

Models and Algorithms for Stochastic Network Design and Flow Problems: Applications in Truckload Procurement Auctions and Renewable Energy

by

Richard Li-Yang Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2010

Doctoral Committee:

Associate Professor Amy E.M. Cohn, Chair

Professor Mark Stephen Daskin

Assistant Professor Damian R. Beil

Assistant Professor Amitabh Sinha

Assistant Professor Duncan S. Callaway, University of California – Berkeley

ACKNOWLEDGEMENTS

My thanks go first to my advisor, Amy Cohn, for her guidance and support these past five years. She's the nicest advisor you can ask for and truly one of the smartest people I know. I would also like to thank my collaborators and/or dissertation committee members, Mark Daskin, Damian Beil, Amitabh Sinha, Duncan Callaway, and Shervin Ahmadbeygi, without them this dissertation would not have been possible. Next, I would like to thank my family (especially, my kids) and friends for making the last five years so memorable and fun. Finally, to my lovely wife Stephanie, thank you for your patience, your encouragement, your love, and most importantly, your extraordinary efforts to keep the kids out of the office when I needed to work.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF FIGURES	vi
LIST OF TABLES	viii
 CHAPTER	
I. Introduction	1
2.1 Solving Truckload Procurement Auctions Over an Exponential Number of Bundles	5
2.2 Stochastic Multicommodity Flow Approach to Combinatorial Truckload Auctions	7
2.3 Including Wind in Power System Siting and Capacity Expansion Models	8
2.4 Iterative Test-and-Prune: Designing Wind Farms with Probabilistic Constraints	9
2.5 Outline for the Remainder of the Dissertation	10
 II. Solving Truckload Procurement Auctions Over An Exponential Number of Bundles	 12
2.1 Introduction and Motivation	12
2.2 Combinatorial Auctions for Truckload Procurement	18
2.2.1 Computing Bundle Bids	19
2.2.2 Traditional Winner Determination Problem	21
2.3 Implicit Truckload Combinatorial Auctions	23
2.3.1 The Bid-Generating Function	24
2.3.2 The Implicit Winner Determination Problem	27
2.3.3 Operational Considerations	32
2.3.4 Privacy Issues of the Implicit Bidding Approach	34
2.4 Computational Experiments	36
2.4.1 Tractability of I-WDP	37
2.4.2 Impact of Operational Considerations	39
2.4.3 Impact of Instance Characteristics	40
2.4.4 Comparison to Bidding Methods in Practice	46
2.5 Conclusions and Future Research	54
 BIBLIOGRAPHY	 57
 III. A Stochastic Multicommodity Flow Approach to Combinatorial Truckload Auctions	 60
3.1 Introduction	60

3.2	CTPA	62
3.2.1	Computing Bundle Bids	62
3.2.2	Traditional WDP	69
3.3	IBA for S-CTPAs	71
3.3.1	WDP with No Recourse	72
3.3.2	WDP with Recourse	73
3.4	An L-shaped method for R-WDP	74
3.5	Accelerating L-shaped method	78
3.5.1	Embedding Network Capacities and Costs within RMP	79
3.5.2	Knapsack Constraints	80
3.5.3	Multi-Cut Generation	81
3.6	Computational Experiments	82
3.6.1	Tractability of R-WDP and performance benefits of A-LSM	83
3.6.2	Value of the Stochastic Solution	85
3.7	Generalization	87
3.7.1	Deterministic Integer MFP	88
3.7.2	Two-Stage Integer MFP	89
3.7.3	Another Application Area	92
3.8	Summary and Conclusion	93

BIBLIOGRAPHY 96

IV. Including Wind in Power System Siting and Capacity Expansion Models 98

4.1	Introduction	98
4.1.1	Background	100
4.1.2	Goals and Contributions	103
4.2	Problem Formulation	103
4.2.1	Sets	103
4.2.2	First Stage Parameters	104
4.2.3	First Stage Variables: Capacity Expansion and Siting	105
4.2.4	First Stage Variables: Network Flows	105
4.2.5	Second Stage Parameters	105
4.2.6	Second Stage Variables: Network Flows	106
4.2.7	Transmission Losses	110
4.2.8	Full Second Stage Model With Quadratic Loss	112
4.3	Benders Decomposition Algorithm	113
4.4	Accelerating Benders Decomposition	115
4.4.1	Network Connectivity	115
4.4.2	Demand Fulfillment	116
4.4.3	Valid Inequalities	118
4.4.4	Multicut Generation	119
4.5	Computational Experiments	120
4.5.1	Test System Generation	121
4.5.2	Computational Results	123
4.6	Summary and Conclusions	125
4.7	Test System Parameters	128

BIBLIOGRAPHY 134

V. An Iterative Test-and-Prune Algorithm for Wind Farm Network Design with a Probabilistic Reliability Constraint 137

5.1	Introduction	137
-----	------------------------	-----

5.2	Simplified Problem	137
5.2.1	Challenges of Mathematical Programming Approach	139
5.2.2	Computational Experiments	141
5.3	Test-and-Prune	142
5.4	Iterative Test-and-Prune	144
5.4.1	Computational Results	147
5.5	Summary and Conclusion	149
	BIBLIOGRAPHY	151
VI.	Conclusions	152
6.1	Conclusions	152

LIST OF FIGURES

Figure

2.1	The benefits of complementarities demonstrated through a two load example. . . .	14
2.2	A cost-effective tour covering bid loads 1, 2, and 3 that leverages a carrier’s pre-existing contracts and spot market opportunities.	20
2.3	Estimate of carrier k ’s repositioning capacity and price (cost plus markup) from city i to city j using a 4 tiered step-function.	21
2.4	Effect of lane quantities on solution time.	42
2.5	Effect of lane quantities on the number of carriers assigned.	42
2.6	Effect of lane quantities on empty movement ratio.	43
2.7	The network on the left is a hub-and-spoke network with a single hub located in Detroit. All bid lanes in this network originate from the hub. The network on the right is a uniform network with bid lanes uniformly dispersed throughout the entire network.	44
2.8	Impact of different network structures on average solution time, empty movement ratio, number of assigned carriers and assigned carriers per lane.	45
2.9	Case 2 permits single bid lane bundles. These bundles may leverage repositioning lanes to form efficient continuous moves.	48
2.10	Case 3 (left) permits two bid lane bundles. Case 4 (right) permits three bid lane bundles.	48
2.11	Projected run-times for evaluating all double bid lane bundles (left, in hours) and all triple bid lane bundles (right, in days).	50
2.12	Optimality gap between various bundles selection criteria vs. implicit bidding approach.	52
3.1	A carrier’s estimated repositioning capacities and prices from city i to city j	64
3.2	A carrier’s repositioning capacity and price from node i to node j in scenario $s = 1$ and $s = 2$	65

4.1	Normalized power output for the same amount of capacity. The top figure shows power output distribution when all turbines are installed at a single (windiest) site. The bottom figure shows power output distribution when turbines are uniformly distributed across twenty sites.	99
4.2	Conversion from wind speed to power.	102
4.3	Interconnecting wind farms to a common interconnection point.	122
4.4	TS1 - Optimality gap as a function of iteration count.	124
4.5	TS2 - Optimality gap as a function of iteration count.	125
4.6	TS3 - Optimality gap as a function of iteration count.	126
5.1	Test-and-Prune Flowchart.	143
5.2	Solving T&P with coarse granularity.	145
5.3	I-T&P FlowChart.	147

LIST OF TABLES

Table

2.1	I-WDP solution characteristics for various auctions sizes.	39
2.2	Average solution times for I-WDP with constraints on total load volume, number of assigned carriers and number of assigned carriers per lane.	40
2.3	Comparison of explicit bidding approaches to the implicit bidding approach.	49
3.1	Test Set Characteristics for Various Auctions Sizes	83
3.2	S-LSM and A-LSM Solution Characteristics for Various Test Sets	85
3.3	Value of the Stochastic Solution	86
4.1	Algorithmic Performance	123
4.2	TS1 - Generation and Load Parameters	128
4.3	TS1 - Transmission Parameters	129
4.4	TS2 - Generation and Load Parameters	130
4.5	TS2 - Transmission Parameters	131
4.6	TS3 - Generation and Load Parameters	132
4.7	TS3 - Transmission Parameters	133
5.1	MIP performance	141
5.2	Comparison of MIP performance to I-T&P performance	147

CHAPTER I

Introduction

This dissertation presents novel mathematical models and algorithms for stochastic network design and flow (SNDF) problems: the optimal design and flow of a network under uncertainty to meet specific requirements while minimizing expected total cost. These types of network problems arise in a wide variety of domains. For example, in telecommunication network design, the objective may be to determine the amount of capacity to install on each link (arc) of a telecommunication network such that there is sufficient capacity to carry an uncertain amount of traffic (telephone calls or data transmissions) simultaneously between various source and sink nodes. In disaster relief management, the network is defined by a set of nodes representing emergency-supply sources, sinks or transshipment points. The goal may be to allocate emergency-supplies in a pre-emergency stage such that demand in the post-emergency stage can be met at the least possible cost. Uncertainties in node and arc capacities exist due to potential damage to shipment routes and facilities during the event of a disaster.

In this dissertation, we focus on two other important applications: truckload procurement auctions and wind farm network design. In truckload procurement, carriers minimize their costs to ship loads by creating efficient tours, continuous movements

with minimal empty mileage. This is often accomplished by combining loads they are bidding on with their repositioning capacities. However, these repositioning capacities are not known with certainty at the time of the procurement auction, as they may depend on future contracts and spot market opportunities.

In recent years, wind farm network design has garnered much interest because of federal and state initiatives to reduce greenhouse gas emissions in an effort to forestall global warming. In wind farm network design, a primary concern is the impact on overall system reliability of integrating a large amount of intermittent wind resources (turbines). This problem is highly stochastic because wind speed is stochastic and varies spatially and temporally across sites, future demand is not fully known, and both generators and transmission lines are subject to random failures.

SNDF problems, such as those above, often have characteristics that render them difficult to model and computationally challenging to solve, including:

Nonlinear costs: There are often fixed costs associated with network expansion decisions. In power system transmission expansion planning, for example, there are fixed costs associated with adding new transmission capacity in each corridor. These fixed costs may reflect one-time regulatory fees, land leases, et cetera. These are traditionally modeled using Big-M constraints, which state that capacity cannot be added on a specific corridor unless the fixed cost is assessed. These types of constraints are well known to be computationally challenging because they lead to weak linear programming (LP) relaxations and fractional capacity expansion decisions. Weak LP relaxations lead to poor lower bounds, and fractional expansion decisions lead to large branch-and-bound trees and thus to significantly increased computational times.

Nonlinear flows: The flows of commodities on arcs may be subject to nonlinear

losses. In power systems, for example, electricity flow on a transmission line is subject to “ i^2R ” transmission losses, and in communication networks, packets are increasingly (at an exponential rate) dropped as network congestion rises. Nonlinear flow decisions like these are unfortunately quite common in SNDF problems. These challenges are further magnified because often a subset of the decision variables must take on integral values. In such cases, the resulting model is nonlinear and discrete.

Probabilistic reliability requirements: In addition to minimizing total cost, there may be an additional requirement on the reliability of network designs. For example, in WFND, a key consideration is the impact of intermittent wind power on overall system reliability. When such a reliability requirement exists, network reliability requirements can be enforced either as an objective penalty, penalizing the system for unmet demand, or as a hard constraint on the minimum probability of meeting demand. The second case is especially challenging. In WFND, for example, the conversion from wind speed to power output is highly nonlinear (and non-convex) and discontinuous; moreover, integrating a multivariate probability distribution for multi-area wind speeds within this power curve poses even greater challenges.

Stochastic parameters: In SNDF problems, network parameters are not known with certainty. For example, in combinatorial truckload procurement auctions (CTPA), carriers do not have complete information about their future repositioning capacities and costs at the time of the auction, as these capacities may depend on other contracts they may be awarded and/or spot market opportunities at the time of operation. In CTPA, carriers’ repositioning capacities and prices, represented by arc capacities and costs in the network model, are discrete random variables with some joint distribution. To minimize the expected cost for procuring truckload services, each scenario (which represents a realization of the random variables and its asso-

ciated probability) must be considered explicitly. Incorporating these uncertainties into the model results in an extremely large SNDF problem and thus poses significant computational challenges.

In all of the aforementioned applications, additional application-specific requirements are added to the nominal network design and flow problem, while simplifying steps are typically taken to achieve tractability at the expense of solution quality. The goal of this research is, therefore, to capture as much of the complexities within these problems as possible. We focus on truckload procurement auction and wind farm network design problems characterized by uncertainties in node supplies and/or demands and in arc capacities and/or costs.

This dissertation presents models and algorithms to address all of these challenges for truckload procurement auctions and wind farm network design. We develop models and algorithms for these specific applications; whenever possible, we also extend our results to a more general class of network flow problems.

In the remainder of the introduction, we provide motivation, describe challenges, and summarize our contributions to SNDF problems within the context of our research on CTPAs and WFNDs. In §1.1, we introduce a novel implicit bidding approach (IBA) that permits the solution of fully-enumerated combinatorial auctions in a single round, something that was not possible with preexisting approaches. By using IBA, we can circumvent the main computational challenges of CTPAs by reposing this problem as an integer multicommodity flow problem of polynomial size (with respect to the auction parameters). In §1.2, we describe the extension of CTPAs to consider network uncertainties. Specifically, we show that this problem can be formulated as a more general two-stage multicommodity flow problem (TS-MFP) in which there are uncertainties in the costs and/or capacities of a subset of the arcs.

In §1.3, we present our work on WFND problems subject to penalties for unmet demand. An accelerated Benders decomposition algorithm is developed to solve this problem. In §1.4, we consider an alternative model for WFND problems in which reliability is enforced via a constraint on loss-of-load expectation. We conclude the introduction in §1.5 with an outline of the remainder of the dissertation.

1.1 Solving Truckload Procurement Auctions Over an Exponential Number of Bundles

Truckload carriers provide hundreds of billions of dollars worth of services to shippers in the United States alone each year. Internet auctions provide these shippers with a fast and easy way to negotiate potential contracts with a large number of carriers. Combinatorial auctions have the added benefit of allowing multiple lanes to be considered simultaneously in a single auction. This capacity is important because it enables carriers to connect multiple lanes (where a lane is defined by an origin, a destination, and a volume indicating the number of loads) in continuous moves or tours, decreasing the empty mileage that must be driven and therefore increasing overall efficiency. However, achieving full economies of scope and scale in combinatorial auctions requires bidding on an exponential number of bundles, which is not tractable except for very small auctions. In most real-world auctions, bidding is instead typically limited to a very small subset of the potential bids. We present a new bidding framework, the *Implicit Bidding Approach*, for combinatorial truckload procurement auctions that enables the complete set of all possible bids to be considered implicitly, without placing the corresponding burden of an exponential number of bids on the bidders or the auctioneer. This approach leverages the fact that there is a known and amenable structure underlying the cost of servicing a given set of bid lanes. Specifically, the least-cost tour (or set of tours) needed to cover a set of lanes

can be computed by solving a minimum-cost flow problem. We therefore propose to embed this underlying cost structure (which we refer to as a bid-generating function) directly into WDP. This approach eliminates the need for the bidder to compute and communicate an exponential number of bids. Furthermore, we will show that the resulting WDP is a special case of an integer multi-commodity flow (MCF) problem of polynomial size. We then provide extensive computational results to demonstrate the increased tractability that our approach provides. Finally, we conclude with numerical analyses to assess the quality of the solutions that are generated and to demonstrate the benefits of our approach over existing bidding methods in practice.

The contributions of this research are in:

1. presenting a new implicit bidding approach for combinatorial auctions that enables the complete set of all possible bids to be considered implicitly, and thus achieves full economies of scope;
2. developing tractable models to solve a basic truckload procurement auction to optimality, in a single round, fully considering (implicitly) the exhaustive set of all possible bids;
3. showing how the power of mathematical programming can enable this basic problem to be extended to include additional important real-world operational considerations; and
4. taking advantage of this new capability to solve fully-enumerated truckload procurement auctions as a tool for conducting numerical analysis on the characteristics of CTPA solutions.

1.2 Stochastic Multicommodity Flow Approach to Combinatorial Truckload Auctions

In this research, we consider a fully-enumerated stochastic combinatorial truckload procurement auction (S-CTPA) characterized by uncertainties in carriers' repositioning capacities and costs. Typically, carriers participating in combinatorial truckload auctions estimate repositioning capacities and costs for different opportunities. However, shipments awarded on the basis of such estimates are often sub-optimal in expectation. To rectify this problem, we leverage the implicit bidding approach for truckload procurement auctions to derive a tractable winner determination problem that is fully expressive and completely captures carriers' uncertain repositioning capacities and costs. We then present an accelerated decomposition algorithm for solving the resulting winner determination problem and extensive computational studies to demonstrate its efficacy. Finally, we generalize the models and algorithms presented for stochastic combinatorial truckload procurement auctions to a class of stochastic network flow problems, which we call Two-Stage Multicommodity Flow problems (TS-MFPs), and demonstrate the applicability to a variety of other problems, such as vaccine distribution and emergency relief.

The contributions of this research are in:

1. presenting models and decomposition algorithms for fully-enumerated S-CTPAs, where carriers have uncertain repositioning capacities and costs;
2. proposing procedures to accelerate the decomposition algorithm for solving S-CTPAs and demonstrating their efficacy for solving practically-sized instances;
3. taking advantage of this new capability to solve fully-enumerated S-CTPAs to demonstrate the value of the stochastic solution over the deterministic solution,

obtained by solving the expected value problem (a related deterministic model that uses the expected values of uncertain parameters);

4. generalizing the stochastic model and algorithmic approach presented for S-CTPAs to a more general stochastic network flow problem and demonstrating its applicability to a variety of other problems.

1.3 Including Wind in Power System Siting and Capacity Expansion Models

Wind-generated electricity is widely regarded as the most promising way to reduce pollution and greenhouse gas emissions, but it also presents new challenges not found in the design of conventional (i.e. nuclear or fossil-fuel based) generation networks. We propose a new model, Wind Farm Network Design (WFND), for generation- and transmission- expansion planning that integrates both wind-based and conventional power generation. Because transmission losses grow nonlinearly with distance, and wind generation is often located far from demand points, adequate modeling of transmission losses becomes particularly important when including wind-based generation in network design problems. We consider linear and quadratic transmission loss models for WFND and present a Benders Decomposition (BD) algorithm, whose integer master problem prescribes a network design, and whose linear subproblems convey information about the operating and loss-of-load costs (LOLC). However, the standard BD algorithm performs poorly because of the lack of network information in the master problem and the weakness of standard optimality cuts. Accordingly, we enforce necessary conditions regarding network connectivity and demand fulfillment in the master problem along with iteratively adding sets of valid inequalities and adopting a scenario aggregation strategy (to generate multiple optimality cuts per Benders Iteration) to improve the convergence of BD. Finally, we present a compu-

tational study involving three test systems to demonstrate the tractability of WFND and the efficacy of our proposed solution strategy.

The contributions of this research are in:

1. presenting a new model for the design of wind farm networks in a multi-area power system;
2. modeling an integrated generation and transmission expansion problem with explicit considerations for system uncertainties, fixed-siting costs and nonlinear transmission losses;
3. introducing an accelerated Benders decomposition algorithm (A-BD) that efficiently solves WFND problems with a large number of scenarios.

1.4 Iterative Test-and-Prune: Designing Wind Farms with Probabilistic Constraints

In this research, we consider the question of how to extend WFND problems to incorporate a probabilistic constraint on the loss-of-load expectation (LOLE). We explore the fundamental challenges traditional mathematical programming (MP) approaches encounter in solving WFND problems with LOLE-constraints and demonstrate these difficulties via computational experiments. We then present a novel algorithmic approach, which we call *Iterative Test-and-Prune* (I-T&P), for solving LOLE-constrained WFND problems. I-T&P is a hybrid algorithm that leverages the power of mathematical programming (and other approaches) to solve a series of easy feasibility problems within a larger meta-search algorithm. We present computational results for a simplified version of LOLE-constrained WFND problems and demonstrate the greater efficacy of I-T&P over standard mathematical programming approaches.

The contributions of this research are in:

1. presenting a new model for the design of wind farm networks that incorporates probabilistic constraints on LOLE;
2. developing a hybrid algorithm, Iterative Test-and-Prune, for solving WFND problems with a LOLE constraint and demonstrating the algorithm's efficacy via computational experiments.

1.5 Outline for the Remainder of the Dissertation

The remainder of the dissertation is organized as follows. In Chapter II, we present a novel approach for combinatorial auctions, the implicit bidding approach, that simultaneously addresses the two main hurdles of combinatorial auctions. We demonstrate the viability of IBA for CTPAs by presenting extensive computational results. In Chapter III, we consider fully-enumerated CTPAs under uncertain network capacities and costs. Specifically, we consider the case where carriers' repositioning capacities and costs are subject to uncertainty. To the best of our knowledge, we are the first to propose a model and solution algorithm for fully-enumerated stochastic CTPAs. Finally, we generalize our model and algorithm for Stochastic CTPAs to a broader class of network flow problems, which we called two-stage multicommodity flow problems. In Chapter IV, we present models and algorithms for the integrated generation and transmission expansion planning problem with wind resources, which we called *wind farm network design* (WFND). We present an efficient decomposition algorithm for solving WFND and extensive computational results that demonstrate the effectiveness of our approach. In Chapter V, we consider WFND problems with a probabilistic constraint on LOLE. We demonstrate that this model is extremely challenging and that direct applications of mathematical programming approaches

are not viable. We present a hybrid algorithm, which we called *Iterative Test-and-Prune* (I-T&P), that leverages mathematical programming (and other approaches) to solve a series of easy feasibility problems within a larger meta-search algorithm. Chapter VI concludes with a discussion of the contributions of this dissertation and suggestions for future research.

CHAPTER II

Solving Truckload Procurement Auctions Over An Exponential Number of Bundles

2.1 Introduction and Motivation

U.S. freight transportation expenditures in 2005 exceeded \$700 billion. Of this amount, \$300 billion was accounted for by the truckload segment [3]. In many corporations, transportation expenditures can be as high as 30% of the overall cost of goods sold [7]; furthermore, trucking is often the dominant cost. Therefore, reducing trucking expenditures can greatly reduce a shipper's cost of goods sold and improve profitability.

Typically, shippers estimate their freight to be shipped in an upcoming year based on the prior year's shipments [28]. When contracting out truckload services, a shipper puts forth a request for quotes (RFQ) for a network of lanes. Traditionally, *carriers* (i.e. trucking companies) have submitted quotes for individual lanes in the RFQ. This is akin to a *single-item reverse auction*, where each lane is awarded independently to a single carrier using a single criterion, usually price [49].

Today internet auctions provide shippers with a fast and easy way to simultaneously negotiate multiple potential contracts with a large number of carriers. The use of the internet as an auction medium has the benefit of decreasing information-gathering, participation, and transaction costs, as well as increasing geographic and

temporal conveniences [38]. Large corporations such as The Home Depot, Walmart Stores and Staples Inc. rely on applications from software providers to procure \$billions worth of services annually via internet auctions [19], [35], and [11]. Prominent providers of such software currently include CombineNet and Manhattan Associates.

Over the past decade, many of these auctions have allowed bidders to bid on combinations of lanes instead of bidding only on individual lanes. Such auctions, called *combinatorial auctions*, have three stages. First, the auctioneer (on behalf of the shipper) announces *multiple* lanes for bid (henceforth, *bid lanes*) in the auction. Second, the bidders (here, the carriers) submit bids for sets of bid lanes (*bundles*), rather than bidding on each bid lane individually. Third, the auctioneer determines the best set of bundles that collectively cover each bid lane, and awards contracts for these bundles (rather than awarding individual bid lanes) to the corresponding bidders.

An important benefit of combinatorial auctions is that they often make it possible to capture the benefits of *substitution effects* and *complementarities*, in which the value of a set is not simply the sum of its parts. Using a combinatorial auction in such cases allows bidders to express their true preferences, with the goal of finding better allocations. This is the case in truckload shipping, due, for example, to the fact that carriers must not only transport the bid lanes that they have been awarded, but must also return drivers home. If carriers can string together multiple loads to form a *continuous move* (tour), then they can decrease their empty mileage and thereby reduce cost.

We illustrate this in Figure 1 with a simple example with two loads. Here, we see that the bidder's bid price of transporting a load from A to B is $l \cdot x$ (the direct movement price per mile (l) times the distance (x) to move the load from A to B)

plus $e \cdot x$ (the empty movement price per mile (e) times the distance (x) to return the driver home from B to A). The price of transporting a load from B to A is computed similarly. The price of transporting *both* loads, however, is not $2(l \cdot x + e \cdot x)$ but instead only $2(l \cdot x)$, because the two loads can be combined to form a single tour, without any empty movements. If the loads were auctioned individually, in two separate auctions, then the bidders would face the following dilemma when bidding on the first load. If they bid high presuming the full price (including the empty return trip), they might lose the auction for bidding too high a price. But if they bid low presuming only the direct movement price and then did not win the second load, they could lose money. A combinatorial auction ameliorates this by allowing three bids: one for winning only load A, one for winning only load B, and one for winning both.

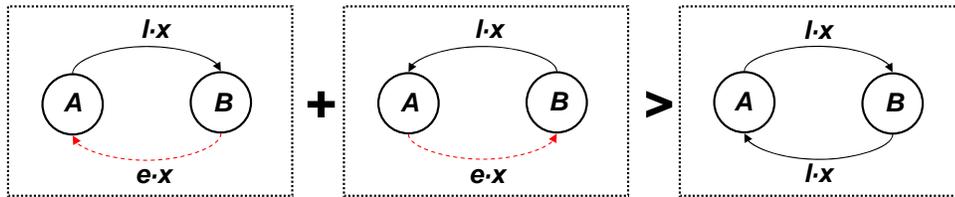


Figure 2.1: The benefits of complementarities demonstrated through a two load example.

More broadly, combinatorial auctions allow carriers to bid on *bundles* of lanes to produce more efficient movements. The efficiency gained in combinatorial bidding, in turn, allows carriers to submit more aggressive bids, thereby reducing transportation costs for the shipper.

There is a substantial stream of literature on the explicit computation of bundle bids for truckload transportation auctions. [54], [36] and [4] provide methods for identifying bundles that are likely to be good to bid on, and efficiently computing the bids for those bundles. [25] study a dynamic setting: they show how to compute

non-combinatorial bids for entire contracts in an environment where contracts are put up for bid sequentially over time. All of these works recognize that truckload transportation services is an area where the potential for economies of scale and scope is particularly rich. In fact, there is opportunity to leverage the complementarities in lanes among different shippers even before the auction begins [20], [21]. Likewise, there is also an opportunity to leverage synergies in lanes among different carriers to improve efficiencies in the transportation of contracted lanes [24], [50], [31], and [34].

However, two major hurdles remain that prevent the full realization of the benefits of combinatorial auctions. The first is *bid expression and communication*: to fully express economies of scale and scope among all items being auctioned, bidders must construct and submit bids for an exponential number of subsets of these items ($2^n - 1$ for an n -item auction). This is clearly intractable for all but the smallest instances. The second hurdle is in solving the *winner determination problem* (WDP), typically formulated as a *set partitioning problem* [6], to select the least-cost set of bundles such that each item is in exactly one bundle. WDP is an integer program with an exponentially-large number of binary variables and thus also intractable for all but the smallest instances.

[18] presented the state of knowledge for solving combinatorial auctions and suggest the use of an “oracle” to alleviate the burden of expressing and communicating an exponential number of bids. The auctioneer invokes the appropriate oracle at any stage of an auction to determine the bid for a particular bundle. Alternatively, an auctioneer may specify a bidding language [46], [40], [47], [1], [18], [42] and [30] to be used by all bidders. Bidding languages specify ways in which bids must be restricted to a subset of the potential bundles. Alternatively, these bidding languages may allow for full expression of preferences, provided that the preferences have some

special structure.

Next, assuming one can overcome the difficulties of bid expression and communication, the auctioneer is still left with solving an exponentially-large WDP. [46] and [18] showed that WDP is computationally manageable if the structure of allowable bids permits decomposition into disjoint groups, yields a tractable number of combinations, or results in constraint matrices with integral extreme points. Another strategy is to shift the computational burden of solving WDP to the bidders. [8], [9], and [32] proposed mechanisms that allow bidders to iteratively submit improvement bids. Finally, [46], [27], [56], [47], [48], and [30] presented algorithms and heuristics to efficiently solve WDP under certain conditions. Readers interested in a more comprehensive examination of the theory and applications of combinatorial auctions should refer to a recent book edited by [17].

In practice, the hurdles of exponential bidding and an exponentially-large WDP are sometimes circumvented by using iterative bidding, by restricting bidding to a small number of bundles, and by using exact and approximation algorithms, as discussed in the references above. However, bidding on only a small subset of lanes prevents the full realization of the benefits of a combinatorial auction, and incentive compatibility and individual rationality of the auction might be compromised if the auction is not solved to optimality [45].

No single generalized approach can find optimal solutions to fully-enumerated combinatorial auctions for all classes of problems. [41] discussed why, in the worst case, a general problem may require exponential communication. Therefore, as highlighted in the preceding paragraphs, most research focuses on exploiting problem structure to find acceptable solutions for specific types of auctions. The goal of our research is to show that the underlying structure of a truckload procurement problem

can be exploited similarly, enabling us to find solutions to fully-enumerated auctions in practical time frames.

This research extends the existing literature on combinatorial truckload procurement auctions (CTPA). For example, [49] and [12] presented the state of knowledge for CTPAs. [35] and [19] described early uses of combinatorial auctions for truckload procurement. [52] and [44] addressed carrier bidding strategies in multi-round auctions. [29] extended the carrier assignment models used in WDP to include shipper non-price objectives and carrier transit point costs.

Nevertheless, the full benefits of combinatorial auctions for truckload procurement have not yet been achieved in practice. One recent study [43] showed that only 28 percent of carriers submit bids of more than one lane in combinatorial auctions and the majority of these carriers only submit 2-7 multi-lane bundles due to practical constraints on bid preparation time, computational resources and technical expertise at their disposal.

We propose an *implicit bidding approach* to truckload procurement auctions that *can* (implicitly) capture the full, exponential set of bundles. This approach leverages the fact that there is a known and amenable structure underlying the cost of servicing a given set of bid lanes. Specifically, the least-cost tour (or set of tours) needed to cover a set of lanes can be computed by solving a *minimum-cost flow* problem. We therefore propose to embed this underlying cost structure (which we refer to as a *bid-generating function*) directly into WDP. This eliminates the need for the bidder to compute and communicate an exponential number of bids. Furthermore, we will show that the resulting WDP can be re-formulated as a *multi-commodity flow* (MCF) problem of polynomial size. Our computational results demonstrate the practical performance of the implicit bidding approach.

The contributions of this research are in:

1. presenting a new implicit bidding approach for combinatorial truckload procurement auctions that enables the complete set of all possible bids to be considered implicitly, and thus achieves full economies of scope;
2. developing tractable models to solve a basic truckload procurement auction to optimality, in single round, fully considering (implicitly) the exhaustive set of all possible bids;
3. showing how the power of mathematical programming can enable this basic problem to be extended to include additional important real-world operational considerations; and
4. taking advantage of this new capability to solve fully-enumerated truckload procurement auctions as a tool for conducting numerical analysis on the characteristics of CTPA solutions.

The remainder of this chapter is organized as follows. In §2.2, we formally present combinatorial auctions for truckload procurement. In §2.3, we introduce the implicit bidding approach for combinatorial truckload procurement auctions. In §2.4, we present computational experiments focusing on the tractability of the implicit bidding approach, solution characteristics under a variety of conditions, and performance comparison to bidding methods in practice. We conclude in §2.5 with a summary of our contributions and our suggestions for future research.

2.2 Combinatorial Auctions for Truckload Procurement

In a basic truckload procurement auction, the auctioneer specifies a set of bid lanes, each defined by an origin, a destination and a volume (typically corresponding

to the expected number of loads). Given a bundle of bid lanes, carriers determine the least-cost set of tours to serve these bid lanes, then use this cost in computing their bid price for this particular bundle. For example, in a first-price auction, carriers typically bid true-cost plus a percentage-based markup [51]. Throughout the manuscript, we will assume a first-price auction with a percentage-based markup for the sake of exposition. However, our approach is applicable to other auction mechanisms as well, for example, forward auctions where items are sold. Finally, the auctioneer solves a WDP to select bundles and allocate the corresponding lanes to the winning carriers.

2.2.1 Computing Bundle Bids

In order to understand how carriers compute their bids, we must first understand their cost structure. The carriers' cost of service can be decomposed by individual movements. The obvious cost is the *direct movement cost*, associated with actually moving a load from lane origin to destination. This cost is well understood by the carrier and is largely a function of distance (fuel, equipment depreciation, driver's wage, tolls, et cetera).

In addition, there is the *repositioning cost* associated with moving a truck from the destination of one lane to the origin of the next, so as to form tours. To minimize cost, and thus improve the probability of winning bids, carriers must try to build efficient continuous movements with minimal empty mileage. This can be accomplished not only by combining bid lanes, but also by taking advantage of a carrier's pre-existing contracted lanes and opportunities on the *spot market*. For example, Figure 2.2 shows a sequence of movements for efficiently transporting three bid loads.

These *repositioning opportunities* are the key to determining the actual bid price for any bundle of lanes to be served, resulting in significant synergies and comple-

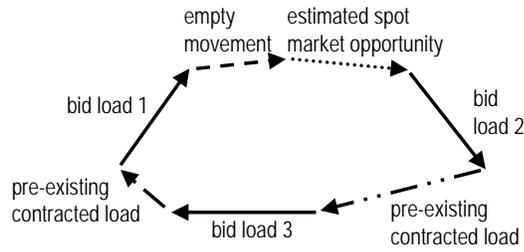


Figure 2.2: A cost-effective tour covering bid loads 1, 2, and 3 that leverages a carrier’s pre-existing contracts and spot market opportunities.

mentarities, as seen in the example above. Because repositioning opportunities are often not known with certainty at the time of the auction, carriers typically estimate these opportunities for each directed city pair (i, j) in the network. One way to represent this is with an n -tiered step function, where each tier represents capacity and cost estimates of a different type of repositioning opportunity. For example, one tier might represent the expected number of movements from a pre-existing contracted lane (with other shippers), which can be used “for free”, as these movements represent hired rather than empty movements. Another tier might represent the potential for *partial connections*: pre-existing lanes that require the driver to travel empty from i to some nearby location before picking up the load and delivering it to some location near j , thereby incurring limited empty mileage costs. Estimates of spot market opportunities would be represented by additional tiers as well. Finally, the highest cost tier, with infinite capacity, represents empty movements from i to j .

Figure 2.3 provides an example of such a step function. More generally, high-traffic city pairs would have high-capacity, low-cost tiers because of the abundance of backhaul opportunities, while low-traffic city pairs would have lower-capacity, higher-cost tiers representing the decreased likelihood of finding complementary lanes.

Finally, we reiterate that given a set of bid lanes, direct movement cost and

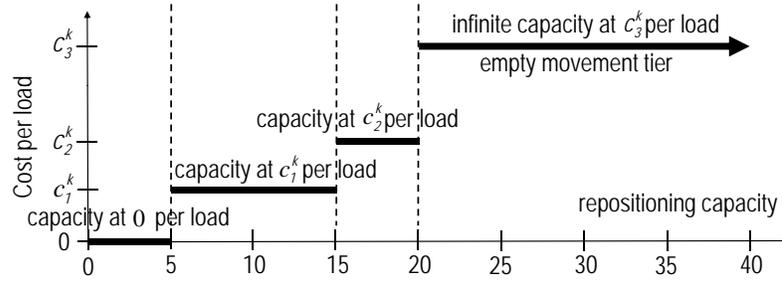


Figure 2.3: Estimate of carrier k 's repositioning capacity and price (cost plus markup) from city i to city j using a 4 tiered step-function.

repositioning opportunities (costs and capacities), carriers determine the least-cost set of tours to serve these bid lanes, then use this cost in computing the bid price, typically, cost plus a percentage-based markup.

2.2.2 Traditional Winner Determination Problem

Once the bids have been submitted, the auctioneer then solves the WDP to select bundles and allocate lanes to winning carriers. The traditional winner determination formulation (T-WDP) is as follows:

Sets

- \mathcal{K} is the set of carriers and $K = |\mathcal{K}|$
- $\mathcal{A}^{\mathcal{L}}$ is the set of arcs representing bid lanes and $L = |\mathcal{A}^{\mathcal{L}}|$
- \mathcal{S}^k is the set of bundles submitted by carrier k

Parameters

- D_a is the expected volume of bid lane a , $\forall a \in \mathcal{A}^{\mathcal{L}}$
- δ_{sa}^k is the number of lane a movements in bundle s of carrier k , $\forall a \in \mathcal{A}^{\mathcal{L}}$, $k \in \mathcal{K}$, $s \in \mathcal{S}^k$

- b_s^k is carrier k 's bid price (cost plus a percentage-based markup) for bundle s , $\forall k \in \mathcal{K}, s \in \mathcal{S}^k$

Variable

- x_s^k is a binary variable that takes value 1 if carrier k is awarded bundle s and 0 otherwise, $\forall k \in \mathcal{K}, s \in \mathcal{S}^k$

$$(2.1a) \quad (\text{T-WDP}) \quad \min \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}^k} b_s^k x_s^k$$

$$(2.1b) \quad \text{s.t.} \quad \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}^k} \delta_{sa}^k x_s^k = D_a \quad \forall a \in \mathcal{A}^{\mathcal{L}}$$

$$(2.1c) \quad \sum_{s \in \mathcal{S}^k} x_s^k \leq 1 \quad \forall k \in \mathcal{K}$$

$$(2.1d) \quad x_s^k \in \{0, 1\} \quad \forall k \in \mathcal{K}, s \in \mathcal{S}^k$$

The objective (2.1a) is to minimize the total cost the shipper pays for procuring truckload services for all lanes in $\mathcal{A}^{\mathcal{L}}$. Constraint set (2.1b) states that bundles must be chosen such that each lane in $\mathcal{A}^{\mathcal{L}}$ is fully covered by the awarded bundles. In a fully-enumerated CTPA, an additional constraint set (2.1c) stating that each carrier can be awarded at most one bundle is imposed (note that each bundle might contain more than one tour). This constraint set is needed to ensure that we do not select a combination of bundles for a given carrier such that, in total, the combination of bundles violates some of the carrier's operational constraints (for example, using more repositioning capacity than there exists).

We conclude this section by re-iterating the fact that, for practically-sized truckload procurement auctions with thousands of lanes, it is of course not possible to explicitly enumerate all bundle bids. For each lane $a \in \mathcal{A}^{\mathcal{L}}$, a carrier may bid a volume of zero up to D_a , so, the total number of distinct combinations is equal to

$$\Lambda = \prod_{a \in \mathcal{A}^c} (D_a + 1).$$

Even in the case of a regional carrier who bids only on a subset of a few dozen lanes, this would still entail millions of bundle combinations. Instead, carriers submit bids for only a small subset of the bundles, due to practical constraints on bid-preparation time, computational resources, and technical expertise available at their disposal [43]. As a result, the solution quality of CTPAs in practice is often compromised.

2.3 Implicit Truckload Combinatorial Auctions

In the majority of the combinatorial auction literature, bundle prices are assumed to be *exogenously endowed*. For CTPAs, although most of the literature continues with this exogenous endowment of bundle bids, there is more recent literature on the identification and pricing of bundles. For instance, [54], and [36] show how carriers can efficiently identify promising bundles to bid on as well as the bid price for those bundles. [25] also explore how synergies impact the pricing of bids for bundles in a dynamic environment, where demand for loads to be served arrives over time.

We follow a similar approach for the purpose of establishing how carriers generate bids for specific bundles. The key idea, as recognized by the papers cited above, is that bundle prices depend critically on the network structure, existing commitments, and potential future commitments of the carrier. We model these by a *bid-generating function* (BGF).

More importantly, the structure of this BGF can now be exploited using an *implicit bidding approach* to solve WDP using BGFs directly, in lieu of the actual bids. This enables the exhaustive set of bundles to be considered implicitly without sacrificing tractability.

2.3.1 The Bid-Generating Function

Given a bundle of bid lanes s , carrier k 's price to service these lanes is comprised of the individual direct movement and repositioning movement prices. For a given set of lanes, the direct movement prices are fixed and known in advance. The repositioning movement prices, on the other hand, depend on the continuous moves that the carrier constructs to minimize the total price of the bundle. These repositioning moves enable carriers to exploit synergies and complementarities that exist in serving lanes of a bundle. As such, the price of a bundle may be significantly different from the sum of bid prices of individual lanes in that bundle, as discussed in §2.2.1.

For a given carrier k , the problem of determining the least-cost set of continuous moves (and thus, the value of the corresponding bids) to serve a set of bid lanes can be computed by solving a network flow problem on a directed graph, $G(\mathcal{N}, \mathcal{A})$. In this graph, node set \mathcal{N} represents origin, destination and/or transshipment cities and arc set \mathcal{A} represents direct movement and repositioning movement lanes with associated arc prices (cost plus profit markup) and capacities. Repositioning movement arcs may include the carrier's estimates of opportunities from pre-existing contracted lanes, anticipated opportunities on the spot market, and empty movements. In particular, we construct one arc for each tier of repositioning capacity between a directed city pair (as described in §2.2.1).

The problem is then to create the least-cost set of tours in this network such that each bid lane in s is covered. The notation and formulation for this BGF (which we denote by f^k) are as follows.

Sets

- \mathcal{N} is the set of nodes corresponding to arc origins or destinations

- \mathcal{A}^k is the set of arcs representing carrier k 's estimated repositioning opportunities, one for each tier of a directed city pair

Parameters

- $\mathcal{O}(a)$ is the origin of arc a , $\forall a \in \mathcal{A}^{\mathcal{L}} \cup \mathcal{A}^k$
- $\mathcal{D}(a)$ is the destination of arc a , $\forall a \in \mathcal{A}^{\mathcal{L}} \cup \mathcal{A}^k$
- p_a^k is carrier k 's price for a unit movement on arc a , $\forall a \in \mathcal{A}^{\mathcal{L}} \cup \mathcal{A}^k$
- u_a^k is carrier k 's estimated repositioning capacity on arc a , $\forall a \in \mathcal{A}^k$
- \mathbf{x}^{ks} is a vector of carrier k 's bid volume, where element x_a^{ks} represents the bid volume of lane $a \in \mathcal{A}^{\mathcal{L}}$ in bundle s

Observe that the price of a unit movement on arc a , denoted p_a^k , is strictly a function of the carrier's cost to complete a movement on this arc plus a profit markup. In turn, the profit markup can account for factors such as competitive strategy, geographic location of depots, transshipment centers, et cetera. Carriers can also use the parameter p_a^k to indicate an undesirable lane (e.g. a lane outside their region of coverage) by setting a very high value. Lastly, in the context of pricing bundle s , the quantity \mathbf{x}^{ks} is a parameter of the BGF f^k and not a variable.

Variable

- y_a^k is the number of repositioning movements on arc a made by carrier k , $\forall a \in \mathcal{A}^k$

The BGF can now be represented as the following integer program:

$$(2.2a) \quad f^k(\mathbf{x}^{ks}) = \sum_{a \in \mathcal{A}^c} p_a^k x_a^{ks} + \min \sum_{a \in \mathcal{A}^k} p_a^k y_a^k$$

$$(2.2b) \quad \text{s.t.} \quad \sum_{a \in \mathcal{A}^c: \mathcal{O}(a)=i} x_a^{ks} + \sum_{a \in \mathcal{A}^k: \mathcal{O}(a)=i} y_a^k = \sum_{a' \in \mathcal{A}^c: \mathcal{D}(a')=i} x_{a'}^{ks} + \sum_{a' \in \mathcal{A}^k: \mathcal{D}(a')=i} y_{a'}^k \quad \forall i \in \mathcal{N}$$

$$(2.2c) \quad y_a^k \leq u_a^k \quad \forall a \in \mathcal{A}^k$$

$$(2.2d) \quad y_a^k \in \mathbb{Z}^+ \quad \forall a \in \mathcal{A}^k$$

The objective function (2.2a) states that carrier k 's price for bundle s is the sum of direct movement prices (which depends solely on s and is known in advance) and the repositioning movement prices (which depends on the chosen routing). Because these sums are over distinct sets of movement arcs, the carrier can submit different prices for bid arcs and repositioning arcs over the same origin-destination pairs. Constraint set (2.2b) ensures flow conservation at each node in the network; that is, the number of movements into a node must be equal to the number of movements out of the node. This ensures that the resulting solution is a set of tours covering lanes in s . Constraint set (2.2c) states that the repositioning capacity used must be less than or equal to the available capacity.

$f^k(\mathbf{x}^{ks})$ has two important structural characteristics that have significant impact on tractability. First, the integrality restrictions (2.2d) can be replaced by non-negativity constraints $y_a^k \geq 0, \forall a \in \mathcal{A}^k$ because the constraint matrix of this problem is totally unimodular. Secondly, BGF can be reformulated as a *circulation problem* via a simple variable redefinition. Circulation problems, which are special cases of minimum cost flow problems, are well known to be easy to solve. For examples of polynomial time algorithms for the circulation problem please refer to [2].

2.3.2 The Implicit Winner Determination Problem

Using the traditional auction mechanism described in §2.2.2, each carrier must solve a BGF (2.2) to obtain a bid price for each bundle of interest. For real-world truckload procurement auctions with thousands of lanes, constructing bid prices for the full exponential set of bundles is not possible. Furthermore, even if carriers could compute and communicate bids for all bundles, the auctioneer could not solve the corresponding exponentially-large WDP. We show that these hurdles can be overcome by using an implicit bidding approach, which directly embeds a carrier’s BGF into WDP. The resulting polynomially-sized (with respect to bid lanes, number of carriers and number of nodes) model is solution-equivalent to the fully-enumerated T-WDP but, in contrast, is tractable for practically-sized instances.

The thrust of this implicit bidding approach is the following. Rather than submit an exponential number of bundle-price pairs, each carrier k instead submits the parameters of the BGF, f^k . These parameters are simply a list of all the arcs with corresponding prices (including any profit markups) and capacities. Note, of course, that a carrier choosing not to bid on particular lanes (e.g. those outside their geographic region of coverage) would simply not include those arcs in their parameters. Additionally, the shipper may choose to include a “dummy” carrier with a bid generating function corresponding to the spot market. In this case, the dummy carrier’s cost function for each directed city pair (i, j) in the network would be a trivial single-tiered step function.

The auctioneer can then imbed f^k directly into WDP, which can be reformulated as:

$$\begin{aligned}
(2.3a) \quad & \min \sum_{k \in \mathcal{K}} f^k(\mathbf{x}^k) \\
(2.3b) \quad & \text{s.t.} \quad \sum_{k \in \mathcal{K}} x_a^k = D_a \quad \forall a \in \mathcal{A}^{\mathcal{L}} \\
(2.3c) \quad & x_a^k \in \mathbb{Z}^+ \quad \forall a \in \mathcal{A}^{\mathcal{L}}, k \in \mathcal{K}
\end{aligned}$$

Observe that (2.3) implicitly captures substitution effects and complementarities, resulting in a fully-enumerated truckload procurement auction where each winner is awarded exactly one bundle (possibly empty). This bundle is described by the single vector of decision vector \mathbf{x}^k (with one element per bid lane), taking the place of the *set* of vectors \mathbf{x}^{ks} (with one vector per bundle) previously defined in §2.3.1. For each $a \in \mathcal{A}^{\mathcal{L}}$, x_a^k is the volume of bid lane a assigned to carrier k . Note, however, that in place of $K \cdot \Lambda$ binary variables, there are now only $K \cdot L$ integer variables in the model described by (2.3). As an example, an auction with 10 carriers and 100 bid lanes (each with a volume of 10) translates to a reduction from over 1.37×10^{105} binary variables to only 1,000 integer variables.

Of course, even with this reduction in size, the new formulation may still be quite difficult to solve, depending on the structure of f^k . As we have noted in §2.3.1, however, f^k is simply a circulation problem. Thus, embedding (2.2) in place of the function f^k leaves us with the following mixed integer program, which we denote by Implicit Winner Determination Problem (I-WDP).

$$(2.4a) \quad \min \sum_{k \in \mathcal{K}} \left[\sum_{a \in \mathcal{A}^{\mathcal{L}}} p_a^k x_a^k + \sum_{a \in \mathcal{A}^k} p_a^k y_a^k \right]$$

$$(2.4b) \quad \text{s.t.} \quad \sum_{k \in \mathcal{K}} x_a^k = D_a \quad \forall a \in \mathcal{A}^{\mathcal{L}},$$

$$(2.4c) \quad \sum_{a \in \mathcal{A}^{\mathcal{L}}: \mathcal{O}(a)=i} x_a^k + \sum_{a \in \mathcal{A}^k: \mathcal{O}(a)=i} y_a^k = \sum_{a' \in \mathcal{A}^{\mathcal{L}}: \mathcal{D}(a')=i} x_{a'}^k + \sum_{a' \in \mathcal{A}^k: \mathcal{D}(a')=i} y_{a'}^k$$

$$\forall i \in \mathcal{N}, k \in \mathcal{K},$$

$$(2.4d) \quad y_a^k \leq u_a^k \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^k,$$

$$(2.4e) \quad x_a^k \in \mathbb{Z}^+ \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^{\mathcal{L}},$$

$$(2.4f) \quad y_a^k \in \mathbb{Z}^+ \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^k.$$

Note that x is no longer a fixed parameter in I-WDP, but now a vector of decision variables. I-WDP has two sets of variables, x and y , representing bid lane assignments and the usage of carrier's repositioning capacities, respectively. The objective function (2.4a) minimizes the total price attributed to direct movements and repositioning movements. The lane cover constraint set (2.4b) stipulates that all bid lanes must be covered by selected carriers. Constraint set (2.4c) ensures flow conservation of nodes for each carrier; that is, the number of movements into a node must be equal to the number of movements out of the node, thereby ensuring that the resulting allocation defines a set of continuous moves (tours). Constraint set (2.4d) states that the repositioning capacities used to complete the tours must be less than or equal to the capacities available.

So, rather than solve an exponentially-sized (with respect to the number of bid lanes, number of carriers and number of nodes) T-WDP, we can instead solve a polynomially-sized I-WDP. I-WDP is solution equivalent to a fully-enumerated T-WDP, as formally stated by Proposition II.1.

Proposition II.1. *Consider an auction with a set of carriers, bidding for a set of bid lanes $\mathcal{A}^{\mathcal{L}}$. For each carrier $k \in \mathcal{K}$, if the price of a specific bundle s is given by the solution to f^k (defined by 2.2), then I-WDP is solution-equivalent to a fully-enumerated T-WDP.*

Proof: We will show that an optimal solution to T-WDP is a feasible solution to I-WDP, with equivalent cost, and vice versa. Let \mathcal{S}^* be the set of bundles that correspond to an optimal solution for T-WDP and let $z_{\text{T-WDP}}(\mathcal{S}^*)$ be the total cost. Let $(\mathbf{x}^*, \mathbf{y}^*)$ be the set of vectors that correspond to an optimal solution for I-WDP and let $z_{\text{I-WDP}}(\mathbf{x}^*, \mathbf{y}^*)$ be the total cost.

Claim 1: $z_{\text{T-WDP}}(\mathcal{S}^*) \geq z_{\text{I-WDP}}(\mathbf{x}^*, \mathbf{y}^*)$

Proof: For each $s^k \in \mathcal{S}^*$, define \mathbf{x}^k to be a vector of size L , where $x_a^k = w_a$ (the bid volume of lane a in s^k) $\forall a \in \mathcal{A}^{\mathcal{L}}$. Since p_s^k (price of bundle $s^k \in \mathcal{S}^*$) is obtained by solving $f^k(\mathbf{x}^k)$, there exist vectors \mathbf{y}^k corresponding to the minimum price set of repositioning moves used in s^k . Observe that $(\mathbf{x}^k, \mathbf{y}^k)$ satisfies constraints (2.4c)-(2.4f) of I-WDP. If we let (\mathbf{x}, \mathbf{y}) be defined as the concatenation of $(\mathbf{x}^k, \mathbf{y}^k) \forall s^k \in \mathcal{S}^*$, then (\mathbf{x}, \mathbf{y}) also satisfies (2.4b) and is a feasible solution to I-WDP. Since the cost coefficients of $f^k \forall k \in \mathcal{K}$ and I-WDP are identical, $z_{\text{T-WDP}}(\mathcal{S}^*) = z_{\text{I-WDP}}(\mathbf{x}, \mathbf{y})$. Finally, the optimal solution of I-WDP can only be better, thus we must have $z_{\text{T-WDP}}(\mathcal{S}^*) = z_{\text{I-WDP}}(\mathbf{x}, \mathbf{y}) \geq z_{\text{I-WDP}}(\mathbf{x}^*, \mathbf{y}^*)$.

Claim 2: $z_{\text{T-WDP}}(\mathcal{S}^*) \leq z_{\text{I-WDP}}(\mathbf{x}^*, \mathbf{y}^*)$

Proof: $(\mathbf{x}^*, \mathbf{y}^*)$ can be decomposed into $(\mathbf{x}^k, \mathbf{y}^k)$ for each carrier $k \in \mathcal{K}$. If $p^k(\mathbf{x}^k, \mathbf{y}^k)$ is the total price to serve the bundle defined by \mathbf{x}^k using repositioning movements corresponding to \mathbf{y}^k , then $z_{\text{I-WDP}}(\mathbf{x}^*, \mathbf{y}^*) = \sum_{k \in \mathcal{K}} p^k(\mathbf{x}^k, \mathbf{y}^k)$. Observe that repositioning movements \mathbf{y}^k satisfies (2.2b)-(2.2d), and thus $(\mathbf{x}^k, \mathbf{y}^k)$ is a feasible solution of f^k . Since the optimal solution of $f^k(\mathbf{x}^k)$ can only do better, we must have $f^k(\mathbf{x}^k) \leq$

$p^k(\mathbf{x}^k, \mathbf{y}^k) \forall k \in \mathcal{K}$. This implies $z_{\text{I-WDP}}(\mathcal{S}^*) = \sum_{s^k \in \mathcal{S}^*} f^k(\mathbf{x}^k) \leq \sum_{k \in \mathcal{K}} p^k(\mathbf{x}^k, \mathbf{y}^k) = z_{\text{I-WDP}}(\mathbf{x}^*, \mathbf{y}^*)$. Claims 1 and 2 together imply that $z_{\text{T-WDP}}(\mathcal{S}^*) = z_{\text{I-WDP}}(\mathbf{x}^*, \mathbf{y}^*)$. \square

Finally, as proven by Proposition II.2, we observe that this formulation is a special case of the multi-commodity flow problem.

Proposition II.2. *I-WDP can be reformulated as a multi-commodity flow problem.*

Proof: The proof is by construction. For each bid lane $a \in \mathcal{A}^{\mathcal{L}}$ with corresponding origin i and destination j , define $D_{ij} \equiv D_a$ and let $\mathcal{A}^{\mathcal{B}}$ represent this set of movements. For each carrier $k \in \mathcal{K}$ and bid lane $a \in \mathcal{A}^{\mathcal{L}}$ with corresponding origin i and destination j , define an arc (i, j) with per unit cost $c_{ij}^k \equiv p_a^k$, lower bound of $l_{ij}^k \equiv 0$, upper bound of $u_{ij}^k \equiv D_a$. For each carrier $k \in \mathcal{K}$ and repositioning lane $a \in \mathcal{A}^k$ with corresponding origin i and destination j , define an arc (i, j) with per unit cost $c_{ij}^k \equiv p_a^k$, lower bound $l_{ij}^k \equiv 0$ and upper bound $u_{ij}^k \equiv u_a^k$; let \mathcal{S}_r^k represent this set of arcs. Finally, let $\mathcal{A}^k \equiv \mathcal{A}^{\mathcal{B}} \cup \mathcal{S}_r^k$. Letting x_{ij}^k represent the amount of flow of commodity (i.e. carrier) $k \in \mathcal{K}$ on arc $(i, j) \in \mathcal{A}^k$, I-WDP can be written as:

$$(2.5a) \quad \min \sum_{k \in \mathcal{K}} \sum_{(i,j) \in \mathcal{A}^k} c_{ij}^k x_{ij}^k$$

$$(2.5b) \quad \text{subject to: } D_{ij} \leq \sum_{k \in \mathcal{K}} x_{ij}^k \leq D_{ij} \quad \forall (i, j) \in \mathcal{A}^{\mathcal{B}}$$

$$(2.5c) \quad \sum_{k \in \mathcal{K}} \sum_{j: (i,j) \in \mathcal{A}^k} x_{ij}^k - \sum_{k \in \mathcal{K}} \sum_{j: (j,i) \in \mathcal{A}^k} x_{ji}^k = 0 \quad \forall k \in \mathcal{K}, i \in \mathcal{N}$$

$$(2.5d) \quad l_{ij}^k \leq x_{ij}^k \leq u_{ij}^k \quad \forall k \in \mathcal{K}, (i, j) \in \mathcal{A}^k$$

(2.5) is a multi-commodity flow problem. \square

Although theoretically difficult [22], the multi-commodity flow problem is known to be easy to solve in practice for many real-world instances [2]. This is the case for

the truckload procurement auctions, as we will demonstrate through computational results in §2.4.1.

An implicit assumption in our model is that bidders (carriers) are easily able to compute the p_a^k values. While sometimes computing p_a^k may be as straightforward as adding variable costs (fuel, wages, etc.), amortized fixed costs (equipment depreciation, for example) and a profit margin, carriers may wish to incorporate other considerations such as competition on the lane and expectations of future traffic. We refer the reader to [12] for a discussion on how carriers price individual lanes. There is also a stream of literature (e.g. [25], [52]) on pricing bundles of lanes, in static or dynamic environments; these approaches also provide insight on pricing single lanes. Our work only requires bid prices to be computed for single lanes rather than multi-lane bundles, and then uses the implicit bidding approach to obtain a WDP that is substantially easier to solve.

2.3.3 Operational Considerations

In addition to the basic constraints shown in (2.4), shippers and carriers may have other operational considerations to take into account. Although all models are, of course, simplifications of the real world, we are able to expand our formulation to capture some of the most natural operational considerations. We provide several examples as follows.

First, we begin by defining auxiliary variables q_a^k and q^k for use in defining constraints corresponding to these operational considerations. Let q_a^k be a binary variable that takes value 1 if carrier k is awarded at least one load in lane a and 0 otherwise and q^k be a binary variable that takes value 1 if carrier k is awarded at least one load in *any* lane and 0 otherwise. The following relationships are then helpful in adding operational constraints.

$$(2.6a) \quad q_a^k \leq x_a^k \leq D_a q_a^k \quad \forall a \in \mathcal{A}^{\mathcal{L}}, k \in \mathcal{K}$$

$$(2.6b) \quad x_a^k \leq D_a q_a^k \quad \forall a \in \mathcal{A}^{\mathcal{L}}, k \in \mathcal{K}$$

$$(2.6c) \quad q^k \leq \sum_{a \in \mathcal{A}^{\mathcal{L}}} x_a^k \quad \forall k \in \mathcal{K}$$

- **Load volume:** The shipper and carriers may want to restrict the load volume (across all bid lanes) that a carrier can be awarded. A minimum load volume ensures that carriers are awarded at least a threshold volume. A maximum load volume ensures that carriers' transportation capacities are observed and there is a manageable number of shipper-carrier relationships.

These operational considerations can be modeled as follows where, $\underline{\alpha}^k$ is the minimum number of loads carrier k must win (or nothing), $\bar{\alpha}^k$ is the maximum number of loads carrier k can win.

$$(2.7a) \quad \underline{\alpha}^k q^k \leq \sum_{a \in \mathcal{A}^{\mathcal{L}}} x_a^k \leq \bar{\alpha}^k \quad \forall k \in \mathcal{K}$$

Constraints (2.7a) say that the total volume of loads awarded to carrier k must be between $\underline{\alpha}^k$ and $\bar{\alpha}^k$ (inclusive) or zero.

- **Number of assigned carriers:** The shipper might prefer to award bid lanes to no fewer than $\underline{\beta}$ carriers and to no more than $\bar{\beta}$ carriers, thus ensuring a manageable number of vendor relationships and adequate spreading of risk. We can model such restrictions as follows:

$$(2.8a) \quad \underline{\beta} \leq \sum_{k \in \mathcal{K}} q^k \leq \bar{\beta}$$

Constraints (2.8a) say that the total number of assigned carriers must be between $\underline{\beta}$ and $\bar{\beta}$ (inclusive).

- **Number of assigned carriers per lane:** The shipper may prefer to award a lane to no fewer than $\underline{\gamma}_a$ carriers and no more than $\bar{\gamma}_a$ carriers to ensure operational efficiency.

$$(2.9a) \quad \underline{\gamma}_a \leq \sum_{k \in \mathcal{K}} q_a^k \leq \bar{\gamma}_a \quad \forall a \in \mathcal{A}^{\mathcal{L}}$$

Constraints (2.9a) say that for each lane, $a \in \mathcal{A}^{\mathcal{L}}$, the total number of carriers assigned must be between $\underline{\gamma}_a$ and $\bar{\gamma}_a$ (inclusive).

- **Favoring of incumbents and performance measures:** There is a cost to the shipper to start a new relationship with a carrier. In practice, incumbents are often favored by 3-5 percent – especially for service-critical or time-sensitive lanes [11]. Similarly, the level of service (such as percentage on-time, claims performance, acceptance rate, et cetera) provided by a carrier can also be taken into consideration. These operational considerations can be accounted for by simply adjusting a new carrier’s price coefficients by a constant or multiplicative factor.

2.3.4 Privacy Issues of the Implicit Bidding Approach

We have tacitly presumed that carriers would be willing to submit separate price bids for each possible bundle, and that it is not a privacy concern that prevents them from doing so, but rather a practical one — it simply is not tractable to compute and submit such a large number of bids. However, submitting price bids for every possible bundle – either explicitly through enumerative bidding or implicitly through a bid-generating function — transmits a substantial amount of information to the auctioneer. Sharing such a large amount of information may naturally raise privacy concerns for the carriers. Broadly speaking, there will always be tension between the perceived risk of providing information versus the opportunities to be gained by

leveraging synergies (which can benefit both the carriers and the auctioneer).

Thus it is worth noting that our implicit bidding approach will be subject to such tensions. For example, transmitting tiered pricing for different repositioning-arc volumes in the bid-generating function implicitly reveals capacity information about the carrier. On the other hand, there are substantial gains to be made by an efficient allocation of lanes to carriers, resulting in a win-win situation for both the auctioneer and the carriers. For the auctioneer, the gains are clear in that the overall price of procuring the transportation services will be lower. For the carriers, the efficient allocation resulting from our approach means that overall, the empty movement by carriers is much lower than in an inefficient allocation. Therefore, carriers are able to better serve lanes that they win (i.e. serve with lower empty/wasteful repositioning movement), and thereby use their excess capacity to earn more revenues from other markets. While it is of course true that some individual carriers may earn lower profit from this approach when compared to some other, less expressive approach, our approach drives overall inefficiencies resulting from unprofitable repositioning moves out of the system. From a longer-term perspective, greater truckload efficiency makes the truckload market more competitive versus alternative shipment modes like rail and air, which has implications for the long-term viability of carriers.

We believe, given the benefits of achieving full economies of scope in a fully-enumerated CTPA, that carriers and shipper have significant incentives to overcome these concerns. Consider an analogous example, Vendor Managed Inventory (VMI) systems [14], in which a retailer provides its suppliers with direct visibility to inventory levels. Like our proposed approach, VMI raises concerns about information privacy and security. However, VMI systems are widely used today by large corporations such as Walmart and The Home Depot because of the substantial benefits they

provide [37]. The practice of CPFR (collaborative planning, forecasting and replenishment) in the manufacturing and distribution sectors requires even more extensive sharing of information [5]. The reason different agents (manufacturers, distributors, retailers, etc.) participate in these systems despite concerns about sharing critical information about capabilities and forecasts is that these systems allow all parties to benefit from the realized efficiencies.

In practice, third party service providers such as Manhattan Associates and Ariba can provide services such as bidder pre-qualification and transaction confidentiality to improve information security and privacy and to limit the risk of information leakage. Additionally, emerging research on cryptographically-secured auctions [33], [26] provides an additional way to protect information. We believe our proposed method provides significant incentives for its use and as such may galvanize deployment of existing, or development of new, infrastructures.

2.4 Computational Experiments

We conducted a set of computational experiments to assess the overall effectiveness of the implicit bidding approach for CTPAs. Specifically, we have focused on:

- the **tractability of I-WDP** and the impact of instance size (number of bid lanes and load volume) on solution time;
- the **impact of operational constraints** (load volume, assigned carriers, assigned carriers per lane) on solution time;
- the **impact of instance characteristics** (repositioning capacity and network structure) on solution characteristics (solution time, number of assigned carriers, and empty movement ratio);

and

- a **comparison to bidding methods in practice** in terms of solution times and allocation costs.

Computational experiments were conducted on a Sun x4600-M2 with 8 AMD Opteron 8218 processors and 64 GB of RAM. The test machine was running Red Hat Enterprise Linux 4. Models and algorithms were coded using C++ and ILOG Concert Technology and solved using ILOG CPLEX 10.0. Parameter files for all computational instances in §2.4 are available online [15].

2.4.1 Tractability of I-WDP

We evaluate the performance of I-WDP on randomly generated instances representing various sized auctions. Random instances are controlled by the following parameters: number of nodes (cities), number of carriers, number of bid lanes, repositioning capacity per carrier, and carriers' price structures (represented by pairs of direct movement and empty movement price-per-mile).

We generated five sets of experiments representing auctions of size 1000, 2000, 3000, 4000 and 5000 bid lanes on a network with 100 nodes representing the 100 most populous cities in the United States. There are 50 carriers (bidders) bidding in each auction. For each set of auctions, we randomly generated 10 instances and report cumulative statistics. The volume of each bid lane is selected uniformly between 50 and 200 loads. A carrier's repositioning capacity is represented by a set of capacitated, preexisting contracted lanes (that can be used for "free") and a set of uncapacitated empty movement lanes. The number of preexisting contracted lanes per carrier is selected uniformly between five and fifteen percent of the number of bid lanes, with each pre-existing contracted lane volume selected uniformly between

10 and 100 loads.

Carriers' movement prices are generated by multiplying travel distance and a per-mile movement price to ensure that the triangle inequality is satisfied. A carrier's price to serve an additional load in a bid lane is equal to the distance from bid lane origin to bid lane destination times the carrier's direct movement price-per-mile, generated using a Normal distribution, $N(1.10, 0.05^2)$. Similarly, a carrier's price to move empty between any city pair is equal to the distance between the city pair times the carrier's empty movement price-per-mile generated using a Normal distribution, $N(0.80, 0.05^2)$.

Results

Solution characteristics for the five auction sizes are shown in Table 2.1. The median times reported are substantially lower than the averages which indicates that average solution times are skewed by a few long running instances. It is interesting to note that, generally, average solution times are inversely proportional to the size (number of bid lanes) of the auction. All else being equal, increasing the number of bid lanes in the auction actually *improves* solution time. Intuitively, given a fixed-size network with uniformly distributed lanes, increasing the number of lanes in the auction improves the probability of finding complementary lanes. Therefore, for large auctions, the price of the auction is primarily dominated by direct movement prices and the majority of bid lanes are allocated to a smaller number of low-cost carriers.

The solution times suggest that we can in fact solve to optimality fully-enumerated CTPAs of up to 5000 bid lanes (over 600,000 bid loads) with relative ease. We contrast this again with the traditional approach, which would require each carrier to compute and submit an exponential number of bundle bids and the auctioneer to solve a T-WDP with a corresponding number of binary variables - clearly, an

No. bid lanes	Avg. no bid loads	Avg. repos. capacity (loads)	Solution Time (sec.)			Avg. no. carriers assigned	Avg. no. of B&B nodes
			Avg.	Median	Std. Dev.		
1,000	124,425	276,780	181	91	183	32	6
2,000	249,665	551,792	140	99	137	22	2
3,000	376,260	820,569	123	99	93	15	5
4,000	500,493	1,078,834	147	89	123	13	10
5,000	623,654	1,370,068	127	81	122	11	2

Table 2.1: I-WDP solution characteristics for various auctions sizes.

intractable task.

2.4.2 Impact of Operational Considerations

We next consider the impact on performance of imposing constraints on load volume, number of assigned carriers and number of assigned carriers per lane. We again consider auctions with 1000, 2000, 3000, 4000 and 5000 bid lanes, using the basic instances generated in §2.4.1 as the baseline. In addition, we conducted the following sets of experiments: basic problem with additional constraints on load volume (constraints 2.6b-2.6c, 2.7a), basic problem with additional constraints on the number of assigned carriers (constraints 2.6b-2.6c, 2.8a) and basic problem with constraints on the number of assigned carriers per lane (constraints 2.6a, 2.9a). We constrained the total load volume awarded to any carrier to be between 50 loads and 40 percent of the total load volume. The number of assigned carriers was constrained to be at least 5 and at most 20. Lastly, we constrained the number of assigned carriers per lane to be at most ten.

Results

As expected, solution time increased with additional constraints. These operational considerations depend on imposing constraints (2.6a) and (2.6b) to define the auxiliary variables q^k and q_a^k . Constraint set (2.6a) and (2.6b) are generally very weak

Number of bid lanes	Number of bid loads	Repositioning capacity (loads)	Avg. solution time (sec.)			
			Basic	Load volume	Assigned carriers	Carriers per lane
1,000	124,507	276,569	181	1,671	3,886	643
2,000	249,675	551,522	140	1,361	1,841	530
3,000	374,999	817,088	123	840	676	900
4,000	499,878	1,087,532	147	1,135	1,012	624
5,000	624,412	1,384,599	127	774	1,784	773

Table 2.2: Average solution times for I-WDP with constraints on total load volume, number of assigned carriers and number of assigned carriers per lane.

because in most cases the entire lane volume is not assigned to a single carrier. These types of “big M” constraints typically lead to a weak linear programming (LP) relaxation, which is well known to be computationally undesirable [39]. There is potential here for future research to develop “stronger” alternative formulations and cutting plane algorithms for improving LP relaxations of constrained I-WDP.

Nonetheless, constrained I-WDPs remain tractable for auctions with up to 5000 bid lanes (over 600,000 bid loads), in some cases with improving tractability as the number of bid lanes in the auction grows. The results represented in Table 2.2 were obtained using default CPLEX solver settings and no preprocessing routines. This also leads us to believe that further improvements in solution times of I-WDP with operational constraints are attainable.

2.4.3 Impact of Instance Characteristics

In the preceding sections, we demonstrated the viability of the implicit bidding approach by presenting computational results for CTPAs with up to 5000 bid lanes. Furthermore, we showed that these models can be extended to account for some practical considerations and still maintain tractability.

Now that we have a tractable way to solve, in a single round, fully-enumerated CTPAs to optimality (which was not possible in the past) we can also conduct numerical analysis to better understand the performance and characteristics of practical

CTPAs. To our knowledge, in the literature there has not appeared such a study of fully-enumerated CTPA outcomes.

In particular, we consider the following questions:

- How does the number of lanes affect solution characteristics?
- How do differences in network structure affect solution characteristics?

Effects of Repositioning Lane Quantity

We first consider how varying carriers' repositioning capacities impacts solution time, the number of assigned carriers, and empty movement ratio, defined as the ratio of empty movement distance to total (direct plus empty) movement distance. In this numerical experiment, we again consider auctions of 1000, 2000, 3000, 4000 and 5000 bid lanes. With the exception of the number of repositioning lanes, parameter settings are identical to those described in §2.4.1. For each of these auction sizes, we hold the number of bid lanes constant and vary the number of repositioning lanes (of each carrier) as a percentage of the number of bid lanes, ranging from zero to 100 percent.

Results

In Figures 2.4, 2.5 and 2.6, we present results for auctions with repositioning capacities for each carrier varying from zero to 50 percent of the number of bid lanes. Computational results showed that these trends continue to hold beyond these ranges up to 100 percent of the number of bid lanes.

Our computational results show that CTPAs have special properties at two extremes: when carriers have no repositioning capacities or very large repositioning capacities. At these extremes, the I-WDP is extremely tractable as evident by the small solution times in Figure 2.4. Furthermore, at these extremes the number of

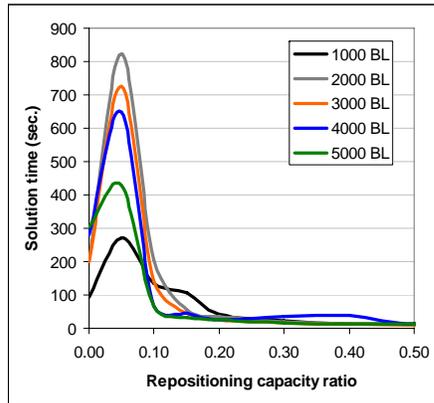


Figure 2.4: Effect of lane quantities on solution time.

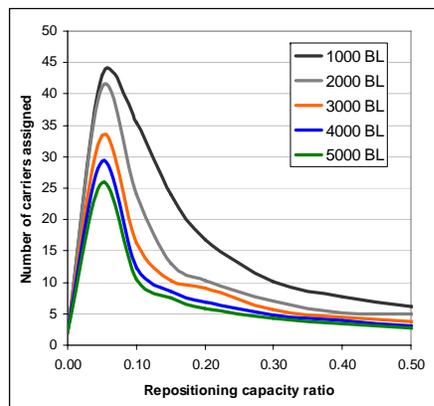


Figure 2.5: Effect of lane quantities on the number of carriers assigned.

assigned carriers is relatively small (Figure 2.5). Intuitively, when carriers have no repositioning capacity, the cost of the auction is dominated by the carriers' direct and empty movement costs, therefore, a small number of low-cost carriers typically win the majority of the bid lanes. As carriers' repositioning capacities increase, the likelihood of finding cost-effective connections also increases, leading to decreases in empty movements (Figure 2.6). As this happens, the majority of continuous movements are formed by combining bid lanes with carriers' pre-existing contracted lanes.

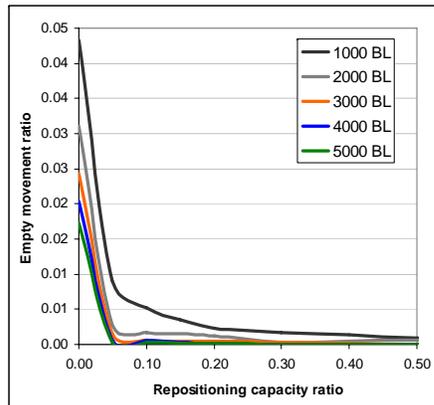


Figure 2.6: Effect of lane quantities on empty movement ratio.

In this case, the cost of the auctions is dominated by direct movement cost and a small number of carriers with the lowest direct movement costs again typically win out.

Effects of Network Structure

Next, we evaluate how solution characteristics change as the structure of the network varies. Specifically, what is the impact on solution time, number of assigned carriers, and empty movement ratio? We again consider a network with 100 nodes, which we now divide into six regions. Each bidder is defined to be either a national or regional carrier. National carriers have repositioning capacities that are uniformly dispersed throughout the entire network, while regional carriers have repositioning capacities that are concentrated in one specific region. Bid lanes are generated in either a uniform network, where bid lanes have randomly selected origin and destination cities, or in a hub-and-spoke network, where bid lanes originate from one of three hubs (selected a priori) and terminate at a random node in an adjacent region. Examples of a hub-and-spoke network and a uniform network are provided in Figure 2.7.

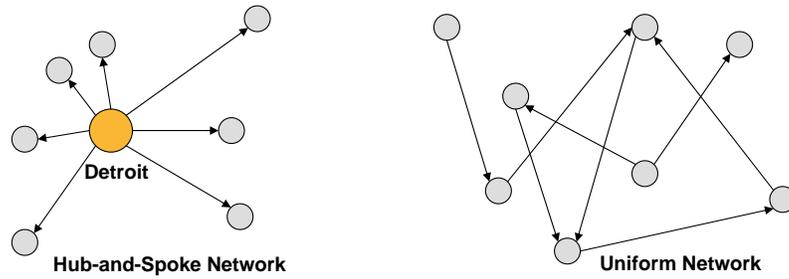


Figure 2.7: The network on the left is a hub-and-spoke network with a single hub located in Detroit. All bid lanes in this network originate from the hub. The network on the right is a uniform network with bid lanes uniformly dispersed throughout the entire network.

The computational results presented below are based on auctions with a total load volume of 100,000 and 50 carriers. Carriers are assigned 100 repositioning (pre-existing contracted) lanes, each with volume 100. We compare computational results for the following four network structures:

1. Network One consists of national carriers bidding on 1000 bid lanes (with average lane volume of 100) uniformly generated on the network. For each carrier, we uniformly generate 100 repositioning lanes on the network.
2. Network Two consists of regional carriers bidding on 1000 bid lanes uniformly generated on the network. For each carrier, we randomly generate repositioning lanes within the carrier's pre-assigned region.
3. Network Three consists of national carriers bidding on 50 bid lanes (with average lane volume of 2000) generated on a hub-and-spoke network. For each carrier, we uniformly generate 100 repositioning lanes on the network.
4. Network Four consists of regional carriers bidding on 50 bid lanes (with average lane volume of 2000) generated in a hub-and-spoke network. For each carrier, we randomly generate repositioning lanes within the carrier's pre-assigned region.

Results

Figure 2.8-a shows that in a very unstructured network (Network One), with national carriers and uniform bid lanes, the average solution time is 202 seconds. In contrast, with a very structured network (Network Four) consisting of only regional carriers and hub-and-spoke bid lanes the average solution time is only 13 seconds. In Network One, there is significantly more fractionality; carrier characteristics, in terms of costs and repositioning capacity, are very homogenous and bid lanes are uniformly generated throughout the network. In Network Four, there is less fractionality as carrier characteristics are more heterogeneous; each regional carrier has repositioning capacity that is concentrated in a specific region of the network. Observe that the computational results presented earlier in §2.4.1 and §2.4.2 are based on the least tractable setup, with national carriers and uniform bid lanes. As such, we can expect computational performance of our approach to be even better in real-world networks with some structure.

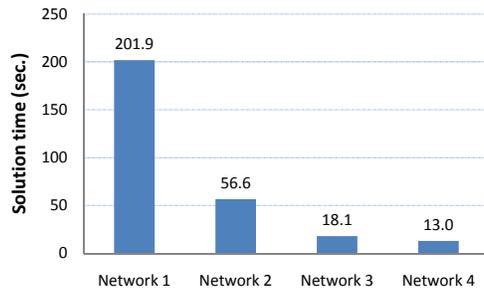


Figure 2.8-a

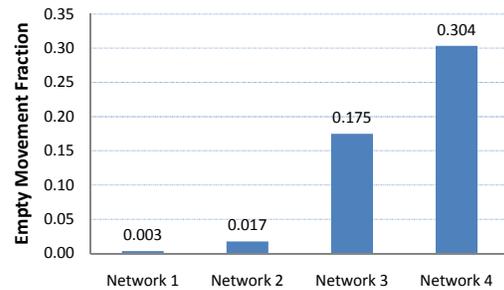


Figure 2.8-b

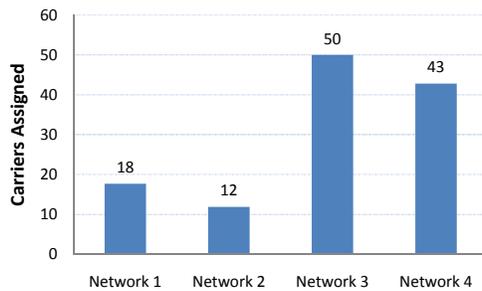


Figure 2.8-c

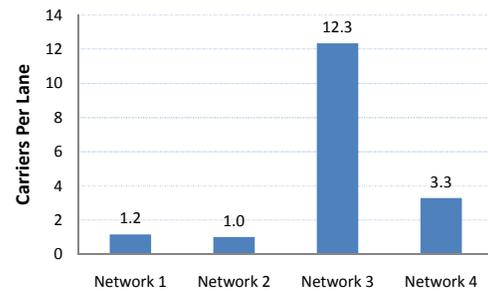


Figure 2.8-d

Figure 2.8: Impact of different network structures on average solution time, empty movement ratio, number of assigned carriers and assigned carriers per lane.

Figure 2.8-b shows that CTPAs on networks with hub-and-spoke bid lanes (Networks Three and Four) result in higher empty movements. This is as expected, since carriers must return empty to hubs (bid lane origin) more often to pick up a bid load. With respect to carrier types, national carriers are better able to exploit complementarities between their repositioning capacities and bid lanes and hence, service bid lanes more efficiently (0.175 empty movement fraction for national carriers compared to 0.304 empty movement fraction for regional carriers).

With respect to the number of carriers assigned (Figure 2.8-c) and number of carriers assigned per lane (Figure 2.8-d), less structure implies fewer carriers assigned and more structure implies more carriers assigned. On less structured networks, a few of the lower-cost carriers typically dominate, while on a more structured network, the unique set of repositioning lanes that each carrier brings to the auction plays a key role in forming efficient movements and so more carriers are likely to be allocated lanes.

2.4.4 Comparison to Bidding Methods in Practice

The research outlined in this chapter is premised on two key ideas. First, the solution quality in a CTPA can improve significantly as the number of combinations bid upon grows. Second, the run time in a CTPA can worsen significantly as the number of combinations bid upon grows. In this section, we focused on reconciling these two conflicting issues. We now focus on a comparison between bidding methods in practice and our implicit bidding approach.

Enumerative Bidding Approaches for Low Cardinality Bundles

We use this section to show the impact on solution quality and run time of combinatorial bidding. A detailed study [43] of carriers' bidding behavior showed that

in practice, only 28 percent of carriers participating in a combinatorial auction submit multi-lane bids. Furthermore, among those carriers submitting multi-lane bids, most submit only two to seven multi-lane bids of low cardinality (with a median and a mode of two lanes per bundle). In the following experiments, we show that the transition from single-lane bidding to bidding on bundles with even just two or three lanes can dramatically increase solution quality, but also lead to prohibitive increases in run time. We contrast this with our approach, which can exhaustively (implicitly) consider all bundles of any size, while maintaining tractability for the same problem instances.

For all of the experiments in this section, we assume a small auction in which 5 carriers bid for lanes (each with volume one) randomly generated in a network of 100 cities. Each carrier has 10 repositioning lanes (each with volume one) that can be used as part of a continuous move for free. Each carrier's price structure is randomly generated as described in §2.4.1.

We begin by considering an instance in which 100 bid lanes are being auctioned off by the shipper. We compute the outcome of this auction for five cases:

1. Case One permits only single-lane bids with empty returns. In this case, all bid lanes will be allocated to the lowest price (with respect to direct and empty movement prices) carrier(s).
2. Case Two permits only single-lane bids (see Figure 2.9). However, we allow these bids to reflect the opportunity each carrier has for efficiencies associated with using repositioning lanes in their network for backhaul. Specifically, for each bid lane $a \in \mathcal{A}^{\mathcal{L}}$, each carrier computes their lowest price for transporting that lane, taking into consideration the option to use repositioning lanes, and submits the corresponding bid for lane a . In addition, to account for the fact

that a carrier cannot use a repositioning lane (with volume one) in more than one bundle, we impose a constraint such that at most one bundle, among those that share a repositioning lane, can be chosen. Finally, we must also include one bid for each lane associated with returning empty, so as to ensure feasibility.

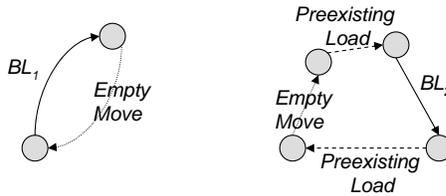


Figure 2.9: Case 2 permits single bid lane bundles. These bundles may leverage repositioning lanes to form efficient continuous moves.

- Case Three permits carriers to combine two bid lanes together whenever they create a efficient continuous move (see Figure 2.10). Specifically, in addition to the bids from Case Two, carriers also bid on pairs of bids lanes, finding the cheapest continuous move that covers both of these lanes (again, using repositioning lanes as well whenever beneficial), and submitting this bid price for the pair of lanes. In addition, to account for the fact that a carrier cannot use a repositioning lane in more than one bundle, we impose a constraint such that at most one bundle, among those that share a repositioning lane, can be chosen.

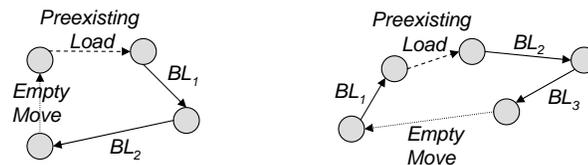


Figure 2.10: Case 3 (left) permits two bid lane bundles. Case 4 (right) permits three bid lane bundles.

- Case Four extends Case Two by also allowing all sets of three bid lanes to be

combined in continuous moves (see Figure 2.10). In addition, to account for the fact that a carrier cannot use a repositioning lane in more than one bundle, we impose a constraint such that at most one bundle, among those that share a repositioning lane, can be chosen.

5. Case Five uses the implicit bidding approach to consider all possible bids of any number of bid lanes.

Case	Number of Bundles	MCF (sec.)	SPP (sec.)	Total (sec.)	Total Cost	Cost Differential
Singles - Empty Returns	100	0	0	0	216,785	51.23%
Singles	600	21	2	23	192,331	34.17%
Singles+Doubles	25,350	976	12	988	149,469	4.27%
Singles+Doubles+Triples	833,850	31,923	681	32,604	145,212	1.30%
I-WDP	-	-	-	4	143,349	-

Table 2.3: Comparison of explicit bidding approaches to the implicit bidding approach.

Table 2.3 shows the results of these auctions. We present the average results for ten randomly generated instances. Column one shows the case, column two shows the number of bundles that was considered, and column three shows the total time spent pricing bundles by solving individual minimum cost flow problems. Column four shows the time spent solving T-WDP to determine the allocation of bundles to carriers. Column five shows the total time to solve the model. Column six shows the total cost. The final column gives the ratio of overall cost to the optimal cost to be found if all bundles, regardless of size, are considered. Of course, this is in fact the solution obtained by solving the I-WDP.

Observe that moving from single to double bids improves the solution quality dramatically (29.90%). The reduction from double to triple (2.97%) is less dramatic but still significant. Although it would appear that moving from triples to all lanes

does not improve solution quality by a large margin, a 1.30% improvement is still quite meaningful in the trucking industry, where profit margins range from 2% to 4% [16]. Perhaps more importantly, these cost improvements associated with moving from single- to double- to triple- lane bidding come at a substantial cost in terms of run-time, from 23 to 988 to 32,604 seconds (in contrast with the 4 second run time for I-WDP using the implicit bidding approach). Furthermore, this example is small in size. As the number of bid lanes increases run-time increases prohibitively, as seen in the next experiment.

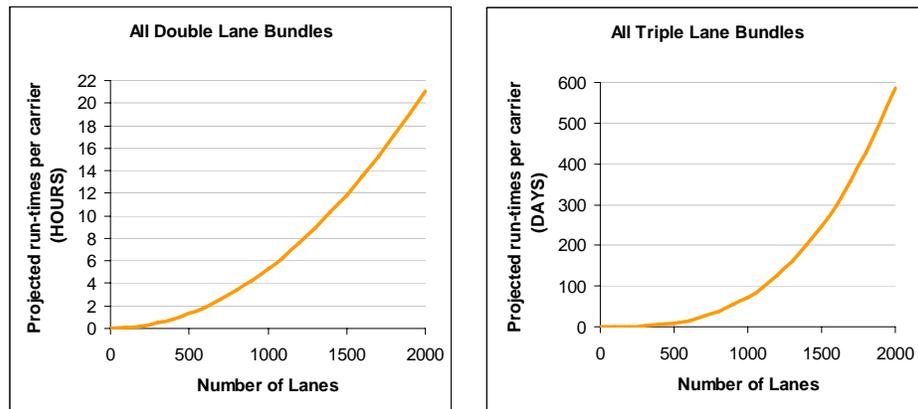


Figure 2.11: Projected run-times for evaluating all double bid lane bundles (left, in hours) and all triple bid lane bundles (right, in days).

Figure 2.11 projects run times for generating all double- and triple- lane bids as a function of the number of bid lanes. These were computed by extrapolating from the results in Table 2.3. On average each minimum cost flow problem takes 0.038 seconds. We multiply 0.038 seconds by the total number of possible double- and triple- lane bids to project run-times. These projections are actually underestimates, for three reasons. First, as the networks increase in size, the individual minimum cost flow problems take longer to solve (i.e. the run time per bid increases). Second,

if we consider lanes with volume greater than one load, then the number of possible double- and triple- lane bids is even larger. Third, as the number of bids grows very large, computational performance will be impacted by computer memory limitations. Finally, we observe that in the examples shown here, the run-time challenge is a function of the bid generation, with the actual WDP solving quite quickly. This will not always remain true – as the number of bids grows the impact on IP performance will begin to show.

Selective Bidding Approaches for Low Cardinality Bundles

The results in §2.4.4 suggest that even bidding on all bundles of size three will not be tractable for most truckload auctions, thereby missing substantial opportunities for cost efficiency. A logical counter-argument is to suggest not bidding on all bundles of multiple lanes but only bidding on “good” bundles. However, this idea presents two major challenges. First, what defines a “good” bundle? Second, how can we find these “good” bundles?

To address these questions, we conducted the following experiments. We identified three metrics that could be used to evaluate bundles. These are:

1. *Absolute empty mileage* - quality of a bundle is measured by the total empty miles traveled. Lower absolute empty mileage is more desirable.
2. *Empty mileage ratio* - quality of a bundle is measured by the ratio of empty miles traveled to total miles traveled. Lower ratio of empty mileage to total mileage is more desirable.
3. *Random* - bundles are randomly selected from the set of all single, double, and triple bid lane bundles.

For each of these metrics, we ran a separate auction. In each case, we included all

single-lane bids (both with and without the use of repositioning lanes), as in Case 2 of §2.4.4. In addition, we enumerated all double- and triple- bid lane bundles and, for each carrier, selected the best bids relative to the metric and included these in the auction. We then solved the auction and reported the final cost. Figure 2.12 shows the outcomes for each of these three metrics, as the number of bids per carrier varies. These are reported as percentages relative to the optimal value, computed by our implicit bidding approach.

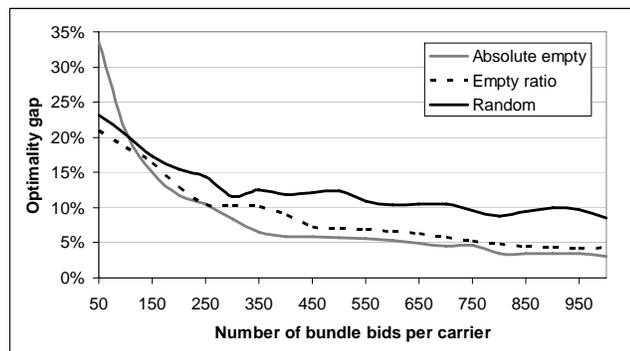


Figure 2.12: Optimality gap between various bundles selection criteria vs. implicit bidding approach.

Observe, first, that the optimal solution relative to bidding the exhaustive set of doubles and triples provides a natural lower bound, which in turn has a non-zero gap relative to the optimal – in other words, the optimal solution contains some bundles with four or more bid lanes. Note also that although metrics 1 and 2 appear to dominate metric 3 for all but small numbers of bids, it is not always obvious which metric would lead to the better solution.

Furthermore, we see substantial improvement as the number of included bids increases. It is interesting to observe that this does not reflect an increase in the number of multi-lane bids being included in the final solution. Given 100 bid lanes, for example, at most 33 bundles of size three could ever be included in the final solution. Rather, what we are observing is the fact that the most desirable bundles

are *not* those that best satisfy the given metrics. The reason for this is that we are not concerned with bundles in isolation, but rather with how they fit together with other bundles to complete the auction. As an extreme case, suppose that two carriers shared a common bid lane in their “best” bundles. Because that lane can only be awarded once, only one of those bids could be chosen. Had one of the carriers bid their “second best” bundle, not containing that lane, both bids might be chosen.

Finally, we note that although the run-time for the T-WDP of these instances is certainly shorter than when including all double- and triple- bid load bundles, because of the decrease in size of the IP, we do not know of any efficient way to find these bundles – in our case, we resorted to enumeration, which is no faster than the run time of Case Three.

Observations

We conclude this section by summarizing our observations. First, including multi-lane bundles, even of small size, can greatly improve solution quality, but at a tremendous impact on run time. Even small auctions become intractable when including just triple-bid lane bundles, never mind bundles of larger cardinality. Second, we cannot overcome this by including only those bundles which are “good bundles” – both because of the computational challenges associated with identifying these bundles and also because the best bundles, in isolation, might not form the best combination of bids. Finally, we observe that our computational results here are only from auctions of a limited size. Although the implicit bidding approach enables us to solve much larger auctions to optimality, we cannot compare the results to the traditional approach, as the traditional approach cannot be solved to optimality (even across all double- and triple- bid lane bundles), except for very small instances. However, as the number of lanes in the auction grows, there may be substantial new

opportunities for combining larger sets of lanes (i.e. bundles of size four or more) to incur improved savings, suggesting even more improvement in solution quality as the number of lanes grows.

2.5 Conclusions and Future Research

In this chapter, we introduced an implicit bidding approach to solve CTPAs in a single round while implicitly considering the exhaustive set of all possible bundles. This approach directly addresses the two main challenges of combinatorial auctions: bidding on an exponentially-large set of bundles and solving the corresponding exponentially-large WDP. Using the implicit bidding approach, instead of submitting an exponential number of bundles, each carrier simply submits a BGF, which is embedded directly into the WDP. We showed that in truckload transportation, a carrier’s BGF is a circulation problem and the resulting I-WDP is a multi-commodity flow problem, which is generally known to be tractable in practice. Tractability was demonstrated through extensive computational experiments for auctions with up to 5000 bid lanes and over 600,000 loads. Furthermore, I-WDP can be extended to include additional operational considerations while preserving tractability. In short, we presented a new approach and models for solving CTPAs to optimality that are computationally efficient, consider the exhaustive set bundles, and achieve full economies of scope, which is not possible with current approaches.

We also took advantage of this new capability to solve fully-enumerated CTPAs to optimality as a tool for conducting numerical analysis on the quality and characteristics of solutions. We showed that, using this approach, shippers can conduct numerical experiments to assess how CTPA characteristics (e.g., solution time, number of carriers assigned, empty movement ratio, et cetera) are likely to change with

important problem parameters (e.g., number of carriers, number of lanes, carriers' repositioning capacities, et cetera). Additionally, the shipper can use our approach to perform what-if analysis to assess the cost impact of imposing various operational constraints before finalizing contracting decisions. Lastly, we compared the implicit bidding approach to bidding methods in practice and showed its benefits both in terms of solution quality and runtime.

In terms of future work, we envision two types of research. First, extensions are possible for our work in CTPAs. For instance, additional operational considerations could be addressed, such as regional coverage requirements, backup carrier bids, and maximum tour length constraints. Maximum tour length constraints are applicable because drivers and equipment must be returned home within a limited time window. This constraint set can only be addressed with explicit knowledge of bid lane allocations and the tours constructed to cover these allocated lanes. We are currently addressing this problem using column generation to solve a tour based model, where each variable represents a viable tour or set of tours.

Furthermore, now that we can solve CTPAs in a reasonable amount of time, we can use this tool to assess the quality of various auction mechanisms for truckload procurement. Specifically, how would different auction mechanisms (first price, second price, et cetera) perform under various procurement settings? Even with just a single item, revenue-or cost-equivalence between standard auction formats fails if bidders are asymmetric.

Additionally, uncertainties in the cost parameters exist due to spot market variability, carriers' uncertainties about their existing and future networks, and timing effects; detailed modeling of such uncertainties and development of appropriate solution approaches are interesting, but challenging, directions for future research.

Secondly, future work could extend the use of the implicit bidding approach to other application domains. Of particular interest is the identification of domains for which the bid-generating approach appears amenable. A sample of potential domains include wireless spectrum auctions, energy auctions, and procurement auctions with capacity-constrained suppliers.

BIBLIOGRAPHY

- [1] Abrache, J., B. Bourbeau, T. G. Crainic, M. Gendreau. 2004. A new bidding framework for combinatorial e-auctions. *Computers and Operations Research* **31**(8) 1177-1204.
- [2] Ahuja, R., T. Magnanti, J. Orlin. 1993. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, NJ.
- [3] American Trucking Association. 2007. US Freight Transportation Forecast to 2017, Arlington, VA
- [4] Ausubel, L., P. Milgrom. 2006. The lovely and lonely Vickrey auction. P. Cramton, R. Steinberg, Y. Shoham, eds., *Combinatorial Auctions*. MIT Press.
- [5] Aviv, Y. 2004. Collaborative forecasting and its impact on supply chain performance. D. Simchi-Levi, S.D. Wu, Z.-J. Shen, eds., *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*. Kluwer Academic Publishers.
- [6] Balas, E., M. Padberg. 1976. Set Partitioning: A Survey. *SIAM Review* **18**(4) 710-760.
- [7] Ballou, R. H. 1992. *Business Logistics Management*. Prentice Hall, Inc. Englewood Cliffs, NJ.
- [8] Banks, J., J. Ledyard, D. Porter. 1989. Allocating uncertain and unresponsive resources: an experimental approach. *RAND Journal of Economics*. **20**(1), Spring 1989.
- [9] Bykowsky, M., Cull, R., Ledyard, J. 2000. Mutually destructive bidding: the FCC auction design problem. *Journal of Regulatory Economics*, Springer, **17**(3), 205-28, May.
- [10] Caplice, C. 1996. Optimization-based bidding, a new framework for shipper-carrier relationships. Ph.D. thesis, Massachusetts Institute of Technology.
- [11] Caplice, C., Y. Sheffi. 2003. Optimization based-procurement for transportation services. *Journal of Business Logistics*. **24**(2) 109-128.
- [12] Caplice, C., Y. Sheffi. 2006. Combinatorial auctions for truckload transportation. P. Cramton, R. Steinberg, Y. Shoham, eds., *Combinatorial Auctions*. MIT Press.
- [13] Chang, T. 2008. Decision support for truckload carriers in one-shot combinatorial auctions. *Transportation Research Part B* - article in press
- [14] Chopra, S., P. Meindl. 2007. *Supply Chain Management: Strategy, Planning and Operation*. Pearson Prentice Hall.
- [15] Cohn, A. 2009. *Parameter Files*. <http://www-personal.umich.edu/~amycohn/papers.html>.
- [16] Costello, B. 2003. *Motor carrier update – technical report*, American Trucking Associations. Summer Transportation Fuels Outlook Conference.
- [17] Cramton, P., R. Steinberg, Y. Shoham. 2006. *Combinatorial Auctions*. MIT Press.

- [18] de Vries, S., R. Vohra. 2003. Combinatorial auctions: a survey. *INFORMS Journal on Computing*. **15**(3) 284-309.
- [19] Elmaghraby, W., P. Keskinocak. 2002. Combinatorial auctions in procurement. *Technical Report, The Logistics Institute, Georgia Institute of Technology, Atlanta, GA*.
- [20] Ergun, O., G. Kuyzu, M. Savelsbergh. 2007a. Reducing truckload transportation costs through collaboration. *Transportation Science* **41**(2) 206-221.
- [21] Ergun, O., G. Kuyzu, M. Savelsbergh. 2007b. Shipper collaboration. *Computers and Operations Research* **34** 1551-1560.
- [22] Even, S., A. Itai, A. Shamir. 1976. On the complexity of timetable and multicommodity flow problems. *SIAM Journal on Computing* **5**(4) 691-703.
- [23] Figliozzi, M., H. Mahmassani, P. Jaillet. 2002. A framework for the study of carrier strategies in an auction based transportation marketplace. *TRB Annual Meeting*.
- [24] Figliozzi, M., H. Mahmassani, P. Jaillet. 2006. Analysis and evaluation of incentive-compatible dynamic mechanisms for carrier collaboration. *Transportation Research Record* **1996** 34-40.
- [25] Figliozzi, M., H. Mahmassani, P. Jaillet. 2007. Pricing in dynamic vehicle routing problems. *Transportation Science* **41**(3) 302-318.
- [26] Franklin, M., M. Reiter. 1996. The design and implementation of a secure auction service. *IEEE Trans. Soft. Eng.* **22**(5) 302-312.
- [27] Fujishima, Y., K. Leyton-Brown, Y. Shoham. 1999. Taming the computational complexity of combinatorial auctions: optimal and approximate approaches. *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*.
- [28] Foster J., Strasser S. 1991. Carrier/modal selection factors: the shipper/carrier paradox. *Transportation Research Forum*. **31** 206-212.
- [29] Guo, Y., A. Lim, B. Rodrigues, Y. Zhu. 2006. Carrier assignment models in transportation procurement. *Journal of Operational Research Society*. **57**(12) 1504-1510.
- [30] Günlük, O., L. Ladányi, S. de Vries. 2005. A branch-and-price algorithm and new test problems for spectrum auctions. *Management Science*. **51**(3) 391-406.
- [31] Houghtalen, L., O. Ergun and J. Sokol, Designing mechanisms for the management of carrier alliances. Working Paper, Georgia Institute of Technology, 2008.
- [32] Kelly, F., R. Steinberg. 2000. A combinatorial auction with multiple winners for universal service. *Management Science*. **46**(4) 586-596.
- [33] Kudo, M. 1998. Secure electronic sealed-bid auction protocol with public key cryptography. *IEICE Trans. Fund. Electron. Commun. Comput. Sci.* **E81-A**(1) 20-27.
- [34] Krajewska, M., H. Kopfer. 2006. Collaborating Freight Forwarding Enterprises. *OR Spectrum* **28** (301-317).
- [35] Ledyard, J., M. Olson, D. Porter, J Swanson, D. Torma. 2002. The first use of a combined-value auction for transportation services. *Interfaces* **32**(5) 4-12.
- [36] Lee, C.-G., R.H. Kwon, Z. Ma. 2007. A carrier's optimal bid generation problem in combinatorial auctions for transportation procurement. *Transportation Research Part E* **43** 173-191.
- [37] Lee, H., V. Padmanabhan, S. Whang. The bullwhip effect in supply chains. *MIT Sloan Management Review* **38**(3) 93-102.

- [38] Lucking-Reiley, D. 2000. Auctions on the internet: what's being auctioned, and how? *Journal of Industrial Economics* **48**(3) 227-252.
- [39] Nemhauser, G., L. Wolsey. 1999. *Integer and Combinatorial Optimization*. John Wiley & Sons, Inc., New York, NY.
- [40] Nisan, N. 2000. Bidding and allocation in combinatorial auctions. *Proceedings of the 2nd ACM Conference on Electronic Commerce, 2001* 1-12.
- [41] Nisan, N., I. Segal, 2006. The communication requirements of efficient allocations and supporting prices. *Journal of Economic Theory* **129**(2006) 192-224.
- [42] Pekec, A., M. Rothkopf. 2003. Combinatorial auction design. *Management Science* **49**(11) 1485-1503.
- [43] Plummer, C. 2002. Bidder response to combinatorial auctions in truckload procurement. Master thesis, Massachusetts Institute of Technology.
- [44] Raychaudhuri S., D. Veeramani. 2006. Bidding strategies in multi-round combinatorial auctions for transportation services. *Technical Report. Department of Industrial Engineering, University of Wisconsin, Madison, MD*.
- [45] Ronen, A. 2006. Incentive compatibility in computationally feasible combinatorial auctions. P. Cramton, R. Steinberg, Y. Shoham, eds., *Combinatorial Auctions*. MIT Press.
- [46] Rothkopf M., A. Pekec, R. Harstad. 1998. Computationally manageable combinatorial auctions. *Management Science* **44**(8) 1131-1147.
- [47] Sandholm, T. 2002. Algorithm for optimal winner determination in combinatorial auctions. *Artificial Intelligence* **153** 1-54.
- [48] Sandholm, T., S. Suri, D. Gilpin, D. Levine. 2005. CABOB: A fast optimal algorithm for winner determination in combinatorial auctions. *Management Science* **51**(3) 374-390.
- [49] Sheffi, Y. 2004. Combinatorial auctions in the procurement of transportation services. *Interfaces* **34**(4) 245-252.
- [50] Song, J., A. Regan. 2003. An auction based collaborative carrier network. *Transportation Research Record* **43** 173-191.
- [51] Song, J., A. Regan. 2004. Combinatorial auctions for transportation service procurement: the carrier perspective. *Transportation Research Record* **1833** 40-46
- [52] Song, J., A. Regan. 2005. Approximation algorithms for the bid construction problem in combinatorial auctions for the procurement of freight transportation constructs. *Transportation Research Part B* **39** 912-933
- [53] Standard & Poor: 2001. Industry survey in transportation: commercial **3**(M-Z).
- [54] Wang, X., M. Xia. 2005. Combinatorial bid generation problem for transportation service procurement. *Transportation Research Record* **1923** 189-198.
- [55] Wolsey, L. 1998. *Integer Programming*. John Wiley & Sons, New York, NY.
- [56] Zurel E., Nisan N. 2001. An efficient approximate allocation algorithm for combinatorial auctions. *Proceedings of the 3rd ACM Conference on Electronic Commerce, 2001*.

CHAPTER III

A Stochastic Multicommodity Flow Approach to Combinatorial Truckload Auctions

3.1 Introduction

In this chapter, we consider a *stochastic combinatorial truckload procurement auction* (S-CTPA) where carriers have uncertain backhaul capacity and costs. Carriers participating in truckload auctions typically assume estimated repositioning capacities and costs, because the actual capacities and costs in the future are not known with certainty. However, allocating bid lanes to carriers on the basis of such estimates may not be optimal in expectation. For example, a carrier may be awarded a set of bid lanes assuming the availability of a complementary set of repositioning opportunities. However, if these repositioning opportunities are not fully available or are available at higher costs than anticipated, the carrier may not receive sufficient revenue to make a profit, or even to cover the operating cost required to serve these bid lanes. On the other hand, if more repositioning opportunities exist or are available at lower than estimated cost, the carrier may be awarded a payment above what is efficient, at the expense of increased cost to the shipper.

Given the challenges of exponential bidding and carriers' uncertain repositioning capacities and costs, we propose a method to simultaneously resolve these two problems. We introduce two bid generating functions, one in which carriers use expected

values for uncertain repositioning capacities and costs and one in which carriers consider all repositioning capacity and cost scenarios. In the former, the resulting BGF in a minimum cost flow problem structurally identical to the deterministic BGF introduced in Chapter II. In the later, the resulting BGF, which we call a BGF-S, is a large-scale minimum cost flow problem in which constraints have a block structure, one block for each scenario. We use the implicit bidding approach to circumvent the computational challenges of traditional CTPAs by embedding carriers' BGF-S directly into the WDP. This results in a tractable implicit WDP that is a two-stage stochastic integer program.

The contributions of this research are in:

1. presenting models and decomposition algorithms for *fully-enumerated*, stochastic combinatorial truckload procurement auctions (S-CTPAs), where carriers have uncertain repositioning capacities and costs;
2. proposing procedures to accelerate the decomposition algorithm for solving S-CTPAs and demonstrating their efficacy for solving practically-sized instances;
3. taking advantage of this new capability to solve fully-enumerated S-CTPAs to demonstrate the value of the stochastic solution over the deterministic solution, obtained by solving the expected value problem (a related deterministic model that uses the expected values of uncertain parameters);
4. generalizing the stochastic model and algorithmic approach presented for S-CTPAs to a more general stochastic network flow problem, which we call the two-stage stochastic multicommodity flow problem (TS-MFP) and demonstrating its applicability to other problems.

The outline for the rest of the chapter is as follows. In §3.2, we formally present

stochastic combinatorial auctions for truckload procurement. In §3.3, we briefly review IBA and develop two tractable winner determination problems for fully-enumerated S-CTPAs. §3.4 presents a cutting plane algorithm, based on the L-shaped method, for solving S-CTPAs. §3.5 introduces various modeling and algorithmic procedures to improve the efficacy of the proposed solution algorithm. Computational experiments and results are presented in §3.6. §3.7 presents a generalization of the models and algorithms for S-CTPA to TS-MFP and other applications of TS-MFP. We conclude in §3.8 with a summary of our contributions and our suggestions for future research.

3.2 CTPA

The basic truckload procurement auction is conducted by an auctioneer (on behalf of the shipper), with a set \mathcal{K} of prospective bidders (carriers). First, the auctioneer announces a set of *bid lanes* \mathcal{A}^ℓ , with $L = |\mathcal{A}^\ell|$, being auctioned, where each bid lane $a \in \mathcal{A}^\ell$ is defined by an origin, a destination, and a volume d_a , corresponding to the number of loads in that bid lane. A *bundle* or combination is a subset of \mathcal{A}^ℓ , and different carriers are interested in selling different bundles to the shipper at different prices. Each carrier k determines a set of bundles of interest (\mathcal{S}^k) and submits a bid ($b_s^k \geq 0, s \in \mathcal{S}^k$) to the auctioneer for each of these bundles. Finally, the auctioneer, on behalf of the shipper, solves a *winner-determination problem* (WDP), to partition the bid lanes in \mathcal{A}^ℓ and award bundles to carriers in such a way that the total cost is minimized.

3.2.1 Computing Bundle Bids

To understand how carriers determine the bid-price for a bundle b , we must first understand how carriers determine their cost. Given a bundle of bid lanes, carriers

must determine the least-cost way to serve these bid lanes and use this cost to derive a bid-price. A carrier’s cost can be decomposed into two types, *direct movement-* and *repositioning movement-* cost. Direct movement cost is the cost associated with moving a bid load from load origin to load destination. This cost is largely known and is primarily a function of distance.

Repositioning movement cost is the cost associated with moving a truck from the destination of the current bid load to the origin of the subsequent bid load. This cost will vary depending on the chosen routing and the carrier’s *backhaul capacity*. In order to maximize the probability of winning business, a carrier must find efficient *tours* (continuous movements) that minimizes the total repositioning movement cost incurred to cover the bid lanes in bundle b . This can be accomplished by combining bid lanes in bundle b with the carrier’s backhaul capacity, which represents opportunities from existing contracted loads, future contracts, *spot market* opportunities, and empty movements. The efficiencies of these repositioning movements are the key to determining the actual bid price for any bundle b .

In practical CTPAs, carriers determine the types of repositioning movements available to them for *each* directed city pair (i, j) and derive an estimated capacity and cost for each of these types. We represent a carrier’s repositioning opportunities for *each directed city pair* (i, j) in the network using a piecewise-linear function. An example of this function is shown in Figure 3.1.

In this example, each of the five line segments represents a different repositioning opportunity type from city i to city j . The first segment (leftmost) might represent the number of movements from preexisting contracted lanes and, because these movements represent hired rather than empty movements, they can be used “free.” The second, third, and fourth segments, each with different prices and capacities,

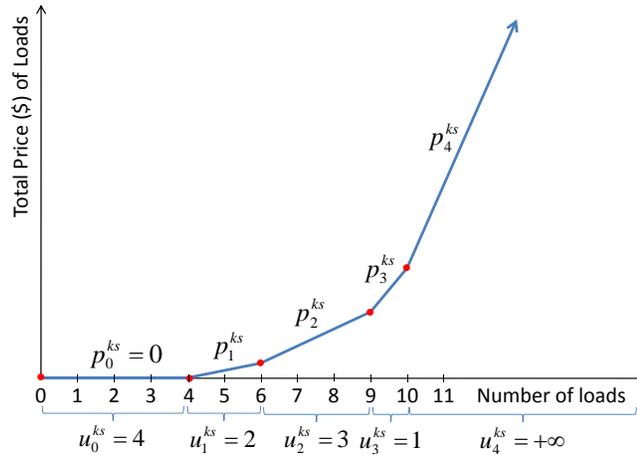


Figure 3.1: A carrier's estimated repositioning capacities and prices from city i to city j

may represent spot market opportunities, estimated opportunities from future contracts, and partial connections (preexisting lanes that require the driver to move empty from city i to a location nearby before picking up a load and delivering that load to a location near city j , thereby incurring a limited amount of empty mileage costs). The fifth (last) segment, with the *highest price* and *infinite capacity*, represents empty movements from city i to city j . In the context of a network, each segment of a piecewise linear function would be represented by an arc with associated arc cost and capacity corresponding to the segment's slope and length.

Despite the fact that most, if not all, CTPAs assume these carriers have deterministic repositioning capacities and costs, in practice carriers do not have full information about their future repositioning capacities and cost. Under the latter condition, a model that considers carriers' uncertain repositioning costs and capacities may perform better than one that does not. We refer to such a CTPA, with uncertain carrier capacities and costs, as a *stochastic combinatorial truckload procurement auction* (S-CTPA).

In a S-CTPA, we assume that the set of bid lanes \mathcal{A}^ℓ to be auctioned is known with certainty. That is, the shipper determines the lanes *a priori* and corresponding volume for which to procure services. We assume that carrier k 's repositioning prices (cost plus profit markup) and capacities for each arc in \mathcal{A}^k are discrete random parameters with a known joint distribution and finite support. Let ξ^k represent a discretely distributed random price-capacity data vector with finite support Ξ^k . In this problem a scenario s represents a realization of arc capacity and price for each arc $a \in \mathcal{A}^k$. Let $s = 1, \dots, |\Xi^k|$ be the index of scenarios and ρ^s represent the realization probability of scenario s .

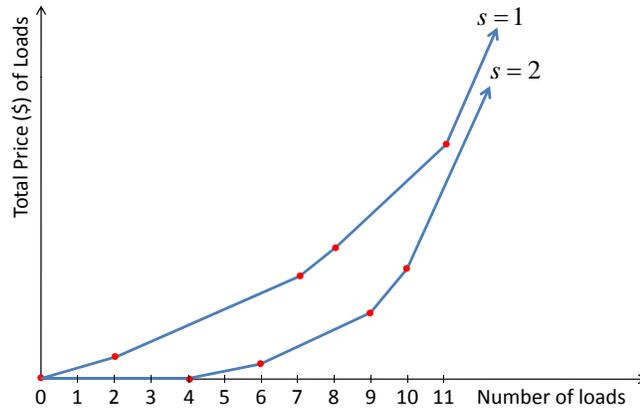


Figure 3.2: A carrier's repositioning capacity and price from node i to node j in scenario $s = 1$ and $s = 2$.

Then the repositioning capacities and costs for each directed node pair (i, j) can be described by a *set* of piecewise linear functions, one function for each scenario $s = 1, \dots, |\Xi^k|$. Figure 3.2 depicts a carrier's repositioning opportunities from city i to city j under scenario 1 and scenario 2. Each segment of a piece-wise linear function in 3.2 represents capacity and price realization for a specific repositioning opportunity type under a particular scenario realization.

We reiterate that, given a bundle b , direct movement cost, and repositioning

movement opportunities (cost and capacity), carriers determine the least cost set of tours to serve the bid lanes in b and then use this cost to determine the bid price, typically, cost plus a percentage markup. In the following sections, we present two bid-generating functions for S-CTPAs. In the first, we adopt a common approach for dealing with stochastic models by assuming *expected* repositioning costs and capacities. This problem is referred as the *expected value problem* (EVP). In the second, we consider the full distribution of repositioning capacities and costs and evaluate the bid price under each scenario, $s = 1, \dots, |\Xi^k|$, to derive an *expected bid price* over all the scenarios.

Pricing a Bundle of Bid Lanes with No Recourse

In the context of pricing bundle b , the vector $\mathbf{x}_b^k = (x_{b1}^k, \dots, x_{ba}^k, \dots, x_{bL}^k)$ is a parameter that describes the composition of bundle b and not a vector of variables. A component of \mathbf{x}_b^k , x_{ba}^k , refers to the number of loads in bid lane $a \in \mathcal{A}^\ell$ bid upon. $\bar{p}_a^k = \left\lceil \sum_{s=1}^{|\Xi^k|} \rho^{ks} p_a^{ks} \right\rceil$ is the ceiling of the *expected price* of a unit movement on arc a and $\bar{u}_a^k = \left\lfloor \sum_{s=1}^{|\Xi^k|} \rho^{ks} u_a^{ks} \right\rfloor$ is the floor of the *expected capacity* of arc a . In defining these parameters, we make mild assumptions that carriers estimate repositioning capacities and prices in integral values. Specifically, carriers estimate repositioning movement prices in whole dollars and repositioning movement capacities in integral number of movements. These rounding rules provide a lower bound estimate on a carrier's expected repositioning movement capacities and an upper bound estimate on expected repositioning movement costs. In making these estimates, carriers can also employ other types of rounding rules. Finally, variable y_a^k ($\forall a \in \mathcal{A}^k$) is an integer variable indicating the number of repositioning movements taken over arc $a \in \mathcal{A}^k$.

Given the aforementioned parameters and variables, the carrier's bundle pricing

(MCF) problem and solvable in polynomial time ([1]).

Pricing a Bundle of Bid Lanes with Recourse

Although solving the expected value pricing problem is commonly used in practice because it is easy to compute, doing so can have unfavorable results, as we will show in §3.6.2. In this section, we consider the BGF with recourse; that is, we consider the problem of finding the expected price to serve bundle b over all scenarios $s = 1, \dots, |\Xi^k|$ by determining the optimal set of tours to serve bundle b under each scenario separately and then weighting the total price of each scenario to obtain the expected bid-price for the bundle.

$$(3.2a) \quad (\text{BGF-S}) \quad f^k(\mathbf{x}_b^k) = \sum_{a \in \mathcal{A}^\ell} p_a^k x_{ba}^k + \min \sum_{s=1}^{|\Xi^k|} \rho^{ks} \left[\sum_{a \in \mathcal{A}^k} p_a^{ks} y_a^{ks} \right]$$

$$(3.2b) \quad \text{s.t.} \quad \sum_{a \in \mathcal{A}^\ell: \mathcal{O}(a)=i} x_{ba}^k + \sum_{a \in \mathcal{A}^k: \mathcal{O}(a)=i} y_a^{ks} = \sum_{a \in \mathcal{A}^\ell: \mathcal{D}(a)=i} x_{ba}^k + \sum_{a \in \mathcal{A}^k: \mathcal{D}(a)=i} y_a^{ks} \\ \forall i \in \mathcal{N}, s = 1, \dots, |\Xi^k|,$$

$$(3.2c) \quad y_a^{ks} \leq u_a^{ks} \quad \forall a \in \mathcal{A}^k, s = 1, \dots, |\Xi^k|,$$

$$(3.2d) \quad y_a^{ks} \geq 0 \quad \forall a \in \mathcal{A}^k, s = 1, \dots, |\Xi^k|.$$

Carrier k 's objective is to determine the expected price to cover each bid lane in bundle b . This can be accomplished by determining the minimum cost set of tours to serve each bid lane in b weighted over all scenarios, $s = 1, \dots, |\Xi^k|$. The objective function (3.2a) states that carrier k 's price for bundle b is the sum of direct movement prices (which depend only on bundle b and is known in advance) and expected repositioning movement prices weighted over all scenarios, which depend on the chosen routing. Constraints (3.2b) are flow balance constraints, one for each scenario-node pair. Constraints (3.2c) are repositioning capacity constraints, one for each arc in \mathcal{A}^k and scenario $s = 1, \dots, |\Xi^k|$.

Since scenarios are independent and there are no arc capacity constraints across scenarios, we can decompose (3.2) by scenarios. Consequent to that decomposition, the bid-price of bundle b can be obtained by solving $|\Xi^k|$ independent MCF problems, one for each scenario $s = 1, \dots, |\Xi^k|$, rather than a single extremely large MCF problem (3.2), which may not be directly solvable.

Although MCF problems are computationally easy to solve, submitting a single bundle bid now requires a carrier to solve $|\Xi^k|$ MCF problems, one for each scenario $s = 1, \dots, |\Xi^k|$. Given that carriers' repositioning capacities and costs are highly stochastic, the number of scenarios $|\Xi^k|$ may be large, thereby, making explicit pricing of individual bundles extremely challenging computationally.

3.2.2 Traditional WDP

Once carriers price each bundle of interest and submit bundle bids, the auctioneer then solves the following mathematical program, denoted (T-WDP), where the T stands for "traditional." The parameter δ_{ba}^k represents the number of lane a movements in bundle b of carrier k , and the binary variable x_b^k takes the value 1 if carrier k is awarded bundle b , and 0 otherwise.

$$\begin{aligned}
 (3.3a) \quad & \text{(T-WDP)} \quad \min \sum_{k \in \mathcal{K}} \sum_{b \in \mathcal{B}^k} p_b^k x_b^k \\
 (3.3b) \quad & \text{s.t.} \quad \sum_{k \in \mathcal{K}} \sum_{b \in \mathcal{B}^k} \delta_{ba}^k x_b^k = d_a \quad \forall a \in \mathcal{A}^\ell, \\
 (3.3c) \quad & \sum_{b \in \mathcal{B}^k} x_b^k \leq 1 \quad \forall k \in \mathcal{K}, \\
 (3.3d) \quad & x_b^k \in \{0, 1\} \quad \forall k \in \mathcal{K}, b \in \mathcal{B}^k.
 \end{aligned}$$

The objective (3.3a) is to minimize the total cost the shipper pays for procuring truckload services for all bid lanes in \mathcal{A}^ℓ . Constraints (3.3b) specify that each bid lane

$a \in \mathcal{A}^\ell$ must be fully allocated among the awarded bundles. In the case of a *fully-enumerated* S-CTPA, where each carrier bids on every distinct bundle, additional constraints (3.3c), restricting each carrier to at most one winning bundle, are required to ensure that a carrier’s repositioning capacity is not violated.

In the context of a traditional S-CTPA, each carrier k must solve bid-generating function (3.1) or (3.2) for each bundle $b \in \mathcal{S}^k$. Although these pricing problems may be computationally tractable, the traditional approach to S-CTPAs suffers as a result of the combinatorial explosion in the number of bundles, presenting a significant hurdle to its practical implementation. For each bid lane $a \in \mathcal{A}^\ell$, a carrier may bid a volume of zero up to d_a ; as such the total number of distinct combinations (and hence, number of distinct bundles) is:

$$(3.4) \quad M = \prod_{a \in \mathcal{A}^b} (d_a + 1).$$

For practically-sized CTPAs, with thousands of bid lanes, the number of combinations M can be enormous. Additionally, if carriers bid using BGF-S then pricing *each* bundle require solving $|\Xi^k|$ individual MCF problems, one for each scenario $s = 1, \dots, |\Xi^k|$. Therefore, explicitly pricing each bundle is clearly intractable even for the smallest size auctions in practice. In practice, carriers typically submit bids for only a very small subset of the distinct bundles [19]. As a result, the full benefits of S-CTPAs in practice cannot be achieved via the traditional bidding approach.

Researchers have suggested several approaches to overcome this hurdle, as discussed in Chapter II. These suggested approaches, however, either sacrifice capturing the exhaustive set of bundles (losing optimality) and/or require several auction rounds (resulting in an increase in procurement cycle times). In the next section, we describe IBA as it is applied to S-CTPAs and highlight its many benefits.

3.3 IBA for S-CTPAs

The key idea of IBA is that carriers, rather than having to solve multiple BGF or BGF-S for a large number of bundles, can simply transmit the parameters of their bid-generating function f^k to the auctioneer as their “bid.” These parameters are simply a list of arcs with associated arc prices and capacities. The auctioneer then simultaneously, in a single round, solves one large optimization problem that determines the allocation of bid lanes to carriers as well as the price to the carriers of the bundles awarded to them.

The auctioneer embeds the parameters of f^k directly into the WDP to get the following optimization problem:

$$\begin{aligned}
 (3.5a) \quad & \min \sum_{k \in \mathcal{K}} f^k(x^k) \\
 (3.5b) \quad & \text{s.t.} \quad \sum_{k \in \mathcal{K}} x_a^k = d_a \quad \forall a \in \mathcal{A}^\ell, \\
 (3.5c) \quad & x_a^k \in \{0, 1\} \quad \forall a \in \mathcal{A}^\ell, k \in \mathcal{K}.
 \end{aligned}$$

Solving (3.5) is equivalent to solving a fully-enumerated S-CTPA, where each carrier submits a bid for each of the M distinct bundles and the auctioneer solves an exponentially-sized T-WDP (Proposition II.1). The optimal solution \mathbf{x}^* awards exactly one bundle (possibly empty) to each carrier and \mathbf{x}^{k*} (with one element per bid lane $a \in \mathcal{A}^\ell$) corresponds to the bundle awarded to carrier k . The problem (3.5) has only $K \times L$ integer variables, compared to $K \times M$ binary variables in T-WDP. As an example, in a 10 carrier and 10 bid lanes (each with a volume of 10 truckloads) S-CTPA, this translates to a reduction from more than 2.5×10^{11} binary variables in T-WDP to only 100 integer variables in (3.5). Of course, the tractability of (3.5) depends greatly on the structure of the bid-generating function f^k . In the

next section, we describe two resulting *implicit* WDPs, the first using BGF as the bid-generating function and the second using BGF-S as the bid-generating function.

3.3.1 WDP with No Recourse

In the case where the carrier's bid-generating function is given by (3.1), the resulting pricing problem is a MCF problem. Substituting (3.1) directly into (3.5) results in the following *expected value winner determination problem* (EV-WDP):

$$(3.6a) \quad \min \sum_{k \in \mathcal{K}} \left[\sum_{a \in \mathcal{A}^\ell} p_a^k x_a^k + \min \sum_{a \in \mathcal{A}^k} \bar{p}_a^k y_a^k \right]$$

$$(3.6b) \quad \text{s.t.} \quad \sum_{k \in \mathcal{K}} x_a^k = d_a \quad \forall a \in \mathcal{A}^\ell$$

$$(3.6c) \quad \sum_{a \in \mathcal{A}^\ell: \mathcal{O}(a)=i} x_a^k + \sum_{a \in \mathcal{A}^k: \mathcal{O}(a)=i} y_a^k = \sum_{a \in \mathcal{A}^\ell: \mathcal{D}(a)=i} x_a^k + \sum_{a \in \mathcal{A}^k: \mathcal{D}(a)=i} y_a^k \quad \forall i \in \mathcal{N}, k \in \mathcal{K},$$

$$(3.6d) \quad y_a^k \leq \bar{u}_a^k \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^k,$$

$$(3.6e) \quad x_a^k \in \mathbb{Z}^+ \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^\ell,$$

$$(3.6f) \quad y_a^k \in \mathbb{Z}^+ \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^k.$$

In EV-WDP, the objective (3.6a) is to minimize the total cost, consisting of direct- and repositioning- movement costs incurred for procuring truckload services for all lanes in \mathcal{A}^ℓ . Constraints (3.6b) ensure that bid lane volumes are fully allocated among the set of carriers. Constraints (3.6c) enforce flow balance constraints for each carrier; that is, for each node $i \in \mathcal{N}$ the number of movements into node i must be equal to the number of movements out of node i . These constraints force all nodes $i \in \mathcal{N}$ to be transshipment and thus ensure that the resulting solution is a set of tours. Constraints (3.6d) enforce capacity constraints on each carrier's repositioning arcs. Finally, constraints (3.6e) and (3.6f) are variable integrality constraints.

We observe that in this formulation the integrality of y_a^k variables can be relaxed.

Once integer variables x_a^k are fixed (i.e. bid lanes have been allocated), the rest of the problem decomposes by carriers. This simplifies the problem to a collection of MCF problems, one per carrier. As stated earlier, MCF problems have TU constraint matrices; thus, integrality constraints (3.6f) can be replaced by a corresponding set of non-negativity constraints.

3.3.2 WDP with Recourse

The difference between WDP with recourse and WDP with no recourse is that a carrier may change repositioning movements after information about repositioning capacities and costs is revealed. In such a case, the carrier's bundle pricing problem is given by BGF-S (3.2), an extremely large MCF problem that is decomposable into $|\Xi^k|$ smaller MCF problems, one per scenario. Substituting BGF-S (3.2) directly into (3.5) results in the following *recourse winner determination problem* (R-WDP):

$$(3.7a) \quad \min \sum_{k \in \mathcal{K}} \left[\sum_{a \in \mathcal{A}^\ell} p_a^k x_a^k + \sum_{s=1}^{|\Xi^k|} \rho^{ks} \sum_{a \in \mathcal{A}^k} p_a^k y_a^{ks} \right]$$

$$(3.7b) \quad \text{s.t.} \quad \sum_{k \in \mathcal{K}} x_a^k = d_a \quad \forall a \in \mathcal{A}^\ell,$$

$$(3.7c) \quad \sum_{a \in \mathcal{A}^\ell: \mathcal{O}(a)=i} x_a^k - \sum_{a \in \mathcal{A}^k: \mathcal{O}(a)=i} y_a^{ks} = \sum_{a \in \mathcal{A}^\ell: \mathcal{D}(a)=i} x_a^k - \sum_{a \in \mathcal{A}^k: \mathcal{D}(a)=i} y_a^{ks} \quad \forall k \in \mathcal{K}, i \in \mathcal{N}, s = 1, \dots, |\Xi^k|,$$

$$(3.7d) \quad y_a^{ks} \leq u_a^{ks} \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^k, s = 1, \dots, |\Xi^k|,$$

$$(3.7e) \quad x_a^k \in \mathbb{Z}^+ \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^\ell,$$

$$(3.7f) \quad y_a^{ks} \in \mathbb{Z}^+ \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^k, s = 1, \dots, |\Xi^k|.$$

In this problem, the flow decisions \mathbf{x} , corresponding to bid lane allocations to carriers, and thus are made in advance of the realization of random events. These decisions are referred to as *first-stage* decisions. Flow decisions \mathbf{y} , corresponding

to carriers' repositioning movements to form a minimum-cost set of tours, can be deferred until after the realization of random events and are referred to as *second-stage* or *recourse* decisions. This type of stochastic model, with two stages of integer decision variables, are known as two-stage stochastic integer programs. For a detail overview of stochastic programming please refer to [4].

In solving EV-WDP and R-WDP, we observed a fundamental trade-off between solution quality and computational difficulty. EV-WDP, as described by (3.6), is a moderately-sized integer MFP that is solvable directly using a standard MIP solver. We demonstrate the tractability of this problem for various size auctions in §3.6.2. In contrast, R-WDP, as described by (3.7), is a large-scale integer MFP that is not solvable directly. In §3.4 and §3.5, we describe an algorithmic framework based on the L-shaped method for effectively solving R-WDP. However, despite the increased computational difficulties, the *expected* result of using the stochastic solution given by R-WDP is superior to that of the expected value solution give by EV-WDP. This result, for general two-stage stochastic programs, was established by [18]. In §3.6.2, we demonstrate the value of the stochastic solution over the expected value solution via computational results for S-CTPAs of varying sizes.

3.4 An L-shaped method for R-WDP

R-WDP, as presented in (3.7), is the *extensive form* of a two-stage stochastic integer program. Before presenting the decomposition algorithm for solving R-WDP, we define the *deterministic equivalent* of (3.7) as follows:

$$\begin{aligned}
(3.8a) \quad & \min \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}^\ell} p_a^k x_a^k + \mathcal{Q}(\mathbf{x}) \\
(3.8b) \quad & \text{s.t.} \quad \sum_{k \in \mathcal{K}} x_a^k = d_a \quad \forall a \in \mathcal{A}^\ell \\
(3.8c) \quad & x_a^k \in \mathbb{Z}^+ \quad \forall a \in \mathcal{A}^k, k \in \mathcal{K}
\end{aligned}$$

In the first stage, we fix bid lane allocations \mathbf{x} to carriers. Then the *expected recourse function* $\mathcal{Q}(\mathbf{x})$ represents the expected repositioning movement costs given bid lane allocations \mathbf{x} and is given by

$$(3.9) \quad \mathcal{Q}(\mathbf{x}) = \sum_{k \in \mathcal{K}} \sum_{s=1}^{|\Xi^k|} \rho^{ks} \mathcal{Q}^k(\mathbf{x}, s)$$

and $\mathcal{Q}^k(\mathbf{x}, s)$ represents carrier k 's minimum repositioning movement costs in scenario s given bid lane allocation \mathbf{x} .

$$\begin{aligned}
(3.10a) \quad & \mathcal{Q}^k(\mathbf{x}, s) = \min \sum_{a \in \mathcal{A}^k} p_a^{ks} y_a^{ks} \\
(3.10b) \quad & \text{s.t.} : \quad \sum_{a \in \mathcal{A}^k: \mathcal{O}(a)=i} y_a^{ks} - \sum_{a \in \mathcal{A}^k: \mathcal{D}(a)=i} y_a^{ks} = b_{i,\mathbf{x}}^{ks} \quad \forall i \in \mathcal{N}, \\
(3.10c) \quad & y_a^{ks} \leq u_a^{ks} \quad \forall a \in \mathcal{A}^k, \\
(3.10d) \quad & y_a^{ks} \in \mathbb{Z}^+ \quad \forall a \in \mathcal{A}^k.
\end{aligned}$$

We observe that in the context of problem (3.10), the bid lane allocations \mathbf{x} is a parameter and not a variable. The node (supply/demand) parameters $b_{i,\mathbf{x}}^{ks}$ given by

$$b_{i,\mathbf{x}}^{ks} = \sum_{a \in \mathcal{A}^\ell: \mathcal{D}(a)=i} x_a^k - \sum_{a \in \mathcal{A}^\ell: \mathcal{O}(a)=i} x_a^k \quad \forall k \in \mathcal{K}, i \in \mathcal{N}$$

represent the net demand (flow in minus flow out) at each city $i \in \mathcal{N}$ as a function of the bid lane allocation \mathbf{x} . Since $b_{i,\mathbf{x}}^{ks}$ are integers and (3.10) is a MCF similar to the

bid generation function described by (3.1), we can relax the integrality requirements on \mathbf{y} (3.10d) and replace them by nonnegativity constraints. Before presenting the decomposition algorithm for R-WDP, we state the dual to the linear relaxation of (3.10), which we call DSP.

$$(3.11a) \quad \mathcal{Q}_D^k(\mathbf{x}, s) = \max \sum_{i \in \mathcal{N}} b_{i,\mathbf{x}}^{ks} \alpha_i^{ks} + \sum_{a \in \mathcal{A}^k} u_a^{ks} \beta_a^{ks}$$

$$(3.11b) \quad \text{s.t.} \quad \alpha_{\Theta(a)}^{ks} - \alpha_{\mathcal{D}(a)}^{ks} + \beta_a^{ks} \leq p_a^{ks} \quad \forall a \in \mathcal{A}^k,$$

$$(3.11c) \quad \alpha_i^{ks} \text{ free} \quad \forall i \in \mathcal{N},$$

$$(3.11d) \quad \beta_a^{ks} \leq 0 \quad \forall a \in \mathcal{A}^k.$$

We solve R-WDP using a variant of the L-shaped method, whose integer master problem allocates bid lanes to carriers and whose linear MCF subproblems convey information about a carrier's repositioning movement costs under each scenario, $s = 1, \dots, |\Xi^k|$. The steps of the L-Shaped method are as follows:

1. **Initialization:**

Set lower bound LB=0, upper bound UB=+ ∞ , and iteration counter $I = 0$. Let ε be the predefined optimality gap at termination.

2. **Solve the following *restricted master problem* (RMP):**

Let θ be an auxiliary variable that defines the second-stage cost in RMP.

$$(3.12a) \quad \min \quad \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}^\ell} p_a^k x_a^k + \theta$$

$$(3.12b) \quad \text{s.t.} \quad \sum_{k \in \mathcal{K}} \sum_{s=1}^{|\Xi^k|} \sum_{i \in \mathcal{N}} a_{it}^{ks} \left[\sum_{a \in \mathcal{A}^\ell: \mathcal{O}(a)=i} x_a^k - \sum_{a \in \mathcal{A}^\ell: \mathcal{D}(a)=i} x_a^k \right] + \sum_{k \in \mathcal{K}} \sum_{s=1}^{|\Xi^k|} d_t^{ks} \leq \theta$$

$$\forall t = 1, \dots, I,$$

$$(3.12c) \quad \sum_{k \in \mathcal{K}} x_a^k = d_a \quad \forall a \in \mathcal{A}^\ell,$$

$$(3.12d) \quad x_a^k \in \mathbb{Z}^+ \quad \forall a \in \mathcal{A}^\ell, k \in \mathcal{K}.$$

Let $\bar{\mathbf{x}}$ be the incumbent bid lane allocation and update LB to be the optimal solution of RMP.

3. Check optimality gap:

If $(\text{UB} - \text{LB})/\text{LB} \leq \varepsilon$, exit and return the incumbent solution $\bar{\mathbf{x}}$, else, continue to Step 4.

4. Solve dual subproblems (DSP):

For each carrier $k \in \mathcal{K}$ and scenario $s = 1, \dots, |\Xi^k|$,

(a) Solve $\mathcal{Q}_D^k(\bar{\mathbf{x}}, s)$ and let $\bar{\alpha}$ and $\bar{\beta}$ represent its solution.

(b) Using $\bar{\alpha}$ and $\bar{\beta}$, compute the following RMP cut coefficients for iteration t :

$$(3.13a) \quad a_{it}^{ks} = \rho^{ks} \bar{\alpha}_i^{ks} \quad \forall i \in \mathcal{N},$$

$$(3.13b) \quad d_t^{ks} = \rho^{ks} \sum_{a \in \mathcal{A}^k} u_a^{ks} \bar{\beta}_a^{ks}.$$

5. Update ub:

$$(3.14) \quad \text{Let } q(\bar{\mathbf{x}}) = \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}^\ell} p_a^k \bar{x}_a^k + \sum_{k \in \mathcal{K}} \sum_{s=1}^{|\Xi^k|} \rho^{ks} \mathcal{Q}_D^k(\bar{\mathbf{x}}, s)$$

If $q(\bar{\mathbf{x}}) < \text{UB}$, let $\bar{\mathbf{x}}$ be the incumbent solution and let $\text{UB} = q(\bar{\mathbf{x}})$

6. Generate RMP optimality cut:

Let $I = I + 1$. Use cut coefficients generated in Step (4b) to create RMP cut (3.20) for iteration I , and return to Step 2.

By taking advantage of carriers' independent repositioning capacities and costs, we avoid the need to solve a large MCF that sums across commodities, and instead solve a set of smaller MCFs, one for each commodity. This decomposition has a significant impact of the runtime of step (4a), which represents the bottleneck operation for problems with a large number of scenarios, as is the case for S-CTPAs.

Although cutting-plane algorithms like the L-shaped method terminate in a finite number of iterations, the number of iterations needed to converge to the desired optimality gap may be extremely large for problems (like R-WDP) that have a network flow second-stage problem. In the next section, we introduce additional network structure to the RMP, valid inequalities, and multiple optimality cuts to improve algorithmic performance.

3.5 Accelerating L-shaped method

R-WDP is a large-scale, two-stage stochastic integer program, and initial computational results show that a direct application of the L-shaped method performs poorly because of limited network information in the initial RMPs and the weakness of the traditional Benders Cut due to commonly observed degeneracy in the optimal solution of network flow subproblems. In this section, we discuss in detail the sources of these problems and present enhancements to both the model and the L-shaped method to accelerate the convergence of the algorithm.

3.5.1 Embedding Network Capacities and Costs within RMP

One of the main weaknesses of cutting plane algorithms, such as the L-shaped method, when it is applied to network design problems, is that a large number of iterations is required to produce enough optimality cuts to relate adequate information about the network flow costs in the second stage. Because of the lack of network information in RMP, initial iterations tend to allocate bid lanes “greedily” to a small number of low cost carriers. To remedy this deficiency, we propose the addition of a set of network flow constraints to the RMP to provide a lower bound on repositioning movement costs for any bid lane allocation \mathbf{x} .

For each carrier k and repositioning arc $a \in \mathcal{A}^k$, let $\check{p}_a^k = \min \{p_a^{ks} | s = 1, \dots, |\Xi^k|\}$ represent the minimum unit movement cost on arc a over all scenarios and let $\hat{u}_a^k = \max \{u_a^{ks} | s = 1, \dots, |\Xi^k|\}$ represent the maximum arc a capacity over all scenarios.

$$(3.15a) \quad \tilde{\mathcal{Q}}^k(\mathbf{x}) = \min \sum_{a \in \mathcal{A}^k} \check{p}_a^k f_a^k$$

$$(3.15b) \quad \text{s.t.} \quad \sum_{a \in \mathcal{A}^k: \mathcal{O}(a)=i} f_a^k - \sum_{a \in \mathcal{A}^k: \mathcal{D}(a)=i} f_a^k = b_{i,\mathbf{x}}^k \quad \forall i \in \mathcal{N},$$

$$(3.15c) \quad f_a^k \leq \hat{u}_a^k \quad \forall a \in \mathcal{A}^k,$$

$$(3.15d) \quad f_a^k \geq 0 \quad \forall a \in \mathcal{A}^k.$$

Objective (3.15a) defines a lower-bound approximation of total repositioning movement cost. Constraints (3.15b) ensure that for each carrier k the number of movements into a node i is equal to the number of movements out of i . Constraints (3.15c) restrict the usage of carriers’ repositioning movements to the maximum available capacity over all scenarios.

We observe that $\tilde{Q}^k(\mathbf{x})$ is a MCF problem structurally identical to $Q^k(\mathbf{x}, s)$. The two problems differ only with respect to arc prices \mathbf{p} and arc capacity upper bounds \mathbf{u} . As stated by Proposition III.1, constraints defined by (3.15) provides a lower-bound on repositioning costs as a function of bid lane allocation \mathbf{x} .

Proposition III.1. $\sum_{k \in \mathcal{K}} \tilde{Q}^k(\mathbf{x}) \leq Q(\mathbf{x})$.

Proof: Given a feasible bid lane allocation \mathbf{x} , for any carrier $k \in \mathcal{K}$ and scenario $s = 1, \dots, |\Xi^k|$, the inequality $\tilde{Q}^k(\mathbf{x}) \leq Q^k(\mathbf{x}, s)$ holds because:

(I) $u_a^{ks} \leq \hat{u}_a^k \forall a \in \mathcal{A}^k$ implies that if \mathbf{y}^* is an optimal solution of $Q^k(\mathbf{x}, s)$ then it is a feasible solution of $\tilde{Q}^k(\mathbf{x})$;

(II) $\hat{p}_a^k \leq p_a^{ks} \forall a \in \mathcal{A}^k$ implies $\sum_{a \in \mathcal{A}^k} \hat{p}_a^k y_a^* \leq \sum_{a \in \mathcal{A}^k} p_a^{ks} y_a^*$.

Then the following relationships follows directly,

$$\sum_{k \in \mathcal{K}} \tilde{Q}^k(\mathbf{x}) = \sum_{k \in \mathcal{K}} \left[\sum_{s=1}^{|\Xi^k|} \rho^{ks} \tilde{Q}^k(\mathbf{x}) \right] \leq \sum_{k \in \mathcal{K}} \left[\sum_{s=1}^{|\Xi^k|} \rho^{ks} Q_p^k(\mathbf{x}, s) \right] = Q(\mathbf{x}). \quad \square$$

By Proposition III.1, we can impose the following set of constraints on RMP to provide a lower bound approximation of carriers' repositioning capacities and costs:

$$(3.16a) \quad \sum_{a \in \mathcal{A}^k} \check{p}_a^k f_a^k \leq \theta$$

$$(3.16b) \quad \sum_{a \in \mathcal{A}^k: \mathcal{O}(a)=i} f_a^k - \sum_{a \in \mathcal{A}^k: \mathcal{D}(a)=i} f_a^k = b_{i,\mathbf{x}}^k \quad \forall k \in \mathcal{K}, i \in \mathcal{N},$$

$$(3.16c) \quad f_a^k \leq \hat{u}_a^k \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^k,$$

$$(3.16d) \quad f_a^k \geq 0 \quad \forall k \in \mathcal{K}, a \in \mathcal{A}^k.$$

3.5.2 Knapsack Constraints

In this section, we introduce valid inequalities derived from optimality cuts generated during each iteration of the L-shaped method and the incumbent upper bound

UB. These cuts are similar in spirit to those introduced by [21] for supply chain planning problems. Let

$$(3.17) \quad \sum_{k \in \mathcal{K}} \sum_{s=1}^{|\Xi^k|} \sum_{i \in \mathcal{N}} a_{it}^{ks} \left[\sum_{a \in \mathcal{A}^\ell: \mathcal{O}(a)=i} x_a^k - \sum_{a \in \mathcal{A}^\ell: \mathcal{D}(a)=i} -x_a^k \right] + \sum_{k \in \mathcal{K}} \sum_{s=1}^{|\Xi^k|} d_t^{ks} \leq \theta$$

be the optimality cut generated in step (6) of the t^{th} iteration of the L-shaped method.

Since UB is the best upper bound, the following inequality holds,

$$(3.18) \quad \text{UB} \geq \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}^\ell} p_a^k x_a^k + \theta$$

Substituting the left-hand-side of (3.17) for θ in (3.18) and isolating the first-stage decision variables \mathbf{x} to one side yields the following valid inequality:

$$(3.19) \quad \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}^\ell} \left[p_a^k + \sum_{s=1}^{|\Xi^k|} \left(a_{\mathcal{O}(a),t}^{ks} - a_{\mathcal{D}(a),t}^{ks} \right) \right] x_a^k \leq \left[\text{UB} - \sum_{k \in \mathcal{K}} \sum_{s=1}^{|\Xi^k|} d_t^{ks} \right]$$

These type of knapsack constraints in conjunction with the standard optimality cut can have a significant impact on solution quality in subsequent iterations. As the L-shaped method progresses, UB decreases and the right-hand-side of (3.19) continues to decrease, thereby, tightening these knapsack inequalities.

3.5.3 Multi-Cut Generation

Traditionally, in the L-shaped method all s realizations of the second-stage programs are solved to obtain their optimal simplex multipliers. These multipliers are then aggregated into a single cut as shown in (3.20). However, the block structure of the two-stage stochastic program allows for multiple cuts to be generated simultaneously. Multiple cuts present more information to the first stage problem (RMP) and as such may improve the convergence property of the L-shaped method. However, this improvement comes at a cost of a larger first stage problem which

may negatively impact RMP runtime. Although this is not always true, in practice adding multiple optimality cuts often decreases the number of L-shaped iterations and overall runtime.

[5] describes a multicut L-shaped method for two-stage stochastic linear programs where an optimality cut is added for each scenario s . However, since $|\Xi^k|$ is extremely large in R-WDP problems, a direct application of the multicut approach will quickly render the RMP intractable. In solving R-WDP, we take a hybrid approach in which subsets of the realizations are aggregated to produce a reduced number of optimality cuts. In computational experiments in §3.6, we generate a single optimality cut for each carrier that is awarded a bid lane in the current master iteration. Specifically for each carrier $k \in \mathcal{K}$ awarded a least one bid load in the current iteration, we solve DSP for each scenario s to obtain simplex multipliers and then construct a single optimality cut that aggregates these multipliers across all $|\Xi^k|$ scenarios. These optimality cuts are given by

$$(3.20) \quad \sum_{s=1}^{|\Xi^k|} \sum_{i \in \mathcal{N}} a_{it}^{ks} \left[\sum_{a \in \mathcal{A}^\ell: \mathcal{O}(a)=i} x_a^k - \sum_{a \in \mathcal{A}^\ell: \mathcal{D}(a)=i} x_a^k \right] + \sum_{s=1}^{|\Xi^k|} d_t^{ks} \leq \theta^k.$$

3.6 Computational Experiments

In this section, we describe the computational experiments performed to assess the viability of S-CTPA using IBA. Specifically, we demonstrate the tractability of R-WDP, the performance benefits of the accelerated L-shaped method (A-LSM) over the standard L-shaped method (S-LSM), and the benefit of the stochastic solution (obtained by solving R-WDP) over the expected value solution (EV-WDP). We begin in §3.6.1 by first describing how the test instances for these studies were generated, and then compare the computational characteristics of A-LSM and S-LSM in

solving these instances. Finally, in §3.6.2, we assess the value of stochastic solutions over expected value solutions. All computational experiments were conducted on a Sun SunFire x4600 server, using a single AMD Opteron 8218 processor and 3.0 GB of RAM. We use the CPLEX 11.0 callable library to solve the mixed integer RMP and the linear DSP. We restrict all tests to a maximum runtime of 24 hours (1,440 minutes).

3.6.1 Tractability of R–WDP and performance benefits of A–LSM

We evaluated the tractability of R–WDP on randomly generated instances representing various-sized S–CTPAs. Random instances are controlled by the following parameters: number of nodes (cities), number of bidders (carriers), number of bid lanes (and corresponding volumes), number of repositioning lanes (and corresponding volumes) per carrier, and carriers’ price structures (represented by pairs of direct-movement and empty-movement price-per-mile).

Table 3.1: Test Set Characteristics for Various Auctions Sizes

Set No	No. Carriers	No. Nodes	No. of Bid Lanes	Avg. No. of Bid Loads	Range of Repositioning Lanes Per Carrier	Avg. No. of Repositioning Loads
1	20	40	100	12,279	50-100	88,100
2	30	50	300	37,386	100-300	456,793
3	40	60	500	62,427	200-400	900,678
4	50	100	500	62,956	200-400	1,128,370
5	50	100	1000	124,937	200-400	1,115,635
6	50	100	2000	250,877	200-400	1,103,918

We generated the six test sets, representing auctions of various sizes, shown in Table 3.1. Bid lane and existing lanes (per carrier) are generated on a network with up to 100 nodes representing the 100 most populous cities in the United States. There are 20 to 50 bidders bidding in each auction. For each set of auctions, we randomly generated four instances and report cumulative statistics. The volume of each bid lane is selected uniformly between 50 and 200 loads. A carrier’s repositioning capacity is represented by a set of capacitated, preexisting contracted lanes (that can

be used for “free”) and a set of uncapacitated empty movement lanes. The number of preexisting contracted lanes (and corresponding lane volume) per carrier is selected uniformly between the ranges specified in Table 3.1. Since the capacities of these pre-existing contracted lanes may be correlated and are not known with certainty at the time of the auction, we randomly classify carriers into two categories: correlated and uncorrelated. If a carrier has correlated pre-existing contracted lanes, then the availabilities of these lanes are correlated within a given scenario. For example, if a carrier has 5 pre-existing contracted lanes in a particular scenario, then either all lanes (and associated volume) are available for repositioning movement or none are available. On the other hand, if a carrier has uncorrelated pre-existing contracted lanes, for each lane and each scenario we uniformly chose the lane volume to be between 0 and the maximum value.

Carriers’ movement prices are generated by multiplying travel distance and a per-mile movement price to ensure triangle inequality is satisfied. A carrier’s price to serve an additional load in a bid lane is equal to the distance from the bid lane origin to the bid lane destination times the carrier’s direct movement price-per-mile, generated using a Normal distribution, $N(1.10, 0.05)$. Similarly, a carrier’s price to move empty between any city pair is equal to the distance between the city pair times the carrier’s empty movement price-per-mile generated using a Normal distribution, $N(0.80, 0.05)$.

Results

Solution characteristics for the six auction sizes are shown in Table 3.2. This table represents the aggregated results of 24 test instances (four per set). The result for each test set (represented by a row) is the average of 4 randomly generated instances. All times are reported in minutes.

Table 3.2: S-LSM and A-LSM Solution Characteristics for Various Test Sets

Set No.	S-LSM					A-LSM				
	No. itr.	Opt. Gap %	MP time	SP time	Total time	No. itr.	Opt. Gap %	MP time	SP time	Total time
1	294	NA	40	1,037	1,047	11	< 1	1	35	36
2	63	< 1	2	291	293	11	< 1	1	41	42
3	49	< 1	2	360	362	9	< 1	2	54	56
4	98	NA	7	1,173	1,180	10	< 1	10	63	112
5	56	< 1	12	782	794	8	< 1	11	109	120
6	30	< 1	3	368	371	4	< 1	12	64	76

Using A-LSM, we were able to solve all 24 test instances within the 24 hour (1440 minutes) runtime limit. S-LSM was also largely tractable. With the exception of one instance in test set 1 and two instances in test set 4, we were able to solve all test instances within the 24 hour runtime limit.

However, we observed a substantial performance difference between S-LSM and A-LSM. Using A-LSM, the average run times varies from a little over 30 minutes to a maximum of two hours. This is in stark contrast to S-LSM, where average run times varies from more than four hours to approximately 24 hours.

As expected, as auction size increases runtime also increases. However, these increases appear to be fairly moderate as evident by the difference between the run time of set 1 and set 4. However, it is interesting to note that, when everything else is equal, increasing the number of bid lanes in the auction actually improves solution time, as evident by computational results for sets 4 to 6. Intuitively, given a fixed-sized network with uniformly distributed lanes, increasing the number of lanes in the auction improves the probability of finding complementary lanes.

3.6.2 Value of the Stochastic Solution

In this section, we compare the solutions of the EV-WDP and R-WDP to assess the value of the stochastic solution. The comparison is performed as follows:

1. Solve the expected value winner determination problem (EV-WDP) by using

expected values for uncertain repositioning price and cost parameters, and let $\bar{\mathbf{x}}$ be the optimal bid lane allocation;

2. For each carrier k and scenario $s = 1, \dots, |\Xi^k|$, find the minimal repositioning movement cost to cover bid lanes in $\bar{\mathbf{x}}$ by solving $Q^k(\bar{\mathbf{x}}, s)$;
3. Compute the *expected value* of the expected value solution (EES) as follows:

$$(3.21) \quad \text{EES} = \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}^k} p_a^k \bar{\mathbf{x}}_a^k + \sum_{k \in \mathcal{K}} \sum_{s=1}^{|\Xi^k|} \rho^{ks} Q^k(\bar{\mathbf{x}}, s)$$

The value of the stochastic solution (VSS) is simply the difference between EES and the stochastic solution (SS), given by the solution of R-WDP. In Table 3.3, we present the value of the stochastic solution for all six test sets. Column 1 describes the set number. Column 2 describes the average optimality gap % which we computed as $(\text{EES} - \text{SS})/\text{SS}$. Columns 3 and 4 indicate the average number of winning carriers for EV-WDP and R-WDP respectively.

Table 3.3: Value of the Stochastic Solution

Set No	Opt. Gap %	No. of Carriers Assigned	
		EV-WDP	R-WDP
1	1.48	12	11
2	2.44	8	8
3	1.75	10	8
4	1.11	12	11
5	0.48	14	17
6	0.64	15	13

We observed that moving from EV-WDP to R-WDP improves the solution quality by 0.48 to 2.44 percent. Although it would appear that moving from the computationally easy EV-WDP to the computationally more challenging R-WDP does not improve solution quality by a large margin, a 0.48 to 2.44 percent improvement is quite meaningful in the trucking industry, where profit margins range from 2% to 4% ([8]). But perhaps, just as important, using A-LSM these cost improvements can be attained at a very manageable increase in runtime (from 32 to 105 minutes).

3.7 Generalization

In many real world application contexts, a set of physical commodities, each limited by their own network flow constraints, may share an underlying network structure. The objective of these applications is to minimize the total cost of flowing commodities on this network from source nodes to demand nodes in a way that minimizes total cost, while observing flow capacity constraints. Since arc capacities are limited, commodities interact when they flow on the same arcs. This type of problem is commonly referred to as a *multicommodity flow problem* (MFP).

In this section, we generalize R-WDP to a stochastic network flow problem, which we call a *two-stage integer multi-commodity flow problem* (TS-MFP), in which there is uncertainty in the cost and/or capacity of some of the arcs. Flow decisions over all arcs are made in two stages; *first-stage* flow decisions correspond to flows over deterministic arcs with known capacity and cost, and *second-stage* flow decisions correspond to flows over stochastic arcs with uncertain capacity and/or cost.

MFPs are commonly used to model transportation networks, telecommunication networks, network interdiction problems, and multi-product distribution systems. In most MFP literature, the network structure and cost are assumed to be deterministic. Deterministic integer and linear MFP have been well studied. Two early papers by [3] and [16] describe various solution methodologies for linear MFP. In recent years, integer MFPs have received increased interest due to their relevance to various application domains ([12], [9], and [13]).

However, the literature on stochastic linear MFPs is limited to a few instances [2], [5] and [24]. The literature that incorporates stochastic elements in *integer* MFPs is even more limited. The closest related work is that of [25] in which they proposed an

approximate dynamic programming-based methodology for multi-period stochastic integer MFP with uncertain demand, arising in fleet management.

3.7.1 Deterministic Integer MFP

Let us begin with the deterministic mathematical formulation of an integer MFP. Let $\mathcal{G}(\mathcal{N}, \mathcal{A})$ be a directed network defined by a set of nodes \mathcal{N} and a set of directed arcs \mathcal{A} . Each arc has an associated cost c_a^k denoting the cost per unit flow of commodity $k \in \mathcal{K}$ on that arc. Using this notation, the *arc-chain* formulation of integer MFP can be described as follows:

$$(3.22a) \quad \min \quad \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}} c_a^k x_a^k$$

$$(3.22b) \quad \text{s.t.} \quad \sum_{a \in \mathcal{A}: \mathcal{O}(a)=i} x_a^k - \sum_{a \in \mathcal{A}: \mathcal{D}(a)=i} x_a^k = b_i^k \quad \forall i \in \mathcal{N}, k \in \mathcal{K},$$

$$(3.22c) \quad \ell_a \leq \sum_{k \in \mathcal{K}} x_a^k \leq u_a \quad \forall a \in \mathcal{A},$$

$$(3.22d) \quad \ell_a^k \leq x_a^k \leq u_a^k \quad \forall a \in \mathcal{A}, k \in \mathcal{K},$$

$$(3.22e) \quad x_a^k \in \mathbb{Z}^+ \quad \forall a \in \mathcal{A}, k \in \mathcal{K}.$$

This formulation has a collection of $|\mathcal{N}| + |\mathcal{K}|$ flow balance constraints (3.22b), one for each commodity-node pair, modeling the flow of each commodity. Since commodities interact and compete for the used of finite arc capacities, we have a constraint (3.22c) that restricts the total flow, across *all* commodities, on each arc $a \in \mathcal{A}$ to be between the lower bound ℓ_a and the capacity limit u_a . Additionally, we also restrict the total flow of *each* commodity k on each arc $a \in \mathcal{A}$ to be between the lower bound ℓ_a^k and capacity limit u_a^k (3.22d). Lastly, we constrain the flow on each arc to be integer values (3.22e).

Observation III.2. *EV-WDP is a special case of integer MFP.*

Proof: The proof is by construction. EV-WDP can be transformed into an integer MFP (3.22) via the following variable, parameter and set redefinitions: (1) let $\ell_a = u_a = d_a \forall a \in \mathcal{A}^\ell$, (2) let $\ell_a^k = 0 \forall k \in \mathcal{K}, a \in \mathcal{A}^k$, (3) let $u_a^k = \bar{u}_a^k$ if $a \in \mathcal{A}^k$ and 0 otherwise $\forall k \in \mathcal{K}, a \in \mathcal{A}^k$, (4) let the vector of flow variables \mathbf{x} be the concatenation of flow variables \mathbf{x} and \mathbf{y} in EV-WDP, and finally, (5) let $\mathcal{A} = \mathcal{A}^\ell \bigcup_{k \in \mathcal{K}} \mathcal{A}^k$. Then the resulting formulation is an integer MFP of the form (3.22). \square

3.7.2 Two-Stage Integer MFP

In many practical applications, all network attributes of an MFP may not be known with certainty; for example, a carrier's backhaul capacity in truckload transportation and transportation capacity during emergency response depends on the availability of transportation arcs, which are contingent on the realization of random events. We now introduce a special form of MFP, which we call TS-MFP, where arc capacities and demands are uncertain. Specifically, we focus on the case where the following are true:

1. Arcs $a \in \mathcal{A}$ are partitioned into two disjoint sets \mathcal{A}^1 and $\mathcal{A}^k \forall k \in \mathcal{K}$, where arcs in \mathcal{A}^1 have known capacity and cost and arcs in \mathcal{A}^k have uncertain capacity and cost;
2. Commodities $k \in \mathcal{K}$ interact and compete for the use of finite arc capacities for arcs in \mathcal{A}^1 ;
3. Commodities do *not* interact or compete for the use of arc capacities in $\mathcal{A}^k \forall k \in \mathcal{K}$;
4. Node demands depend on several unknown factors and as such are not known

with certainty when first-stage flow decisions ($a \in \mathcal{A}^1$) are made.

In TS–MFP, the first stage variables allocate capacity of “shared” arcs in \mathcal{A}^1 among the different commodities $k \in \mathcal{K}$, after which the second-stage problems determine the optimal flow of each commodity k on arcs $a \in \mathcal{A}^k$ to meet commodity-specific demands. The resulting master problem is an integer program, while the resulting second stage problem is a MCF problem. For clarity of exposition, we define two sets of variables x_a^k and y_a^k , corresponding to first-stage flows on arcs $a \in \mathcal{A}^1$ and second-stage flows on arcs $a \in \mathcal{A}^k \forall k \in \mathcal{K}$ respectively.

Since uncertain parameters are associated with individual commodities, let ξ^k represent a discretely distributed random vector with finite support Ξ^k . In this problem a scenario $\xi^k = (\mathbf{b}, \mathbf{l}, \mathbf{u}, \mathbf{c})$ is a vector that represents a realization of *commodity-specific* node demands and arc capacity and cost for *all* arcs in \mathcal{A}^k . Index the scenarios by $s = 1, \dots, |\Xi^k|$ and let ρ^{ks} represent the realization probability of scenario s of commodity k . Then for each scenario s , c_a^{ks} is the cost of flowing one unit of commodity k on arc a , ℓ_a^{ks} and u_a^{ks} are flow upper- and lower- bounds of commodity k on arc a , and b_i^{ks} represents the demand for commodity k in node i . The first-stage decision variables x_a^k represent the flow of commodity k on arcs in \mathcal{A}^1 and the second-stage decision variables y_a^{ks} represent the flow of commodity k on arc $a \in \mathcal{A}^k$ in scenario s .

Then the deterministic equivalent of TS–MFP is given by

$$\begin{aligned}
(3.23a) \quad & \min \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}^1} c_a^k x_a^k + \mathcal{Q}(x) \\
(3.23b) \quad & \text{s.t. } \ell_u \leq \sum_{k \in \mathcal{K}} x_a^k \leq u_a \quad \forall a \in \mathcal{A}^1, \\
(3.23c) \quad & \ell_a^k \leq x_a^k \leq u_a^k \quad \forall a \in \mathcal{A}^1, k \in \mathcal{K}, \\
(3.23d) \quad & x_a^k \in \mathbb{Z}^+ \quad \forall a \in \mathcal{A}^1, k \in \mathcal{K}.
\end{aligned}$$

Given first-stage flow decisions \mathbf{x} and our assumption that commodities do not interact or compete for the use of arc capacities in $\mathcal{A}^k \forall k \in \mathcal{K}$, the *expected recourse* function $\mathcal{Q}(x)$ can be decomposed by commodities and is defined as follows:

$$(3.24) \quad \mathcal{Q}(\mathbf{x}) = \sum_{k \in \mathcal{K}} \sum_{s=1}^{|\Xi^k|} \rho^{ks} \mathcal{Q}^k(\mathbf{x}, s)$$

where $\mathcal{Q}^k(\mathbf{x}, s)$ is the recourse function of commodity k in scenario s and is given by

$$\begin{aligned}
(3.25a) \quad & \mathcal{Q}^k(\mathbf{x}, s) : \min \sum_{a \in \mathcal{A}^k} c_a^{ks} y_a^{ks} \\
& \text{s.t. } \sum_{a \in \mathcal{A}^1: \mathcal{D}(a)=i} x_a^k + \sum_{a \in \mathcal{A}^k: \mathcal{D}(a)=i} y_a^{ks} - \sum_{a \in \mathcal{A}^1: \mathcal{O}(a)=i} x_a^k - \sum_{a \in \mathcal{A}^k: \mathcal{O}(a)=i} y_a^{ks} \\
(3.25b) \quad & = b_i^{ks} \quad \forall i \in \mathcal{N}, \\
(3.25c) \quad & \ell_a^{ks} \leq y_a^{ks} \leq u_a^{ks} \quad \forall a \in \mathcal{A}^k, \\
(3.25d) \quad & y_a^{ks} \in \mathbb{Z}^+ \quad \forall a \in \mathcal{A}^k.
\end{aligned}$$

Observation III.3. *R-WDP is a special case of TS-MFP.*

Proof The proof is by construction. R-WDP can be converted to a TS-MFP (3.23) via the following variable, parameter and set redefinitions: (1) let $\mathcal{A}^1 = \mathcal{A}^\ell$,

(2) let $c_a^k = p_a^k \forall k \in \mathcal{K}$, $a \in \mathcal{A}^\ell$, (3) let $c_a^{ks} = p_a^{ks} \forall k \in \mathcal{K}$, $a \in \mathcal{A}^k$, $s = 1, \dots, |\Xi^k|$,
 (4) let $l_a = u_a = d_a \forall a \in \mathcal{A}^\ell$, (5) let $l_a^k = 0 \forall k \in \mathcal{K}$, $a \in \mathcal{A}^\ell$, (6) let $l_a^{ks} = 0 \forall k \in \mathcal{K}$, $a \in \mathcal{A}^k$, $s = 1, \dots, |\Xi^k|$, and finally, let $b_i^{ks} = 0 \forall k \in \mathcal{K}$, $s = 1, \dots, |\Xi^k|$, $i \in \mathcal{N}$.

Then the resulting formulation is a TS-MFP of the form (3.23). \square

We conclude this section by noting that the algorithmic enhancements proposed in §3.5 are applicable to TS-MFP. In the following section, we demonstrate the wide applicability of TS-MFP to another important problem, emergency response planning.

3.7.3 Another Application Area

In this section, we describe another important problems that can be modeled as TS-MFP and solved using the accelerated decomposition algorithm proposed in §3.5.

Natural Disaster Planning

In urban areas that are prone to natural disasters such as earthquakes, hurricanes, floods, droughts, et cetera., it is imperative that public officials and aid organizations make adequate preparation and planning to respond to the occurrence of natural disasters. Two very important questions to address in natural disaster planning are (1) how to allocate available emergency supplies among candidate warehouses prior to the occurrence of a natural disaster, and (2) following the occurrence of a natural disaster, how to transport critical commodities, such as food, medicine, water, and clothing, to disaster areas such that loss-of-life is minimized and efficiency of rescue operations is maximized. A multi-modal, deterministic version of such a problem was studied by [11].

This type of planning problem can be modeled as a TS-MFP. In this problem,

first-stage decisions correspond to preparatory (re)allocation of emergency commodities, such as food, medicine, water, and clothing, prior to the occurrence of a natural disaster. Second-stage decisions correspond to recourse decisions made after the occurrence of a natural disaster. Immediately after the occurrence of a natural disaster, transportation networks may be disrupted and emergency supplies may be damaged at certain locations. As such, the second-stage decision is to determine the optimal allocation and transportation of available emergency supplies from warehouses to disaster areas such that the overall system objective is optimized (e.g. minimizing the loss-of-life).

3.8 Summary and Conclusion

In this chapter, we consider a stochastic combinatorial truckload procurement auction problem that generalizes the deterministic combinatorial truckload procurement auction presented in Chapter II. We begin by describing carriers' pricing problems when repositioning capacities and costs are uncertain. We showed that using IBA (introduced in Chapter II), we can overcome the two main computational challenges of stochastic combinatorial truckload procurement auction: pricing an exponential number of bundles and solving the corresponding exponentially-sized winner determination problem. Using IBA, we can instead solve a single integer MFP. In the case when carriers use expected values for uncertain parameters, the resulting implicit winner determination is a moderately-sized integer MFP, which we call EV-WDP. This problem is solvable using standard commercial mixed integer programming solvers.

If carriers consider the full distribution of uncertain parameters, the resulting winner determination problem is a two-stage integer MFP, which we call R-WDP. We de-

scribed a decomposition framework based on the L-shaped method and then present model and algorithmic enhancements to improve the convergence of this algorithm. We begin by introducing network structures to the master problem to provide a good lower bound approximation on repositioning movement costs. Next, we present valid inequalities for the master problem and a hybrid approach in which subsets of the realizations are aggregated to produce a reduced number of combination optimality cuts. We then examine the effectiveness of our proposed methodologies on CTPAs of various sizes to demonstrate the benefits of our accelerated L-shaped method, versus the standard L-shaped method. We conclude the computational section with a comparison of the stochastic solution to the expected value solution to demonstrate the value of the stochastic solution.

We then present generalizations of our model and algorithmic approach to encompass a broader class of stochastic MFP, which we call two-stage multicommodity flow problem (TS-MFP), and demonstrate its applicability to an important disaster planning problem.

Several avenues for future research exist. In S-CTPA, a natural extension would be to consider the shipper's uncertainty in bid lane volumes. The additional modeling challenge, however, would be substantial because in this case the bid lane allocation (first-stage) decision is no longer deterministic. Another interesting S-CTPA extension would be the inclusion of operational constraints introduced in §2.3.3. Finally, it is worthwhile to note that we assumed risk neutral carriers, as such, their objective is to minimize *expected* cost. In practice, some carriers may be risk averse. For such carriers, a bid generating function objective other than minimizing the expected price may be more appropriate (e.g. looking at the 90th or 95th percentile of expected cost) to protect against catastrophic losses. Alternatively, other risk measures such

as conditional value at risk function may also be applicable.

In algorithmic development, we observed that the majority of computational time was spent on solving dual subproblems to obtain simplex multipliers for the optimality cuts. It would be interesting to investigate how we can speed up the subproblem solution time. Many subproblems are parametrically very close (the uncertain parameter values are similar); it may be interesting to see if we can take advantage of these similarities to reduce subproblem runtime. Finally, stochastic programs with network recourse often have degenerate optimal DSP solutions. It would be of interest and benefit to develop an algorithm to *efficiently* find the best (strongest) optimality cut among the set of optimality cuts corresponding to degenerate optimal DSP solutions.

BIBLIOGRAPHY

- [1] Ahuja, R., T. Magnanti, J. Orlin. 1993. Network Flows: Theory, Algorithms, and Applications. Prentice Hall, Englewood Cliffs, NJ.
- [2] Aneja, Y., K. Nair. 1982. A Multicommodity Network Flows with Probabilistic Losses. *Management Science* **28** 1080-1086.
- [3] Assad, A. 1978. Multicommodity network flows - a survey. *Networks* **8**(1) 37-91.
- [4] Birge, J., F. Louveaux. 1997. Introduction to Stochastic Programming. Springer-Verlag, New York, NY.
- [5] Barbarosoglu, G., Y. Arda. 2004. A two-stage stochastic programming framework for transportation planning in disaster response. *Journal of Operational Research Society* **55** 43-53.
- [6] Bureau of Economic Analysis. 2005. BEA News, U.S. Department of Commerce, news release August 31, 2005. <http://www.bea.gov/bea/rels.htm>.
- [7] Caplice, C., Y. Sheffi. 2006. Combinatorial auctions for truckload transportation. P. Cramton, R. Steinberg, Y. Shoham, eds., *Combinatorial Auctions*. MIT Press.
- [8] Costello, B. 2003. Motor carrier update. Technical report, NASEO Summer Transportation Fuels Outlook Conference, Washington, DC.
- [9] Crainic, T., M. Gendreau, P. Soriano, M. Toulouse. 1991. A tabu search procedure for multicommodity location/allocation with balancing requirements. *Annals of Operations Research* **41**(4) 359-383.
- [10] Figliozzi, M., H. Mahmassani, P. Jaillet. 2003. A framework for the study of carrier strategies in an auction based transportation marketplace. *TRB Annual Meeting*.
- [11] Haghani, A., S.C. Oh. 1996. Formulation and solution of a multi-commodity, multi-modal network flow model for disaster relief operations. *Transportation Research A* **30** 231-250.
- [12] Hane, C., C. Barnhart, E. Johnson, R. Marsten, G. Nemhauser, B. Sigismondi. 1995. The fleet assignment problem: Solving a large-scale integer program. *Mathematical Programming* **70**(3) 211-232.
- [13] Holmberg, K., Yuan D. 1998. A Lagrangean Approach to Network Design Problems. *International Transactions in Operational Research* **5**(6) 529-539.
- [14] Huff, A. 2006. Freight by wire. <http://www.etrucker.com/apps/news/article.asp?id=12683>. Retrieved May 2006.
- [15] Internet Truckstop. 2006. The Internet Truckstop. <http://www.truckstop.com>. Retrieved May 2006.
- [16] Kennington, J. 1978. A Survey of Linear Cost Multicommodity Network Flows. *Operations Research* **26**(2) 209-236.

- [17] Ledyard, J., M. Olson, D. Porter, J Swanson, D. Torma. 2002. The first use of a combined-value auction for transportation services. *Interfaces* **32**(5) 4-12.
- [18] Madansky, A. 1960. Inequalities for Stochastic Linear Programming Problems. *Management Science* **6**(2) 197-204.
- [19] Plummer, C. 2003. Bidder response to combinatorial auctions in truckload procurement. Master's thesis, Massachusetts Institute of Technology.
- [20] Sandholm, T., C. Boutilier. 2006. Preference elicitation in combinatorial auctions. P. Cramton, R. Steinberg, Y. Shoham, eds., *Combinatorial Auctions*. MIT Press.
- [21] Santoso, T., S. Ahmed, M. Goetschalckx, A. Shapiro. 2005. A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research* **167** 96-115.
- [22] Sheffi, Y. 2004. Combinatorial auctions in the procurement of transportation services. *Interfaces* **34**(4) 245-252.
- [23] Song, J., A. Regan. 2003. Combinatorial auctions for transportation service procurement: The carrier perspective. *TRB Annual Meeting*.
- [24] Soroush, H., P.B. Mirchandani. 1990. The stochastic multicommodity flow problem. *Networks* **20** 121-155.
- [25] Topaloglu, H., W.B. Powell. 2006. Dynamic programming approximations for stochastic, time-staged integer multicommodity flow problems. *INFORMS Journal on Computing*, **18**, 31-42.

CHAPTER IV

Including Wind in Power System Siting and Capacity Expansion Models

4.1 Introduction

Wind power is the fastest growing source of electricity in the U.S. Currently, over half the states in the country have passed so-called Renewable Portfolio Standards (RPS), which require utilities to procure a significant share (ranging from 5 to 25%) of their electricity from renewable resources in the near future [40]. Because wind is almost always the most cost-competitive renewable electricity source, it is expected to comprise the overwhelming majority of capacity installed to comply with these laws.

In response, there has been a proliferation of supporting policies to facilitate wind development, and these are likely to result in massive infrastructure expenditures. For example, states are building new transmission lines to areas with high-quality wind resources, expecting that wind developers will build new generation capacity in response [19],[7]. Moreover, the North American Reliability Corporation (NERC) is requiring its regional members to analyze scenarios with up to 15% of generation capacity coming from renewable sources, in part to assess how much new conventional generation must also be built so as to ensure reliability (for example, during periods of low wind speed [27]).

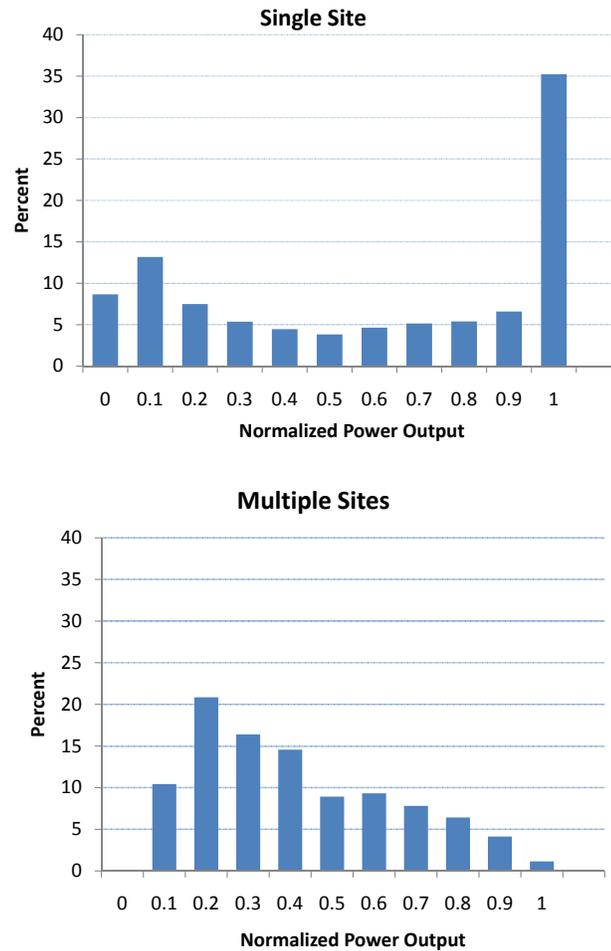


Figure 4.1: Normalized power output for the same amount of capacity. The top figure shows power output distribution when all turbines are installed at a single (windiest) site. The bottom figure shows power output distribution when turbines are uniformly distributed across twenty sites.

Unlike conventional sources of generation, which can usually be located near demand points or embedded in existing transmission networks so as to reduce transmission costs and losses, wind power varies with wind speed, which in turn varies significantly over both time and space. In particular, the aggregate output from a spatially-diverse collection of wind sites is far less variable than the output from the same capacity installed in a concentrated area. As illustrated in Figure 4.1 [19], a spatially diverse network of wind farms has lower variance (more reliable) and can

thus be integrated more easily with existing infrastructure. However, a spatially diverse network also requires higher siting costs and transmission costs and are likely to produce lower output (compared to installing all turbines at the windiest site), as illustrated in Figure 4.1 .

Current system planning studies that consider wind resources focus on *site-level* optimization, for example, selecting candidate sites on the basis of average wind speed [19],[7]. This approach, however, neglects the effect of spatio-temporal correlation across wind sites on power system reliability and operations.

Therefore, in this chapter we develop a model and solution framework to find the optimal wind farm network design (WFND). Specifically, we present a model for integrated transmission and generation expansion planning (TGEP) in which we simultaneously consider transmission capacity expansion and generation capacity expansion, for both wind resources as well as conventional generators (nuclear and fossil-fuel based). Additionally, we assume that a significant share of wind capacity must be installed to meet renewable portfolio standards (RPS) requirements, individual state’s commitment to meeting a portion of electricity demands using renewable resources.

4.1.1 Background

There is a large body of work that proposes models and optimization techniques for solving the Generation Expansion Planning (GEP) problem. Exact approaches based on mathematical programming [6],[14],[15] and dynamic programming [15],[25] have been proposed for variations of the GEP problem. Metaheuristic techniques, such as Genetic Algorithms [29], Evolutionary Programming [30], Differential Evolution [35], et cetera, have also been successfully applied to GEP.

However, most of this work, like [18],[10],[39], does not consider system reliability

measures and stochasticity in generation capacity availability, transmission capacity availability and area loads. Additionally, none of this work considers an integrated TGEP. Typically, GEP and Transmission Expansion Planning (TEP) are solved separately due to computational difficulties. For conventional power systems this may be adequate, as most conventional generation expansion planning involves the addition of generation or transmission capacity to existing networks. On the other hand, when considering wind generation, these two sets of decisions must be taken as a whole, as candidate wind resources may be located far from existing transmission infrastructure.

To the best of our knowledge, there has been very little work that considers the integrated TGEP problem, with the lone exception of recent work by [16]. This work presents an integrated generation and transmission expansion planning model and proposes a solution framework based on Benders Decomposition (BD) and Sample-Average Approximation (SAA) [20],[37]. However, [16] considers only conventional generation, uses a simplified lossless transmission model, and ignores fixed-siting costs. For conventional power systems with limited spatial coverage, a lossless transmission model is a reasonable simplification. However, for wind networks that may stretch thousands of miles, it is critical to consider the full quadratic line loss model, as line loss over great distance may have a significant impact on power flow solutions [8].

Despite proposing an integrated TGEP model, [16] presents computational results for the GEP problem and TEP problem separately. The integrated TGEP problem, a two-stage stochastic integer program, is computationally challenging, and the direct application of BD often performs poorly due to limited network information in the master problem and weakness of the standard optimality cut. We discuss some of

these issues that are relevant to WFND in Section 4.4 and present model extensions and procedures to improve the convergence of our proposed solution approach.

As noted in the previous section, wind-based power systems are characteristically different from conventional generation systems, both because wind speed is highly stochastic and location-dependent, and also because of the nonlinear and discontinuous power curve that converts wind speed to power output (Fig. 4.2).

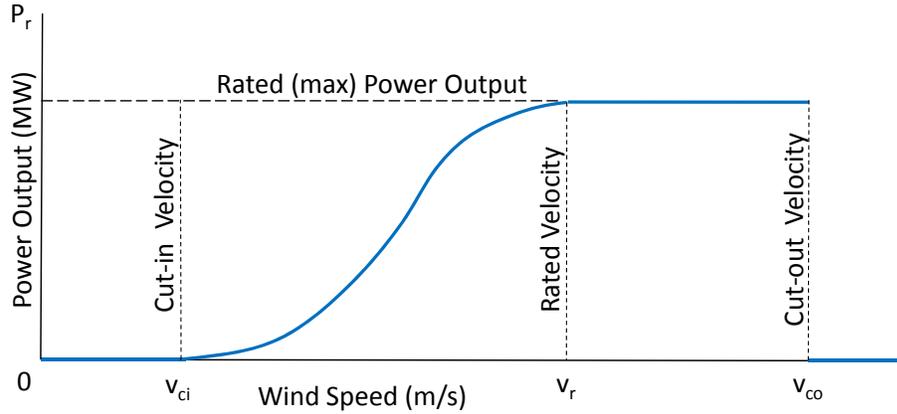


Figure 4.2: Conversion from wind speed to power.

Yet despite these differences, and despite the importance of capacity expansion for wind networks, relatively little has been done in this specific area. The most closely related efforts are those of [23] and [24], and the National Renewable Energy Laboratory’s Wind Deployment System Model, WinDS [26]. Both [23] and [24] use a heuristic dynamic search algorithm to choose wind capacity at several geographically-dispersed locations, assuming a pre-determined level of capacity, no fixed siting costs, and no penalties for violating reliability targets. WinDS, which has been used for nationwide wind deployment scenario analysis, is a large-scale linear programming model for generation and transmission expansion. The model minimizes the cost of meeting demand subject to constraints on emissions and reserve margins, which require that total generation capacity exceed maximum forecasted demand by some

pre-defined percentage.

4.1.2 Goals and Contributions

In this chapter, we present a mixed-integer stochastic programming approach to determine the optimal WFND. The contributions of this research are thus:

1. presenting a new model for the design of wind farm networks in a multi-area power system;
2. modeling an integrated generation and transmission expansion problem with explicit considerations for system uncertainties, fixed-siting costs and nonlinear transmission losses;
3. introducing an accelerated Benders decomposition algorithm (A-BD) that efficiently solves WFND problems with a large number of scenarios.

4.2 Problem Formulation

We begin by introducing the nomenclature for the WFND problem.

4.2.1 Sets

- \mathcal{N} is the set of all nodes (indexed by i and j).
- \mathcal{D} is the set of all demand nodes (indexed by d), $\mathcal{D} \subset \mathcal{N}$.
- D is the artificial demand-sink node, $D \notin \mathcal{D}$.
- \mathcal{G} is the set of all generator types (indexed by g , $g = 0$ is wind).
- \mathcal{T} is the set of transmission line types (indexed by t).
- ξ is the set of all scenarios (indexed by ξ).

4.2.2 First Stage Parameters

- ξ is a wind speed, multi-area load, generation- and transmission- availability scenario.
- h_i is the fixed cost of adding generation capacity to node i .
- h_{ij} is the fixed cost of adding transmission capacity between node i and node j .
- c_i^g is the cost of adding a type g generator at node i .
- c_{ij}^t is the cost of adding a type t transmission line connecting nodes i and j .
- Δ_{RPS} is the minimum amount of installed wind capacity required to meet renewable portfolio standard (RPS) requirements (in MW).
- ρ_i is the wind capacity factor at node i (i.e. fraction of installed wind capacity at node i that is credited towards meeting RPS).
- M_i^g is the maximum number of type g generators that can be installed at node i .
- M_{ij}^t is the maximum number of type t transmission lines connecting node i and node j that can be installed.
- σ is the scaling factor to make siting and investment costs comparable to operating cost (OC) and loss-of-load cost (LOLC).
- b_d is the minimum generation capacity that must be transmittable to demand node d (in MW).
- b_D is the minimum generation capacity that must be transmittable to demand-sink node D (in MW).
- S is the number of scenarios, $S = |\xi|$.

4.2.3 First Stage Variables: Capacity Expansion and Siting

- x_i^g is the number of type g generators to install at node i .
- x_{ij}^t is the number of type t transmission lines connecting i and node j to install.
- z_i is the binary variable that takes value 1 if node i is selected for generation expansion and 0 otherwise.
- z_{ij} is the binary variable that takes value 1 if a new line connecting nodes i and j is added and 0 otherwise.

4.2.4 First Stage Variables: Network Flows

- p_i^d is the generation capacity in node i transmittable to demand node d .
- p_i^D is the generation capacity in node i transmittable to demand-sink node D .
- f_{ij}^d is the potential power flow from node i to node j intended for demand node d .
- f_{ij}^D is the potential power flow from node i to node j intended for demand-sink node D .

4.2.5 Second Stage Parameters

- n_i^g is the marginal operating cost (OC) of a type g generator at node i .
- l_i is the loss-of-load-cost (LOLC) (per MWh) at node i .
- $e_i^{g\xi}$ is the number of existing type g generators available at node i in scenario ξ .
- $e_{ij}^{t\xi}$ is the number of existing type t lines connecting nodes i and j available in scenario ξ .
- m^t is the capacity of a type t transmission line.

- $m_i^{g\xi}$ is the capacity of a type g generation at node i in scenario ξ .
- $\kappa_i^{g\xi}$ is the total capacity of type g generators at node i in scenario ξ .
- $\kappa_{ij}^{t\xi}$ is the total capacity of type t transmission lines connecting node i and node j in scenario ξ .
- d_i^ξ is the demand at node i in scenario ξ .
- λ_{ij}^t is the linear loss coefficient of a type t line connecting node i and node j .
- μ_{ij}^t is the quadratic loss coefficient of a type t line connecting node i and node j .
- L is the number of blocks in piecewise linearization of quadratic line loss.
- $m^{t\ell}$ is the slope of the ℓ^{th} linearization interval of a type t line.
- $\kappa_{ij}^{t\ell\xi}$ is the total capacity of the ℓ^{th} interval of a type t transmission lines connecting node i and node j in scenario ξ .

4.2.6 Second Stage Variables: Network Flows

- p_i^g is the power generated at node i using type g generator(s).

- f_{ij}^t is the power flow from node i to node j on type t line(s).
- $g_{ij}^{t\ell}$ is the power flow from node i to node j on the ℓ^{th} linearization interval of type t line(s).
- s_i is the loss-of-load-cost at node i .

WFND is a two-stage, stochastic integer program where the *first-stage variables* correspond to generation and transmission siting and to capacity expansion decisions that must be made prior to the realization of random variables: which are transmission availability, generation availability, and area load. The objective is to minimize transmission and generation investment cost and *expected* operating cost (OC) and loss-of-load-cost (LOLC) subject to meeting minimum RPS and system reliability requirements. A weighting factor σ is used to scale the expected OC and LOLC to net present cost over the entire planning horizon. In WFND, overall system reliability is measured using expected LOLC, which inherently captures both the magnitude and frequency of load loss. We now present the stochastic programming formulation of WFND.

$$\begin{aligned}
(4.1a) \quad \min \quad & \sum_i \left((h_i z_i + \sum_g c_i^g x_i^g) + \sum_{j:i < j} (h_{ij} z_{ij} + \sum_t c_{ij}^t x_{ij}^t) \right) + \sigma \mathbb{E}_\xi [\mathcal{Q}(\xi, \mathbf{x})] \\
(4.1b) \quad & \text{s.t.} \quad \sum_i \rho_i x_i^0 \geq \Delta_{\text{RPS}} \\
(4.1c) \quad & x_i^g \leq M_i^g z_i, \quad \forall i, g \\
(4.1d) \quad & x_{ij}^t \leq M_{ij}^t z_{ij}, \quad \forall i, j : i < j, t \\
(4.1e) \quad & x_i^g \in \mathbb{Z}^+, \quad \forall i, g \\
(4.1f) \quad & x_{ij}^t \in \mathbb{Z}^+, \quad \forall i, j : i < j, t \\
(4.1g) \quad & z_i \in \{0, 1\}, \quad \forall i \\
(4.1h) \quad & z_{ij} \in \{0, 1\}, \quad \forall i, j : i < j, t
\end{aligned}$$

The objective (4.1a) represents the sum of generation siting, generation expansion, transmission siting, transmission expansion, and expected OC and LOLC. Constraint (4.1b) represents the share of RPS requirements to be met using wind resources. Constraints (4.1c) and (4.1d) enforce generation and transmission expansion limits and siting costs. Constraints (4.1e)–(4.1h) are variable integrality constraints. Since transmission lines are undirected, we only define transmission siting and expansion variables in one direction (for $i < j$).

In WFND, we assume that wind speed at each site, available generation and transmission capacity, and area loads are stochastic parameters with a known joint distribution. Then the *second-stage variables* correspond to power flow decisions under a specific realization of our uncertain parameters. Specifically, let $\boldsymbol{\xi}$ represent the random data vector and ξ represent a particular realization. Then given ξ and network design \mathbf{x} , the total OC and LOLC, which we represent by $\mathcal{Q}(\xi, \mathbf{x})$, is the optimal value of the following linear program:

$$\begin{aligned}
(4.2a) \quad \mathcal{Q}(\xi, \mathbf{x}) &= \min \sum_i \sum_g n_i^g p_i^g + \sum_i l_i s_i \\
(4.2b) \quad \text{s.t.} \quad & \sum_g p_i^g + \sum_j \sum_t (\mathcal{L}(f_{ji}^t) - f_{ij}^t) = d_i^\xi - s_i, \quad \forall i \\
(4.2c) \quad & f_{ij}^t \leq \kappa_{ij}^{t\xi}, \quad \forall i, j, t \\
(4.2d) \quad & p_i^g \leq \kappa_i^{g\xi}, \quad \forall i, g \\
(4.2e) \quad & p_i^g \geq 0, \quad \forall i, g \\
(4.2f) \quad & f_{ij}^t \geq 0, \quad \forall i, j, t \\
(4.2g) \quad & s_i \geq 0, \quad \forall i
\end{aligned}$$

For clarity of exposition, we have eliminated the index ξ from the second-stage decision variables (p, f, s) . The objective function (4.2a) is the sum of OC and LOLC. Constraints (4.2b) enforce flow balance at each node. Specifically, the total power produced, plus loss-adjusted power inflow $(\mathcal{L}(f_{ji}^t))$, minus power outflow, must be equal to the load at node i minus any loss-of-load. Constraints (4.2c) enforce power flow limits on each transmission line type, where $\kappa_{ij}^{t\xi} = m^t \cdot (e_{ij}^{t\xi} + x_{ij}^t)$ represents the total transmission capacity of type t lines connecting node i to node j under scenario ξ and network design \mathbf{x} . Constraints (4.2d) enforce generation capacity limits, where $\kappa_i^{g\xi} = m_i^{g\xi} \cdot (e_i^{g\xi} + x_i^g)$ represents the total capacity of type g generators at node i under scenario ξ and network design \mathbf{x} . Note that the capacity of a generator $m_i^{g\xi}$ is scenario dependent; for example, wind turbine power output at node i is a function of scenario-dependent wind speed. Constraints (4.2e)–(4.2g) are variable non-negativity constraints. Within the context of a two-stage stochastic program, the network design vector \mathbf{x} is a parameter of the second stage problem described by (4.2a)–(4.2g).

We have also chosen to present the second stage problem $\mathcal{Q}(\xi, \mathbf{x})$ as a network flow problem. However, if so desired, a linearized DC power flow model with quadratic loss, such as those present in [1], can also be use in place of (4.2).

4.2.7 Transmission Losses

As has been shown in [19] and Figure 4.1, the aggregate output of a collection of spatially-diverse wind sites is far less variable than the same capacity installed in a concentrated area. However, this reliability benefit comes at an increase in transmission investment cost and, potentially, a decrease in total output, as the windiest (highest average wind speed) site may not be a good candidate from a reliability (low variability in power output) standpoint. Because candidate wind farms may be located in very spatially-diverse areas, far away from load centers, it is critical to account for the impact of transmission losses across great distances. In this section, we present two alternative enhancements to the model to account for transmission losses.

Linear Line Loss

In this approximation, the fraction of power lost on a type t line connecting node i and node j is strictly a function of distance and the line type. We model this loss as a linear function of line flow f_{ij}^t , where $\lambda_{ij}^t \in (0, 1)$ represents the fraction of power that is transmitted to node j .

$$(4.3) \quad \mathcal{L}(f_{ij}^t) = \lambda_{ij}^t f_{ij}^t \quad \forall i, j, t$$

Quadratic Line Loss

In this approximation, the fraction of power lost on a type t line connecting node i and node j is now a function of distance, line type, and power flow. We model this

loss as a quadratic function of the total flow f_{ij}^t and line parameter μ_{ij}^t . Observe that μ_{ij}^t is a function of both the line type and distance.

Then the quadratic loss-adjusted inflow $\mathcal{L}(f_{ij}^t)$ is given by:

$$(4.4) \quad \mathcal{L}(f_{ij}^t) = f_{ij}^t - \mu_{ij}^t \left(\frac{f_{ij}^t}{e_{ij}^{t\xi} + x_{ij}^t} \right)^2 (e_{ij}^{t\xi} + x_{ij}^t)$$

where $\left(\frac{f_{ij}^t}{e_{ij}^{t\xi} + x_{ij}^t} \right)$ represents the total flow from node i to node j on a *single* type t line in scenario ξ . For brevity we rewrite (4.4) as follows:

$$(4.5) \quad \mathcal{L}(f_{ij}^t) = f_{ij}^t - \frac{\mu_{ij}^t}{e_{ij}^{t\xi} + x_{ij}^t} (f_{ij}^t)^2$$

Constraint (4.5) contains a quadratic term $(f_{ij}^t)^2$, which we approximate by using L line segments. Since the flow on a *single* type t line is bounded within the interval $[0, m^t]$, we partition this interval into L smaller intervals using the break points $a_0 = 0 < a_1 < \dots < a_{L-1} < a_L$. Then the length (which represents maximum flow) of the ℓ^{th} interval is given by $\kappa_{ij}^{t\ell\xi} = (a_\ell - a_{\ell-1}) \cdot (e_{ij}^{t\xi} + x_{ij}^t)$, and the corresponding slope of the function can be computed by:

$$(4.6) \quad m^{t\ell} = \frac{(a_\ell^2 - a_{\ell-1}^2)(e_{ij}^{t\xi} + x_{ij}^t)}{a_\ell - a_{\ell-1}}, \quad \forall \ell$$

Using the above definitions, the quadratic term $(f_{ij}^t)^2$ in (4.5) can be approximated as follows:

$$(4.7a) \quad f_{ij}^t = \sum_{\ell} g_{ij}^{t\ell}, \quad \forall i, j, t$$

$$(4.7b) \quad (f_{ij}^t)^2 \approx \sum_{\ell} m^{t\ell} g_{ij}^{t\ell}, \quad \forall i, j, t$$

$$(4.7c) \quad g_{ij}^{t\ell} \leq \kappa_{ij}^{t\ell\xi}, \quad \forall i, j, t, \ell$$

$$(4.7d) \quad g_{ij}^{t\ell} \geq 0, \quad \forall i, j, t, \ell$$

(4.7a) specifies that the flow magnitude summed across all L intervals must be equal to f_{ij}^t . (4.7b) approximates the quadratic term $(f_{ij}^t)^2$ using L line segments, where $m^{t\ell}$ and $g_{ij}^{t\ell}$ represent the slope and flow magnitude of the ℓ^{th} interval of type t lines respectively. (4.7c) restricts the flow on the ℓ^{th} interval of type t lines to be at most $\kappa_{ij}^{t\ell\xi}$.

For the remainder of this chapter, we will focus on the quadratic transmission loss model, which most accurately reflects quadratic power loss behavior of transmission lines.

4.2.8 Full Second Stage Model With Quadratic Loss

Using the quadratic definition of $\mathcal{L}(f_{ij}^t)$ in (4.5) and constraints (4.7a)–(4.7d), we replace all f_{ij}^t terms with the RHS of (4.7a) and substitute $(f_{ij}^t)^2$ with the RHS of (4.7b). Then the complete second stage problem with quadratic line losses can be formulated as the following linear program:

$$(4.8a) \quad Q(\xi, \mathbf{x}) = \min \sum_i \sum_g n_i^g p_i^g + \sum_i l_i s_i$$

$$(4.8b) \quad \text{s.t.} \sum_g p_i^g + \sum_j \sum_t \sum_\ell r_{ji}^{t\ell} g_{ji}^{t\ell} - g_{ij}^{t\ell} = d_i^\xi - s_i, \quad \forall i \quad (\alpha)$$

$$(4.8c) \quad \sum_\ell g_{ij}^{t\ell} \leq \kappa_{ij}^{t\xi}, \quad \forall i, j, t \quad (\beta)$$

$$(4.8d) \quad p_i^g \leq \kappa_i^{g\xi}, \quad \forall i, g \quad (\gamma)$$

$$(4.8e) \quad g_{ij}^{t\ell} \leq \kappa_{ij}^{t\ell\xi}, \quad \forall i, j, t, \ell \quad (\pi)$$

$$(4.8f) \quad p_i^g \geq 0, \quad \forall i, g$$

$$(4.8g) \quad s_i \geq 0, \quad \forall i$$

$$(4.8h) \quad g_{ij}^{t\ell} \geq 0, \quad \forall i, j, t, \ell$$

where $r_{ij}^{t\ell} = 1 - \frac{\mu_{ij}^t (a_\ell^2 - a_{\ell-1}^2)}{a_\ell - a_{\ell-1}}$. The dual variables corresponding to primal constraints

(4.8b) – (4.8e) are shown in parentheses immediately following these constraints.

Observe that in approximating quadratic line losses, each flow variable f_{ij}^t in the original second stage problem has been replaced by a set of L smaller flow variables $g_{ij}^{t\ell}$, where each $g_{ij}^{t\ell}$ is the flow on a new arc in our reformulated network.

4.3 Benders Decomposition Algorithm

We solve WFND using a variant of BD commonly known as the L-shaped method with integer first-stage variables [34],[38]. Below we briefly state the BD algorithm as it is applied to WFND.

1. Initialization:

Set lower bound $LB=0$, upper bound $UB=+\infty$, and iteration counter $I = 0$. Let \bar{z} and \bar{x} represent the incumbent first-stage solution (initially undefined).

2. Solve the following *restricted master problem* (RMP):

$$(4.9a) \quad LB = \min \sum_i \left(h_i z_i + \sum_g c_i^g x_i^g + \sum_{j:i<j} (h_{ij} z_{ij} + \sum_t c_{ij}^t x_{ij}^t) \right) + \frac{\sigma}{S} \theta$$

$$(4.9b) \quad \text{s.t. Constraints (4.1b) – (4.1h)}$$

$$(4.9c) \quad \sum_{\xi} \sum_i \left(\sum_g a_{ik}^{g\xi} x_i^g + \sum_{j:i<j} \sum_t (a_{ijk}^{t\xi} + a_{jik}^{t\xi}) x_{ij}^t + b_k^{\xi} \right) \leq \theta,$$

$$\forall k = 1, \dots, I$$

$$(4.9d) \quad \theta \geq 0$$

and update LB to be the optimal solution of RMP.

3. Check optimality gap:

If $(UB - LB)/LB \leq \varepsilon$, exit and return the incumbent solution.

4. Solve dual subproblems (DSP):

For each scenario $\xi = 1, \dots, S$

(a) Solve the dual second-stage problem:

$$(4.10a) \quad Q^D(\xi, \bar{\mathbf{x}}) = \max \sum_i \left(d_i^\xi \alpha_i + \sum_g \kappa_i^{g\xi} \gamma_i^g \right. \\ \left. + \sum_j \sum_t (\kappa_{ij}^{t\xi} \beta_{ij}^t + \sum_\ell \kappa_{ij}^{t\xi} \pi_{ij}^{t\ell}) \right)$$

$$(4.10b) \quad \alpha_i + \gamma_i^g \leq n_i^g, \quad \forall i, g$$

$$(4.10c) \quad \alpha_i \leq l_i, \quad \forall i$$

$$(4.10d) \quad r_{ij}^{t\ell} \alpha_j - \alpha_i + \beta_{ij}^t + \pi_{ij}^{t\ell} \leq 0, \quad \forall i, j, t, \ell$$

$$(4.10e) \quad \beta_{ij}^t \leq 0, \quad \forall i, j, t$$

$$(4.10f) \quad \gamma_i^g \leq 0, \quad \forall i, g$$

$$(4.10g) \quad \pi_{ij}^{t\ell} \leq 0, \quad \forall i, j, t, \ell$$

(b) Define and compute the following RMP cut coefficients:

$$(4.11a) \quad a_{ik}^{g\xi} = m_i^{g\xi} \gamma_i^g, \quad \forall i, g$$

$$(4.11b) \quad a_{ijk}^{t\xi} = m^t \beta_{ij}^t + \sum_\ell \Delta^{t\ell} \pi_{ij}^{t\ell}, \quad \forall i, j, t$$

$$(4.11c) \quad b_k^\xi = \sum_i d_i^\xi \alpha_i + \sum_i \sum_g m_i^{g\xi} e_i^{g\xi} \gamma_i^g \\ + \sum_i \sum_j \sum_t \left(m^t e_{ij}^{t\xi} \beta_{ij}^t + \sum_\ell \Delta^{t\ell} e_{ij}^{t\xi} \pi_{ij}^{t\ell} \right)$$

5. Update ub:

$$(4.12) \quad q(\bar{\mathbf{z}}, \bar{\mathbf{x}}) = \min \sum_i \left(h_i z_i + \sum_g c_i^g x_i^g \right) \\ + \sum_{j:i < j} \left(h_{ij} z_{ij} + \sum_t c_{ij}^t x_{ij}^t \right) + \frac{\sigma}{S} \sum_\xi Q^D(\xi, \bar{\mathbf{x}})$$

If $q(\bar{\mathbf{z}}, \bar{\mathbf{x}}) < \text{UB}$, let $\bar{\mathbf{x}}$ be the incumbent solution and let $\text{UB} = q(\bar{\mathbf{z}}, \bar{\mathbf{x}})$

6. Generate optimality cut:

Let $I = I + 1$, use cut coefficients generated in step (4b) to create a RMP cut

(4.9c) for iteration I , and return to step 2.

Although cutting plane algorithms like BD terminate in a finite number of iterations, the number of iterations needed to converge to the desired optimality gap may be extremely large for network design problems like WFND. In the next section, we introduce additional network structure to RMP, valid inequalities, and multiple optimality cuts to improve algorithmic performance.

4.4 Accelerating Benders Decomposition

WFND is a large-scale, two-stage stochastic integer program, and in our initial computational experiments we observed (as is often the case) that a direct application of BD performs poorly due to certain structural properties of the model. In this section we discuss the sources of these problems and present enhancements to both the model and BD to accelerate the convergence of our solution approach.

4.4.1 Network Connectivity

Unlike conventional TGEP, where much of the transmission infrastructure is already in place, in WFND problems candidate wind farms may be located in remote areas far away from existing transmission infrastructure. As such, transmission expansion and generation expansion decisions must be made simultaneously.

In early iterations of BD, transmission capacity and generation capacity are added piecemeal, resulting in network designs that are often unconnected. This leads to large LOLC in the second-stage and a weak UB. Performance is further degraded by the fact that these initial cuts increase the size of the integer RMP and, consequently, its run-time.

We remedy this problem by enforcing network connectivity in RMP. That is, in a valid network design for each generation node (new or existing), there must exist at least one path from that generation node to a demand node. This augmented RMP

network structure is defined as follows.

Let y_{ij} be the flow of electricity from node i to node j , D be an artificial demand-sink node, and y_{iD} be the flow of electricity from node i to demand-sink node D .

Then RMP's network connectivity constraints can be defined as follows:

$$(4.13a) \quad \sum_j y_{ij} - \sum_j y_{ji} = z_i, \quad \forall i$$

$$(4.13b) \quad \sum_i y_{iD} = \sum_i z_i$$

$$(4.13c) \quad y_{ij} \leq N \sum_t (x_{ij}^t + e_{ij}^t), \quad \forall i, j$$

$$(4.13d) \quad y_{ij} \geq 0, \quad \forall i, j$$

Constraints (4.13a) are flow balance constraints that specify that for each node where generation capacity is installed, a net of one unit of electricity must flow out of that node. Constraint (4.13b) specifies that the total flow into node D must be equal to the number of nodes with new generation capacity. This constraint ensures that a path exists from each new generation node to a demand node. Constraints (4.13c) specify that power can only flow on existing or newly installed lines. Observe that for any pair of nodes (i, j) that has existing transmission capacity $\sum_{\forall t} e_{ij}^t > 0$, constraints (4.13c) are never binding and can therefore be eliminated. Constraints (4.13d) enforce non-negative flows.

4.4.2 Demand Fulfillment

The aforementioned network connectivity constraints ensure that generation capacity is always *connected* to demand points. However, that alone is not sufficient to ensure that *sufficient* generation and transmission capacity is added to meet demand in all areas. As a consequence, initial network designs consist of very little additional generation and transmission capacity, which corresponds to lower objective values. This again results in large LOLC in the second stage and a weak UB.

To remedy this problem, we introduce additional network structure in RMP to ensure that *adequate* transmission and generation capacity is installed to meet (I) area loads and (II) total system load.

To enforce (I), we define flow balance, generation, and transmission constraints for *each* demand node $d \in \mathcal{D}$, as follows:

$$(4.14a) \quad p_i^d + \sum_j f_{ji}^d - \sum_j f_{ij}^d = 0, \quad \forall i \neq d$$

$$(4.14b) \quad p_i^d + \sum_j f_{ji}^d - \sum_j f_{ij}^d = b_d, \quad i = d$$

$$(4.14c) \quad p_i^d \leq \sum_g m_{\text{cap}}^g \cdot (e_i^g + x_i^g), \quad \forall i$$

$$(4.14d) \quad f_{ij}^d \leq \sum_t m_{\text{cap}}^t \cdot (e_{ij}^t + x_{ij}^t), \quad \forall i, j$$

Constraints (4.14a) enforce flow balance for each non- d nodes; that is, the total power produced, plus power inflow, minus power outflow, must be equal to zero. These constraints in essence force all non- d nodes to be trans-shipment nodes. Constraints (4.14b) enforce flow balance at demand node d ; that is, the total power produced, plus power inflow, minus power outflow, must be equal to the demand fulfilment parameter b_d . Constraints (4.14c) and (4.14d) enforce generation- and transmission-capacity limits based on the total amount of existing- and newly-installed- capacity. Constraints (4.14a) - (4.14d) together ensure that there is sufficient generation- and transmission- capacity to deliver b_d units of electricity to demand node d in isolation (not considering demands at other nodes).

To enforce (II), we define a new set of variables, $f_{iD} \forall i$, to indicate the amount of electricity transmittable from node i to demand-sink node D and the following sets

of flow balance, generation, and transmission constraints:

$$(4.15a) \quad p_i + \sum_j f_{ji} - \sum_j f_{ij} = f_{iD}, \quad \forall i$$

$$(4.15b) \quad \sum_d f_{dD} \geq b_D,$$

$$(4.15c) \quad p_i \leq \sum_g m_{\text{cap}}^g \cdot (e_i^g + x_i^g), \quad \forall i$$

$$(4.15d) \quad f_{ij} \leq \sum_t m_{\text{cap}}^t \cdot (e_{ij}^t + x_{ij}^t), \quad \forall i, j$$

Constraints (4.15a) enforce flow balance for all nodes $i \in \mathcal{N}$; that is, the total power produced, plus power inflow, minus power outflow, must be equal to the flow to demand-sink node D . Constraint (4.15b) specifies at least b_D units of installed capacity must be transmittable to demand-sink node D from demand nodes $d \in \mathcal{D}$. Finally, constraints (4.15c) and (4.15d) enforce generation- and transmission- capacity limits based on the amount of existing and newly installed capacity. Constraints (4.15a) - (4.15d) together ensure that there is sufficient generation- and transmission- capacity to deliver b_D units of electricity to demand-sink node D . Since there is a directed arc from each demand node d to D , these constraints ensure that there is sufficient generation- and transmission- capacity to deliver b_D units of electricity to the demand nodes.

Although these two sets of demand fulfillment constraints cannot guarantee that there is adequate generation- and transmission capacity to satisfy demand in each scenario ξ , they do prevent the selection of poor (inadequate) network designs in the early iterations of Benders Decomposition.

4.4.3 Valid Inequalities

In this section, we introduce valid inequalities derived from optimality cuts generated during each iteration of BD and the incumbent upper bound UB. These cuts

are similar in spirit to those introduced by [32] for supply chain planning problems.

Let

$$(4.16) \quad \sum_{\xi} \sum_i \left(\sum_g a_{ik}^{g\xi} x_i^g + \sum_{j:i<j} \sum_t a_{ijk}^{t\xi} x_{ij}^t \right) + b_k \leq \theta \quad \forall k = 1, \dots, I$$

be the optimality cut derived at the end of step (4) of iteration I . Since UB is the best upper bound, we have

$$(4.17) \quad \text{UB} \geq \sum_i \left(h_i z_i + \sum_g c_i^g x_i^g + \sum_{j:i<j} (h_{ij} z_{ij} + \sum_t c_{ij}^t x_{ij}^t) \right) + \frac{\sigma}{S} \theta$$

Substituting the LHS of (4.16) for θ in (4.17) and isolating the first-stage decision variables \mathbf{x} and \mathbf{z} to one side, we get the following valid inequality:

$$(4.18) \quad \sum_i \left(\lfloor h_i \rfloor z_i + \sum_g \lfloor u_{ik}^g \rfloor x_i^g + \sum_{j:i<j} (\lfloor h_{ij} \rfloor z_{ij} + \sum_t \lfloor u_{ijk}^t \rfloor x_{ij}^t) \right) \leq \lfloor \text{UB} - d_k \rfloor$$

where

$$(4.19a) \quad \begin{aligned} u_{ik}^g &= c_i^g + \frac{\sigma}{S} \sum_{\xi} a_{ik}^{g\xi}, \quad \forall i, g \\ u_{ijk}^t &= c_{ij}^t + \frac{\sigma}{S} \sum_{\xi} (a_{ijk}^{t\xi} + a_{jik}^{t\xi}), \quad \forall (i, j) : i < j, t \end{aligned}$$

$$(4.19b) \quad d_k = \frac{\sigma}{S} b_k.$$

These type of knapsack constraints in conjunction with the standard optimality cut can have a significant impact on solution quality in subsequent iterations. As the BD algorithm progresses, UB decreases and the right-hand-side of (4.18) continues to decrease, thereby, tightening these knapsack inequalities.

4.4.4 Multicut Generation

Traditionally, in BD all S realizations of the second-stage programs are solved to obtain their optimal simplex multipliers. These multipliers are then aggregated

into a single cut as shown in (4.9c). However, the block structure of the two-stage stochastic program allows for multiple cuts to be generated simultaneously. Multiple cuts present more information to the first stage problem. Although this is not always true, in practice adding multiple optimality cuts often decreases the number of Benders iterations.

[5] describes a multicut BD algorithm for two-stage stochastic linear programs where an optimality cut is added for each scenario ξ . However, since S is extremely large in WFND problems, a direct application of the multicut approach will quickly render the RMP intractable. In solving WFND, we take a hybrid approach in which subsets of the realizations are aggregated to produce a reduced number of optimality cuts. In computational experiments in §4.5, we generate six optimality cuts per master iteration. Specifically, we solve DSP for each scenario ξ to obtain simplex multipliers and then construct a single optimality cut that aggregates these multipliers across scenarios representing each two month period (approximately 1,464 hourly scenarios). These optimality cuts are given by

$$(4.20) \quad \sum_{\xi=s \times 1,464}^{(s \times 1,464)+1,463} \sum_i \left(\sum_g a_{ik}^{g\xi} x_i^g + \sum_{j:i < j} \sum_t (a_{ijk}^{t\xi} + a_{jik}^{t\xi}) x_{ij}^t + b_k^\xi \right) \leq \theta^s$$

$$\forall s = 0, \dots, 5.$$

4.5 Computational Experiments

In this section, we describe the computational experiments performed to assess the tractability of WFND and the performance of our proposed solution framework. We begin in Section 4.5.1 by describing how the three test systems were generated for this study and then demonstrate the results of our proposed A-BD in Section 4.5.2. All computational experiments were conducted on a Sun SunFire x4600 server, using

only a single AMD Opteron 8218 processor and 1.5 GB of RAM. We use the CPLEX 11.0 callable library to solve the mixed integer RMP and the linear DSP.

4.5.1 Test System Generation

To assess the tractability of WFND and the computational efficacy of our proposed solution framework, we developed three test systems: Test System 1 (TS1), Test System 2 (TS2), and Test System 3 (TS3) using a combination of historical data given in the literature and randomly generated data. TS1 is a network consisting of 18 nodes (representing demand, supply, and transmission interconnection nodes) and 25 arcs. TS2 is a network with 26 nodes and 33 arcs. Finally, TS3 is a network with 34 nodes and 38 arcs. The number of candidate wind farms (supply nodes) in TS1, TS2, and TS3 are 12, 17, and 21 respectively.

Demand nodes in the networks represent five large metropolitan areas on the West coast. We used hourly load data obtained from FERC [9] from January 1, 2004 to December 31, 2004. Therefore, each test system contains 8,784 scenarios representing each hour in 2004.

To capture the correlations between candidate wind farm sites and area loads, we use coincidental wind speed data obtain from NREL's Western Wind Data Set [41] for the same time period. In total, seventy wind sites were randomly selected out of 32,043 candidate locations. For each of our test systems, we randomly select the desired number of candidate wind farm sites from our list of seventy. Wind speed data for these sites were simulated (by 3TIER) for turbines at 100 meters above ground level. The turbine power curve is based on the 3MW Vestas V-90. Readers interested in a more detailed description of the Western Wind Data Set methodology can refer to [42].

Interconnecting multiple wind farms to a common interconnection point and then

connecting that point to a far-away load center can allow the long-distance transmission portion of transmission capacity to be reduced [2]. To generate wind farm transmission networks for our test systems, we follow the network structure proposed by [2]. A simplified version of this network structure is illustrated in Figure 4.3.

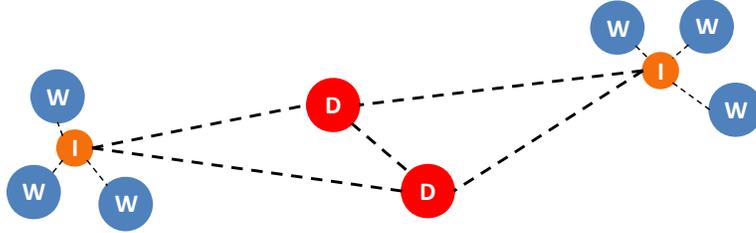


Figure 4.3: Interconnecting wind farms to a common interconnection point.

We begin by randomly selecting the desired number of wind farms (nodes “W”) and then grouping these candidate wind farms based on geographical proximity. For each wind farm within a group, we generate a candidate arc from that wind farm to an interconnection point (nodes “I”) located near the center of mass. These interconnection points are then connected to load centers (nodes “D”).

For the three test systems, we assume there are two types of additional generators: a 200MW conventional generator at a cost of \$100 million each and a 3MW wind turbine generator at a cost of \$3 million each. The RPS requirement Δ_{RPS} is set to be ten percent of the average multi-area load. The generation siting cost is assumed to be \$20 million for all sites. It should be noted, however, that siting costs will typically vary with the location and the size of the site.

It is assumed that each new transmission line has a capacity of 300MVA. Additionally, transmission siting costs and line costs are proportional to the distance between the pair of nodes to be connected. Specifically, the transmission siting cost between nodes i and j is given by $h_{ij} = (1.0 \times 10^6) \times d_{ij}$, and, similarly, the trans-

mission line cost is given by $c_{ij}^t = (5.0 \times 10^5) \times d_{ij}$. Transmission loss parameters are also proportional to distance. In our computational experiments, we assume a single line type with the following loss parameters: linear loss $\lambda_{ij}^t = (6.3 \times 10^{-5}) \times d_{ij}$ and quadratic loss $\mu_{ij}^t = (3.0 \times 10^{-6}) \times d_{ij}$.

Detail generation and transmission parameters for the three test systems can be found in Tables 4.2 to 4.7 in §4.7.

4.5.2 Computational Results

We solve the three test systems using two approaches: a standard application of Benders Decomposition (S-BD) and the A-BD framework proposed in Section 4.4. For each instance we restrict runtime to 24 hours and solve to within a 1% optimality gap. The results of these three test systems are summarized in Table 4.1.

Table 4.1: Algorithmic Performance

	TS1		TS2		TS3	
	S-BD	A-BD	S-BD	A-BD	S-BD	A-BD
Iteration	263	33	338	49	531	235
Time (min.)	600	86	806	118	1,440	728
Opt. Gap %	0.73	0.13	0.83	0.29	7.20	0.62

In Table 4.1, the first column describes the performance metrics. Each subsequent pair of column presents a comparison of S-BD to A-BD for a test system along three metrics: iteration count, solution time (in minutes), and optimality gap. The last row, labeled “Opt. Gap,” is the optimality gap (in percent) at the time of termination, which is computed as $((UB - LB) \times 100)/LB$.

The results in Table 4.1 clearly show that applying S-BD to WFND is not efficient. For all test systems, a large number of iterations and hence, a long runtime, is required for convergence. In Figures 4.4 to 4.6, we make a detailed comparison of the convergence properties of S-BD versus A-BD. These figures show the optimality gap (vertical axis) of S-BD and A-BD as a function of the number of Benders Iterations

(horizontal axis).

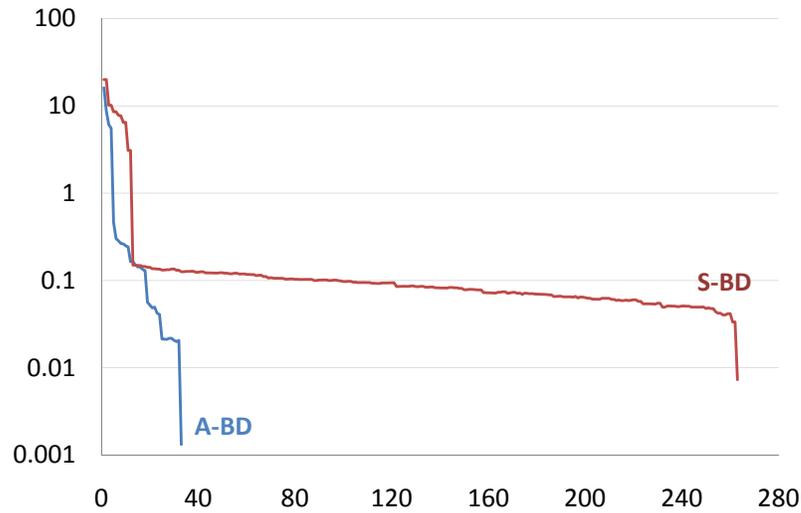


Figure 4.4: TS1 - Optimality gap as a function of iteration count.

In all three test systems, A-BD significantly outperforms S-BD in both solution quality and runtime. Although we were able to solve TS1 and TS2 using S-BD within the 24 hour runtime limit, convergence is slow and several hundred Benders Iterations were needed. This is in contrast to A-BD, which required less than 50 iterations for both of these instances and used only a small fraction of the runtime limit. For TS3, which corresponds to the largest network, the performance difference is even more substantial. As expected, the computational difficulty increases as the size of the network increases. Using S-BD, the entire 24 hour limit was exhausted, and after more than 500 Benders Iterations, an optimality gap of more than 7% still existed. In contrast, we were able to solve TS3 to within 1% optimality gap using A-BD in approximately 12 hours.

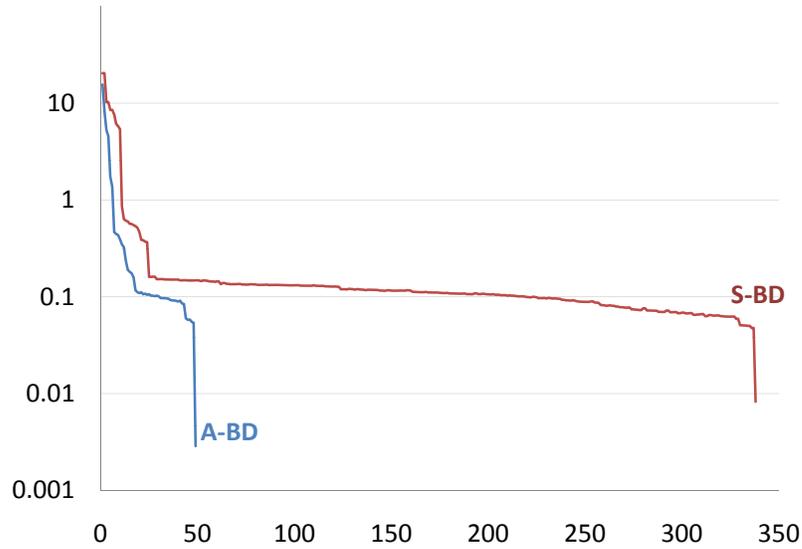


Figure 4.5: TS2 - Optimality gap as a function of iteration count.

4.6 Summary and Conclusions

We introduce a WFND problem that simultaneously considers transmission and generation expansion. Specifically, we focus on a power system expansion problem where wind power makes up a significant portion of new capacity. We begin by presenting WFND as a two-stage stochastic integer program. Next, we consider extensions of the problem to include considerations for transmission losses, which is critical to WFND because areas with good wind regimes may be far away from load centers. Because electricity generated from wind farms may be transmitted across great distance, it is critical that we model transmission losses to accurately represent “ i^2R ” loss behavior.

We present two loss models: linear and quadratic. Quadratic transmission losses are approximated using sets of piecewise-linear functions in order to maintain linearity in the second-stage. Next, we outline a solution framework based on BD and show that this can be applied to WFND with piecewise-linear loss functions.

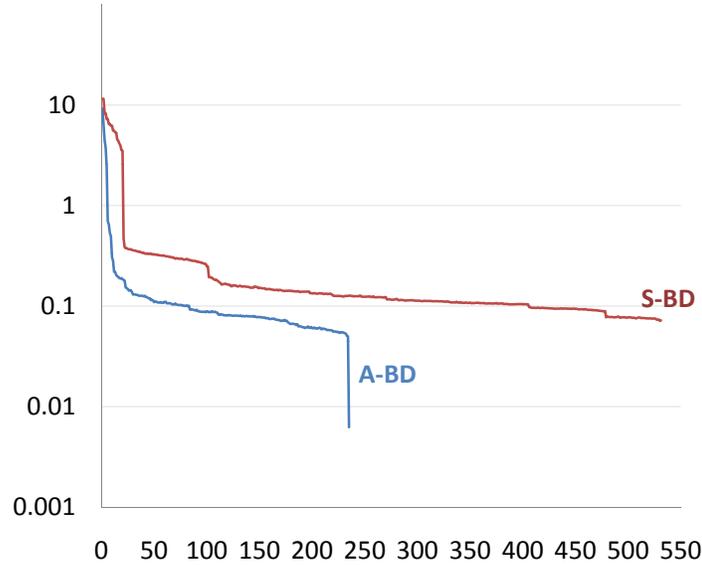


Figure 4.6: TS3 - Optimality gap as a function of iteration count.

Following this, we present procedures and model enhancements to accelerate the convergence of BD. We present two novel techniques for including the second-stage problem’s network structure to RMP via the addition of connectivity and demand fulfillment constraint sets. We also identify a class of valid equalities that can be added to RMP to produce a stronger continuous relaxation. Finally, we use an aggregation procedure to add multiple optimality cuts per Benders Iteration to balance the benefit of adding more second-stage information to RMP and maintaining the tractability of RMP.

We then examine the effectiveness of our methodology on a series of test systems to demonstrate the benefit of our proposed A-BD, versus executing a standard BD algorithm.

Several avenues for future research exist. One important extension would consider the addition of a probabilistic (chance) constraint to capture the loss-of-load-expectation (LOLE) (i.e. hard limits on the *expected* number of hours in which demand can exceed power output). This extension is important because regulatory

agencies may make meeting this type of probabilistic constraint a requirement in the future [31]. However, in the context of mathematical programming, stochastic wind speeds and the nonlinear, discontinuous power curve make closed-form probabilistic constraint formulation intractable. We are currently developing a new hybrid algorithm that leverages mathematical programming within the framework of a directed search. Another interesting avenue for future research would be to generalize the proposed model and algorithm for a broader class of *facility location problems* with uncertain capacity, such as those seen in disaster relief efforts. Finally, our proposed two-stage stochastic integer program can be readily extended to include other form of renewable energy (e.g. solar and geothermal).

4.7 Test System Parameters

Table 4.2: TS1 - Generation and Load Parameters

Area	Peak	Wind	Conv.	Conv.
i	Load (MW)	Max#	Exist. (MW)	Max#
1	3,793	0	4,000	4
2	12,774	0	13,000	13
3	4,024	0	4,000	4
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	400	0	0
8	0	400	0	0
9	0	400	0	0
10	0	400	0	0
11	0	400	0	0
12	0	400	0	0
13	0	400	0	0
14	0	400	0	0
15	0	400	0	0
16	0	400	0	0
17	0	400	0	0
18	0	400	0	0

Table 4.3: TS1 - Transmission Parameters

<i>i</i>	<i>j</i>	Cost (\$m)	Exist#	Max#
1	2	149	0	20
2	3	180	0	20
4	1	77	0	20
4	2	74	0	20
4	3	225	0	20
4	13	89	0	4
4	14	84	0	4
4	15	77	0	4
4	18	18	0	4
5	1	126	0	20
5	2	74	0	20
5	3	145	0	20
5	4	83	0	20
5	6	83	0	20
5	10	91	0	4
5	11	68	0	4
5	16	45	0	4
5	17	34	0	4
6	1	197	0	20
6	2	141	0	20
6	3	76	0	20
6	7	37	0	4
6	8	100	0	4
6	9	41	0	4
6	12	19	0	4

Table 4.4: TS2 - Generation and Load Parameters

Area	Peak	Wind	Conv.	Conv.
<i>i</i>	Load (MW)	Max#	Exist. (MW)	Max#
1	4,605	0	4,000	4
2	15,512	0	13,000	13
3	4,886	0	4,000	4
4	0	0	1,000	2
5	0	0	2,000	2
6	0	0	600	2
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	400	0	0
11	0	400	0	0
12	0	400	0	0
13	0	400	0	0
14	0	400	0	0
15	0	400	0	0
16	0	400	0	0
17	0	400	0	0
18	0	400	0	0
19	0	400	0	0
20	0	400	0	0
21	0	400	0	0
22	0	400	0	0
23	0	400	0	0
24	0	400	0	0
25	0	400	0	0
26	0	400	0	0

Table 4.5: TS2 - Transmission Parameters

i	j	Cost (\$m)	Exist#	Max#
1	2	149	0	25
1	6	33	3	2
2	3	180	0	25
2	5	17	7	2
3	4	7	4	2
7	1	77	0	25
7	2	74	0	25
7	3	225	0	25
7	16	89	0	4
7	17	84	0	4
7	19	77	0	4
7	25	18	0	4
7	26	75	0	4
8	1	126	0	25
8	2	74	0	25
8	3	145	0	25
8	7	83	0	25
8	9	83	0	25
8	13	91	0	4
8	14	68	0	4
8	20	45	0	4
8	21	34	0	4
8	22	128	0	4
8	23	86	0	4
9	1	197	0	25
9	2	141	0	25
9	3	76	0	25
9	10	37	0	4
9	11	100	0	4
9	12	41	0	4
9	15	19	0	4
9	18	59	0	4
9	24	116	0	4

Table 4.6: TS3 - Generation and Load Parameters

Area	Peak	Wind	Conv.	Conv.
i	Load (MW)	Max#	Exist. (MW)	Max#
1	2,832	0	2,000	1
2	3,793	0	3,000	2
3	12,774	0	11,000	6
4	2,846	0	2,400	2
5	4,024	0	3,200	2
6	0	0	1,200	2
7	0	0	1,800	2
8	0	0	5,000	5
9	0	0	2,000	2
10	0	0	1,200	2
11	0	0	0	0
12	0	0	0	0
13	0	0	0	0
14	0	400	0	0
15	0	400	0	0
16	0	400	0	0
17	0	400	0	0
18	0	400	0	0
19	0	400	0	0
20	0	400	0	0
21	0	400	0	0
22	0	400	0	0
23	0	400	0	0
24	0	400	0	0
25	0	400	0	0
26	0	400	0	0
27	0	400	0	0
28	0	400	0	0
29	0	400	0	0
30	0	400	0	0
31	0	400	0	0
32	0	400	0	0
33	0	400	0	0
34	0	400	0	0

Table 4.7: TS3 - Transmission Parameters

i	j	Cost (\$m)	Exist#	Max#
1	7	25	6	2
2	4	18	0	5
2	9	33	7	2
3	8	17	17	4
4	10	43	4	2
5	1	37	0	5
5	6	7	4	2
11	2	77	0	25
11	3	74	0	25
11	4	84	0	25
11	20	89	0	4
11	21	84	0	4
11	22	77	0	4
11	26	18	0	4
11	27	123	0	4
11	28	78	0	4
11	29	78	0	4
12	1	117	0	25
12	2	126	0	25
12	3	74	0	25
12	11	83	0	25
12	17	91	0	4
12	18	68	0	4
12	23	45	0	4
12	25	134	0	4
12	30	9	0	0
12	31	25	0	4
13	1	66	0	25
13	3	141	0	25
13	5	76	0	25
13	12	83	0	25
13	14	37	0	4
13	15	100	0	4
13	16	41	0	4
13	19	19	0	4
13	32	52	0	4
13	33	90	0	4
13	34	37	0	4

BIBLIOGRAPHY

- [1] Alguacil N., A.L. Motto, A.J. Conejo. 2003. Transmission Expansion Planning: A Mixed-Integer LP Approach. *IEEE Transactions on Power Systems*. **18** (3) 1070-1077.
- [2] Archer C., M. Jacobson. 2007. Supplying Baseload Power and Reducing Transmission Planning Requirements by Interconnecting Wind Farms. *Journal of Applied Meteorology and Climatology*. **46** 1701-1716.
- [3] Bahiense L., G. Oliveira, M. Pereira, S. Granville. 2001. A Mixed Integer Disjunctive Model for Transmission Network Expansion. *IEEE Transactions on Power Systems*. **16** (3) 560-565.
- [4] Binato S., M. Pereira, S. Granville. 2001. A New Benders Decomposition Approach to Solve Power Transmission Network Design Problems. *IEEE Transactions on Power Systems*. **16** (2) 235-240.
- [5] Birge J.R., F. V. Louveaux. 1988. A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*. **34** 84-392.
- [6] Bloom J. 1983. Solving an Electricity Generating Capacity Expansion Planning Problem by Generalized Bender's Decomposition. *Operations Research*. **31** (1) 84-100.
- [7] Meza J.L.C., M.B. Yildirim, A.S.M. Masud. 2007. A Model for the Multiperiod Multiobjective Power Generation Expansion Problem, *IEEE Transactions on Power Systems*. **22** (2) 871-878.
- [8] Crevier D. 1972. Approximate Transmission Network Models for Use in Analysis and Design. *MIT Energy Laboratory Report EL-73-008*.
- [9] <http://www.ferc.gov/docsfiling/eforms/form714/data.asp>
- [10] Firmo H.T., L.F.L. Legey. 2002. Generation Expansion Planning: An Iterative Genetic Algorithm Approach. *IEEE Transactions on Power Systems*. **17** (3) 901-906.
- [11] Giorsetto P., K. Utsurogi. 1983. Development of A New Procedure For Reliability Modeling of Wind Turbine Generators. *IEEE Transactions on Power Systems*. **PAS-102** (1) 134-143.
- [12] Hobbs B.F., Y. Ji. 1999. Stochastic Programming-Based Bounding of Expected Production Costs For Multiarea Electric Power Systems, *Operations Research*. **47** (6) 836-848.
- [13] Infanger G. 1993. Planning Under Uncertainty: Solving Large-Scale Stochastic Linear Programs, Danvers, MA: Boyd & Fraser.
- [14] Jirutitijaroen P., C. Singh. 2006. Multi-Area Generation Adequacy Planning Using Stochastic Programming. *IEEE Power Engineering Soc. Power Systems Conf. Expo*.
- [15] Jirutitijaroen P., C. Singh. 2006. Reliability and Cost Tradeoff in Multiarea Power System Generation Expansion Using Dynamic Programming and Global Decomposition, *IEEE Trans. Power Syst.* **21** (3) 1432-1441.

- [16] Jirutitijaroen P., C. Singh. 2008. Reliability Constrained Multi-Area Adequacy Planning Using Stochastic Programming With Sample Average Approximations, *IEEE Trans. PowerSyst.* **23** (2) 504-513.
- [17] Karki R., R. Billinton. 2004. Cost-Effective Wind Energy Utilization for Reliable Power Supply. *IEEE Transactions on Energy Conversion.* **19** (2) 435-440.
- [18] Khodr H.M., J. F. Gmez, L. Barnique, J. H. Vivas, P. Paiva, J. M. Yusta, A. J. Urdaneta. 2002. A Linear Programming Methodology for the Optimization of Electric Power-Generation Schemes. *IEEE Transactions on Power Systems.* **17** (3) 864-869.
- [19] Kirby B. 2008. valuating Transmission Costs and Wind Benefits in Texas: Examining the ERCOT CREZ Transmission Study. DOCKET NO. 33672. PUBLIC UTILITY COMMISSION OF TEXAS.
- [20] Linderoth J.T., A. Shapiro, S. J. Wright. 2006. The Empirical Behavior of Sampling Methods for Stochastic Programming, *Ann. Oper. Res.* **142** 219-245.
- [21] Malcolm S., S. Zenios. 1994. Robust Optimization for Power Systems Capacity Expansion Under Uncertainty. *The Journal of the Operational Research Society.* **45** (9) 1040-1049.
- [22] McCusker S.A., B.F. Hobbs. 2003. A Nested Benders Decomposition Approach to Locating Distributed Generation in a Multiarea Power System, *Networks & Spatial Economics.* **3** (2) 197-223.
- [23] Milligan M., R. Artig. 1999. Choosing wind power plant locations and sizes based on electric reliability measures using multiple year wind speed measurements. Natioal Renewable Energy Laboratory. **NREL/CP-500-26724.** 1-11.
- [24] Milligan M., T. Factor. 2000. Application of a Dynamic Fuzzy Search Algorithm to Determine Optimal Wind Plant Sizes and Locations in Iowa. *Wind Engineering.* **24** (4) 271-290.
- [25] Neelakanta P.S., M. H. Arsal. 1999. Integrated resource planning using segmentation method based dynamic programming. *IEEE Transactions on Power Systems.* **14** (1) 375-385.
- [26] National Renewable Energy Laboratory (NREL). 2009. Wind Deployment System (WinDS) Model. <http://www.nrel.gov/analysis/winds/>.
- [27] North American Electric Reliability Corporation. 2008. Long Term Reliability Assessment.
- [28] Oliveira E., I. Silva, J. Pereira, S. Carneiro. 2005. Transmission System Expansion Planning Using a Sigmoid Function to Handle Integer Investment Variables. *IEEE Transactions on Power Systems.* **20** (3) 1616-1621.
- [29] Park J.B., Y.M. Park, J.R. Won, K.Y. Lee. 2000. An Improved Genetic Algorithm for Generation Expansion Planning, *IEEE Transactions on Power Systems.* **15** (3) 916-922.
- [30] Park Y.M., J.R. Won, J.B. Park, D.G. Kim. 1999. Generation expansion planning based on an advanced evolutionary programming, *IEEE Transactions on Power Systems.* **14** (1) 299-305.
- [31] Pletka, R. 2008. Renewable Energy Transmission Initiative Phase 1A Final Report, prepared for RETI Coordinating Committee. Black and Veatch Report Number 149148.
- [32] Santoso, T., S. Ahmed, M. Goetschalckx, A. Shapiro. 2005. A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research* **167** 96-115.
- [33] Shrestha G.B., P.A.J. Fonseka. 2004. Congestion-driven Transmission Expansion in Competitive Power Markets. *IEEE Transactions on Power Systems.* **19** (3) 1658-1665.

- [34] Van Slyke R., R.J.B. Wets. 1969. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM J Appl Math.* **17** 638-663.
- [35] Storn R., K. Price. 1997. Differential evolution - A Simple and Efficient Heuristic for Global Optimization Over Continuous Space. *J. Global Optimization.* **11** 341-359.
- [36] Vallée F., J. Lobry, O. Deblecker. 2007. Impact of the Wind Geographical Correlation Level for Reliability Studies, *IEEE Transactions on Power Systems.* **22** (4) 2232-2239.
- [37] Verwij B., S. Ahmed, A.J. Kleywegt, G. Nemhauser, A. Shapiro. 2003. The Sample Average Approximation Method Applied To Stochastic Routing Problems: A Computational Study. *Comput. Optim. Appl.* **24** 289-333.
- [38] Wollmer R. 1980. Two-Stage Linear Programming Under Uncertainty with 0-1 Integer First-Stage Variables. *Math Programming.* **19** 279-288.
- [39] Zhu J., M. Chow. 1997. A Review of Emerging Techniques on Generation Expansion Planning. *IEEE Transactions on Power Systems.* **12** (4) 1722-1728.
- [40] U.S. Department of Energy. 2009. Energy Efficiency and Renewable Energy. http://apps1.eere.energy.gov/states/maps/renewable_portfolio_states.cfm.
- [41] <http://www.nrel.gov/wind/westernwind/>
- [42] <http://www.nrel.gov/wind/integrationdatasets/western/methodology.html>

CHAPTER V

An Iterative Test-and-Prune Algorithm for Wind Farm Network Design with a Probabilistic Reliability Constraint

5.1 Introduction

In Chapter IV, we investigated a variant of WFND problem that incorporates transmission losses and a loss-of-load-cost (LOLC) reliability measure. In this chapter, we investigate an important variant of WFND that incorporates a probabilistic constraint on loss-of-load-expectation (LOLE). Although little attention has been directed to the literature on LOLE-constrained generation expansion planning problems, this problem is extremely important because, in practice, regulatory requirements specify reliability as a probabilistic constraint [7]; that is, the expected amount of time that power systems meet demand must exceed a pre-defined threshold Δ_{rps} .

Probabilistically constrained problems are extremely challenging for conventional power systems. The difficulties are even greater for power systems that incorporate wind resources, because conversion from wind speed to power output is highly non-linear and discontinuous (Figure 4.2), and integrating a multivariate wind speed distribution within this power curve is virtually impossible.

5.2 Simplified Problem

In this section, we describe the challenges faced by using traditional mathematical programming (MP) formulations to solve a version of the LOLE-constrained WFND

problem. For the sake of exposition, we reduce the WFND model presented in Chapter IV to a system with decision variables governing only installed wind capacity and associated fixed siting costs. Additionally, we remove from consideration all conventional generation sources and assume infinite transmission capacity with no losses and a single demand site. Lastly, we remove the LOLC penalty from the objective function. These assumptions essentially have the effect of removing the second stage problem, as wind turbines have essentially zero (or nearly so) marginal operating cost, and therefore, the operating cost is zero everywhere. The siting variable z_i takes on the value 1 if wind turbines are installed at candidate wind farm i and 0 otherwise. The generation expansion variable x_i indicates the number of turbines to install at wind farm i . Then the simplified LOLE-constrained WFND problem is given by

$$\begin{aligned}
(5.1a) \quad & \min \quad \sum_{i \in \mathcal{N}} h_i z_i + \sum_{i \in \mathcal{N}} c_i x_i \\
(5.1b) \quad & \text{s.t.} \quad \sum_{i \in \mathcal{N}} \rho_i x_i \geq \Delta_{\text{rps}} \\
(5.1c) \quad & \sum_{i \in \mathcal{N}} \kappa_i^\xi x_i \geq d^\xi I^\xi \quad \forall \xi = 1, \dots, |\Xi| \\
(5.1d) \quad & \sum_{\xi=1}^{|\Xi|} I^\xi \geq \alpha |\Xi| \\
(5.1e) \quad & x_i \leq M_i z_i \quad \forall i \in \mathcal{N} \\
(5.1f) \quad & x_i \in \mathbb{Z}^+ \quad \forall i \in \mathcal{N} \\
(5.1g) \quad & z_i \in \{0, 1\} \quad \forall i \in \mathcal{N} \\
(5.1h) \quad & I^\xi \in \{0, 1\} \quad \forall \xi = 1, \dots, |\Xi|
\end{aligned}$$

The objective (5.1a) is to minimize the total siting and turbine installation cost. Constraint (5.1b) states that enough wind capacity must be installed to meet the

RPS requirement. Constraints (5.1c) state that, for each scenario $\xi = 1, \dots, |\Xi|$, the indicator variable I^ξ can take on the value one only if installed wind capacity exceeds demand for that scenario. Constraint (5.1d) ensures that, across all scenarios, the fraction of scenarios where installed wind capacity exceeds demand is greater than our predefined limit Δ_{rps} . Constraints (5.1e) restrict the number of turbines that can be installed at each site to at most the site limit M_i . Constraints (5.1f)-(5.1h) are variable integrality constraints.

5.2.1 Challenges of Mathematical Programming Approach

Unfortunately, even this simplified formulation of WFND is computationally challenging if it is solved using a MP-based approach. In this section, we highlight some of these challenges and provide motivation for a new meta-search algorithm, which we present in §5.3. The main challenges of LOLE-constrained WFND as described in (5.1) stem from two issues: weak linear programming relaxation and large sample size. We discuss each of these issues in turn.

Weak Linear Programming Relaxation

In a MP-based approach, variable integrality restrictions are typically handled by relaxing these integrality restrictions and solving the resulting linear program within a branch-and-bound framework. This approach works well in practice when problems have strong LP relaxations and little symmetry between variables. However, in LOLE-constrained WFND problems, there are incentives for all three sets of variables $(\mathbf{x}, \mathbf{z}, \mathbf{I})$ to be fractional.

The following simplified examples demonstrate why these three sets of variables have weak LP relaxations. Recall that I^ξ is an indicator variable that takes on the value 1 if the current wind network design \mathbf{x} is able to meet demand in scenario ξ and

0 otherwise. This indicator variable is enforced by the integrality constraints on I^ξ in conjunction with constraint (5.1c). In the LP relaxation of problem (5.1), the binary constraint on each variable I^ξ is replaced by a continuous constraint restricting I^ξ to be in $[0, 1]$. In this case, by setting $I^\xi = \sum_{i \in \mathcal{N}} p_i^\xi x_i$, we can get credit for “partially” meeting demand, for example, if $d^\xi = 100$ and $p_i^\xi = 0.5$ and we installed 100 wind turbines at location i and zero elsewhere. Then by setting $I^\xi = \sum_{i \in \mathcal{N}} p_i^\xi x_i$, a fractional solution is constructed where we get credit for meeting 50% of the total demand (that is, $I^\xi = 0.5$). However, meeting demand at 50 percent for two scenarios is not equivalent to meeting 100 percent of the demand fully in one scenario. \mathbf{z} are “Big-M” variables akin to fixed-costs in facility location problems. The weaknesses of these “Big-M” variables are well known and well studied ([5] and [6]). Lastly, there is also incentive for the \mathbf{x} variables to be fractional: that is, to install exactly the right number (even if it is fractional) of turbines to meet demand exactly.

The result of these difficulties is that a traditional MP approach to LOLE-constrained WFND problems would require a large runtime to converge to a high-quality solution and then an additional large runtime is needed in an effort to prove the quality of these solutions, by reducing the optimality gap via tightening the lower bound.

Large Sample Size

Since each additional sample yields a corresponding indicator variable (which has incentive to be fractional) and an associated “Big-M” constraint, the runtime grows exponentially with the increase in sample size. Additionally, when we branch on one of these I^ξ variables, there is very little impact on the remainder of the solution. In this problem, each indicator variable, I^ξ , is independent of all other indicator variables, unlike some MP formulations where variables are dependent and fixing one will automatically lead to integer variables for all associated variables.

5.2.2 Computational Experiments

In this section, we describe the computational experiments performed to assess the tractability of solving (5.1) using standard MP approaches. We developed a set of test systems using three-and-half years' worth of historical wind speed data at five candidate wind sites (in Minnesota) and coincidental loads. The number of turbines per site is restricted to a maximum of 150 and the LOLE-reliability requirement (Δ_{rps}) is set to 0.95. We solve six separate instances using varying levels of peak loads (from 1MW to 40MW). Per turbine capacity rating is assumed to be 3MW.

All computational experiments were conducted on a Sun SunFire x4600 server, using only a single AMD Opteron 8218 processor and 1.5 GB of RAM. We use the CPLEX 11.0 callable library to solve the MIP (5.1) and restrict runtime to a maximum of 24 hours (86,400 seconds).

We present the results of these computational experiments in Table 5.1

Table 5.1: MIP performance

Peak Load (MW)	MIP		
	time (sec.)	obj. val.	opt. gap %
1	86,400	73	24.9
5	86,400	256	34.1
10	86,400	481	40.3
15	86,400	712	42.7
20	86,400	919	39.3
25	86,400	1,141	34.3
40	86,400	1,926	17.2

Observe that none of the six instances can be solved to within the pre-set optimality gap of one percent in the allocated runtime. Furthermore, after solving each MIP instance for 24 hours, we are still left with an extremely large optimality gap (17% to 42%) due largely to the computational challenges described in §5.2.1. The results presented in Table 5.1 clearly indicate that a traditional MP-based approach is demonstrably not a viable way to solve LOLE-constrained WFND problems.

5.3 Test-and-Prune

In solving LOLE-constrained WFND problems, we observed that many of the computational challenges are tied to the use of a MP-based approach; specifically, these challenges are associated with the linearizations of the fixed-charge cost structure and the use of sampling to linearize the non-convex, discontinuous wind turbine power curve, embedded with a multivariate wind speed distribution function. Using an MP-based approach, the non-linear cost function requires linearization via a set of binary variables and associated “Big-M” constraints. Likewise, the probabilistic LOLE-constraint also leads to a large number of binary variables and associated “Big-M” constraints.

However, we also observed that the number of possible network designs, although it may be large, is finite. Additionally, when a network design vector \mathbf{x} is fixed, the resulting problem simplifies. Computing the cost of a network design can be done via a simple vector multiplication, and the feasibility of a network design can be computed in linear time through sampling.

Accordingly, we extend the use of a computational framework developed by [2] call Test-and-Prune (T&P). Instead of solving one difficult (and often intractable) optimization problem, we instead solve a series of simple feasibility problems within a larger search algorithm. The idea behind T&P originates from recent research ([2],[3]) on integrated resource allocation and utilization problems in automotive planning.

In T&P, we begin by enumerating all valid and viable network designs. Next, for each network design we test its feasibility, in this case simply by sampling through all possible scenarios $\xi = 1, \dots, |\Xi|$ and counting the number of scenarios where

demand is fully satisfied. Finally, we select the lowest-cost WFND solution.

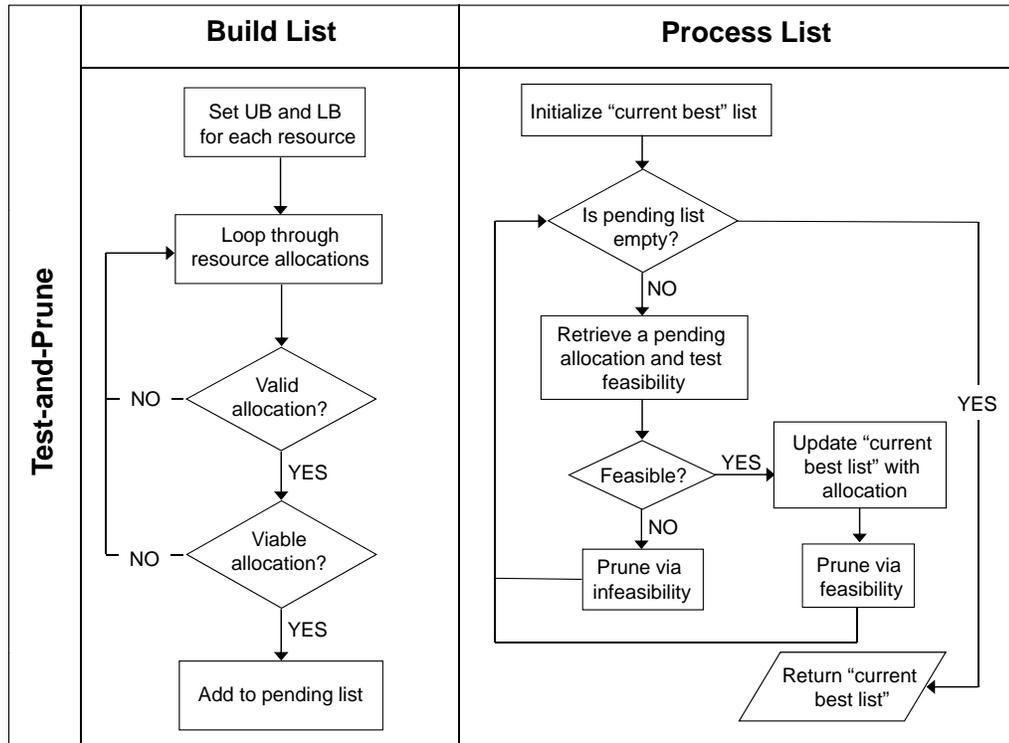


Figure 5.1: Test-and-Prune Flowchart.

T&P is a two-phase algorithm. In the first phase, Build List, we create a pending list of network designs based on the given lower- and upper- bounds on site capacities. Then in the second phase, Process List, we process this list until it is empty, at which time the incumbent solution is optimal. These two phases are illustrated in Figure 5.1 (adopted from [2]).

In the *Build List* phase, we create a pending list of candidate network designs. Using the given limits on upper and lower bounds for each variable \mathbf{x} , we check the validity; that is, we test to ensure that constraint (5.1b) is satisfied. If \mathbf{x} is not valid, we delete this network design. Otherwise, we test its viability, that is, we use bounds and other problem structure to check if \mathbf{x} is clearly infeasible or sub-optimal (with

respect to cost of the network design). If \mathbf{x} is both valid and viable, then it is a candidate network design and we add it to the pending list.

In the *Process List* phase, we begin by selecting a candidate network design \mathbf{x} for the pending list and test its feasibility; specifically, we test \mathbf{x} to see if it satisfies the LOLE constraint. If \mathbf{x} is a feasible solution, we prune the pending list and remove any network design that is within ε (the predefined optimality gap) of the cost of \mathbf{x} and then update the current best list with solution \mathbf{x} . On the contrary, if \mathbf{x} is infeasible, we prune from the pending list any allocation that is dominated by \mathbf{x} , that is, network design \mathbf{x} dominates network design $\hat{\mathbf{x}}$ if $\forall i \in \mathcal{N}, x_i \geq \hat{x}_i$. Since \mathbf{x} is infeasible in this case, any network design that has fewer resources at each site is clearly infeasible as well.

Once \mathbf{x} has been evaluated and the pending list updated, we can select another network design \mathbf{x} from the *reduced* pending list and repeat. The algorithm terminates when the pending list is empty; at that point the incumbent solution is optimal.

5.4 Iterative Test-and-Prune

T&P is appropriate for problems where, given a fixed resource allocation, it is trivial to compute the objective value of the allocation problem and it is easy to test the corresponding utilization problem. In LOLE-constrained WFND, problem 5.1 can be solved by simply enumerating all candidate network designs and testing the feasibility of each design. Then the provably optimal solution is simply the network design with the smallest cost.

However, for WFND problems of practical size, the number of candidate network designs may be too large to handle directly within T&P. For example, with five candidate wind sites, each with a maximum site limit of 200 turbines, there are over

300 billion possible network designs. Clearly, direct application of T&P, even with strong validity and viability tests, may not be sufficient, since the resulting pending list may be enormous.

In preliminary research, we observed high-quality solutions at a coarse granularity, for example, when we consider installing turbines in blocks of ten. In the previous example, with five candidate wind sites, each with a maximum site limit of 200 turbines, using blocks of ten turbines reduces the number of possible network designs from 300 billion to roughly 4 million possible network designs, a much more manageable number. Additionally, solutions obtained using a coarse granularity can be further improved by searching the neighborhood of the incumbent best solution at a finer granularity.

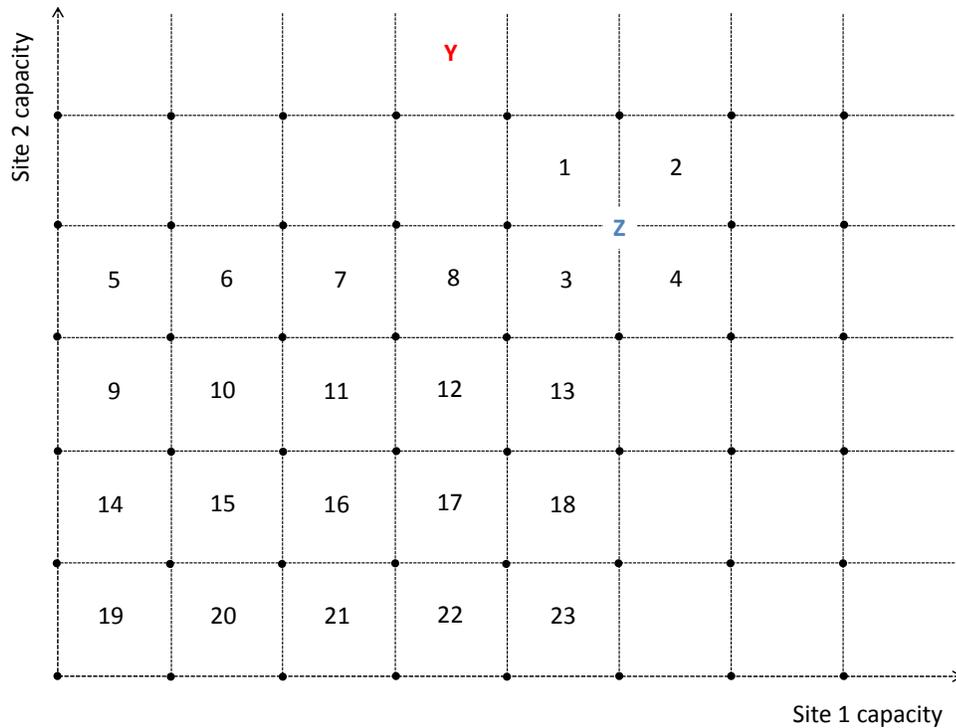


Figure 5.2: Solving T&P with coarse granularity.

However, a neighborhood search around the incumbent solution does not guarantee you will find an optimal solution. For example, consider the simple two-site problem illustrated in Figure 5.2. Suppose \mathbf{z} is the optimal solution obtained by solving the coarse granularity problem defined by the intersections of the grid lines. If \mathbf{z} is the optimal solution, then searching *all* points in the neighborhoods of blocks 1-4 at a granularity of 1 turbine, we will still miss the optimal solution at \mathbf{y} .

Additionally, observe that it is unnecessary to search block 2, or as a matter of fact, any block to the upper right of \mathbf{z} , since any network design in block 2 will have a higher cost. Analogously, we do not have to search blocks 5-23 since the upper right corner (the network design that dominates all other network design within that block) is infeasible. This is because if any network design represented by the upper right corner of blocks 5-23 is feasible, then the optimality of \mathbf{z} (at the current granularity) is contradicted.

Lastly, we make a final observation that the optimal solution at granularity 1 can be found only in one of the unnumbered blocks satisfying the two following “active” conditions:

Active condition 1: The upper right corner of the optimal block is feasible (and thus clearly cannot have a lower cost), and

Active condition 2: The lower left corner of the optimal block has a lower cost than the current best incumbent solution (and thus clearly cannot be feasible).

We refer to blocks that satisfy these two active conditions as “active” blocks. Using these active conditions, we can efficiently search the feasible region for the optimal solution by making recursive calls to T&P at varying granularity levels. This process is describe in Figure 5.3.

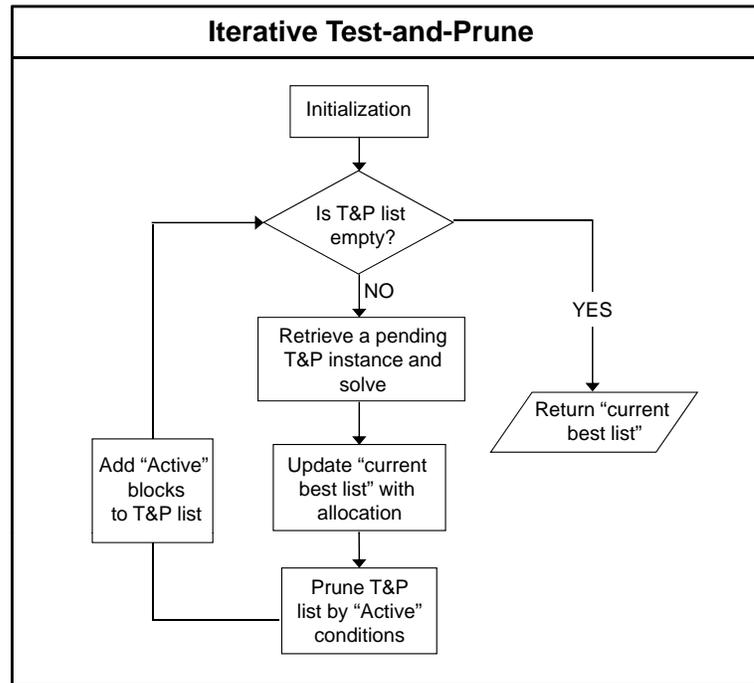


Figure 5.3: I-T&P FlowChart.

5.4.1 Computational Results

In this section, we re-solve the six test systems described in §5.2.2 using I-T&P (using granularities of 25, 5, and 1) and compare these results to those obtained by a traditional MP.

Table 5.2: Comparison of MIP performance to I-T&P performance

Peak Load (MW)	MIP			I-T&P		
	time (sec.)	obj. val.	opt. gap %	time (sec.)	obj. val.	opt. gap %
1	86,400	73	24.9	8	73	< 1
5	86,400	256	34.1	278	256	< 1
10	86,400	481	40.3	364	481	< 1
15	86,400	712	42.7	629	700	< 1
20	86,400	919	39.3	1,025	925	< 1
25	86,400	1,141	34.3	1,119	1,135	< 1
40	86,400	1,926	17.2	102	1,935	< 1

First, observe that using I-T&P all six test systems are solved within the predefined time limit of 24 hours: the fastest instance was solved in a mere 8 seconds and

the slowest instances was solved in less than 20 minutes. This solution time is in stark contrast to the solution times of a traditional MP approach, in which none of the instances is solved within the 24 hour time limit and in which, at termination, a large optimality gap of 17% to 42% remained. Comparing the solution quality of the MP approach to the I-T&P approach, we notice that for all six instances the objective values are comparable. However, using I-T&P, we are able to find provable optimal solutions in a small fraction of the runtime of the MP approach. While the solution qualities of these two approaches are comparable, it is important to note that if we had solved these instances as a standard MP only, we would be unable to know the true optimality gap. Additionally, if we use other test systems, there are no guarantees that the MP solutions would remain competitive with the much faster, and provably optimal, I-T&P solutions.

We conclude this section by summarizing the benefits of I-T&P.

- **explicit calculation of cost:** Since I-T&P calculates cost explicitly, there are no fractionally associated with fixed-costs;
- **explicit calculation of feasibility:** Since feasibility is not imbedded within a mathematical program, we can simply sample each scenario and count the number of scenarios where demand is fully satisfied (this has the added benefit that as sample size increases runtime increases linearly);
- **invariant to adding multiple time periods and stochasticity in conventional generation sources:** Both of these considerations can be handled within the sampling, therefore, no additional constraints are needed to express these considerations within the model.

5.5 Summary and Conclusion

In this chapter, we developed a computationally-efficient means to identify the optimal wind farm network configuration for LOLE-constrained power systems. We discuss the fundamental challenges that traditional mathematical programming (MP) approaches encounter in solving this problem and present computational results to demonstrate these challenges. We present a novel algorithmic framework, which we call *Iterative Test-and-Prune* (I-T&P), for solving discrete optimization problems with non-linear, non-convex, discontinuous, and stochastic constraints and objective functions. Although the motivation for I-T&P is LOLE-constrained WFND problems, this algorithm has broad applicability to other SNDF and resource allocation problems, for example, facility location problems with uncertain capacity such as those encountered in disaster relief efforts ([1], [4], and [8]).

Several avenues for future research exist. One important extension would consider the addition of transmission decisions, multiple demand points, and conventional generators. This is a natural extension since, given a network design, testing its feasibility must be expanded to solving a set of minimum cost flow problems, each of which will be easy to solve and well suited for parallel implementation. Another interesting avenue of future research would be to investigate the structural relationship between I-T&P and Benders Decomposition. As seen in Chapter IV, Benders seems well suited to the LOLC-penalty version of WFND problems but not to the LOLE-constrained version; the opposite is true for I-T&P. Thus, it would be interesting to explore the structural relationship between these two algorithms in the hope of developing a hybridization that can serve as the foundation for solving the real-world applications of WFND problems with *both* a LOLC penalty and a LOLE

constraint.

BIBLIOGRAPHY

- [1] Balcik, B., B. Beamon. 2008. Facility Location in Humanitarian Relief. *International Journal of Logistics: Research and Applications*. **11** (2) 101-121.
- [2] Barlatt, A., A. Cohn, C. Morford, O. Gusikhin, Y. Fradkin. 2007. A Hybridization of Mathematical Programming and Search Techniques for Integrated Operation and Workforce Planning. *Proceedings from the IEEE International Conference on Systems, Man, and Cybernetics*. 632-637.
- [3] Barlatt, A., A. Cohn, Y. Fradkin, O. Gusikhin, C. Morford. 2009. Using Composite Variable Modeling to Achieve Realism and Tractability in Production Planning: An Example from Automotive Stamping. *IIE Transactions*. **41** (5) 421-436.
- [4] Chang, M., Y. Tseng, J. Chen. 2007. A Scenario Planning Approach for the Flood Emergency Logistic Preparation Problem Under Uncertainty. *Transportation Research Part E*. **43** (6) 737-764.
- [5] Drezner, Z., H.W. Hamacher. 2004. *Facility Location: Applications and Theory* Springer-Verlag.
- [6] Melkote, S., M. Daskin. 2001. Capacitated facility location/network design problems. *European Journal of Operational Research*. **129** (3) 481-495.
- [7] Pletka, R. 2008. Renewable Energy Transmission Initiative Phase 1A Final Report, prepared for RETI Coordinating Committee. Black and Veatch Report Number 149148.
- [8] Yi, W., L. Özdamar. 2007. A Dynamic Logistics Coordination Model for Evacuation and Support in Disaster Response Activities. *European Journal of Operational Research*. **173** (3) 1177-1193.

CHAPTER VI

Conclusions

6.1 Conclusions

This dissertation discusses stochastic network design and flow problems in two application contexts: truckload procurement auctions and wind farm network design.

In truckload procurement our contributions are in:

- developing a new implicit bidding approach (IBA) for truckload procurement and other combinatorial auctions;
- developing models to solve a basic CTPA to optimality, in a single round, fully considering (implicitly) the exhaustive set of all possible bids;
- conducting numerical analysis on the characteristics of fully enumerated CTPA solutions;
- developing models and decomposition algorithms for fully-enumerated S-CTPAs, where carriers have uncertain repositioning capacities and costs;
- developing an efficient solution framework for fully enumerated S-CTPAs;
- generalizing the model and algorithmic approach presented for S-CTPAs to a broader class of SNDF problems.

In wind farm network design our contributions are in:

- presenting a new model for the design of wind farm networks in a multi-area power system;
- modeling an integrated generation and transmission expansion problem with explicit considerations for system uncertainties, fixed-siting costs and nonlinear transmission losses;
- introducing an accelerated decomposition algorithm that efficiently solves WFND problems with a large number of scenarios;
- presenting a new model for the design of wind farm networks that incorporates probabilistic constraints on LOLE;
- developing a hybrid algorithm, Iterative Test-and-Prune, for solving WFND problems with a LOLE constraint and demonstrating its efficacy via computational experiments.

In future work, we plan to extend our results for S-CTPAs to encompass uncertainty in the shipper's bid lane volumes. By further analyzing problem structures and parametric uncertainties of the recourse function of the S-CTPA problem, we may be able to reduce subproblem solution times (which is currently the bottleneck operation).

We are also interested in continuing our research on WFND problems. An interesting avenue would be to investigate the structural relationship between I-T&P and Benders Decomposition. As seen in Chapter IV, Benders seems well suited to the LOLC-penalty version of WFND problems but not to the LOLE-constrained version; the opposite is true for I-T&P. Thus, it would be interesting to explore the structural relationship between these two algorithms in the hope of developing a hybridization

that can serve as the foundation for solving WFND problems with *both* a LOLC penalty and a LOLE constraint.