samples can be organized into a small number of groups, on each of which we can build an uncertainty set with a simple structure. Then, we can set the weight coefficient of one uncertainty set to the sum of realization probabilities (the proportion of samples of that group). Hence, instead of computing a large-scale sampling based stochastic programming model, we can solve the compact expanded robust model based on those uncertainty sets, which actually could take significantly less computational time than the stochastic one (see a benchmark result in Table 5.2 in Section 5.4). Moreover, uncertainty sets' weight coefficients can be modified to reflect our concerns on specific situations and to achieve a desired trade-off among computational time, cost, and risk.

Actually, the framework of multiple uncertainty sets is supportive to develop data-driven approaches to fully make use of available data. For example, data within one group, which could be more homogeneous, may allow us to design a structural uncertainty set that is not only computationally friendly but also with quantitative insights. In addition, in an environment where data become available dynamically, the information from new data can be reflected by adjusting weights of existing uncertainty sets, or modifying parameters of a single uncertainty set, or constructing and including a new uncertainty set (and adjusting weights of uncertainty sets), which provides a great flexibility to handle data.

Further, taking a more direct approach, when some prior knowledge is available, e.g., the shape and bounds of the distribution, instead of depending on large-scale sampling to describe a sophisticated random factor, we can use multiple interval uncertainty sets, which have varying sizes and weight coefficients, to capture or approximate that randomness. Note that by considering those uncertainty sets under the two-stage RO umbrella, we can reduce the impact of possible information misrepresentation on the solution's feasibility.

In Appendix E, we present a concrete expanded robust UC model by considering multiple load uncertainty sets. Numerical results of this model and comparison with other existing models are provided in Section 5.4.

5.3.2 The Risk Constrained Robust Unit Commitment Model

In this section, we show how to further extend our modeling capacity. Specifically, based on the nature of random factors and system requirements, we explicitly impose hard constraints to restrict some performance measurements in the worst case situations of uncertainty sets. As a result, any derived solution, if exists, can guarantee its performance with respect to those uncertainty sets. Let γ_k denote our performance restriction in uncertainty sets $\mathbb{U}_k$, $k = 1, \ldots, K$. Also, let $\mathbf{u}_0$ be the nominal situation that the decision maker would like to consider.

The *risk constrained robust unit commitment model* is formulated as

$$\min_{\mathbf{x},\mathbf{y}_0} \mathbf{ax} + \mathbf{g}_0\mathbf{y}_0$$

s.t.

$$\mathbf{Dx} \geq \mathbf{f},\ \ \mathbf{x} \text{ binary}.$$

$$
\begin{aligned}
\mathbf{E}_0\mathbf{y}_0 &\le \mathbf{e}_0, \\
\mathbf{A}_0\mathbf{y}_0 &\le \mathbf{L}_0 - \mathbf{G}_0\mathbf{x} - \mathbf{R}_0\mathbf{u}_0, \\
\mathbf{F}_0\mathbf{y}_0 &= \mathbf{d}_0 - \mathbf{T}_0\mathbf{u}_0, \\
\max_{\mathbf{u}\in\mathbb{U}_k}\ \min_{\mathbf{y}\in\Omega_k(\mathbf{x},\mathbf{u})} \mathbf{g}_k\mathbf{y} &\le \gamma_k,\ \ k = 1,\dots,K \\
\Omega_k(\mathbf{x},\mathbf{u}) &= \{\mathbf{y} : \mathbf{E}_k\mathbf{y} \le \mathbf{e}_k, \\
\mathbf{A}_k\mathbf{y} \le \mathbf{L}_k - \mathbf{G}_k\mathbf{x} - \mathbf{R}_k\mathbf{u}&, \mathbf{F}_k\mathbf{y} = \mathbf{d}_k - \mathbf{T}_k\mathbf{u}\}.
\end{aligned} \tag{5.5}
$$

Again, we mention that the essential solution to the above robust UC model is the first stage $\mathbf{x}$ decisions. It may not be optimal to implement the solution $\mathbf{y}_0$ because the virtual optimal one can be derived after $\mathbf{u}$ is disclosed.

Let $\vec{\gamma}$ denote $\{\gamma_1,\dots,\gamma_K\}$ and $\Phi(\vec{\mathbb{U}},\vec{\gamma})$ be the optimal value of the problem (5.5) with $\vec{\mathbb{U}}$ and $\vec{\gamma}$. When $\vec{\gamma}_1$ ($\vec{\mathbb{U}}_1$, respectively) is component-wise less than or equal to (included in, respectively) $\vec{\gamma}_2$ ($\vec{\mathbb{U}}_2$, respectively), we say $\vec{\gamma}_1 \le \vec{\gamma}_2$ ($\vec{\mathbb{U}}_1$ is included in $\vec{\mathbb{U}}_2$, respectively). Because the consideration of worst case performances is included as constraints, the whole formulation could be infeasible. Under such a situation, following the convention in mathematical programming, we set the optimal value of (5.5) to ∞. Then, we analyze $\Phi(\vec{\mathbb{U}},\vec{\gamma})$ with respect to its input parameters.

Proposition 7 (i) *For a given $\vec{\mathbb{U}}$, the function $\Phi(\vec{\mathbb{U}},\vec{\gamma})$ is non-increasing in $\vec{\gamma}$;* (ii) *for a given $\vec{\gamma}$, the function $\Phi(\vec{\mathbb{U}},\vec{\gamma})$ is non-decreasing in $\vec{\mathbb{U}}$ (in terms of set inclusion).*

Note that if $\mathbb{U}_{k_1} \subseteq \mathbb{U}_{k_2}$, and $\Omega_{k_1} = \Omega_{k_2}$ and $\mathbf{g}_{k_1} = \mathbf{g}_{k_2}$, i.e., recourse problems have the identical structure, it is necessary to set $\gamma_{k_1} < \gamma_{k_2}$. Otherwise, from the proof of Lemma 5.3.1, it is clear that the performance bound constraint over the recourse problem of $\mathbb{U}_{k_2}$ dominates that of $\mathbb{U}_{k_1}$. Therefore, γ_k should be in an increasing order with respect to set inclusion relationship.

We also note that, similar to results in Proposition 6, our risk constrained robust UC resembles the risk constrained stochastic programming models (Li and Shahidehpour (2007), Li et al. (2007), Wu et al. (2008), and Abreu et al. (2012)). In those stochastic UC models, risk is defined by the failure to meet the target performance in scenario s, i.e., $RISK_s =$

$\{0, V_0 - \mathbf{ax} - \mathbf{gy}_s\}^+$ where V_0 be the target performance value and $\mathbf{ax} + \mathbf{gy}_s$ is the performance (to be maximized) in s with $\mathbf{y}_s$ representing the recourse variables. Then, the downside risk can be controlled by adding the following constraint to the stochastic UC model

$$\sum_s p_s RISK_s \leq AL \tag{5.6}$$

where p_s is the realization probability and AL represents the accepted risk level. To build that modeling capability in robust UC, we can first modify Ω_k in (5.5) as

$$\Omega_k(\mathbf{x}, \mathbf{u}) = \{(V, \mathbf{y}) : V \geq 0, V \geq V_0 - \mathbf{ax} - \mathbf{gy},$$
$$\mathbf{E}_k\mathbf{y} \leq \mathbf{e}_k, \mathbf{A}_k\mathbf{y} \leq \mathbf{L}_k - \mathbf{G}_k\mathbf{x} - \mathbf{R}_k\mathbf{u}, \mathbf{F}_k\mathbf{y} = \mathbf{d} - \mathbf{T}_k\mathbf{u}\}.$$

Then, instead of imposing performance bounds on each individual uncertainty sets as in (5.5), we can include a single upper bound on the aggregated performance as follows

$$\sum_k \rho_k (\max_{\mathbf{u} \in \mathbb{U}_k} \min_{(V, \mathbf{y}) \in \Omega_k(\mathbf{x}, \mathbf{u})} V) \leq \sum_k \rho_k \gamma_k.$$

Indeed, if we set recourse problem $\Omega_k = \Omega$, $\gamma_k = AL$, $\rho_k = p_k$, let $\mathbb{U}_k$ and $(V_k, \mathbf{y}_k)$ be single scenarios and recourse decision variables for $k = 1, \ldots, K$, using the same argument of Proposition 6, the aforementioned inequality reduces to (5.6), which again shows the linkage between two-stage RO with multiple uncertainty sets and the scenario based stochastic programming model.

It is worth mentioning that the risk constrained robust UC model (and its generalization - the *risk constrained robust optimization model*) is close to the *globalized affinely adjustable robust counterpart* (Ben-Tal and Nemirovski, 2008) under two stage decision making framework, where violations of constraints due to randomness are bounded to a tolerable level. Nevertheless, our work carries more powerful modeling and solution capabilities given that: (i) recourse problems are fully recoursable and recourse decisions are not required to follow affine rules of random factors; (ii) through considering multiple uncertainty sets, different violation tolerances on different recourse problems can be included; (iii) using the customized column-and-constraint generation method presented in the following subsection, an exact solution can be derived, which does not require the convexity structure of uncertainty sets and can handle discrete contingencies.

In Appendix G, we present a concrete risk constrained UC model by considering G-1 (i.e., one generator in forced outage) and G-2 (i.e., two generators in forced outages) contingencies as our $\mathbb{U}_1$ and $\mathbb{U}_2$. We impose upper bounds on load sheds in worst cases in those two uncertainty sets to control our risks. Numerical results of this model are presented in Section 5.4.

5.3.3 A Solution Procedure

The aforementioned robust UC models can be solved by well-known Benders dual methods, which have been applied to solve classical robust UC models in existing literatures (Bertsimas et al. (2013b), Zhao and Zeng (2012b), Jiang et al. (2011), and Jiang et al. (2012)). A recent *column-and-constraint generation* method has also been developed in Zhao and Zeng (2012b) and Zeng and Zhao (2013) that solves the classical robust UC models. It progressively identifies and includes significant scenarios (and their recourse problems) from the uncertainty set in the derivation of the final solution, which performs an order of magnitude faster than Benders dual methods. A similar strategy is adopted in Bertsimas et al. (2013b), which also includes scenarios and recourse problems to speed up their Benders dual algorithm. Indeed, the basic column-and-constraint generation algorithm can be customized to solve our new robust UC variants. Because the expanded robust UC is similar to the classical robust UC and just needs a few minor modifications, our illustration is within the context of the risk constrained robust UC model.

The column-and-constraint generation method is implemented in a master-subproblem framework. For a given $\mathbf{x}^*$, we define the following subproblem SP_k.

$$\mathcal{Q}_k(\mathbf{x}^*) = \max_{\mathbf{u}\in\mathbb{U}_k} \min_{\mathbf{y}\in\Omega_k(\mathbf{x}^*,\mathbf{u})} \mathbf{g}_k\mathbf{y}$$

s.t.

$$\mathbf{E}_k\mathbf{y} \leq \mathbf{e}_k, \tag{5.7}$$

$$\mathbf{A}_k\mathbf{y} \leq \mathbf{L}_k - \mathbf{G}_k\mathbf{x}^* - \mathbf{R}_k\mathbf{u}, \tag{5.8}$$

$$\mathbf{F}_k\mathbf{y} = \mathbf{d}_k - \mathbf{T}_k\mathbf{u}. \tag{5.9}$$

Although SP_k is a bi-level max-min program, because the inner problem is a linear program, it can be converted into a mixed integer program (MIP) by using classical Karush-Kuhn-Tucker (KKT) conditions (Zeng and Zhao, 2013). Specifically, let $\boldsymbol{\pi}^1$, $\boldsymbol{\pi}^2$, and $\boldsymbol{\pi}^3$ be dual variables for (5.7), (5.8), and (5.9), respectively (satisfying $\boldsymbol{\pi}^1 \leq 0$ and $\boldsymbol{\pi}^2 \leq 0$). Solving SP_k is equivalent to solving the following problem KKT-SP_k.

$$\max_{\substack{\mathbf{u} \in \mathbb{U}_k,\ \mathbf{y} \in \Omega_k(\mathbf{x}^*, \mathbf{u}),\\ \boldsymbol{\pi}^1, \boldsymbol{\pi}^2, \boldsymbol{\pi}^3}} \mathbf{g}_k\mathbf{y}$$

s.t.

$$\text{Constraints (5.7)} - \text{(5.9)},$$

$$\boldsymbol{\pi}^1\mathbf{E}_k + \boldsymbol{\pi}^2\mathbf{A}_k + \boldsymbol{\pi}^3\mathbf{F}_k \leq \mathbf{g}_k,$$

$$\boldsymbol{\pi}^1(\mathbf{E}_k\mathbf{y} - \mathbf{e}_k) = 0,$$

$$\boldsymbol{\pi}^2(\mathbf{A}_k\mathbf{y} - \mathbf{L}_k + \mathbf{G}_k\mathbf{x}^* + \mathbf{R}_k\mathbf{u}) = 0,$$

$$\mathbf{y}(\boldsymbol{\pi}^1\mathbf{E}_k + \boldsymbol{\pi}^2\mathbf{A}_k + \boldsymbol{\pi}^3\mathbf{F}_k - \mathbf{g}_k) = 0.$$

The last three constraints represent complementary slackness and can be transformed into linear constraints using binary variables and big-M linearization method. For example, the n-th constraint in the first group can be replaced by

$$(\mathbf{e}_k - \mathbf{E}_k\mathbf{y})_n \leq Mv_n,\ \boldsymbol{\pi}^1_n \geq -M(1 - v_n),\ v_n \in \{0, 1\}.$$

Next, we provide algorithmic details on solving the risk constrained UC model.

(I) Set $iter = 0$, VIOLATION = VIOL$_k$= FALSE, and $L_k = 0$ for $k = 1, \ldots, K$.

(II) Call an MIP solver to compute the following master problem MP.

$$\min_{\mathbf{x},\mathbf{y}_0} \mathbf{a}\mathbf{x} + \mathbf{g}_0\mathbf{y}_0$$

s.t.

$$\mathbf{D}\mathbf{x} \geq \mathbf{f},\ \ \mathbf{x} \text{ binary},$$

$$\mathbf{E}_0\mathbf{y}_0 \leq \mathbf{e}_0,$$

$$\mathbf{A}_0\mathbf{y}_0 \leq \mathbf{L}_0 - \mathbf{G}_0\mathbf{x} - \mathbf{R}_0\mathbf{u}_0,$$

$$\mathbf{F}_0\mathbf{y}_0 = \mathbf{d}_0 - \mathbf{T}_0\mathbf{u}_0,$$

$$\mathbf{g}_k\mathbf{y}_{k,l} \leq \gamma_k,\ \ \forall k, l = 1, ..., L_k$$

$$\mathbf{E}_k\mathbf{y}_{k,l} \leq \mathbf{e}_k,\ \ \forall k, l = 1, ..., L_k$$

$$\mathbf{A}_k\mathbf{y}_{k,l} \leq \mathbf{L}_k - \mathbf{G}_k\mathbf{x} - \mathbf{R}_k\mathbf{u}_{k,l},\ \ \forall k, l = 1, ..., L_k$$

$$\mathbf{F}_k\mathbf{y}_{k,l} = \mathbf{d}_k - \mathbf{T}_k\mathbf{u}_{k,l},\ \ \forall k, l = 1, ..., L_k$$

Derive an optimal solution $(\mathbf{x}^*, \mathbf{y}_0^*, \mathbf{y}_{1,1}^* \ldots, \mathbf{y}_{1,L_1}^*, \ldots, \mathbf{y}_{K,1}^*, \ldots, \mathbf{y}_{K,L_K}^*)$.

(III) With given $\mathbf{x}^*$, for $k = 1, \ldots, K$, do

(i) call an MIP solver to compute (linearized) KKT-SP$_k$.

(ii) if $\mathcal{Q}_k(\mathbf{x}^*) > \gamma_k$, set VIOLATION = VIOL$_k$ = TRUE, use $\mathbf{u}_{k,L_k+1}$ to record the optimal $\mathbf{u}^*$ and update $L_k = L_k + 1$.

(IV) If VIOLATION = FALSE, return $\mathbf{x}^*$ and terminate. Otherwise, set VIOLATION = FALSE and for $k = 1, \ldots, K$, do

(i) If VIOL$_k$ = TRUE, create variables $\mathbf{y}_{k,L_k}$ and add the following constraints

$$\mathbf{g}_k \mathbf{y}_{k,L_k} \leq \gamma_k$$

$$\mathbf{E}_k \mathbf{y}_{k,L_k} \leq \mathbf{e}_k$$

$$\mathbf{A}_k \mathbf{y}_{k,L_k} \leq \mathbf{L}_k - \mathbf{G}_k \mathbf{x} - \mathbf{R}_k \mathbf{u}_{k,L_k}$$

$$\mathbf{F}_k \mathbf{y}_{k,L_k} = \mathbf{d}_k - \mathbf{T}_k \mathbf{u}_{k,L_k}$$

to MP.

(ii) Set VIOL$_k$= FALSE.

(V) Update $iter = iter + 1$ and go to Step 2.

Given that the second stage recourse problems are linear programs, the convergence and the complexity results follow from the complexity analysis of the column-and-constraint generation method presented in Zeng and Zhao (2013). Let b_k be the number of extreme points of $\mathbb{U}_k$ if it is a polytope (e.g., the uncertainty sets for random loads) or the set cardinality if it is a finite discrete set (e.g., the G-k contingency set). We have

Proposition 8 *The column-and-constraint generation method either terminates with an optimal solution or reports infeasibility of the risk constrained robust UC model (the expanded robust UC model, respectively) in* $O(\prod_{k=1}^{K} b_k)$ *iterations.*

Actually, the computational performance of this method on solving practical problems is drastically better than the theoretical result, which can be seen in the discussion of Fig. 5.4.

We would like to comment two features of the presented column-and-constraint generation procedure. First, with a given $\mathbf{x}^*$, subproblem SP$_k$'s are independent of each other. Hence, it allows us to employ the parallel computing strategy, which is recently adopted to deal with stochastic UC model (Papavasiliou and Oren, 2012), to implement Step 3 for better computational performance. Second, the whole procedure can handle a general polyhedral uncertainty set that can better capture randomness. However, because of the complementary constraints (or linearized ones using big-M and binary variables) in solving SP$_k$, it is computationally very challenging. So, to solve large-scale real problems, we would recommend to use simple polyhedra in the form of (5.3) because SP$_k$ can be reformulated through strong duality into a much simpler MIP for fast computation (Zhao and Zeng (2012b) and Jiang et al. (2011)). The strong duality based reformulation of SP_k is presented in Appendix E and computational comparison between these two solution strategies for SP_k can be found in Appendix H.

5.4 Numerical Examples

In this section, we implement our proposed robust UC models and solution algorithms, numerically investigate their performances, and compare results with those from existing models, along with analysis and discussions to understand their differences and demonstrate advantages of our models.

5.4.1 Data and Experiment Setup

Our numerical study primarily considers to solve a unit commitment problem to minimize its total cost with 11 gas generators (at a single bus with no transmission network) over a time horizon of 24 time periods (hours). A data set with loads of 7 consecutive days is obtained from a utility company in Florida (see specifications on generators and loads in Zhao et al. (2013)). Those load data are used to define uncertainty sets and to generate random scenarios.

Specifically, for the expanded robust UC model presented in Appendix E, following the study in Bertsimas et al. (2013b), we consider the equation (E.6) to define two uncertainty sets for random load $(u_0, \ldots, u_{23})$, i.e., $\mathbb{U}_1$ and $\mathbb{U}_2$. Based on load data over 7 days, $\bar{u}_t$ is set to the average load (over 7 days) in time t, and $\hat{u}_{kt}$ is set to $1.5\sigma_t$ for $k = 1$ (in $\mathbb{U}_1$) and $3\sigma_t$ for $k = 2$ (in $\mathbb{U}_2$), for all t. For stochastic UC model, we randomly generate 500 scenarios to represent random loads, following the normal distribution with parameters $(\bar{u}_t,\, \sigma_t^2)$ for all t. For the risk constrained UC model, we define two uncertainty sets for G-1 and G-2 contingencies, i.e., all possibilities with up to one generator and two generators are down respectively. Load shed restrictions are also explicitly included. Complete formulation of the risk constrained UC model is presented in Appendix G.

To demonstrate the practical usefulness of proposed models, we also perform experiments on larger systems. Numerical results and discussions are presented in Appendix H.

Table 5.1: Expanded Robust UC with $\mathbb{U}_1$ and $\mathbb{U}_2$

Case	ρ_1, ρ_2	Γ_1, Γ_2	iter.	time(s)	Obj.
1	0.86, 0.14	12,12	2	0.869	973087.4
2		12,10	2	0.853	969668.16
3		12,8	2	0.837	966147.13
4		12,6	2	0.869	962286.85
5		12,4	2	0.837	958204.08
6		12,2	2	0.869	954041.31
7		6,6	2	0.79	929118.42
8		6,4	2	0.806	925028.02
9		6,2	2	0.885	920861.62
10	0.6, 0.4	12,12	2	0.79	994916.84
11		12,10	2	0.917	985147.59
12		12,8	2	0.932	975089.39
13		12,6	2	0.917	964060.01
14		12,4	2	0.933	952394.95
15		12,2	2	0.901	940490.81
16		6,6	2	0.837	940911
17		6,4	2	0.885	929224.14
18		6,2	2	0.821	917320.15

Our column-and-constraint generation algorithm is implemented in C++ and tested on a Dell Optiplex 760 desktop computer (Intel Core 2 Duo CPU, 3.0GHz, 3.25GB of RAM) in Windows XP environment. CPLEX 12.5 is adopted as the mixed integer programming solver. The optimality tolerance is set to 10^{-4}.

5.4.2 Expanded Robust UC

In the study of the expanded robust UC, we set parameter Γ_1 of $\mathbb{U}_1$ and Γ_2 of $\mathbb{U}_2$ in a way such that $\Gamma_2 \le \Gamma_1$, knowing that random loads are less likely to reach lower/upper bounds if the uncertain interval is larger. We consider two sets of weight coefficients, i.e., $\rho_1 = 0.86$ and $\rho_2 = 0.14$, and $\rho_1 = 0.6$ and $\rho_2 = 0.4$, in our computation. The first set is selected according to the fact that loads will fall within $\pm 1.5\sigma$ range with probability 0.86 under the normal distribution. The second set is simply selected to emphasize the importance of $\mathbb{U}_2$, which might lead to a more conservative solution. Computational results are presented in Table 5.1. In our discussions and analysis, we use objective values and total costs interchangeably for better exposition.

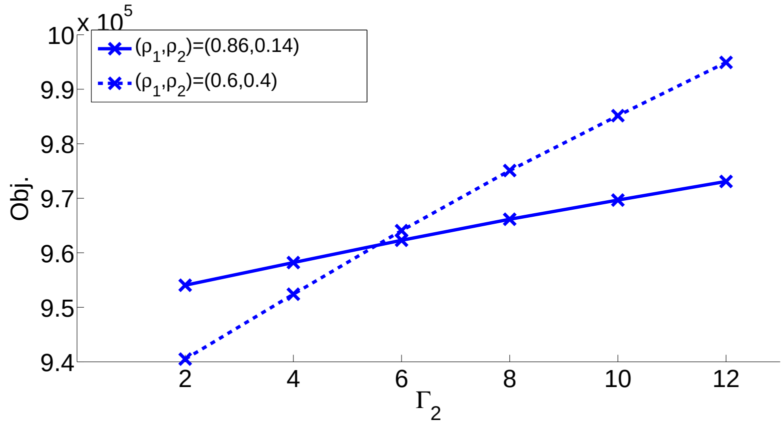


Figure 5.3: Objective Value vs. Γ_2

Similar to the classical robust UC, with Γ_1 and Γ_2 increase, which indicates both $\mathbb{U}_1$ and $\mathbb{U}_2$ become larger, the overall costs increase. Nevertheless, different interaction patterns between uncertainty sets can be observed under different (ρ_1, ρ_2). Figure 5.3 presents the total costs in Table 5.1 with Γ_1 fixed at 12, Γ_2 from 2 to 12, and (ρ_1, ρ_2) set to $(0.86, 0.14)$ and $(0.6, 04)$, respectively. Note that the curve of $(\rho_1, \rho_2) = (0.86, 0.14)$ is much stabler than that of $(\rho_1, \rho_2) = (0.6, 0.4)$, which agrees with the fact that $\mathbb{U}_1$ does not change and its stabilizing effect has a much larger impact when $\rho_1 = 0.86$. It is interesting to note that the curve of $(0.6, 0.4)$ changes from below the other one to clearly above it, with Γ_2 increase. The reason is that when $(\rho_1, \rho_2) = (0.6, 0.4)$, the worst case cost from $\mathbb{U}_2$ has a much larger impact. When Γ_2 is small, the worst case cost from $\mathbb{U}_2$ is low and it causes the total cost small. When Γ_2 is large, the worst case cost from $\mathbb{U}_2$ becomes high, which then dominates that from $\mathbb{U}_1$ and makes the total cost large. Those observations suggest that we can use multiple uncertainty sets (along with their weight coefficients) to model the impact of complicated randomness, which otherwise is challenging to be described by a single uncertainty set.

We also mention that although the expanded robust UC involves multiple uncertainty sets and recourse problems, it can be solved efficiently (within 2 iterations of column-and-constraint generation method). The solution time is comparable to that of the classical robust UC (Zhao and Zeng, 2012b). Given that the classical robust UC is currently evaluated and validated in real systems (Zheng et al. (2012) and Bertsimas et al. (2013b)), the good computational performance of the expanded robust UC ensures its computational feasibility in practice, as an improvement to the classical one.

A computational study of the expanded robust UC with respect to larger systems is presented in Appendix H. Basically, the computational expense could increase with respect to the problem size, which, however, is not drastic. More discussions can be found in Appendix H.

5.4.3 Robust UC and Stochastic UC

In this part, we compare the expanded robust UC with the classical robust UC and the stochastic UC models. We first present their total costs and computational times in Table 5.2 where (ρ_1, ρ_2) is set as $(0.86, 0.14)$ for the expanded robust UC model. When Γ ($= \Gamma_2$) in the classical robust UC is not very small, e.g. ≥ 6, it can be seen that results from the expanded robust UC are between the total cost of the stochastic UC model and those of the classical robust UC. It confirms our understanding that (i) if uncertainty sets are individual scenarios, the expanded robust UC reduces to the stochastic UC model, and (ii) if only one uncertainty presents, the expanded robust UC becomes the classical robust UC; (iii) the expanded model produces balanced solutions between those from the classical robust and stochastic UC models. It is worth mentioning that compared to the stochastic UC model, which takes more than 2000 seconds to compute, the expanded robust UC performs thousands of times faster and has a clear computation advantage. It indicates that a good trade-off among cost, risk, and computational time can be achieved by defining multiple uncertainty sets and by adopting the expanded robust UC.

We further analyze hourly unit commitment decisions made by those three models, which are presented in Table 5.3-5.5 respectively. Note that the unit commitment solution of the expanded robust model is almost the same as that of the stochastic UC model except that generator 1 is turned on for two more hours (hour 9 and 10). Nevertheless, compared with the expanded model, the classical robust model derives a more conservative schedule in which unit 6 has to be on for 23 hours. Those results again confirm the advantage of the expanded robust UC in deriving solutions that balance cost and risk.

Table 5.2: Comparison of Three Models

Stochastic UC		Expanded Robust UC				Classical Robust UC with $\mathbb{U}_2$			
time(s)	Obj.	Γ_1, Γ_2	iter.	time(s)	Obj.	$\Gamma = \Gamma_2$	iter.	time(s)	Obj.
2011.02	880424.3	12,12	2	0.869	973087.52	12	2	0.453	1045220.26
		12,10	2	0.853	969668.72	10	2	0.485	1020833.8
		12,8	2	0.837	966147.02	8	2	0.468	995708.48
		12,6	2	0.869	962286.8	6	2	0.5	968124.63
		12,4	2	0.837	958203.98	4	2	0.438	938907.5
		12,2	2	0.869	954041.22	2	2	0.453	909147.51

Table 5.3: Unit Commitment Decision from Stochastic UC

Unit	**Hour**																							
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 5.4: Unit Commitment Decision from Expanded Robust UC

Unit	**Hour**																							
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 5.5: Unit Commitment Decision of Classical Robust UC

Unit	Hour																							
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

5.4.4 Correlation in Expanded Robust UC

As shown in Zhao and Zeng (2012b) and Jiang et al. (2011), worst case situations for the popular hypercubic set with a budget constraint are those discrete points with u_t at its upper bound or at its nominal value for all t. So, instead of considering $u_t \in [\bar{u}_t - \hat{u}_t, \bar{u}_t + \hat{u}_t]$, we consider a discrete set with $u_t = \bar{u}_t + \hat{u}_t h_t$ where h_t is a binary variable for all t, along with the budget of uncertainty constraint. To eliminate unrealistic worst cases and to capture correlations, we use the load data set to define a $(-1, 0, 1)$ data matrix according to u_ts' behavior of reaching lower bounds (-1), upper bound (1) or otherwise (0). Then, a correlation analysis is done on that $(-1, 0, 1)$ matrix. When a perfect correlation is observed, an equality constraint in the form of (5.4) is included into the discrete version of the uncertainty set. Detailed analysis of $\mathbb{U}_1$ is presented in Appendix F. Because of limited information on u_t reaching bounds of $\mathbb{U}_2$, we keep $\mathbb{U}_2$ in its original formulation.

Numerical results for the expanded robust UC with the aforementioned correlation constraints are presented in Table 5.6, where relative changes of the total cost with respect to those in Table 5.1 are presented. Note that all the total costs are smaller than those in Table 5.1, which indicates that UC solutions are less conservative and also confirms that including correlation constraints eliminates some unrealistic worst case situations. We also observe that those correlation constraints generally reduce the computational time.

Table 5.6: Expanded Robust UC with Correlations

Case	ρ_1, ρ_2	Γ_1, Γ_2	iter.	time(s)	Obj.	ΔObj.(‰)
1		12,12	2	0.328	968499.5	-4.71
2		12,10	2	0.328	965078.38	-4.73
3		12,8	2	0.437	961551.87	-4.76
4		12,6	2	0.438	957664.21	-4.8
5	0.86, 0.14	12,4	2	0.437	953573.81	-4.83
6		12,2	2	0.422	949407.41	-4.86
7		6,6	2	0.625	922849.49	-6.75
8		6,4	2	0.594	918766.72	-6.77
9		6,2	2	0.562	914603.95	-6.8
10		12,12	2	0.438	991802.15	-3.13
11		12,10	2	0.469	982027.5	-3.17
12		12,8	2	0.469	971951.76	-3.22
13		12,6	2	0.438	960844.15	-3.34
14	0.6, 0.4	12,4	2	0.421	949157.3	-3.4
15		12,2	2	0.453	937253.3	-3.44
16		6,6	2	0.468	936545.58	-4.64
17		6,4	2	0.547	924880.51	-4.67
18		6,2	2	0.5	912986.9	-4.72

5.4.5 Risk Constrained Robust UC

Computational results of the risk constrained robust UC model with different γ_1 and γ_2 are presented in Table 5.7, where column *LS in G-1* and *LS in G-2* provide optimal solutions' load sheds in the worst case situations in G-1 and G-2 contingency sets.

Note from Table 5.7 that with γ_1 and γ_2 become more restrictive in G-1 and G-2 contingencies, solutions with higher total costs are derived. When constraints with $\gamma_1 = 200$ and $\gamma_2 = 2000$ are imposed, the model actually becomes infeasible. Hence, we can conclude that if the reliability standard with $\gamma_1 = 200$ and $\gamma_2 = 2000$ is required, the system needs to obtain and operate extra generators. Therefore, this model can also be treated as a decision support tool for system expansion under reliability consideration.

Table 5.7: Results of G-1/G-2 Risk Constrained Model

Case	γ_1, γ_2	iter.	time(s)	Obj.	LS in G-1	LS in G-2
r1	300,3000	5	70.203	885890.9	268.8	2912.15
r2	250,2500	5	44.312	886219.02	225.26	2499.03
r3	200,2000	4	10.35	Infeasible		

Figure 5.4 shows the behavior of our customized column-and-constraint generation method over iterations in solving Case r1. It can be seen that the algorithm starts with a solution

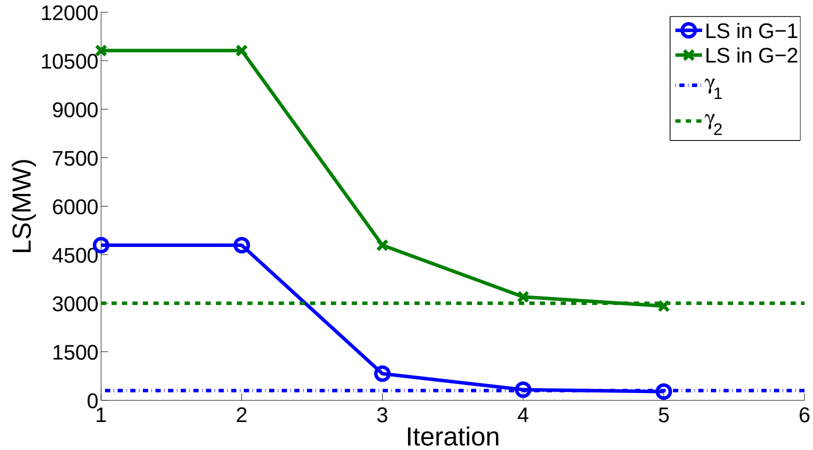


Figure 5.4: Load Sheds over Iterations for Case r1

that seriously violates load shed upper bounds and then progressively produces better ones to meet those requirements. Although there are $11 * 24 = 264$ and $11 \times 24 \times 10 \times 24 = 63360$ different contingencies in G-1 and G-2 sets, the algorithm quickly identifies significant ones (8 contingencies) and converges to an optimal solution in 5 iterations, which demonstrates a superior performance over the theoretical complexity presented in Proposition 8.

We note that, compared to the study for the expanded robust UC, the risk constrained model demands much more computation time. It suggests that imposing bounds on the worst case performances in a two-stage robust optimization model will significantly increase the computational burden. Such a computational challenge is also demonstrated in our numerical study with a system of 36 generators in Table H.3 in Appendix H.

5.5 Conclusion

In this chapter, we explore and extend the modeling capacity of two-stage robust optimization method. We demonstrate the improved capability by presenting two new robust unit commitment models, i.e., the expanded robust unit commitment and the risk constrained robust unit commitment models. We derive some structural properties, show that both models generalize or resemble the scenario based stochastic unit commitment models, and present a customized column-and-constraint generation method. We then perform a set of numerical experiments on those models to illustrate their modeling strength, economic outcomes with respect to related UC models under different uncertainty sets, and the algorithm performance in solving those models.

Although those unit commitment models improve our ability to capture uncertainties and handle risks in power systems, we mention that the presented research is a basic work in exploring robust optimization, a young optimization paradigm that may have many powerful modeling and solution features. For example, a natural extension is to adopt mixed integer recourse programs (Zhao and Zeng, 2012a) that can model quick-start generators and transmission line switching in the second stage (Hedman et al., 2010). Also, advanced methods need to be developed to strengthen uncertainty set description that capture essential dynamics of random renewable energy and also yield computational advantages (Papavasiliou and Oren (2013) and Constantinescu et al. (2011)). In particular, novel methods that analytically make use of existing data, such as the one in Bertsimas et al. (2013a), to construct computationally friendly uncertainty sets in a data-driven fashion are worth of a deep study. Moreover, the concepts of modeling presented in this chapter can be applied into other robust optimization applications, e.g., (Chen et al. (2012) and Zugno and Conejo (2013)), to address practical needs.

6 Conclusion

We successfully solve four reliability issues in transportation networks, distribution networks, and power generation systems in this dissertation.

In Chapter 2 and 3, a reliable design problem for hub-and-spoke air transportation systems subject to random hub failures is studied. At first, single disruption is considered in a compact stochastic model, which is the classical tool to deal with randomness. Due to the complexity of the problem, we applied Lagrangian relaxation and Branch-and-Bound to compute solutions efficiently. A set of experiments that can fully investigate the model performance under complicated condition (multiple correlated hub failures) is also developed. The model is shown to be able to hedge against random hub unavailabilities and significantly improve the served passengers when compared to classical models. Then multiple hub failures are included in a two-stage robust optimization model. Given its modelling advantages, a challenging factor (hub congestions) are also introduced into our formulation. A newly developed column-and-constraint generation method is utilized to obtain optimal solutions. Multiple meaningful insights are discovered from the model that simultaneously includes hub congestion and hub failures.

The projects in Chapter 2 and 3 establish a basis for the studies afterwards. Chapter 4 proposes a reliable p-median facility location problem. The modelling capability of two-stage is further explored and practical features (demand variation and facility capacity) widely neglected by previous studies are integrated into our model. We present a thorough research on the formulation structures followed by insightful computational experiments.

A breakthrough in modelling methodology is proposed in Chapter 5. We develop several variants of the two-stage robust optimization to better describe the randomness in real applications and build a connection between robust optimization and stochastic programming.

Besides the potential future works mentioned in conclusion sections of aforementioned chapters. We emphasize that the contributions made by this dissertation can be utilized to deal with the reliability issues of many other infrastructure systems, including homeland security and hydraulic engineering (Lewis (2006), Lansey et al. (1989), and Su et al. (1987)), and for the theoretical part, the results in Chapter 5 can serve as a foundation for developing approximation algorithms of stochastic programming models (Chen et al. (2007) and Bertsimas and Goyal (2010)).

References

Abreu, L.V., Khodayar, M.E., Shahidehpour, M., Wu, L., 2012. Risk-constrained coordination of cascaded hydro units with variable wind power generation. Sustainable Energy, IEEE Transactions on 3, 359–368.

AirlinesforAmerica, 2014. Annual U.S. Impact of Flight Delays (NEXTOR report). http://www.airlines.org/data/annual-u-s-impact-of-flight-delays-nextor-report/.

Alumur, S., Kara, B.Y., 2008. Network hub location problems: The state of the art. European Journal of Operational Research 190, 1–21.

Amin, M., 2003. North america's electricity infrastructure: Are we ready for more perfect storms? IEEE Security & Privacy 1, 19–25.

An, Y., Zeng, B., Zhang, Y., Zhao, L., 2014. Reliable p-median facility location problem: two-stage robust models and algorithms. Transportation Research Part B 64, 54–72.

An, Y., Zhang, Y., Zeng, B., 2011. The reliable hub-and-spoke design problem: models and algorithms. Technical Report. under reivew, University of South Florida.

Atamturk, A., Zhang, M., 2007. Two-atage robust network flow and design under demand uncertainty. Operations Research 55, 662–673.

Ball, M., Barnhart, C., Nemhauser, G., Odoni, A., 2006. Air transportation: irregular operations and control, in: Handbooks of Operations Research and Management, North-Holland.

Barahona, F., Jensen, D., 1998. Plant location with minimum inventory. Mathematical Programming 83, 101–111.

Baron, O., Milner, J., Naseraldin, H., 2011. Facility location: a robust optimization approach. Production and Operations Management 20, 772–785.

BBC, 2011a. Iceland volcano ash: German air traffic resuming. http://www.bbc.co.uk/news/world-europe-13535054.

BBC, 2011b. Iceland's Grimsvotn volcano starts new eruption. http://www.bbc.co.uk/news/world-europe-13487858.

Ben-Tal, A., El Ghaoui, L., Nemirovski, A., 2009. Robust Optimization. Princeton Univ Press.

Ben-Tal, A., Goryashko, A., Guslitzer, E., Nemirovski, A., 2004. Adjustable robust solutions of uncertain linear programs. Mathematical Programming 99, 351–376.

Ben-Tal, A., Nemirovski, A., 1998. Robust convex optimization. Mathematics of Operations Research 23, 769–805.

Ben-Tal, A., Nemirovski, A., 1999. Robust solutions of uncertain linear programs. Operations Research Letters 25, 1–14.

Ben-Tal, A., Nemirovski, A., 2000. Robust solutions of linear programming problems contaminated with uncertain data. Mathematical Programming 88, 411–424.

Ben-Tal, A., Nemirovski, A., 2008. Selected topics in robust convex optimization. Mathematical Programming 112, 125–158.

Berman, O., Krass, D., Menezes, M., 2007. Facility reliability issues in network p-median problems: strategic centralization and co-location effects. Operations Research 55, 332–350.

Bertsimas, D., Brown, D.B., Caramanis, C., 2011. Theory and applications of robust optimization. SIAM review 53, 464–501.

Bertsimas, D., Goyal, V., 2010. On the power of robust solutions in two-stage stochastic and adaptive optimization problems. Mathematics of Operations Research 35, 284–305.

Bertsimas, D., Gupta, V., Kallus, N., 2013a. Data-Driven Robust Optimization. Technical Report. Massachusetts Institute of Technology.

Bertsimas, D., Litvinov, E., Sun, X., Zhao, J., Zheng, T., 2013b. Adaptive robust optimization for the security constrained unit commitment problem. IEEE Transactions on Power Systems 28, 52–63.

Bertsimas, D., Sim, M., 2003. Robust discrete optimization and network flows. Mathematical Programming 98, 49–71.

Bertsimas, D., Sim, M., 2004. The price of robustness. Operations Research 52, 35–53.

Bratu, S., Barnhart, C., 2006. Flight operations recovery: New approaches considering passenger recovery. Journal of Scheduling 9, 279–298.

BureauofTransportationStatistics, 2014. Bureau of Transportation Statistics T-100 Market data. http://www.transtats.bts.gov/Data_Elements.aspx?Data=1.

de Camargo, R.S., Miranda, Jr., G., Luna, H., 2009a. Benders decomposition for hub location problems with economies of scale. Transportation Science 43, 86–97.

de Camargo, R.S., Miranda Jr, G., Ferreira, R.P.M., Luna, H., 2009b. Multiple allocation hub-and-spoke network design under hub congestion. Computers & Operations Research 36, 3097–3106.

Campbell, J., O'Kelly, M., 2012. Twenty-five years of hub location research. Transportation Science 46, 153–169.

Campell, J., 1994. Integer programming formulations of discrete hub location problems. European Journal of Operational Research 72, 387–405.

Cánovas, L., García, S., Marín, A., 2007. Solving the uncapacitated multiple allocation hub location problem by means of a dual-ascent technique. European Journal of Operational Research 179, 900–1007.

Chaovalitwongse, W., Pardalos, P.M., Prokopyev, O.A., 2004. A new linearization technique for multi-quadratic 0-1 programming problems. Operations Research Letters 32, 517–522.

Chen, C., Li, Y., Huang, G., Li, Y., 2012. A robust optimization method for planning regional-scale electric power systems and managing carbon dioxide. International Journal of Electrical Power & Energy Systems 40, 70–84.

Chen, J., 2007. A hybrid heuristic for the uncapacitated single allocation hub location problem. Omega 35, 211–220.

Chen, Q., Li, X., Ouyang, Y., 2011. Joint inventory-location problem under the risk of probabilistic facility disruptions. Transportation Research Part B 45, 991–1003.

Chen, X., Sim, M., Sun, P., 2007. A robust optimization perspective on stochastic programming. Operations Research 55, 1058–1071.

Conrad, S.H., LeClaire, R.J., O'Reilly, G.P., Uzunalioglu, H., 2006. Critical national infrastructure reliability modeling and analysis. Bell Labs Technical Journal 11, 57–71.

Constantinescu, E.M., Zavala, V.M., Rocklin, M., Lee, S., Anitescu, M., 2011. A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation. Power Systems, IEEE Transactions on 26, 431–441.

Contreras, I., Cordeau, J., Laporte, G., 2011a. Benders decomposition for large-scale uncapacitated hub location. Operations Research 59, 1477–1490.

Contreras, I., Cordeau, J., Laporte, G., 2011b. The dynamic uncapacitated hub location problem. Transportation Science 45, 18–32.

Contreras, I., Cordeau, J., Laporte, G., 2011c. Stochastic uncapacitated hub location. European Journal of Operational Research 212, 518–528.

Contreras, I., Cordeau, J., Laporte, G., 2012. Exact solution of large-scale hub location problems with multiple capacity levels. Transportation Science 46, 439–459.

Cui, T., Ouyang, Y., Shen, Z.M., 2010. Reliable facility location design under the risk of disruptions. Operations Research 58, 998–1011.

Cunha, C.B., Silva, M.R., 2007. A genetic algorithm for the problem of configuring a hub-and-spoke network for a LTL trucking company in brazil. European Journal of Operational Research 179, 747–758.

Daskin, M., 1995. Network and Discrete Location: Models, Algorithms, and Applications. Wiley-Interscience.

Drezner, E., 1995. Facility Location: A Survey of Applications and Methods. Springer.

Drezner, Z., 1987. Heuristic solution methods for two location problems with unreliable facilities. Journal of the Operational Research Society 38, 509–514.

Dueñas-Osorio, L., Vemuru, S.M., 2009. Cascading failures in complex infrastructure systems. Structural safety 31, 157–167.

Elhedhli, S., Hu, F.X., 2005. Hub-and-spoke network design with congestion. Computers & Operations Research 32, 1615–1632.

Elhedhli, S., Wu, H., 2010. A lagrangean heuristic for hub-and-spoke system design with capacity selection and congestion. Informs Journal on Computing 22, 282–296.

Ergun, O., Karakus, G., Keskinocak, P., Swann, J., Villarreal, M., 2010. Operations research to improve disaster supply chain management. Wiley Encyclopedia of Operations Research and Management Science .

Ernst, A.T., Krishnamoorthy, M., 1996. Efficient algorithms for the uncapacitated single allocation p-hub median problem. Location Science 4, 139–154.

Ernst, A.T., Krishnamoorthy, M., 1998a. Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. European Journal of Operational Research 104, 100–112.

Ernst, A.T., Krishnamoorthy, M., 1998b. An exact solution approach based on shortest-paths for p-hub median problems. Informs Journal on Computing 10, 149–162.

Fisher, M.L., 2004. The Lagrangian relaxation method for solving integer programming problems. Management Science 50, 1861–1871.

Flylowcostairlines.com, 2012. Airport Hamburg (HAM).

Frangopol, D.M., Liu, M., 2007. Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost. Structure and infrastructure engineering 3, 29–41.

Gabrel, V., Lacroix, M., Murat, C., Remli, N., 2014. Robust location transportation problems under uncertain demands. Discrete Applied Mathematics 164, 100–111.

Gil, N., Miozzo, M., Massini, S., 2012. The innovation potential of new infrastructure development: An empirical study of heathrow airport's t5 project. Research Policy 41, 452–466.

Grossman, D., 2010. How the Continental-United merger will affect business travelers. http://www.usatoday.com/travel/columnist/grossman/2010-05-04-continental-united-merger_N.htm.

Grove, G.P., O'Kelly, M.E., 1986. Hub networks and simulated schedule delay. Papers of The Regional Science Association 59, 103–119.

Gülpınar, N., Pachamanova, D., Çanakoğlu, E., 2012. Robust strategies for facility location under uncertainty. European Journal of Operational Research 225, 21–35.

Hajimiragha, A.H., Canizares, C.A., Fowler, M.W., Moazeni, S., Elkamel, A., 2011. A robust optimization approach for planning the transition to plug-in hybrid electric vehicles. Power Systems, IEEE Transactions on 26, 2264–2274.

He, X., Chen, A., Chaovalitwongse, W.A., Liu, H.X., 2012. An improved linearization technique for a class of quadratic 0-1 programming problems. Optimization Letters 6, 31–41.

Hedman, K.W., Ferris, M.C., O'Neill, R.P., Fisher, E.B., Oren, S.S., 2010. Co-optimization of generation unit commitment and transmission switching with n-1 reliability. Power Systems, IEEE Transactions on 25, 1052–1063.

Hervet, C., Faye, A., Costa, M.C., Chardy, M., Francfort, S., 2013. Solving the two-stage robust FTTH network design problem under demand uncertainty. Electronic Notes in Discrete Mathematics 41, 335–342.

InteractiveAdvertisingBureau, 2014. IAB Internet Advertising Revenue Report. http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_Report_FY_2011.pdf.

Janic, M., 2005. Modeling the large scale disruption of an airline network. Journal of Transportation Engineering 131, 249–260.

Jergler, D., 2014. Southwest Power Outage Economic Cost Put At $100M. http://www.insurancejournal.com/news/west/2011/09/13/215102.htm.

Jia, H., Ordóñez, F., Dessouky, M., 2007. A modeling framework for facility location of medical services for large-scale emergencies. IIE Transactions 39, 41–55.

Jiang, R., Wang, J., Guan, Y., 2012. Robust unit commitment with wind power and pumped storage hydro. Power Systems, IEEE Transactions on 27, 800–810.

Jiang, R., Zhang, M., Li, G., Guan, Y., 2011. Benders decomposition for the two-stage security constrained robust unit commitment problem. Technical Report. University of Florida.

Kim, H., 2008. Reliable p-hub location problems and protection models for hub network design. Ph.D. thesis. The Ohio State University.

Kim, H., O'Kelly, M.E., 2009. Reliable p-hub location problems in telecommunication networks. Geographical Analysis 41, 283–306.

Lansey, K.E., Duan, N., Mays, L.W., Tung, Y.K., 1989. Water distribution system design under uncertainties. Journal of Water Resources Planning and Management 115, 630–645.

Lewis, T.G., 2006. Critical infrastructure protection in homeland security: defending a networked nation. John Wiley & Sons.

Li, Q., 2011. Decision Support Models for Design of Fortified Distribution Networks. Ph.D. thesis. University of South Florida.

Li, T., Shahidehpour, M., 2007. Risk-constrained generation asset arbitrage in power systems. Power Systems, IEEE Transactions on 22, 1330–1339.

Li, T., Shahidehpour, M., Li, Z., 2007. Risk-constrained bidding strategy with stochastic unit commitment. Power Systems, IEEE Transactions on 22, 449–458.

Li, X., Ouyang, Y., 2010. A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions. Transportation Research Part B 44, 535–548.

Li, X., Ouyang, Y., 2011. Reliable sensor deployment for network traffic surveillance. Transportation Research Part B: Methodological 45, 218–231.

Lim, M., Daskin, M.S., Bassamboo, A., Chopra, S., 2009. A facility reliability problem: Formulation, properties, and algorithm. Naval Research Logistics 57, 58–70.

Lim, M.K., Bassamboo, A., Chopra, S., Daskin, M.S., 2013. Facility location decisions with random disruptions and imperfect estimation. Manufacturing & Service Operations Management 15, 239–249.

Løve, M., Sørensen, K., 2001. Disruption management in the airline industry. Institut for Matematisk Modellering, Danmarks Tekniske Universitet.

Ma, H., Shahidehpour, S., 1999. Unit commitment with transmission security and voltage constraints. Power Systems, IEEE Transactions on 14, 757–764.

Martin, P., Rogers, C.A., 1995. Industrial location and public infrastructure. Journal of International Economics 39, 335–351.

Melo, M., Nickel, S., Saldanha-Da-Gama, F., 2009. Facility location and supply chain management-a review. European Journal of Operational Research 196, 401–412.

MinnesotaInternetTrafficStudies, 2014. Internet Traffic. http://en.wikipedia.org/wiki/Internet_traffic#Internet_backbone_traffic_in_the_United_States.

Moslehi, K., Kumar, R., 2010. A reliability perspective of the smart grid. Smart Grid, IEEE Transactions on 1, 57–64.

Murray, A.T., Grubesic, T., 2007a. Critical Infrastructure. Springer.

Murray, A.T., Grubesic, T.H., 2007b. Overview of reliability and vulnerability in critical infrastructure, in: Critical Infrastructure. Springer, pp. 1–8.

Mygermancity.com, 2012. Bremen Airport (BRE) - Interplanetary Hub For The North.

Nemhauser, G.L., Wolsey, L.A., 1988. Integer and Combinatorial Optimization. Wiley.

NorthAmericanElectricReliabilityCorporation(NERC), 2014. September 2011 Southwest Blackout Event. http://www.nerc.com/pa/rrm/ea/Pages/September-2011-Southwest-Blackout-Event.aspx.

OfficialWebsiteofDepartmentofHomelandSecurity, 2014. Critical Infrastructure. http://www.dhs.gov/critical-infrastructure.

O'Kelly, M.E., 1987. A quadratic integer program for the location of interacting hub facilities. European Journal of Operational Research 32, 393–404.

O'Kelly, M.E., 1992. Hub facility location with fixed costs. Papers in Regional Science 71, 293–306.

O'Kelly, M.E., Kim, H., Kim, C., 2006. Internet reliability with realistic peering. Environment and Planning Part B 33, 325–343.

O'Kelly, M.E., Skorin-Kapov, J., et al., 1996. Hub network design with single and multiple allocation: A computational study. Location Science 4, 125–138.

OnlineCompactOxfordEnglishDictionary, 2014. Infrastructure. http://www.oxforddictionaries.com/us/definition/american_english/infrastructure.

Ostrom, E., Schroeder, L., Wynne, S., et al., 1993. Institutional Incentives and Sustainable Development: Infrastructure Policies in Perspective. Westview Press.

Ostrowski, J., Anjos, M.F., Vannelli, A., 2012. Tight mixed integer linear programming formulations for the unit commitment problem. Power Systems, IEEE Transactions on 27, 39–46.

Palpant, M., Boudia, M., Robelin, C., Gabteni, S., Laburthe, F., 2009. ROADEF 2009 Challenge: Disruption management for commercial aviation. Technical Report. Working paper, Amadeus SAS, Operations Research Division, Sophia Antipolis, France.

Papavasiliou, A., Oren, S.S., 2012. Applying high performance computing to multi-area stochastic unit commitment for renewable penetration, in: FERC technical conference: Increasing Real-Time and Day-Ahead Market Efficiency through Improved Software, Washington DC.

Papavasiliou, A., Oren, S.S., 2013. Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. Operations Research 61, 578–592.

Pederson, P., Dudenhoeffer, D., Hartley, S., Permann, M., 2006. Critical infrastructure interdependency modeling: a survey of us and international research. Idaho National Laboratory , 1–20.

Peng, P., Snyder, L., Lim, A., Liu, Z., 2011. Reliable logistics networks design with facility disruptions. Transportation Research Part B 45, 1190–1211.

Pirkul, H., Schilling, D.A., 1998. An efficient procedure for designing single allocation hub and spoke systems. Management Science 44, S235–S242.

Pita, J., Barnhart, C., Antunes, A., 2012. Integrated flight scheduling and fleet assignment under airport congestion. Transportation Science .

Price, H.J., 2014. Press Release Airline Passenger Travel to Nearly Double in Two Decades. http://www.faa.gov/news/press_releases/news_story.cfm?newsId=13394.

Revelle, C., Eiselt, H., Daskin, M., 2008. A bibliography for some fundamental problem categories in discrete location science. European Journal of Operational Research 184, 817–848.

Rinaldi, S.M., Peerenboom, J.P., Kelly, T.K., 2001. Identifying, understanding, and analyzing critical infrastructure interdependencies. Control Systems, IEEE 21, 11–25.

Saranga, H., Kumar, U.D., 2006. Optimization of aircraft maintenance/support infrastructure using genetic algorithmslevel of repair analysis. Annals of Operations Research 143, 91–106.

Shen, Z., Zhan, R., Zhang, J., 2011. The reliable facility location problem: formulations, heuristics, and approximation algorithms. INFORMS Journal on Computing 23, 470.

Sherali, H.D., Smith, J.C., 2007. An improved linearization strategy for zero-one quadratic programming problems. Optimization Letters 1, 33–47.

Skorin-Kapov, D., Skorin-Kapov, J., O'Kelly, M.E., 1996. Tight linear programming relaxations of uncapacitated p-hub median problems. European Journal of Operational Research 94, 582–593.

Smith, J.C., Lim, C., Sudargho, F., 2007. Survivable network design under optimal and heuristic interdiction scenarios. Journal of Global Optimization 38, 181–199.

Snyder, L., 2006. Facility location under uncertainty: a review. IIE Transactions 38, 547–564.

Snyder, L., Atan, Z., Peng, P., Rong, Y., Schmitt, A., Sinsoysal, B., 2012. OR/MS models for supply chain disruptions: a review. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1689882 (accessed on 1.19.2013).

Snyder, L., Daskin, M.S., 2005. Reliability models for facility location: The expected failure cost case. Transportation Science 39, 400–416.

Sohn, J., Park, S., 1998. Efficient solution procedure and reduced size formulations for p-hub location problems. European Journal of Operational Research 108, 118–126.

Sohn, J., Park, S., 2000. The single allocation problem in the interacting three-hub network. Networks 35, 17–25.

Sridharan, R., 1995. The capacitated plant location problem. European Journal of Operational Research 87, 203–213.

Street, A., Oliveira, F., Arroyo, J.M., 2011. Contingency-constrained unit commitment with n-k security criterion: A robust optimization approach. Power Systems, IEEE Transactions on 26, 1581–1590.

Su, Y.C., Mays, L.W., Duan, N., Lansey, K.E., 1987. Reliability-based optimization model for water distribution systems. Journal of Hydraulic Engineering 113, 1539–1556.

Takriti, S., Birge, J.R., Long, E., 1996. A stochastic model for the unit commitment problem. Power Systems, IEEE Transactions on 11, 1497–1508.

Teo, C., Shu, J., 2004. Warehouse-retailer network design problem. Operations Research 52, 396–408.

Wang, X., Nieuwesteeg, P., Listes, O., Bresler, S., Ogburn, R., 2012. A robust look-ahead unit commitment, in: Power and Energy Society General Meeting, 2012 IEEE, pp. 1–4.

Wang, X., Ouyang, Y., 2013. A continuum approximation approach to competitive facility location design under facility disruption risks. Transportation Research Part B 50, 90–103.

Welman, S., Williams, A., Hechtman, D., 2010. Calculating delay propagation multipliers for cost-benefit analysis. Center for Advanced Aviation System Development and MITRE MP 100039.

Wu, L., Shahidehpour, M., Li, Z., 2008. Genco's risk-constrained hydrothermal scheduling. Power Systems, IEEE Transactions on 23, 1847–1858.

Xiong, P., Jirutitijaroen, P., 2012. An adjustable robust optimization approach for unit commitment under outage contingencies, in: Power and Energy Society General Meeting, 2012 IEEE, IEEE. pp. 1–8.

Zeng, B., An, Y., Zhang, Y., Kim, H., 2010. A reliable hub-spoke model in transportation systems, in: The 4th International Symposium on Transportation Network Reliability (INSTR) Conference Proceedings.

Zeng, B., Zhao, L., 2013. Solving two-stage robust optimization problems using a column-and-constraint generation method. Operations Research Letters 41, 457–461.

Zhao, J., Zheng, T., Litvinov, E., England, I.N., 2010. Enhancing reliability unit commitment with robust optimization, in: FERC technical conference: Enhanced ISO and RTO unit commitment models, Washington DC.

Zhao, L., Zeng, B., 2012a. An exact algorithm for two-stage robust optimization with mixed integer recourse problems. Technical Report. Under Revision, University of South Florida.

Zhao, L., Zeng, B., 2012b. Robust unit commitment problem with demand response and wind energy, in: Power and Energy Society General Meeting, 2012 IEEE, IEEE. pp. 1–8.

Zhao, L., Zeng, B., Buckley, B., 2013. A stochastic unit commitment model with cooling systems. Power Systems, IEEE Transactions on 28, 211–218.

Zheng, T., Zhao, J., Litvinov, E., Zhao, F., 2012. Robust optimization and its application to power system operation, in: Proceedings of CIGRÉ.

Zugno, M., Conejo, A.J., 2013. A robust optimization approach to energy and reserve dispatch in electricity markets. Technical Report. Technical University of Denmark.

Appendices

Appendix A Linearization Techniques and CPLEX Performance

We first introduce the standard linearization of R-SAHMPA, linear reformulation for R-SAHMP is obtained by using standard linearization techniques (Nemhauser and Wolsey, 1988), which is denoted by StdLinear.

$$
\begin{aligned}
\min \; & \sum_{i}\sum_{k\neq i}\sum_{m}\sum_{\substack{j>i\\ j\neq m}} F_{ikmj}w_{ij}(1-q_k-q_m^k)X_{ikmj} \\
& + \sum_{i}\sum_{j>i}\Big(\sum_{m\neq j} F_{iimj}w_{ij}(1-q_m^i)X_{iimj} + \sum_{k\neq i} F_{ikjj}w_{ij}(1-q_k^j)X_{ikjj} + F_{iijj}w_{ij}X_{iijj}\Big) \\
& + \sum_{i}\sum_{k}\sum_{m\neq k}\sum_{j>i}\sum_{n}\rho(F_{inmj}w_{ij}q_kZ^1_{ikmjn} + F_{iknj}w_{ij}q_mZ^2_{ikmjn}) \\
& + \sum_{i}\sum_{k}\sum_{j>i}\sum_{n}\rho F_{innj}w_{ij}q_kZ^1_{ikkj}
\end{aligned} \tag{A.1}
$$

s.t.

$$\text{Constraints (2.2) – (2.10)} \tag{A.2}$$

$$Z^1_{ikmjn} \leq X_{ikmj},\; Z^1_{ikmjn} \leq U_{ijn},\; Z^1_{ikmjn} \geq X_{ikmj} + U_{ijn} - 1 \quad \forall i,k,m,j>i,n \tag{A.3}$$

$$Z^2_{ikmjn} \leq X_{ikmj},\; Z^2_{ikmjn} \leq V_{ijn},\; Z^2_{ikmjn} \geq X_{ikmj} + V_{ijn} - 1 \quad \forall i,k,m,j>i,n \tag{A.4}$$

$$Z^1_{ikmjn}, Z^2_{ikmjn} \geq 0 \quad \forall\; i,k,m,j>i,n. \tag{A.5}$$

Compared with the quadratic form in (2.2)-(2.10), $X_{ikmj}U_{ijn}$ is replaced by Z^1_{ikmjn} and $X_{ikmj}V_{ijn}$ is replaced by Z^2_{ikmjn}. Also, a few sets of constraints are added to enforce that $Z^1_{ikmjn} = X_{ikmj}U_{ijn}$ and $Z^2_{ikmjn} = X_{ikmj}V_{ijn}$. Note that this mixed integer linear reformulation has to deal with a huge number of additional variables and constraints.

Next we give the compact linear reformulation of R-SAHMP. A recent linearization approach and its variants are developed for quadratic 0-1 programs to obtain a compact linear reformulation (see Chaovalitwongse et al. (2004), Sherali and Smith (2007), and He et al. (2012)). While the standard one introduces a quadratic number of extra variables and constraints, this type of linearization method introduces only a linear number of extra variables and constraints. Therefore, we adopt and extend this linearization technique to reformulate our quadratic R-SAHMP model.

First, we point out that $\sum_k \sum_{m \neq k} X_{ikmj} \in \{0,1\}$ for all $i, j > i$. Because this expression appears in the objective function of R-SAHMP, we can treat it simply as a binary variable as a whole and perform linearization with respect to $\sum_k \sum_{m \neq k} X_{ikmj}$ and U_{ijn} ($\sum_{k \neq m} \sum_m X_{ikmj}$ and V_{ijn} can be linearized similarly). We obtain the compact CptLinear formulation as follows.

$$\min \sum_i \sum_{k \neq i} \sum_m \sum_{\substack{j>i \\ j \neq m}} F_{ikmj} w_{ij} (1 - q_k - q_m^k) X_{ikmj}$$

$$+ \sum_i \sum_{j>i} \Big(\sum_{m \neq j} F_{iimj} w_{ij} (1 - q_m^i) X_{iimj} + \sum_{k \neq i} F_{ikjj} w_{ij} (1 - q_k^j) X_{ikjj} + F_{iijj} w_{ij} X_{iijj} \Big)$$

$$+ \sum_i \sum_{j>i} \sum_n (\Omega_{ijn} - \sigma_{ij} U_{ijn})$$

$$+ \sum_i \sum_{j>i} \sum_n (\Theta_{ijn} - \sigma_{ij} V_{ijn})$$

$$+ \sum_i \sum_{j>i} \sum_n (\Gamma_{ijn} - \sigma_{ij} U_{ijn})$$

s.t.

Constraints (2.2) – (2.10)

$$\sum_k \sum_{m \neq k} \rho w_{ij} q_k F_{inmj} X_{ikmj} - s_{ijn} + \sigma_{ij} = \Omega_{ijn} \qquad \forall i, j > i, n \tag{A.6}$$

$$s_{ijn} \leq (\mu_{ij} + \sigma_{ij})(1 - U_{ijn}) \qquad \forall i, j > i, n \tag{A.7}$$

$$\sum_{k \neq m} \sum_m \rho w_{ij} q_m F_{iknj} X_{ikmj} - t_{ijn} + \sigma_{ij} = \Theta_{ijn} \qquad \forall i, j > i, n \tag{A.8}$$

$$t_{ijn} \leq (\mu_{ij} + \sigma_{ij})(1 - V_{ijn}) \qquad \forall i, j > i, n \tag{A.9}$$

$$\sum_k \rho w_{ij} q_k F_{innj} X_{ikkj} - r_{ijn} + \sigma_{ij} = \Gamma_{ijn} \qquad \forall i, j > i, n \tag{A.10}$$

$$r_{ijn} \leq (\mu_{ij} + \sigma_{ij})(1 - U_{ijn}) \qquad \forall i, j > i, n \tag{A.11}$$

$$\Omega_{ijn}, s_{ijn}, \Theta_{ijn}, t_{ijn}, \Gamma_{ijn}, r_{ijn} \geq 0 \qquad \forall i, j > i, n, \tag{A.12}$$

where $\mu_{ij} = \rho w_{ij} \max_{k,m}\{F_{ikmj}\} \max_k \{q_k\}$, and $\sigma_{ij} \geq 0$ is a predetermined coefficient for $i, j > i$. In our numerical study, we set $\sigma_{ij} = 0$ for all i and j. The linearizataion for the quadratic term $\sum_i \sum_k \sum_{m \neq k} \sum_{j>i} \sum_n \rho F_{inmj} w_{ij} q_k X_{ikmj} U_{ijn}$ in R-SAHMP is completed by new variables Ω_{ijn}, s_{ijn} and constraints (A.6) and (A.7). Similarly, Θ_{ijn}, t_{ijn}, (A.8) and (A.9) are used for linearization of $\sum_i \sum_{k \neq m} \sum_m \sum_{j>i} \sum_n \rho F_{iknj} w_{ij} q_m X_{ikmj} V_{ijn}$; Γ_{ijn}, r_{ijn}, (A.10) and (A.11) are for

$$\sum_i \sum_k \sum_{j>i} \sum_n \rho F_{innj} w_{ij} q_k X_{ikkj} U_{ijn}.$$

As a result, following Theorem 1 of He et al. (2012) and using the fact that $\sum_i \sum_{j>i} X_{ikmj}$ takes only a binary value for all k and m, we have

Proposition 9 *The linear CptLinear formulation is equivalent to the quadratic R-SAHMP model. An optimal solution to CptLinear yields an optimal solution to R-SAHMP.*

We mention that if we simply consider the linearization of $X_{ikmj} U_{ijn}$ and $X_{ikmj} V_{ijn}$, the number of additional variables for linearization will be of $O(|\mathbf{N}|^4)$. Nevertheless, due to the variable reduction by considering $\sum_k \sum_{m \neq k} X_{ikmj} / \sum_{k \neq m} \sum_m X_{ikmj}$ as a single binary variable, the number of additional variables in CptLinear is of $O(|\mathbf{N}|^3)$. Given that the number of variables in R-SAHMP is of $O(|\mathbf{N}|^4)$, CptLinear distinguishes itself by the fact that the number of variables does little change.

Appendix B Sample Disruption Probabilities for CAB Data Set

Table B.1: Disruption Probabilities of Potential Hubs in Reliable Model

No.	City	q value	No.	City	q value	No.	City	q value
0	Atlanta	0.023	9	Houston	0.026	18	Phoenix	0.045
1	Baltimore	0.017	10	Kansas City	0.018	19	Pittsburgh	0.012
2	Boston	0.047	11	Los Angeles	0.049	20	St. Louis	0.035
3	Chicago	0.041	12	Memphis	0.024	21	San Francisco	0.043
4	Cincinnati	0.026	13	Miami	0.027	22	Seattle	0.020
5	Cleveland	0.047	14	Minneapolis	0.013	23	Tampa	0.036
6	Dallas-Fort Worth	0.012	15	New Orleans	0.019	24	Washington DC	0.050
7	Denver	0.015	16	New York	0.050			
8	Detroit	0.035	17	Philadelphia	0.024			

Appendix C Disruption Probabilities in Involving Correlations

The probabilities of two nodes $P(D_k = 0, D_m = 0)$ and $P(D_k = 1, D_n = 0)$ are calculated as follows.

Assume that $corr(D_k, D_m) = f(c_{km})$ for any k and m. We let $f(x)$ takes the form $\mathrm{e}^{-\frac{x}{200}}$ so that the corrlation becomes smaller as the distance between two nodes increases.

Since D_k is a binary random variable, $E(D_k) = 0 * P(D_k = 0) + 1 * P(D_k = 1)$, we have

$$E(D_k) = q_k.$$

So,

$$\frac{E[(D_k - q_k)((D_m - q_m)]}{\sigma_{D_k}\sigma_{D_m}} = corr(D_k, D_m),$$

$$\frac{E(D_k D_m) - q_k q_m}{\sigma_{D_k}\sigma_{D_m}} = corr(D_k, D_m).$$

Noting that D_k^2 and $D_k D_m$ are both binary random variable,

$$E(D_k D_m) = P(D_k = 1, D_m = 1),$$

$$\sigma_{D_k}^2 = E(D_k^2) - E^2(D_k) = q_k - q_k^2.$$

Hence we can obtain $P(D_k = 1, D_m = 1)$ by the following equation:

$$P(D_k = 1, D_m = 1) = corr(D_k, D_m)\sqrt{q_k - q_k^2}\sqrt{q_m - q_m^2} + q_k q_m.$$

Then,

$$P(D_k = 1, D_m = 0) = P(D_k = 1) - P(D_k = 1, D_m = 1),$$

$P(D_k = 0, D_m = 1)$ can be obtained similarly.

We also need $P(D_m = 0, D_k = 0)$:

$$P(D_m = 0, D_k = 0) = 1 - P(D_m = 1, D_k = 1) \\ - P(D_m = 0, D_k = 1) - P(D_m = 1, D_k = 0).$$

The probabilities involving three nodes like $P(D_k = 1, D_m = 0, D_n = 0)$ and $P(D_k = 0, D_m = 1, D_n = 0)$ are obtained as follows.

For any different nodes k, m, and n. We need to make further assumptions: fix $D_k = 1$, assume a new probability $P(D_m = 0, D_n = 0|D_k = 1)$:

$$P(D_m = 0, D_n = 0|D_k = 1) = P(D_m = 0, D_n = 0)(1 - \frac{e^{-\frac{c_{km}+c_{kn}}{2}}}{10}).$$

Then, $P(D_m = 0, D_n = 0, D_k = 1) = P(D_m = 0, D_n = 0|D_k = 1)P(D_k = 1)$. $P(D_k = 0, D_m = 1, D_n = 0)$ can be obtained.

Appendix D Nomenclature for Chapter 5

Table D.1: Indices and Sets

i	Generator, $i = 0, 1, ..., I - 1$
t	Planning period, $t = 0, 1, ..., T - 1$
k	Uncertainty set, $k = 1, ..., K$
n	Constraint index
j	Column index of the data matrix (for j-th day)
$\mathbb{U}_k$	The k-th uncertainty set
Ω, Ω_k	Set of feasible solutions of recourse problem

Table D.2: Model Parameters

a_i	Start up cost of unit i
r_i	Running cost of unit i
c_i	Fuel cost of unit i
q_t^+	Purchase price at time t in power market
q_t^-	Sale price at time t in power market
ρ_k	Weight of the worst case cost of the k-th uncertainty set
$\bar{u}_t$	Nominal value of load at time t
$\hat{u}_t(\hat{u}_{kt})$	Bound on load deviation at time t for the (k-th) uncertainty set
σ_t	Standard deviation of the load data at time t
$\Gamma(\Gamma_k)$	the right-hand-side for the budget constraint of the (k-th) uncertainty set
γ_k	the right-hand-side for the risk constraint of the (k-th) uncertainty set
P_i^-	Lower bound output of unit i
P_i^+	Upper bound output of unit i
Δ_+^i	Ramping up limit of unit i
Δ_-^i	Ramping down limit of unit i
m_+^i	Minimum up time limit of unit i
m_-^i	Minimum down time limit of unit i
SR_t	System reserve requirement at time t
d_t	Power demand at time t
p_s	Probability of the scenario s
b_k	Number of extreme points/cardinality of the k-th uncertainty set

Table D.3: Uncertain Factors

$u_t(u_{it})$	Load at time t (forced outage of generator i at time t)

Table D.4: Decision Variables

x_{it}	Binary on/off status of unit i at time t
z_{it}	Binary start up of unit i at time t
y_{it}	Continuous generation of unit i at time t
s_t^+	Purchased power or load shed at time t (continuous)
s_t^-	Sold power at time t (continuous)
w_{it}	Spinning reserve of unit i at time t
h_t, v_n	Auxillary binary variable

Appendix E The Expanded Robust UC Model

In this model, we consider the unit commitment problem with random load, which is captured by multiple uncertainty sets. For simplicity, we do not include spinning reserve constraints and assume a linear fuel cost function. The expanded robust UC is formulated as follows.

$$\min_{\mathbf{x},\mathbf{z}} \sum_{i=0}^{I-1}\sum_{t=0}^{T-1}(r_i x_{it} + a_i z_{it}) + \sum_{k=1}^{K} \rho_k \max_{\mathbf{u}\in\mathbb{U}_k} \min_{(\mathbf{y},\mathbf{s}^+,\mathbf{s}^-)\in\Omega_k(\mathbf{x},\mathbf{z},\mathbf{u})}$$
$$\Big(\sum_{i=0}^{I-1}\sum_{t=0}^{T-1} c_i y_{it} + \sum_{t=0}^{T-1}(q_t^+ s_t^+ - q_t^- s_t^-)\Big) \tag{E.1}$$

s.t.

$$-x_{i(t-1)} + x_{it} - x_{ih} \le 0$$
$$\forall i, t \ge 1, t \le h \le min\{m_+^i + t - 1, T-1\}; \tag{E.2}$$
$$x_{i(t-1)} - x_{it} + x_{ih} \le 1$$
$$\forall i, t \ge 1, t \le h \le min\{m_-^i + t - 1, T-1\}; \tag{E.3}$$
$$-x_{i(t-1)} + x_{it} - z_{it} \le 0 \qquad \forall i, t \ge 1; \tag{E.4}$$
$$x_{it}, z_{it} \in \{0,1\} \qquad \forall i, t; \tag{E.5}$$

where

$$\mathbb{U}_k = \{\mathbf{u} : \bar{u}_t - \hat{u}_{kt} \le u_t \le \bar{u}_t + \hat{u}_{kt} \ \forall t;$$
$$\sum_{t=0}^{T-1} \frac{|u_t - \bar{u}_t|}{\hat{u}_{kt}} \le \Gamma_k\} \qquad k = 1, \ldots, K; \tag{E.6}$$

and

$$\Omega_k(\mathbf{x},\mathbf{z},\mathbf{u}) = \{(\mathbf{y},\mathbf{s}^+,\mathbf{s}^-) : P_i^- x_{it} \le y_{it} \le P_i^+ x_{it} \quad \forall i, t; \tag{E.7}$$
$$\sum_{i=0}^{I-1} y_{it} + s_t^+ - s_t^- = u_t \qquad \forall t; \tag{E.8}$$

$$y_{i(t+1)} \le y_{it} + x_{it}\Delta_{+}^{i} + (1 - x_{it})P_i^+$$
$$\forall i, t = 0, 1, ..., T - 2; \tag{E.9}$$
$$y_{it} \le y_{i(t+1)} + x_{i(t+1)}\Delta_{-}^{i} + (1 - x_{i(t+1)})P_i^+$$
$$\forall i, t = 0, 1, ..., T - 2; \tag{E.10}$$
$$y_{it} \ge 0 \ \forall i, t; s_t^+, s_t^- \ge 0 \ \forall t\}. \tag{E.11}$$

The objective function in (E.1) is to minimize the total cost, including the first stage commitment cost and the second stage economic dispatch cost estimated by the weighted sum of the worst case costs in different uncertainty sets. Constraints in (E.2)-(E.5) are typical commitment constraints that restrict the minimum up and down times, generator status, the start-up decisions, and variable types. $\mathbb{U}_k$ in (E.6) is defined in the same fashion as those in Bertsimas et al. (2013b); Zhao and Zeng (2012b). It uses *a budget constraint* to refine our uncertainty set description. The polyhedral set $\Omega_k(\mathbf{x}, \mathbf{z}, \mathbf{u})$ is the feasible set of the economic dispatch problem, for the fixed $(\mathbf{x}, \mathbf{z}, \mathbf{u})$. $\mathbf{y}, \mathbf{s}^+$, and $\mathbf{s}^-$ represent unit generation, purchased power, and sold power, respectively. Constraints in (E.7) define the lower and upper bounds on generation level. Constraints in (E.8) ensure loads can be satisfied all the time. Constraints (E.9) and (E.10) are ramping up/down limits. Constraints in (E.11) provide variable type restrictions.

In our numerical study, the column-and-constraint generation algorithm is applied while a different strategy is adopted to solve SP_k. Specifically, instead of considering linearized KKT-SP_k, we use strong duality to convert SP_k into the following problem (Dual-SP_k), where $\mathbf{x}^*$ is a given first stage decision, $\boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{\varphi}, \boldsymbol{\omega}$, and $\boldsymbol{\delta}$ are dual variables of constraints (E.7)-(E.10), respectively.

$$\begin{aligned}\max_{\boldsymbol{\lambda},\boldsymbol{\pi},\boldsymbol{\varphi},\boldsymbol{\omega},\boldsymbol{\delta},\boldsymbol{u}} & \sum_{i=0}^{I-1}\sum_{t=0}^{T-1}(P_i^- x_{it}^*\lambda_{it} + P_i^+ x_{it}^*\pi_{it}) + \sum_{t=0}^{T-1} u_t\varphi_t \\ & + \sum_{i=0}^{I-1}\sum_{t=0}^{T-2}(x_{it}^*\Delta_+^i + (1 - x_{it}^*)P_i^+)\omega_{it} \\ & + \sum_{i=0}^{I-1}\sum_{t=0}^{T-2}(x_{i(t+1)}^*\Delta_-^i + (1 - x_{i(t+1)}^*)P_i^+)\delta_{it}\end{aligned} \tag{E.12}$$

s.t.

$$\lambda_{i0} + \pi_{i0} + \varphi_0 - \omega_{i0} + \delta_{i0} \leq c_i \qquad \forall i, t = 0; \tag{E.13}$$

$$\lambda_{it} + \pi_{it} + \varphi_t - \omega_{it} + \delta_{it} + \omega_{i(t-1)} - \delta_{i(t-1)} \leq c_i$$

$$\forall i, t = 1, 2, ..., T-2; \tag{E.14}$$

$$\lambda_{i(T-1)} + \pi_{i(T-1)} + \varphi_{T-1} + \omega_{i(T-1)} - \delta_{i(T-1)} \leq c_i$$

$$\forall i, t = T-1; \tag{E.15}$$

$$\varphi_t \leq q_t^+ \qquad \forall t; \tag{E.16}$$

$$-\varphi_t \leq -q_t^- \qquad \forall t; \tag{E.17}$$

$$\lambda_{it} \geq 0, \pi_{it} \leq 0, \omega_{it} \leq 0, \delta_{it} \leq 0 \ \forall i, t; \varphi_t \ free \ \forall t; \tag{E.18}$$

$$\mathbf{u} \in \mathbb{U}_k. \tag{E.19}$$

Note that Dual-SP$_k$ has linear constraints and a bilinear objective function, which involves product term $u_t\varphi_t$. As proven in Zhao and Zeng (2012b) and Jiang et al. (2011), for the uncertainty set defined in (E.6), regardless of the recourse problem, worst cases will be those with u_t set to $\bar{u}_t$ or $\bar{u}_t + \hat{u}_{kt}$ for all t. So, the continuous uncertainty set is equivalent to a discrete set by simplifying u_t to $\bar{u}_t + \hat{u}_{kt}h_t$ where h_t is a binary variable, which allows us to linearize $u_t\varphi_t$ to convert Dual-SP$_k$ into an MIP formulation. In particular, we mention that, because explicit tight bounds on φ_t can be obtained from (E.16) and (E.17), the resulting MIP has a much simpler structure for fast computing, compared to the linearized KKT-SP$_k$. We refer this method as the strong duality method. Note that such methods is directly applicable when the uncertainty is a binary set, such as G-1 and G-2 sets presented in Appendix G.

Appendix F Correlation Analysis of Load Data

To analyze the behavior of loads reaching lower/upper bounds of an interval, we define a $(-1,0,1)$ *bounding matrix* $\widetilde{\mathbf{H}}$ based on our 24×7 load data. Specifically, for $load_{tj}$ and an interval $[\bar{u}_t - \hat{u}_t, \bar{u}_t - \hat{u}_t]$, we have

$$\widetilde{\mathbf{H}}_{tj} = \begin{cases} -1, & \text{if } \; load_{tj} \le \bar{u}_t - \hat{u}_t; \\ 0, & \text{if } \; \bar{u}_t - \hat{u}_t < load_{tj} < \bar{u}_t + \hat{u}_t; \\ 1, & \text{if } \; load_{tj} \ge \bar{u}_t + \hat{u}_t. \end{cases}$$

In our experiments, we let $\hat{u}_t = 1.5\sigma_t$ to match the uncertainty set $\mathbb{U}_1$. A sample of the resulting bounding matrix for hour 17-23 is presented in Table F.1. Then, we perform correlation analysis on the bounding matrix. The correlation matrix is presented in Table F.2, which shows that loads in hour $\{17, 22, 23\}$ ($\{18, 19\}$ respectively) are perfectly correlated on reaching bounds. As the following example, we can include such equality constraints into $\mathbb{U}_1$ to reflect correlation information on reaching bounds.

$$\frac{u_{17} - \bar{u}_{17}}{\hat{u}_{1,17}} = \frac{u_{22} - \bar{u}_{22}}{\hat{u}_{1,22}} = \frac{u_{23} - \bar{u}_{23}}{\hat{u}_{1,23}}.$$

Those equalities simply reduce the dimension of $\mathbb{U}_1$ and do not incur any additional computation challenge, which is suitable to strengthen the popular uncertainty set, i.e., the hypercubic set with a budget of uncertainty constraint (Bertsimas et al. (2013b), Bertsimas et al. (2011), Zhao and Zeng (2012b), and Jiang et al. (2011)), especially its equivalent discrete set. We mention that when u_ts are continuous, those equalities may not strictly

Table F.1: Bounding Matrix for Hour 17-23

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Hour 17	-1	0	0	0	0	0	0
Hour 18	0	0	0	0	1	0	0
Hour 19	0	0	0	0	1	0	0
Hour 20	0	0	0	0	0	0	0
Hour 21	0	0	0	0	0	0	0
Hour 22	-1	0	0	0	0	0	0
Hour 23	-1	0	0	0	0	0	0

Table F.2: Correlation Matrix for Hour 17-23

	Hour 17	Hour 18	Hour 19	Hour 20	Hour 21	Hour 22	Hour 23
Hour 17	1						
Hour 18	0.167	1					
Hour 19	0.167	1	1				
Hour 20	0	0	0	1			
Hour 21	0	0	0	0	1		
Hour 22	1	0.167	0.167	0	0	1	
Hour 23	1	0.167	0.167	0	0	1	1

hold but they serve as reasonable approximations to capture correlations, given that it is observed in sample data that those loads reach bounds simultaneously.

Appendix G The Risk Constrained Robust UC Model

In this model, we consider the unit commitment problem with bound constraints on load sheds under G-k contingencies (i.e., up to k generators in forced outages). Most notations and variables are identical to those used in the expanded robust UC model. Important differences are: spinning reserve constraints are included in the nominal situation with variable w_{it} representing spinning reserve of generator i at time t and parameter SR_t representing system reserve requirement at time t, $\mathbb{U}_k$ is a finite discrete set to describe the generator outage status, and parameter d_t is used to represent load at time t that is certain. In our experiments, SR_t is set as one tenth of d_t for all t.

$$\min_{\mathbf{x},\mathbf{z},\mathbf{y}_0} \sum_{i=0}^{I-1} \sum_{t=0}^{T-1} r_i x_{it} + a_i z_{it} + c_i y_{0,it} \tag{G.1}$$

s.t.

$$-x_{i(t-1)} + x_{it} - x_{ih} \le 0$$
$$\forall i, t \ge 1, t \le h \le min\{m_+^i + t - 1, T - 1\}; \tag{G.2}$$

$$x_{i(t-1)} - x_{it} + x_{ih} \le 1$$
$$\forall i, t \ge 1, t \le h \le min\{m_-^i + t - 1, T - 1\}; \tag{G.3}$$

$$-x_{i(t-1)} + x_{it} - z_{it} \le 0 \qquad \forall i, t \ge 1; \tag{G.4}$$

$$y_{0,it} \ge P_i^- x_{it} \qquad \forall i, t; \tag{G.5}$$

$$y_{0,it} + w_{it} \le P_i^+ x_{it} \qquad \forall i, t; \tag{G.6}$$

$$\sum_i w_{it} \ge SR_t \qquad \forall t; \tag{G.7}$$

$$y_{0,i(t+1)} \le y_{0,it} + x_{it}\Delta_+^i + (1 - x_{it})P_i^+$$
$$\forall i, t = 0, 1, ..., T - 2; \tag{G.8}$$

$$y_{0,it} \le y_{0,i(t+1)} + x_{i(t+1)}\Delta_-^i + (1 - x_{i(t+1)})P_i^+$$
$$\forall i, t = 0, 1, ..., T - 2; \tag{G.9}$$

$$\sum_{i=0}^{I-1} y_{0,it} - s^-_{0,t} = d_t \qquad \forall t; \tag{G.10}$$

$$\max_{\mathbf{u}\in\mathbb{U}_k} \min_{(\mathbf{y},\mathbf{s}^+)\in\Omega_k(\mathbf{x},\mathbf{z},\mathbf{u})} \sum_t s^+_t \le \gamma_k \quad k = 1,\dots,K \tag{G.11}$$

$$x_{it}, z_{it} \in \{0,1\},\ y_{0,it} \ge 0,\ s^-_{0,t} \ge 0,\ w_{it} \ge 0 \quad \forall i,t; \tag{G.12}$$

where

$$\mathbb{U}_k = \{\mathbf{u} : \sum_i u_{it} \le k \quad \forall t \tag{G.13}$$

$$u_{i(t+1)} \ge u_{it} \quad \forall i, 0 \le t \le T-2; \tag{G.14}$$

$$u_{it} \in \{0,1\} \quad \forall i,t; \} \qquad k = 1,\dots,K; \tag{G.15}$$

$$\tag{G.16}$$

and

$$\Omega_k(\mathbf{x},\mathbf{z},\mathbf{u}) = \{(\mathbf{y},\mathbf{s}^+,\mathbf{s}^-) : s^+_t - s^-_t = d_t - \sum_i y_{it}\ \forall t; \tag{G.17}$$

$$P^-_i x_{it}(1-u_{it}) \le y_{it} \le P^+_i x_{it}(1-u_{it}) \quad \forall i,t; \tag{G.18}$$

$$\begin{aligned} &y_{i(t+1)} \le y_{it} + x_{it}\Delta^i_+ + (1-x_{it})P^+_i \\ &\qquad \forall i, t = 0,1,...,T-2; \end{aligned} \tag{G.19}$$

$$\begin{aligned} &y_{it} \le y_{i(t+1)} + x_{i(t+1)}\Delta^i_- + (1-x_{i(t+1)})P^+_i \\ &+P^+_i u_{i(t+1)} \qquad \forall i, t = 0,1,...,T-2; \end{aligned} \tag{G.20}$$

$$s^+_t \ge 0,\ s^-_t \ge 0,\ \forall t;\ y_{it} \ge 0\ \forall i,t\}. \tag{G.21}$$

The objective function in (G.1) is to minimize the overall cost in the nominal situation. Constraints in (G.2)-(G.10) are the regular unit commitment constrains, along with variable type restrictions in (G.12). Constraints in (G.11) define the different restrictions on the overall load shed in K contingency sets. The contingency set $\mathbb{U}_k$ in (G.13)-(G.15) includes all the contingencies with up to k generator outages. Specifically, constraints in (G.13) ensure that at any time, no more than k generators are in outage. Constraints in (G.14) indicate that once generator i is in outage at time t, i.e., $u_{i,t} = 1$, it remains in outage status.

Finally, the set $\Omega_k(\mathbf{x}, \mathbf{z}, \mathbf{u})$ in (G.17)-(G.21) defines the feasible set of the economic dispatch subject to fixed $(\mathbf{x}, \mathbf{z}, \mathbf{u})$. Note from (G.18) that once generator i is in outage at time t, its generation will be zero. Also, (G.20) ensures that the ramping down constraint is not needed if generator i is in outage at time $t+1$.

Appendix H Numerical Study on Large Systems

To verify their performances and applicability among practical-scale systems, we investigate our proposed models and solution methods on large instances.

Data of 36 thermal units, including generator parameters and hourly load information, from IEEE 118-bus system (Ma and Shahidehpour, 1999), and Zhao and Zeng (2012b) are adopted as the basic testbed. To define load uncertainty sets, we treat the given load as nominal value $\bar{u}_t$ for all t, and let the deviations $\hat{u}_{1t}$ be 15% and $\hat{u}_{2t}$ be 30% of $\bar{u}_t$, respectively. To generate larger instances, we follow the strategy used in Ostrowski et al. (2012) to replicate generators and scale up hourly loads accordingly, which yield systems with 72 and 108 units. For uncertainty sets of the latter two systems, the deviation $\hat{u}_{kt}$ are $\sqrt{2}$ and $\sqrt{3}$ times of their counterparts for the system of 36 units, respectively.

In all numerical experiments, we set weight coefficients (ρ_1, ρ_2) to $(0.6, 0.4)$ as this pair of coefficients may demand more computation time, according to Table 5.1. To evaluate different methods to solve the max-min subproblem SP_k, we implement both KKT condition based method presented in Section 5.3.3 and the strong duality based method presented in Appendix E when Γ_1 and Γ_2 are integral. For KKT condition method, the big-M value used in linearization procedure is set to 10^5. Computational results are presented in Table H.1, where the column "SP time(s)" presents the average solution time for max-min subproblems over iterations.

We first note from Table H.1 that both the number of iterations and computational time, regardless of different methods to solve SP_k, generally increase with respect to the problem sizes. Nevertheless, in most cases the increase is mild, which indicates the expanded robust UC models, along with the customized column-and-constraint generation algorithm, could be used to deal with practical-scale instances. Second, we observe that the customized column-and-constraint generation algorithm with the strong duality method is much faster than the other variant with KKT method. On average, the former one is about 2.5 times faster than the latter one. After taking a closer look at the differences, we notice that the strong duality method can quickly solve SP_k while KKT method takes much longer computational time. On average, the former one is 6.8 times faster than the latter one. For some instances, the

Table H.1: Expanded Robust UC on Large Systems

# of Units	Γ_1, Γ_2	strong duality method			KKT method			Obj.
		iter.	time(s)	SP time(s)	iter.	time(s)	SP time(s)	
36	12,12	2	0.906	0.062	2	2.671	0.473	791604.23
	12,10	2	0.875	0.055	2	2.78	0.5	779675.63
	12,8	3	2.062	0.083	3	3.73	0.323	766767.83
	12,6	3	2	0.081	3	4.41	0.477	752645.03
	12,4	3	1.984	0.091	3	4.19	0.437	735609.83
	12,2	4	7.312	0.16	3	6.11	0.346	718151.83
	6,6	3	2.657	0.104	3	3.75	0.331	723425.63
	6,4	3	2.5	0.122	3	4.38	0.445	706390.43
	6,2	4	8.499	0.121	4	10.84	0.375	688932.43
72	12,12	3	4.749	0.247	3	16.703	1.912	1780542.59
	12,10	2	2.062	0.172	2	6.69	1.266	1746723.41
	12,8	3	3.984	0.24	3	14.83	1.971	1710214.64
	12,6	2	1.829	0.184	2	6.19	1.215	1670269.34
	12,4	3	4.687	0.253	3	12.98	1.534	1608116.59
	12,2	2	4.5	0.242	2	8.56	1.183	1539172.26
	6,6	3	4.875	0.24	3	10.17	1.05	1587624.39
	6,4	4	8.875	0.301	3	12.41	1.406	1525471.64
	6,2	3	15.937	0.253	3	24.11	1.247	1456527.31
108	12,12	4	13.718	0.486	4	64.827	5.775	2913662.29
	12,10	6	43.264	0.535	6	130.83	6.451	2851679.47
	12,8	4	21.234	0.555	5	73.34	4.506	2784353.64
	12,6	2	3.312	0.352	2	13.08	2.703	2710424.36
	12,4	2	3.281	0.34	2	12.39	2.559	2576045.65
	12,2	2	6.797	0.266	2	15.27	2.355	2430536.04
	6,6	3	8.906	0.487	3	20.59	2.331	2558595.9
	6,4	3	9.062	0.513	3	17.58	1.786	2424217.19
	6,2	3	19.687	0.399	3	29.33	1.883	2278707.58
average		3	7.761	0.257	2.963	19.731	1.735	

former one could be 12.1 times faster. Those numerical results confirm the importance of adopting a simple uncertainty set and solving max-min subproblems using the strong duality method.

As in Bertsimas and Sim (2004) and Bertsimas et al. (2013b), the aggregated volatility of $T(=24)$ random inputs is proportional to $\sqrt{T}$. So, we perform a set of experiments on instances with Γ_1 and Γ_2 scaled according to $\sqrt{T}$, for which, because $\sqrt{T}$ is not integral, only KKT method is applied. Numerical results are presented in Table H.2. Basically, all instances can be solved within an acceptable amount of time, comparable to those of KKT method presented in Table H.1. Another observation is that the computational time for max-min subproblem generally reduces when Γ_1 and Γ_2 reduce, which is more obvious when the system has 108 units. It can be explained by the fact that when Γ_1 and Γ_2 reduce, the uncertainty sets become smaller and less algorithmic operations are needed to solve max-min subproblems.

Finally, we investigate our risk constrained robust UC model for large systems. Noting that it is much more difficult to compute, we just present results on the basic system with 36 units. In Table H.3, the restrictions of total load shed over 24 hours are set to 0.1%, 0.5%, 1%, and 2% of total load, whose numerical values are 84.24, 421.18, 842.36, and 1684.72 respectively. Compared to results in Table 5.7, it is clear that risk constrained robust UC model is not only difficult to compute, but also very sensitive to problem size. Hence, more powerful algorithm enhancements are needed.

Table H.2: Expanded Robust UC with $\Gamma \propto O(\sqrt{T})$

# of Units	Γ_1, Γ_2	iter.	time(s)	SP time(s)	Obj.
36	$3\sqrt{T}$, $3\sqrt{T}$	2	3.109	0.563	818211.98
	$3\sqrt{T}$, $2\sqrt{T}$	2	2.95	0.516	789847.92
	$3\sqrt{T}$, $\sqrt{T}$	3	5.19	0.576	754670.3
	$3\sqrt{T}$, $0.5\sqrt{T}$	4	6.53	0.344	733777.07
	$3\sqrt{T}$, $0.1\sqrt{T}$	2	2.02	0.199	708355.59
	$\sqrt{T}$, $\sqrt{T}$	4	7.23	0.491	707014.04
	$\sqrt{T}$, $0.5\sqrt{T}$	5	15.75	0.438	685859.35
	$\sqrt{T}$, $0.1\sqrt{T}$	2	2.28	0.168	661224.89
72	$3\sqrt{T}$, $3\sqrt{T}$	2	7.546	1.441	1855800.68
	$3\sqrt{T}$, $2\sqrt{T}$	2	9.98	2.105	1775495
	$3\sqrt{T}$, $\sqrt{T}$	2	7.19	1.406	1666346.01
	$3\sqrt{T}$, $0.5\sqrt{T}$	3	18.48	1.732	1582627.59
	$3\sqrt{T}$, $0.1\sqrt{T}$	2	10.81	1.445	1497669.3
	$\sqrt{T}$, $\sqrt{T}$	3	12.52	1.419	1531173.16
	$\sqrt{T}$, $0.5\sqrt{T}$	3	16.86	0.886	1447835.32
	$\sqrt{T}$, $0.1\sqrt{T}$	3	20.31	0.948	1364148.11
108	$3\sqrt{T}$, $3\sqrt{T}$	3	41.234	5.695	3052000.24
	$3\sqrt{T}$, $2\sqrt{T}$	5	84.25	5.394	2906694.05
	$3\sqrt{T}$, $\sqrt{T}$	2	11.02	2.168	2691161.8
	$3\sqrt{T}$, $0.5\sqrt{T}$	2	14.27	2.219	2510368.58
	$3\sqrt{T}$, $0.1\sqrt{T}$	2	11.86	1.394	2345691.03
	$\sqrt{T}$, $\sqrt{T}$	3	29.75	3.735	2443532.59
	$\sqrt{T}$, $0.5\sqrt{T}$	3	29.2	1.94	2262739.37
	$\sqrt{T}$, $0.1\sqrt{T}$	3	26.77	1.497	2098061.82
average		2.792	16.546	1.613	

Table H.3: Results of G-1/G-2 Risk Constrained Model with 36 Units

Case	γ_1, γ_2	iter.	time(s)	Obj.	LS in G-1	LS in G-2
r4	84.24, 1684.72	1	321.056	622163.9	0	629
r5	84.24, 842.36	1	321.06	622163.9	0	629
r6	84.24, 421.18	5	1395.721	622483.9	0	413

Appendix I Reprint Permission for Chapter 4

ELSEVIER LICENSE
TERMS AND CONDITIONS

Oct 12, 2014

This is a License Agreement between Yu An ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Yu An
Customer address	12603 Lake Square Circle ORLANDO, FL 32821
License number	3474390789254
License date	Sep 22, 2014
Licensed content publisher	Elsevier
Licensed content publication	Transportation Research Part B: Methodological
Licensed content title	Reliable p-median facility location problem: two-stage robust models and algorithms
Licensed content author	Yu An,Bo Zeng,Yu Zhang,Long Zhao
Licensed content date	June 2014
Licensed content volume number	64
Licensed content issue number	n/a
Number of pages	19
Start Page	54
End Page	72
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Reliable Design and Operations of Infrastructure Systems
Expected completion date	Nov 2014
Estimated size (number of pages)	200

Elsevier VAT number	GB 494 6272 12
Price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	**0.00 USD**

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
“Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER].” Also Lancet special credit - “Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier.”
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will

Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

For journal authors: the following clauses are applicable in addition to the above: Permission granted is limited to the author accepted manuscript version* of your paper.

***Accepted Author Manuscript (AAM) Definition:** An accepted author manuscript (AAM) is the author's version of the manuscript of an article that has been accepted for publication and which may include any author-incorporated changes suggested through the processes of submission processing, peer review, and editor-author communications. AAMs do not include other publisher value-added contributions such as copy-editing, formatting, technical enhancements and (if relevant) pagination.

You are not allowed to download and post the published journal article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Posting to a repository: Authors may post their AAM immediately to their employer's institutional repository for internal use only and may make their manuscript publically available after the journal-specific embargo period has ended.

Please also refer to Elsevier's Article Posting Policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only.. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

Elsevier Open Access Terms and Conditions

Elsevier publishes Open Access articles in both its Open Access journals and via its Open Access articles option in subscription journals.

Authors publishing in an Open Access journal or who choose to make their article Open Access in an Elsevier subscription journal select one of the following Creative Commons user licenses, which define how a reader may reuse their work: Creative Commons Attribution License (CC BY), Creative Commons Attribution – Non Commercial - ShareAlike (CC BY NC SA) and Creative Commons Attribution – Non Commercial – No Derivatives (CC BY NC ND)

Terms & Conditions applicable to all Elsevier Open Access articles:

Appendix J Reprint Permission for Chapter 5

IEEE Requesting permission to reuse content from an IEEE publication

Title: Exploring the Modeling Capacity of Two-Stage Robust Optimization: Variants of Robust Unit Commitment Model

Author: An, Y.; Zeng, B.

Publication: Power Systems, IEEE Transactions on

Publisher: IEEE

Copyright © 1969, IEEE

Logged in as:
Yu An
Account #:
3000836688

LOGOUT

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW