STRATEGIC NETWORK GROWTH

WITH RECRUITMENT MODEL

A Dissertation
Presented to
The Academic Faculty

By

Wuthichai Wongthatsanekorn

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
May 2006

STRATEGIC NETWORK GROWTH

WITH RECRUITMENT MODEL

Approved by:

Dr. Jane C. Ammons, Co-Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Matthew J. Realff, Co-Advisor
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Ozlem Ergun
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Valerie Thomas
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Chelsea C. White III
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: April 10, 2006

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude towards my advisors, Dr. Jane Ammons and Dr. Matthew Realff for their support, patience, and guidance during my time at Georgia Institute of Technology. I would not have come this far without their novel ideas and guidance. During the time I have worked with them, I have learnt a lot from them in academics and everything else.

I am also thankful to my dissertation committee members – Dr. Ozlem Ergun, Dr. Valerie Thomas, and Dr. Chelsea White for their valuable comments. I am also grateful for the generous interaction and guidance provided from many industry experts, including Chuck Boelkins of P2AD and Nader Nejad of Molam. I thank Bob Donaghue for giving me the opportunity to work with P2AD for Georgia on the project to study state of Georgia's E-Scrap handling alternatives. I thank Dr. Christos Alexopoulos and Dr. Leon McGinnis for giving me an opportunity to work on the Supply Chain Game for the Virtual Factory Laboratory during the summer of 2002. I also appreciate the financial support I received from Dr. Faiz Al-khayyal during my second year in the program

I have been helped by many good friends during these years. Thanks to my office mates Tiravat Assavapokee (P'To), Josh Pas, I-Hsuan Ethan Hong, and Manu Sharma. We had a good time together in the office, CRC and tennis court. I would like to specifically thank P'To for being a generous mentor and big brother to me. I also thank my friends in Atlanta. Thank you P'Lek, P'Kong, P'Gnn, N'Moo, N'Vi, Winny, Chompoo, Nan, Kong, and Aung, I also thank all my friends who I met over many years in the U.S. for their support and friendship. Since I would surely miss at least one person if I tried to list them all, I choose to thank them all at once: Thank you. I wish everyone the very best. Also, I want to thank my girlfriend, Wananya (Eh), who always believes in me and encourages me throughout the years.

Finally, and most importantly, I thank my family. To my parents, I thank their support, patience and love. To the rest of my family (my four brothers), thank you for always being there.

Of course, there must necessarily be other people who deserve the gratitude, but whom I have not mentioned here. To them, please accept my sincere apology and thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In order to achieve stable and sustainable systems for recycling post-consumer goods, it is frequently necessary to concentrate the flows from many collection points to meet the volume requirements for the recycler. This motivates the importance of growing the collection network over time to both meet volume targets and keep costs to a minimum. Furthermore, this problem is complicated by the spatial variation of the density of resources and heterogeneous attitudes of agents towards participating in recycling activities. This research addresses a complex and interconnected set of strategic and tactical decisions that guide the growth of reverse supply chain networks over time. It contributes the first models and solution approaches that capture the uncertain, multi-stage aspects of the strategic reverse supply chain design problem.

This dissertation has two major components: a tactical recruitment model and a strategic investment model. These capture the two major decision levels for the system, the former for the regional collector who is responsible for recruiting material sources to the network, the latter for the processor who needs to allocate his scarce resources over time and to regions to enable the recruitment to be effective. First, I develop a recruitment model for the regional collector. A contribution of the model is conceptualizing the individual agent behavior as a Markov Process over various states of "willingness" to join the network. The agent's willingness state is changed by the recruitment actions and budget expenditures of the regional collector. The recruitment model is posed as a stochastic dynamic program. Three solution approaches are

developed to solve this problem in Chapter 4. A numerical study of the solution approaches is performed in Chapter 5.

The second component involves a key set of decisions on how to allocate resources effectively to grow the network to meet long term collection targets and collection cost constraints. The growth occurs not only in a temporal but also in a spatial dimension over regions. The recruitment problem appears as a sub-problem for the strategic model and this leads to a multi-time scale Markov decision problem, which is presented in Chapter 6. A heuristic approach which decomposes the strategic problem is proposed to solve realistically sized problems. Chapter 7 provides numerical valuation of the heuristic approach for small and realistically sized problems.

CHAPTER 1

INTRODUCTION

My dissertation addresses a complex set of decisions that surround the growth of reverse supply chain networks over time. It is motivated by an industrial example associated with a carpet recycling company recruiting carpet retailers to participate in a used carpet collection program coupled with establishing the rest of the necessary collection network infrastructure. A key set of decisions addresses how to recruit retailers effectively to meet long term collection targets and collection cost constraints. The percent participation of the retailers can be raised by aggressive recruitment, but there are limited financial and human resources that can be deployed in any given time period. Furthermore, the problem of network growth is complicated by the spatial variation of the density of used carpet resources and heterogeneous attitudes of retailers towards participating in recycling activities. Modeling this complex multi-stage decision-making situation requires careful design of interacting models that address specific aspects of the decision problem. The complexity of the problem raises the degree of difficulty to achieve the optimal solution for realistically sized problems for the processor. Hence, my dissertation aims to develop a methodology to support decision making for such problems in order to provide insight to the processor and to contribute to the general body of knowledge for supply chain design.

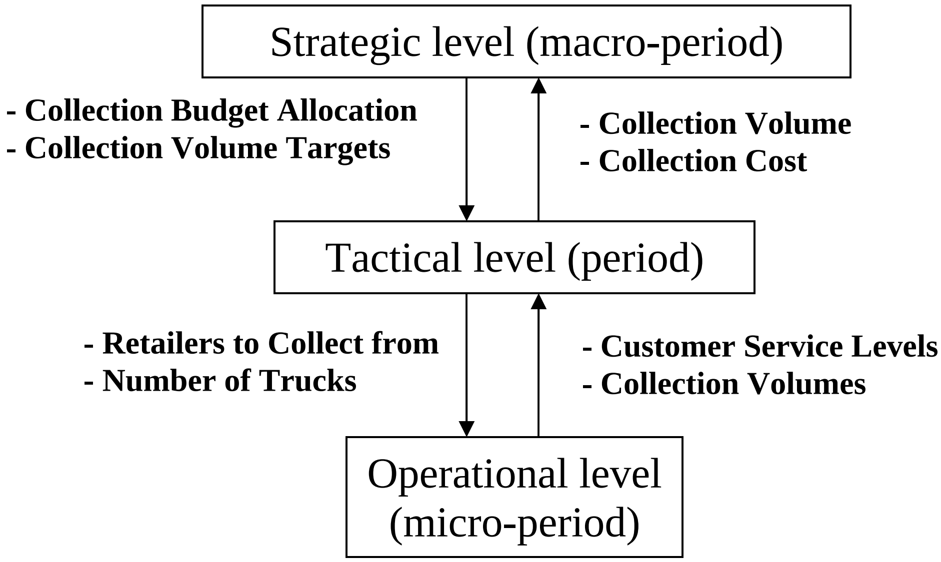

Figure 1.1 Structure of the Decision Levels

I use a spatial and temporal decomposition to model the network growth problem. The decompositions result in several models. Overarching is a processor strategic problem that allocates resources to grow the network over regions and within macro-periods. The use of resources over a finer time discretization, termed periods, and within each region is handled by a tactical decision problem by regional collector agents. Finally, the recruitment problem results in a network that must be serviced by trucks over smaller time increments, or micro-periods. Figure 1.1 illustrates how the decisions from the upper levels are passed onto the lower levels. In addition, it also shows how the results are sent back from the lower levels to the upper levels. In my dissertation, I develop models and methodology for each decision level. Simultaneously, an overall methodology to coordinate and capture decisions and passing information among levels is developed.

This chapter provides background for the general problem and approach. The relevant academic literature is reviewed in Chapter 2. In Chapter 3, a formal definition of the overall problem is provided. Chapter 4 develops the tactical level model and corresponding solution methodologies. Chapter 5 presents a numerical study for the tactical level problem. In Chapter 6, the strategic level model and solution methodology are developed. Chapter 7 presents a numerical study for strategic level problem. The summary, contributions and future extensions are given in Chapter 8.

In this chapter, first I provide the motivation of the problem based on industry examples. Then, I overview insights related in this dissertation in three areas: reverse production systems, hierarchical decision making and simulation-based optimization. These motivation and overviews provide a background for the proposed research.

1.1 Problem Motivation

According to Carpet and Rug Institute (2003), the U.S. carpet industry produced 1.833 billion square yards of carpeting in 1998 alone. These produced carpets, if sold, will be disposed of by approximately 2008 since the average carpet life is about 5-10 years. This presents a potential of over 8 billion pounds of carpets that need to be recycled or disposed of. Note that these figures are just for one year sales alone. Broadloom carpeting forms the majority of carpeting marketed in the U.S. Carpeting itself is composed of 49.2% face fiber, 38.4% latex and calcium carbonate, and the rest is often polypropylene. Face fibers are mostly nylon 6 or nylon 6,6 polymers.

The carpet industry has had limited success with recycling nylon carpets. Post consumer carpet scrap is priced at approximately $0.06a pound[1] for truck load quantity (40,000 pounds or more) while nylon 6 (after processing nylon-6 scrap) is priced at $1.59[2] for truck load quantity. Hence, the nylon fibers produced in 1998 alone are worth a potential one billion U.S. dollars. It is this economic potential in recovering used nylon carpeting and avoiding landfill tipping fees that drives several of the current efforts in carpeting recycling. However, according to the data from Carpet America Recovery Effort's Annual Report in 2004, 4,000 million pounds of used carpet is discarded to the landfill in year 2003 while less than 100 million pounds is recycled.

With the effort of companies belonging to the Carpet and Rug Institute to recycle more carpet, Carpet and Rug Institute (2003) showed how the carpet industry has reduced its environmental footprint over the years. The amount of landfill use, carbon dioxide emissions, energy consumption, water usage and hazardous air pollutants have been reduced by 80% over the past 12 years from 1990. It is crucial to recognize these reductions occurred while production increased by 47% over the same period. Their goal by 2012 is to reduce the environmental footprint an additional 28 percent.

However, in the past decade, two major carpet recycling companies suffered major financial problem that led to the closure of the recycling plants. The first company is Evergreen Nylon Recycling, a joint venture of DSM and AlliedSignal, who built a $100-million plant in Augusta, Georgia. The plant began operation in November 1999 and

---

[1] Canada's Waste Recycling Marketplace (2006)

[2] IDES The Plastics Web (2006)

could chemically reclaim nearly 200 million pound a year of nylon 6 carpet waste (Plastics Technology 2002). In addition to eliminating significant carpet waste, the renewal of nylon 6 will save 4.4 trillion BTUs of energy annually—enough power to heat over 100,000 homes each year (International Fiber Journal 1999). It also was designed to produce 100 million pounds of virgin-quality nylon 6 per year, reduce the dependence on foreign oil, and create new jobs throughout the recycling industry.

However, the company was shut down in 2001, less than two years after it started up (Atlanta Business Chronicle 2001). Higher than expected production costs combined with current business and economic conditions for caprolactam[3] in general led to the shut down. The closure of this facility scuttled numerous carpet recycling operations across the country and left Polyamid 2000 (PA 2000), a German corporation, as the only viable, large scale recycler of post-consumer nylon carpet fiber.

The second company was Polyamid 2000, located in Premnitz, Germany, who set up their plant in 1999 to reclaim post-consumer carpet material from all over Europe using chemical depolymerization of nylon 6. The huge PA 2000 installation cost around $200 million. In 2002, PA 2000 produced about 15 million pound a year of nylon 6 but ran at about 75% of capacity (Plastics Technology 2002). The plant had expected to process roughly 264 million pound a year of carpet into 20 million pound a year of nylon 6. The underlying problem is that the content of nylon in European waste carpets was less than expected. Recyclers had expected 30% of European carpet to be nylon 6 but they only found 20% of nylon 6. PA 2000 then turned to the used carpet markets in U.S.,

[3] Caprolactam is a building block of all Nylon 6 products. It is obtained by depolymerizing the nylon content of the used carpet.

mainly in California. Furthermore, the recyclers faced the problem of low landfill tipping fees in the US. An increase in transportation cost of importing the material from U.S. drove PA 2000 to declare bankruptcy in July 2003 (Carpet America Recovery Effort 2004).

The failure of these two companies raises an important question to the recyclers or processors: what went wrong? One explanation is the high cost of the supply material. In theory, it is not difficult to provide either the Evergreen or Polyamide 2000 plants with the volumes of used carpet they required – which were a small fraction of all used carpet. What was challenging was to provide the volumes at low enough cost to make them viable – witness the incredible lengths Polyamide 2000 went to by importing used carpet from the US, and specifically California!

Therefore the careful planning of the collection network to supply the capital intensive processing plant can be a critical factor in the success or failure of recycling operations. However, the processor is faced with a significant challenge. Processors are typically not familiar with the waste business; the collection of "trash" is not a core competency of their organization, nor do they have existing waste hauling contracts that they can exploit to get the material. This leads to the need to recruit a layer of the supply to the system. In the case of carpet this might be the retailers who sell carpet, as they are the ones to whom the carpet is typically returned by the installers. One solution to this recruitment problem is to subcontract the responsibility of recruiting to a local regional collector. This collector is then allocated a budget with which to recruit the retailers to the network. This budget may be financial incentives in this case. Alternatively, the processor may decide to do the recruitment and collection itself, and in this case, the budget will most likely reflect the amount of time and personnel resources that are devoted to a particular region.

A further issue that is faced by the processor is the need to build the network to meet a significant demand that ramps very rapidly in volume in comparison to the volume collected from any one source. Consumer products that are correlated with population imply numerous collection agents to provide adequate material to support the economy of scale required at the processing site. Thus, it is necessary to concentrate the flows from many collection points to meet the volume requirements of the processor. In the carpet recycling there is a need to collect from many retailer locations to feed one separation and material recovery facility.

The challenging question of how to allocate one's effort in building a collection network over time and in many regions is addressed in this research. I model the decisions required to build a collection network to meet the target demand while minimizing the cost in the long run. The resulting framework can greatly benefit a company who may want to reopen the high-cost processing plant, make them financially viable, and stay in business longer. Consequently, more carpet waste will be kept out of the landfill and the recycling rate will be increased. This also helps Carpet and Rug Institute achieve their recycling rate target in 2012, which is a key sustainability goal for the whole industry.

1.2 Overview of Reverse Production Systems

A reverse production system (RPS) is a network of transportation logistics and processing functions that collect, refurbish, and demanufacture for reuse and/or recycle used or end-of-life e-scrap products. Realff et al. (1999) illustrated the material flows in forward and reverse production system as shown in Figure 1.2. An initial mathematical model (mixed integer linear programming) was developed to plan a reverse production system for carpet recycling by Newton (2000). Assavapokee (2004) then proposed an extension of this to represent reverse logistic infrastructure for both single period and multiple periods timeframes applied to the electronics waste stream. The model has the ability to make the strategic decisions for the location of collecting centers and processing centers, the type of collected materials allowed at each collecting center, the type of processes installed at each processing center, and number of pounds collected, processed and transported throughout the reverse logistics system. Both of these models assumed that the targeted collection amounts would appear at the collection centers without additional effort on the part of the system.

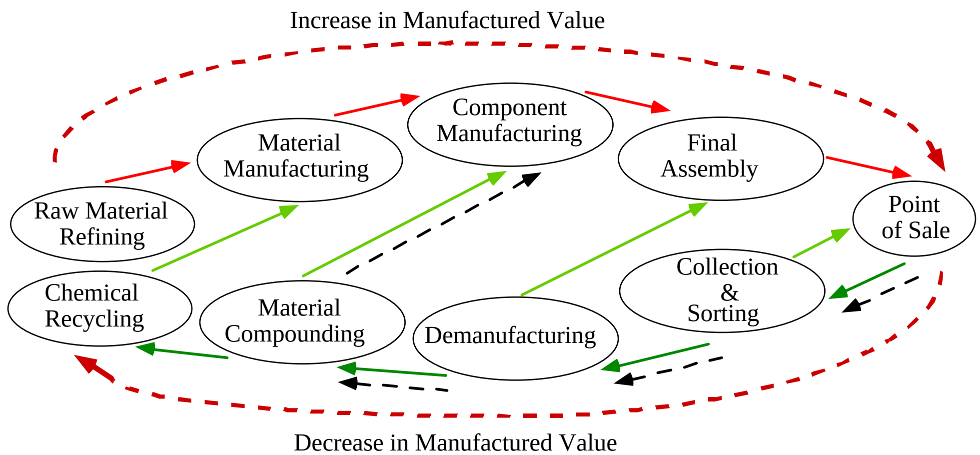

Figure 1.2: Material Flows in Forward and Reverse Production Systems (from Realff et al. 1999)

1.3 Overview of Hierarchical Decision Making

In the collection network growth problem, decisions are made at different hierarchical levels with interaction and feedback between the levels. In the literature, hierarchical decision making problems arises in many applications. One of them is the production planning problem (Bitran et al. 1986, Sethi and Zhang 1994). The general description of the production planning problem is the following. There are two levels: the "marketing management" level and "operational" level. At the marketing management level, the decision maker needs to control which family of products to produce over each decision epoch, where a family is a set of items consuming the same resources and sharing the same setup. At the operational level, the decision maker needs to determine actual production quantities of the items in the family given stochastic demands for the items, production capacity, holding cost, etc. The objective is to develop a two-level production plan to maximize the net revenue of the manufacturing system.

Sethi et al. (1995) further investigate hierarchical decisions by including capacity expansion in the marketing management level. A firm must satisfy a given increasing demand for its product over time to minimize its total discounted costs of investment in new capacity and production, and inventory/shortage over a finite horizon. The firm has a number of existing machines which are failure-prone with given rates of breakdown and repair. As demand increases, the firm has an option to purchase a number of new machines, identical to the existing machines, at a fixed given cost for each machine. In this problem, strategic planning from the management level bases its capacity expansion decisions on some aggregated, rather than detailed, information from the operational level. Subsequently, the operational level makes production planning decision given the capacity decisions made at the upper level.

Complexity of hierarchical decision making also comes from the uncertainty of future outcomes in the operational level and the decisions at each level. Thus, the decision maker at the higher level is confronted with difficult decisions in specifying a multi-period plan that optimizes the system's objective. With these complexities, the hierarchical decision making approach requires careful design to account for the interactions of each level.

1.4 Overview of Control Optimization Area

For control optimization problems, a familiar control objective is to guide a system from an initial state so that the output optimizes a predefined performance measure of the system. Specialized solution methods exist for optimal control problems involving linear systems and quadratic cost functions, and methods based on the calculus of variations can yield close-form solution for restricted classes of problems. A well-known general framework is the dynamic programming (DP) method (Bertsekas 1987, Bellman 1957, Howard 1960, Puterman 1994, and Karin 1955).

From the perspective of control optimization, DP usually is confronted by curses of modeling and dimensionality. A very small problem may contain one-thousand states but one may have to store one million transition probabilities just for one action (Bertsekas 1995). In recent years, simulation-based optimization has been a rapidly evolving subdiscipline in the simulation and optimization research areas to solve control optimization problems. Engineers and scientists have always wanted the ability to optimize the systems using simulation models. Computational capacity has increased tremendously over the years, this has served as one of the driving forces for research in this area along with fruitful interactions between the disciplines of operations research, artificial intelligence and statistics that have led to new algorithmic insights into how to solve large problems approximately.

Modeling a complex system using declarative mathematical programming constructs is difficult when significant portions of system behavior are procedural in nature. In addition, mathematical programming models typically scale badly as the problem size increases, both in terms of the length of the declaration (number of constraints and variables) and in required solution time. This is a particular problem when the application requires a plan that stretches across multiple periods and which must consider a number of different possible scenarios driven by the uncertainty in values of the underlying problem data. This leads to a combinatorial explosion in the variables and constraints required to represent the problem. It creates awkward representational choices in capturing the relationships between the scenarios.

Some real world applications do not have the luxury of waiting for optimal results and the need to evaluate alternative models and data streams to gain insight is lost if each computation is very time consuming. The tradeoff is often between solving an approximation exactly, or a more exact representation approximately. A particularly common approximation is to solve the problem assuming there is no uncertainty in the model or data. This assumption dramatically simplifies the problem statement and solution methods, but uncertainty is a significant factor in many strategic system design problems. A faster, approximate, approach might be a more effective decision support approach for these applications. However, the decrease in the solution quality needs to be balanced against the gain in computational efficiency.

One alternative to declarative mathematical programming is a simulation based optimization approach. Simulation can act as a function evaluation in complex systems. Since the objective function is not required to be in a closed form, as well as naturally imposing the system constraints, it can capture more realistic situations without the representational burden of declarative formulations. A simulation model can include some crucial random events such as allowing breakdowns of machines or rush orders from customers. The solution obtained from running the simulation for many replications is a sample mean. By the strong law of large numbers, as the number of replications increases, the sample mean will converge to the true mean. However, depending on the specification of the problem, the run time of the simulation might be so extensive that using simulation also can be computationally expensive. This is a significant drawback when a solution with good estimates of higher order moments of the performance distribution is required. The overriding problem is that simulation may not select specific sample paths that have particularly bad implications for certain decisions, and hence estimation and probabilistic guarantees replace concrete proofs of validity or worse case performance.

The main reason a simulation-based optimization approach is used in this research is its ability to model complex systems close to the real world processes without having to make oversimplifying assumptions. In particular, the decisions have to be made in the face of uncertainty and over multiple time periods. In my research, the uncertainty comes from the recruitment outcome of the agents to the recycling network.

The objective of my dissertation is to develop a new approach for supporting decisions for supply chain infrastructure design and growth over time for a realistic size problem. This approach composes of three levels of decisions: strategic decisions, recruitment decisions and servicing decisions. The methodology to obtain the decisions in the first two levels is developed in this dissertation. These components of the approach have been prototyped and tested using small and large examples.

This dissertation provides a better understanding of the collection network growth over time. At the tactical level, the model helps the regional collectors to understand the recruitment activity better and provides a methodology to assist them in reaching the target goal with high confidence. This is done by introducing a new model to capture the heterogeneous attitudes of the retailers towards recruitment. At the strategic level, a modified multi-time scale Markov Decision Process model (Chang et al. 2003) is developed. Furthermore, the methodology for solving the model yields a solution that can have a significant impact for the processor. The processor can incorporate the expected amount of collection volume and cost into his or her long term and short term investment decision making in order to be financially viable. However, the number of available retailers for the processor is so large that this makes the collection network growth problem large-scale and difficult to solve. Hence, this dissertation aims at proposing the solution approach that can solve a realistic sized problem rather than an exact method that is limited to a small problem.

In the next chapter, I provide a review of relevant literature for this dissertation. This supporting literature can be classified into three main areas: collection within reverse production system, hierarchical decision making and control optimization.

CHAPTER 2

LITERATURE REVIEW

This chapter reviews the literature and background related to this research. Overviews are presented of the following three main areas: Collection within Reverse Production Systems, Hierarchical Decision Making, and Control Optimization. These are covered in sections 2.1, 2.2, and 2.3 respectively.

2.1 Literature Review of Collection within Reverse Production Systems

Flapper (1995, 1996) and Fleishmann et al. (1997, 2000) give systematic overviews of the logistic components of reuse and recycling. In the reverse supply chain literature, special attention has been paid to the design of RPS infrastructure (Barros et al. 1998, Krikkee 1998, Realff et al. 2004, Pochampally and Gupta 2003). The design specifies where to locate the collection and processing sites in order to benefit the entire system's objective. Within the framework of reverse production infrastructure design, Fleischmann et al. (2000) characterize the main activities in the product recovery networks as collection, selection, re-processing, disposal, and re-distribution.

This research focuses on building a collection network over time instead of the entire and complete RPS design by concentrating on the collection, sorting and delivery to a given end use with a specific target capacity. The importance of collection system is addressed by Biehl et al. (2005) and Fleischmann (2000). In the context of RPS, collection includes all activities rendering used products available and physically moving them to some point where further treatment is accomplished. Collection of used carpet from carpet dealerships (Realff et al. 1999) and take-back of used copiers from customers (Krikke 1998) are typical examples from the above case studies.

Fleishmann et al. (1997) describe reverse distribution as the collection and transportation of used products and packages. Reverse distribution can take place through the original forward channel, through a separate reverse channel (Caruso et al. 1993, Kroon and Vrijens 1995, Barros 1998, Spengler et al. 1997), or through combinations of the forward and reverse channels (Salomon et al. 1996, Del Castillo and Cochran 1996). Guiltinan and Nwokoye (1975) provide one of the first analyses of reverse distribution networks. Figure 2.1 shows a framework for reverse distribution combining the forward flow from producer to user, and the reverse flow from user to producer. This research specifically examines the process of transporting the materials from the consumers to the collectors and from the collectors to the recycler. This portion of the supply chain is highlighted within the solid rectangular box in Figure 2.1.

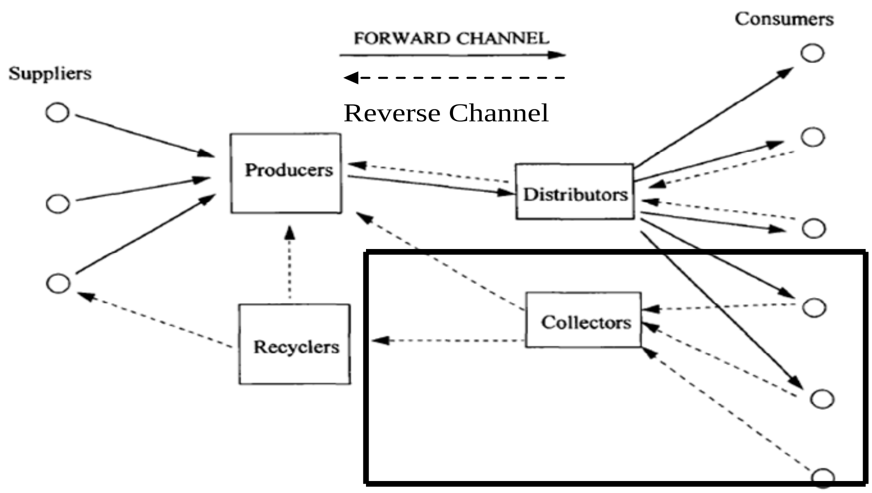

Figure 2.1: Framework Reverse Distribution (from Fleischmann et al. 1997)

In general, the decisions for collection system design include what kind of products to collect, which processing options to choose for incoming products, where to locate the collection and consolidation points (Jayaraman 1996, Solomon et al. 1994, 1996, Spengler et al. 1997), and how to batch the collected end-of-life products. In this research, the decisions do not focus on determining where to locate the collection and consolidation points for a given product and processing option. These decisions presuppose that the demand for collection is known and that the minimization of transportation cost at the local level is the key driving force. The problem addressed in this thesis is to strategically design the growth of the collection system used in the reverse production system over time and under uncertainty. There is considerable freedom in which subset of sources is used to provide the material, but the behavior of these sources is not under the direct control of the planner. The sources must be recruited to the network.

Several authors have proposed modifications of traditional facility location models (Mirchandani and Francis 1989, Flapper et al. 1996, Louwers et al. 1999) for the design of reverse distribution networks. In addition, Berman et al. 2001 and Berman and Huang 2004 examine the collection problem (such as garbage collection) as a collection depots location problem. One special characteristic to be taken in to account is the convergent structure of the network from many sources to a few demand points or collection points (Ginter and Starling 1978). Such ‘many to few’ problems have also been studied in the hazardous waste disposal literature (e.g. Batta and Chiu 1988 and Erkut 1996). By contrast, traditional location models typically consider a divergent network structure from few sources or distributors to many demand points.

Building collection networks has also been studied as part of a solid waste management problem: how to effectively allocate recycling drop-off stations, of appropriate size; and how to design efficient collection-vehicle routing and scheduling in the solid waste network. Various types of mathematical programming models have been developed for this problem. The methodologies include linear programming (Hsieh and Ho 1993, Lund and Tchobanoglous 1994; Huang et al. 1992, 1993), mixed integer programming (Anderson 1968, Marks et al. 1970, Chang et al. 1993, Chang and Wang 1995), and dynamic programming (Baetz 1990, Huang et. al. 1994). Some researchers study this problem from the multi-objective programming standpoint (Caruso et al. 1993, Chang and Wei 1999).

However, for the solid waste management problem, the focus is the operational level (designing the collection route) and the tactical level (designing drop-off locations). In this research, the focus is on the strategic and tactical levels with different objectives at each level. In addition, most collection problems, such as the solid waste collection problem, assume that there are many collection points of sources and each source has different generation rates. There is also a constraint that all collection generated must be collected for process or disposal (Huang et al. 1995, Chang et al. 1997, Chang and Wei 1999, Realff et al. 2004). However, it does not address how to physically or commercially grow the network of source of collection to reach a long term target in order to feed a recycling facility.

Another particular feature of reverse distribution networks is their high degree of uncertainty in supply both in terms of quantity and quality of used products returned by the consumers. Both are important determinants for a suitable network structure since, for example, high quality products may justify higher transportation costs (and thus a more centralized network structure), whereas extensive transportation of low value products is uneconomical.

Under solid waste management planning, conventional mathematical programming approaches for dealing with uncertainties may be classified by the following three methods: (1) stochastic programming approaches, (2) fuzzy integer programming, and (3) scenario analysis. Stochastic programming approaches can effectively deal with various probabilistic uncertainties in decision making and are particularly useful when the values of system components vary but their probabilistic descriptions are known (Yudi and Tsoy 1974, Glover 1976, Kunsch 1990). However, the increased data requirements for specifying the parameters' probability distributions affect the method's applicability. The stochastic programming approach may lead to large or complicated models that are difficult to solve in practical applications (Rockafellar and Wets 1991).

In the past decade, fuzzy sets theory and interval programming technique have received wide attention in the field of planning solid waste management systems. For example, Koo et al. (1991) accomplished the site planning of a regional hazardous waste treatment center in Korea using a fuzzy multiobjective programming algorithm. Huang et al. (1992, 1993, 1994) developed a grey linear programming, grey fuzzy linear programming, and grey fuzzy dynamic programming approaches to deal with a hypothetical solid waste management in Canada. Chang et al. (1995, 1996) apply fuzzy goal programming in dealing with several issues in the integrated solid waste management in Taiwan.

In scenario analysis (Rockafellar and Wets 1991, Assavapokee et al. 2005), the uncertainty of system components is modeled by a small number of subproblems derived from the underlying optimization, which corresponds to different scenarios. The model of the collection system in this research uses a stochastic dynamic programming model to address uncertainty in the collection system.

In addition to focusing on building a collection network, this research implements a recruitment model for collection points within the collection system. Recruitment models in the literature focus on employment recruitment, human resource management, and physiological models in medical research (Darmon 2003, Treven 2006, Hawkins 1992, Georgiou and Tsantas (2002). Mehlmann (1980) use a recruitment concept for a long-term manpower planning problem. Coughlan and Grayson (1998) examine the problem where the individual distributors play two key roles in network marketing organizations (e.g. Amway, Mary Kay and NuSkin): they sell product, and they recruit new distributors. They develop a model of network marketing organization network growth that shows how compensation and other network characteristics affect growth and profitability of the distributor. In their context, one distributor recruits others by socially interacting with them in one form or another. They represented this process by adapting a diffusion model formulation to the recruitment process (Bass 1969). This model allows for network growth via both inherent attraction (the innovation effect) and the spread of word-of-mouth (the imitation effect). They introduce a recruitment function which includes innovation and imitation terms. This thesis includes the notion of the recruitment of the collectors, instead of the distributors. Furthermore, the recruitment process is represented in a more complex form, not just a closed-form function.

Overall, this research examines at the reverse distribution problem with a focus on the collection component and features the issue of growing the collection network. The strategic and tactical decisions are key, rather than the operational decisions. The decisions required for the strategic and tactical levels are different, and made under differing time scale and uncertainty. The next section provides an overview of how hierarchical decision making has been approached in the published literature.

2.2 Literature Review of Hierarchical Decision Making

The collection network growth problem in this research is to integrate overall plans to establish and grow infrastructure, and to determine effective tactics to operate the infrastructure. The overall plans include 1) building the infrastructure, 2) determining the recruitment process to grow the network over time in order to achieve the plan, and 3) servicing the given network. However, these plans are decided at different hierarchical decision levels and different time scales.

The hierarchical decision making problem arises in many applications such as semiconductor fabrication (Panigrahi and Bhatnagar 2004, 2006), production planning (Bitran et al. 1986), shop floor control (Qiu and Joshi 1999) and network traffic applications (Tuan and Park 1999). Decision making in collection network growth incorporates hierarchical decision making in the sense that the decisions for each level have impacts on lower levels and decisions at lower levels feedback to decisions at the higher level for subsequent decision periods.

The Markov Decision Process (MDP) (Chang et al. 2003, Panigrahi and Bhatnagar 2004, 2006) has been introduced to model the hierarchical decision making problem. Chang et al. (2003) propose a model called the Multi-time scale Markov Decision Process (MMDP) for hierarchically structured sequential decision making processes. The decisions in each hierarchical level are made at different discrete time-scales. They also present an exact MMDP solution approach and study some approximation methods, along with an heuristic sampling-based scheme. In addition, McDonnel et al. (2004) examine shop floor control systems and model them using game theory, as set-up games. They also propose a heuristic approach to solve the set-up game model.

Panigrahi and Bhatnagar (2004) and Chang et al. (2003) propose a simulation-based approach, Q-learning, to solve the MMDP problem. Panigrahi and Bhatnagar (2006) and Borkar (2005) consider the same problem and develop a simulation based two-timescale actor-critic algorithm in a general framework. Hauskrecht et al. (1998) introduce the concept of macro actions and propose the solution method via these macro actions.

In this research, a hierarchical decision making model is introduced based on the MMDP model (Chang et al. 2003) but uses different assumptions on the form of the interactions between the hierarchical levels. A heuristic method is also proposed to solve the multi-level decision making problem. The heuristic employs methodologies from dynamic programming and simulation-based optimization combined with mathematical programming and model predictive control. In next section, I provide the basic concept and literature review of these areas.

2.3 Literature Review of Control Optimization

The objective of control optimization is to determine a set of actions to be taken in the different system states in order to optimize a predefined system performance measure. In general, mathematical programming methods such as linear, non-linear and integer programming can be applied to parametric optimization (Pham and Karaboga 1998). For control optimization, dynamic programming is the classical tool (Bellman 1957, Puterman 1994, Sethi and Thompson 2000).

The MDP is a specific subset of control optimization problems. According to Puterman (1994), a MDP model is used by a decision maker who is confronted with the problem of influencing the behavior of a probabilistic system as it evolves through time. The decision maker has to choose the actions through time, and its goal is to choose a sequence of actions that lead the system to perform optimally with respect to some predetermined performance criteria. Basically, the decision maker wishes to find an optimal *policy* – defined as a sequence of actions in time. When deciding which action to take, he or she must anticipate the available opportunities and costs (or rewards) associated with future system states. Many problems of practical importance have been formulated as MDP. An extensive investigation of the theory and application of this framework can be found in many books such as those by Bertsekas (1987) and Ross (1983).

There are many classes of MDPs such as the Finite-Horizon MDP, the Infinite-horizon MDP and the Discounted MDP. This dissertation employs a finite horizon, discrete-time MDP. The theory and computation of this type of problem uses backward induction (Dynamic Programming or DP) to recursively evaluate expected reward. Bellman (1957) presents the optimality equations and the Principle of Optimality in his book which introduced and illustrated many of the key ideas of DP. Karlin (1955), Howard (1960), and Hinderer (1970) provide extensive introductions to the area.

However, the DP algorithm frequently is confronted by curses of modeling and dimensionality. A very small real-world problem may contain one-thousand states but one may have to store one million transition probabilities just for one action (Bertsekas 1995). Because of the high dimensionality of this information it is difficult to store and access, it is also hard to generate a solution.

Recently, a methodology called Reinforcement Learning (RL) has been introduced into the simulation-based optimization literature (Kumar 1985, Kumar and Variya 1986, Kaebling et al. 1996) to overcome dimensionality challenge. The concept of simulation-based optimization is also reviewed by Fu (2001), Law and McComas (2002), and Shapiro (1996). The RL approach is essentially a form of simulation-based dynamic programming used primarily to solve Markov decisions problems. The RL approach can generate optimal or near-optimal solutions without having to compute or store transition probabilities (Sutton 1992). In this research, an RL-based approach called the Q-Learning algorithm (Watkins 1989, Rummery and Niranjan 1994, Bertsekas and Tsisikilis 1995, and Sutton 1992) is applied to the hierarchical decision making model. Gosavi (2003) and Kaebling et al. (1996) give an extensive description of this method.

Overall, my dissertation focuses on the collection network growth problem within the context of reverse production system design. The addition of the recruitment concept and the uncertainty of the response to recruitment from agents increase the complexity of the model. The overall objective is to integrate overall plans to grow the collection network and to determine effective tactics to operate the network over time. The comprehensive plans include decisions in three levels made at different time scales.

At the tactical level, I propose an innovative way to examine more sophisticated behaviors of the agents in the recruitment process. To the best of my knowledge, no recruitment has been investigated in the supply chain area. This results in a new recruitment model for the supply chain application. My recruitment model is developed using a stochastic dynamic programming framework. An exact and two heuristics methods are proposed.

In order to handle the hierarchical decision making, a novel strategic model is developed to incorporate the recruitment model. The strategic model offers a robust way to integrate decisions from other levels in order to achieve the objectives that are different from most research on collection within reverse production systems. This strategic model is developed based on the Multi-time scale Markov Decision Process (MMDP) framework (Chang et al. 2003). Different assumptions in the strategic model lead to different MMDP model from the previous work. To the best of my knowledge, there is no published work examining the hierarchical decision making with the objectives proposed in this research. Furthermore, the heuristic for the strategic model is developed for multi-level decision making in order to solve realistically sized problems incorporating the fact that the heuristic method for the tactical level problem can provide a solution quickly. The heuristic includes the methodologies from dynamic programming, Q-Learning, and integer programming (IP). This model can aid processors with high set-up costs to plan the collection network more efficiently in order to be financially viable. Next chapter will cover more details of problem description in three levels: strategic, tactical and operational levels.

CHAPTER 3

PROBLEM DEFINITION

In this chapter the overall nature and structure of the decision models for strategic reverse production system growth problems are presented. I divide the strategic network growth problem into three parts: strategic, tactical and operational problems. The strategic problem involves designing a long-term collection network over a horizon of several macro-periods and a regional marketing plan to grow required collection volume capacity and minimizing the total collection costs. The strategic problem operates at a multi-region geographic scale and over several periods. The strategic plan is revised at the end of each period to reflect the feedback from the regions.

The tactical problem provides a recruitment plan to achieve the collection volume target set by the strategic decisions for the given marketing budget allocation. The operational problem schedules the pick-up service to the recruited retailers. The recruitment problem and pick-up service problem are decided on different timescales than the strategic problem, referred to as periods and micro-periods respectively. I provide a detailed description of problem statement in section 3.1, and I describe an important assumption for the operational level in section 3.2.

3.1 Problem Statement

In Chapter 1, I motivated the problem with a carpet industry example of a failure of a recycling system to reach its goals. On September 1, 2005, Shaw Industries Group, Inc., agreed to purchase the U.S. nylon carpet fibers business from Honeywell International (Shaw Floors: Honeywell 2005). Shaw will also acquire Honeywell's 50 percent stake in Evergreen Nylon Recycling based in Augusta. The company plans to reopen the recycling plant owned by Evergreen. However, as discussed in Chapter 1, operating the large nylon 6 recycling plant is a challenging problem. Without a collection strategy, Shaw will face the same possibility for failure as Evergreen Nylon Recycling in 2001 and Polyamid 2000 in 2003.

The high cost and large scale of the processing site drives the company to have enough supply at low cost to be financially viable. Shaw will change its operation from Evergreen, scaling back the facility throughput to about half that of the original 200 million lbs of nylon 6, but it will need 100 million pounds of used nylon 6 carpet per year. The large number of small firms and low landfill tipping fee can make collecting a large amount of post-consumer carpet difficult. There are roughly 7,000[4] small firms out of 8,000 firms in U.S. who can potentially supply the used carpet and these firms currently send the used carpet to the landfill.

In addition to the collection problem, Shaw has to compete with the companies in China for supply as well as other domestic recyclers. Currently these companies are buying a significant amount of used carpet from U.S. sources and capitalizing on cheap

[4] U.S. Census Bureau (2003)

backhaul transportation costs to bring it to China for recycling. They are simultaneously building nylon 6 polymerization capacity and the caprolactam plants to produce the monomer, which has long term implications for the demand for used carpet that could destabilize the recycling industry. To survive and thrive, over time Shaw must find a way to grow capacity and find/expand markets while being robust to uncertainties or changes in transportation costs and technology.

The collection system must manage network growth by investing wisely in the recruitment of collection sites and their retention. Recruitment decisions play out over time, regions, and market segments (e.g., targeting multi-family dwellings for used nylon 6 carpet collection due to higher facility turnover and use of this carpet type in this segment). Recruitment strategies range from adding individual sites (e.g., a carpet sales location) to large scale additions (e.g., adding a retailer system, like Home Depot or StarNet or Carpet One).

From Shaw's perspective, I derive the following problem statement. A carpet recycling company wants to build a network infrastructure to collect, process, and transport used carpet to the company's processing site for recycling such that the total cost of the entire system is minimized and collection volume targets are met. Hence, the collection network growth can be viewed as a problem for a single decision maker. A collection center can perform the collection tasks as well as certain processing tasks. Processing tasks may include sorting and packaging (baling). Only the processor's site can perform a more complex chemical recycling process (Mihut et al. 2001). Figure 3.1 depicts how the processes are linked in the carpet recycling industry.

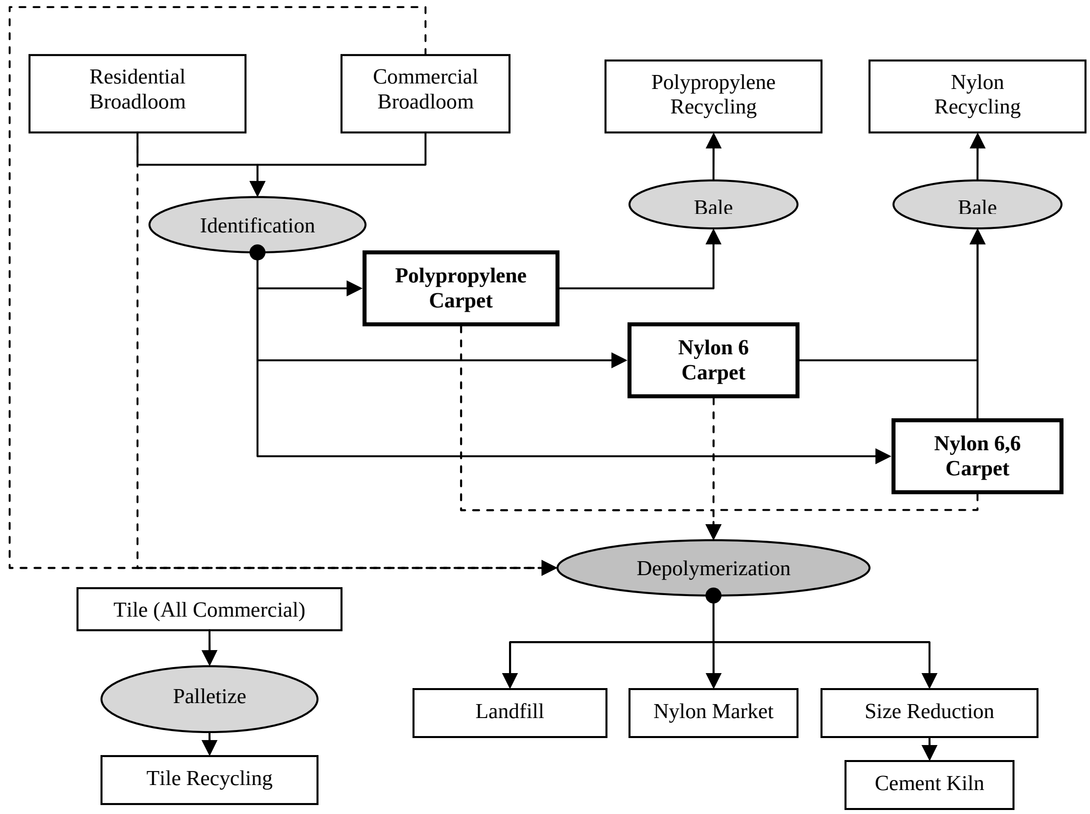

Figure 3.1: An Idealized Carpet Recycling Process Diagram

To justify the large capital cost of the depolymerization plant, the processor, or recycling company, must find a way to collect a sufficient amount of used carpets to supply the processing site. I take the perspective of the processor.

In Figure 3.2, I summarize overall inputs and the outputs of the strategic network growth problem followed by a discussion on how this problem may be subdivided into three levels: strategic, tactical and operational.

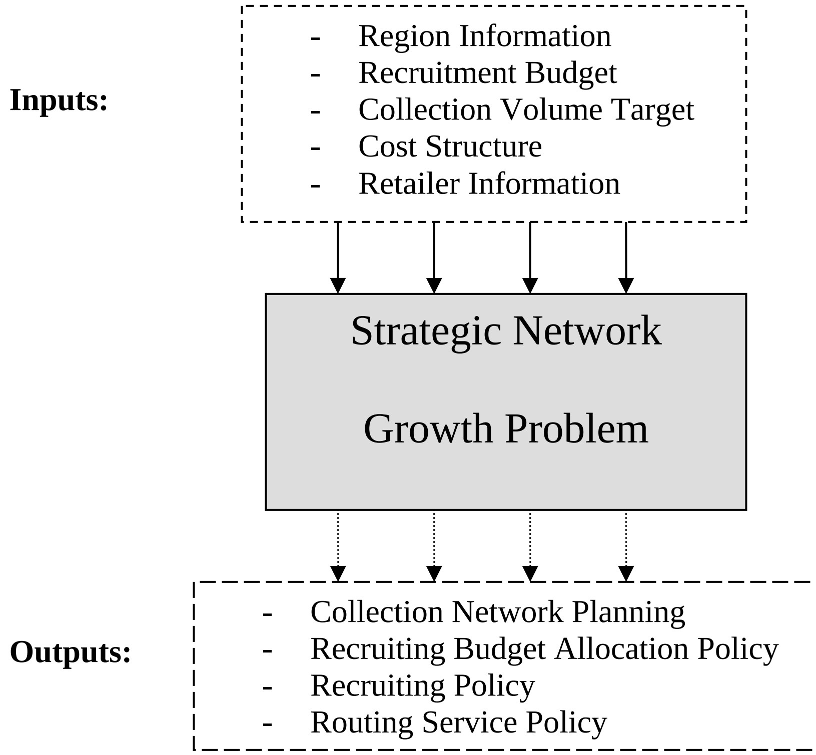

Figure 3.2: Inputs and Outputs of the Strategic Network Growth Problem

Collection network planning consists of deciding how much to collect in each region and how to grow collection volume over time. These decisions are critical *strategic decisions*. I consider this overall planning problem at a national scale, for which I have a set of regions in which to build collection networks. Each region consists of a collection center and a set of retailers that can be recruited to the network. For simplicity, I assume that there are no value added processes such as sorting at the collection center and there is no capacity limit at the collection center. Each regional collector operates its own collection center in its region. Overall, the entities in the reverse supply chain include the processors, the regional collectors, and suppliers (retailers). They are connected as shown in Figure 3.3. For simplicity, throughout this dissertation, I focus on a single type of material to be collected. The costs associated with strategic planning in this reverse supply chain include:

- Fixed and variable collection costs, and
- Fixed and variable opening costs for collection centers.

In addition to these costs, the processor spends a marketing budget over the planning macro-periods. In the strategic problem, for a specific budget, I assume that I can obtain an estimated supply quantity from each region based on the result from the recruitment problem. I assume that the collectors make their own decisions on which retailers to recruit based on the recruiting budget given to them. The resulting supply quantity estimation for each region depends on the region's recruiting budget and characteristics. One region may consist of many small retailers who are more than willing to supply the source to the processor while another region may consist of retailers who have already agreed to send their source to other processors.

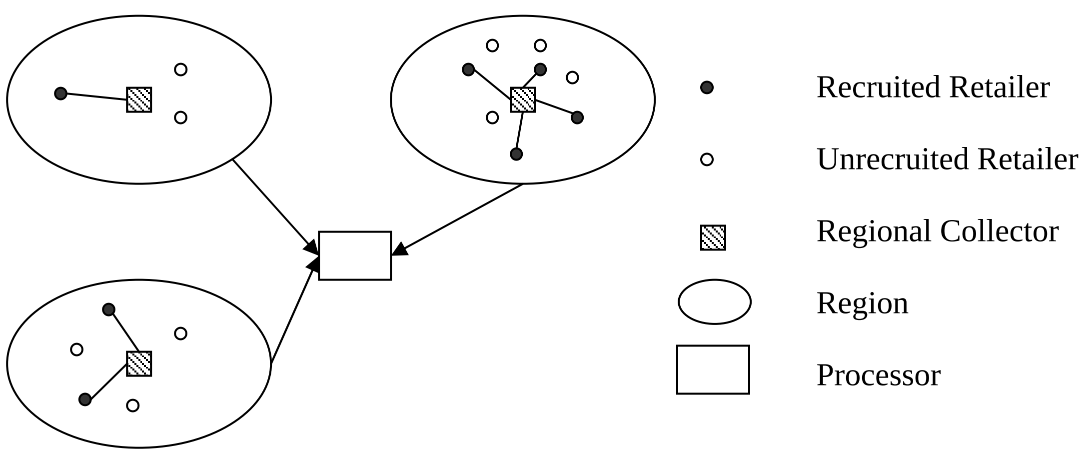

Figure 3.3: Network Growth Infrastructure

At the strategic level, I am not concerned with how regional retailer recruitment is planned by the collector in each region and how to physically collect the resource from various sources surrounding the regional collection center. All amounts collected from every region are transported to the final customer location, the processor's site. In the strategic level, there are two main objectives. The first objective is a capacity target for collection volume to be satisfied at the processor's site at the end of planning horizon. This target is a total collection *capacity*, not a total collection amount. The second one is that the total associated costs must be minimized. The processor collection capacity target and total recruiting budget are provided as inputs to the strategic problem. I adopt a decision horizon of multiple macro-periods with decisions made every macro-period. In summary, the design of the strategic infrastructure of the reverse supply chain system attempts to answer the following questions:

- In which regions, and when, should collection centers be opened?
- How much should be invested towards recruitment in each specific region?

When a regional collection center is opened, a regional recruiting budget may or may not have been allocated. If there is no recruiting budget allocated to the region, the regional collector performs the collection from existing recruited retailers only. Alternatively, if there is a recruiting budget, the collector uses it to recruit new retailers while performing collection on the existing network of retailers. This reflects the notion that an existing collection network is in place, and may require augmenting with new retailers and/or regions.

Given the answers to these strategic questions, many *tactical decisions* need to be made. These decisions are determined by the collector. I adopt a decision horizon of one macro-period with decisions made every period. These decisions include:

- Given the recruitment budget to be spent over one macro-period, how much of the budget should be allocated in a given period?
- Which subset of retailers should be targeted for recruitment during each period?

I define the objective of the tactical problem for a regional collector as how to achieve the specified regional collection target for the collection center. The collector cannot collect materials from a retailer unless both parties agree to the arrangement. Both parties reach agreement through the recruitment process. The collector serves as a recruiter and the retailer acts as a recruitee. The goal is to recruit the retailers in order to have access to their material by using the recruiting budget efficiently to reach the target amount. The retailer may or may not accept the offer depending upon its recruitment allocation. Therefore, a key modeling assumption predicts how the retailer behaves with respect to the offered incentives (Guide and Van Wassenhove 2001). According to the Florida Department of Environmental Protection (2000), providing the incentives to the collectors is one of the recommendations in the carpet pilot collection program done in Sarasota County in Florida.

Given the answers from the recruitment problem, many *operational decisions* need to be answered in order to operate the regional collection network. In a given period, the collector has to schedule transportation to pick up the recruited material from the recruited retailers in its regional network. In actual applications, each retailer generates the source material in random quantities per time period. I adopt an operational decision horizon of one period with decisions made every micro-period. Some of these operational decisions include:

- How many trucks are required used to service the retailers over the period? (I assume that this decision is made once at the beginning of the period and every truck has the same capacity.)
- Which retailers to be served by which truck in the specific time period?

The costs associated with the operational problem include:

- Fixed and variable transportation cost,
- Penalty cost if there are failures in service, and
- Penalty cost if the target collection amount does not satisfy the overall network goals.

Combining all the decisions together, the strategic, tactical and operational problems are depicted in Figure 3.4. Strategic level decisions have an impact on the decisions in the tactical level which also affects the decisions in the operational level.

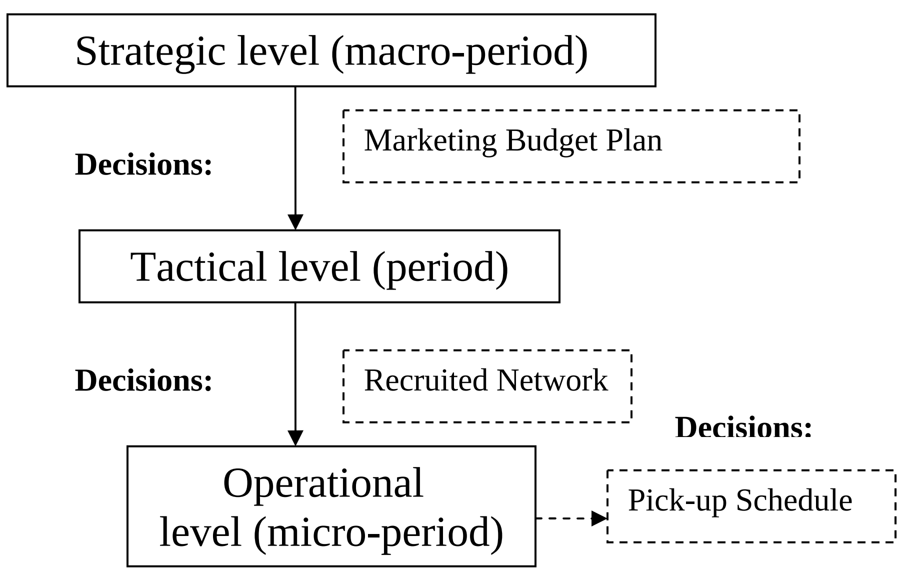

Figure 3.4: Information Passed between the Decision Levels

Not only in the carpet RPS, the collection network growth problem also occurs in other industries such as recycling electronics scrap, materials, and plastic bottles. For example, in the plastic bottle industry, recycling bottle is not an attractive business for the major beverage makers such as Coke and Pepsi although 2 out of 3 soda bottles (15 million) were not recycled in 1998 (Container Recycling Institute 2005). In April 2000, Coke announced that they have committed to a goal of using 10% recycled content in its plastic bottles by the end of this year. That goal is only 2.5% of the total amount of bottles the company sold in 1993. This implies that plastics-recycling firms also demand enough of supply with low cost to justify the cost of the high-technology processing site and operation. The company in this industry also faces another competitor from China. The U.S. plastics recycling industry experiences a shortage because so much is being exported to China for recycling there (MSNBC U.S. News Environment 2005).

The strategic and tactical problems are designed to understand, model, and support decision making in growing carpet-recycling supply chain networks. To be more specific, I have provided the relationship between the timescale and decision levels in Figure 3.5. For example, if the strategic planning horizon has a macro-period of five years, the network structure solutions and target quantity decisions in year one are passed onto tactical level problem in order to solve the recruitment problem over a number of periods, in this case months. The recruitment is determined monthly to identify which retailers are to be recruited and which retailers are to be served for that month. These decisions are acknowledged by the operational level, which constructs a policy for each micro-period, one day, for the truck operation plan based on the information on the state of the retailers in the network.

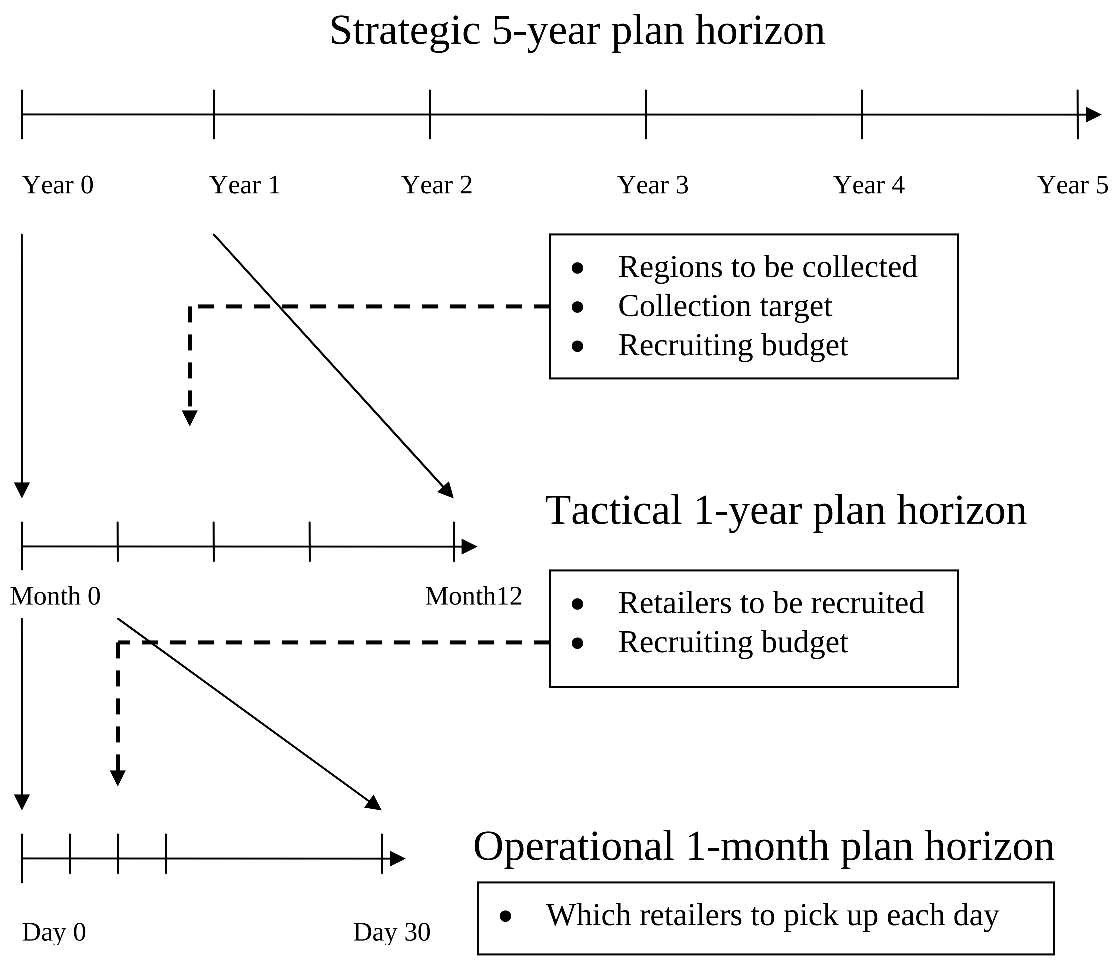

Figure 3.5: Diagram Showing the Connections between the Strategic, Tactical and Operational Levels

3.2 Assumptions for the Operational Level

My dissertation focuses on models for the tactical and strategic level problems and their associated solution methodologies. For simplicity, I assume that each retailer's generation rate is deterministic and that the pick-up schedule in the operational level can be solved optimally. Hence, recruited retailers form the collection network and provide a fixed amount of resource to be delivered to the collection center at the end of the month. This implies that the pick-up service has satisfactory performance. Therefore, the recruited retailers never leave the network. The only cost incurred is the transportation cost from collection center to the retailer. This assumption aims to reduce the computation burden to obtain the solution for the strategic and tactical level problems. The goal of this research is to be able to approach the solution of a problem of realistic size. Furthermore, the operational level problem can be viewed as a scheduling problem in solid waste management where many researchers (Hsieh and Ho 1993, Chang et al. 1993, Huang et al. 1993) have already proposed many methodologies for this type of problem.

In the collection network growth problem, a recycling company with large capacity and expensive processing facility requires enough material with low collection cost in order to be financially viable. This problem is a long term problem where the collection strategy can be changed yearly depending on the outcome of the previous year. My research provides the multi-time model that captures the decision making in each level and time period in order to grow the collection network wisely. The solution approach for this model can assist the company to make better decisions to meet the target collection volume with the least costs and eventually thrive financially. The end result is environmentally invaluable because less amount of used carpet will end up in the landfill.

With the overall problem description and the assumptions on the operational level, I discuss the model and approach for the tactical level in Chapter 4 problem and for strategic level problem in Chapter 6. A numerical study of tactical approach is presented in Chapter 5 and a numerical study of strategic approach is displayed in Chapter 7.

CHAPTER 4

MODEL AND SOLUTION APPROACH FOR THE TACTICAL LEVEL PROBLEM

In this chapter a model for the tactical recruitment of sources is developed. This model builds upon the solution from the operational level and the budget allocation from the strategic level. The tactical problem addresses the regional collector's problem of recruiting the retailers to join its network. This chapter provides the modeling details for the recruitment problem, a general framework for the recruitment process, and a methodology to solve the recruitment problem. A crucial element of the model is the individual retailer state behavior, which is assumed to follow a Markov process. The recruitment actions and budget expenditures move the retailers into more favorable states. Thus, the overall problem is to balance the short term recruitment with the longer term actions to increase the future cumulative recruitment.

This chapter is organized as follows. In section 4.1, the description of tactical level problem is discussed. Then I introduce the general framework for the recruitment problem in section 4.2. In section 4.3, the tactical problem is formulated with a stochastic dynamic programming framework. In section 4.4, an exact method is developed to solve the tactical problem. In the next two sections, two heuristics are proposed. In section 4.5, I develop a simulation based optimization methodology, the Q-Learning Based Heuristic. In section 4.5, I propose a rolling horizon based methodology, the Rolling IP with DP Heuristic.

4.1 Tactical Level Problem

According to Chapter 3, the input from the strategic level informs the tactical level about which regions to recruit, how much budget can be spent in each recruiting region, and how much regional collection volume should be targeted in one macro-period. The output from the tactical level informs the operational level about recruiting retailers and the corresponding budget allocation for each period. The operational level then follows the recruitment plan from the tactical level, updates the recruiting network in each month, and performs the collection accordingly. The information flows for tactical level are displayed in Figure 4.1.

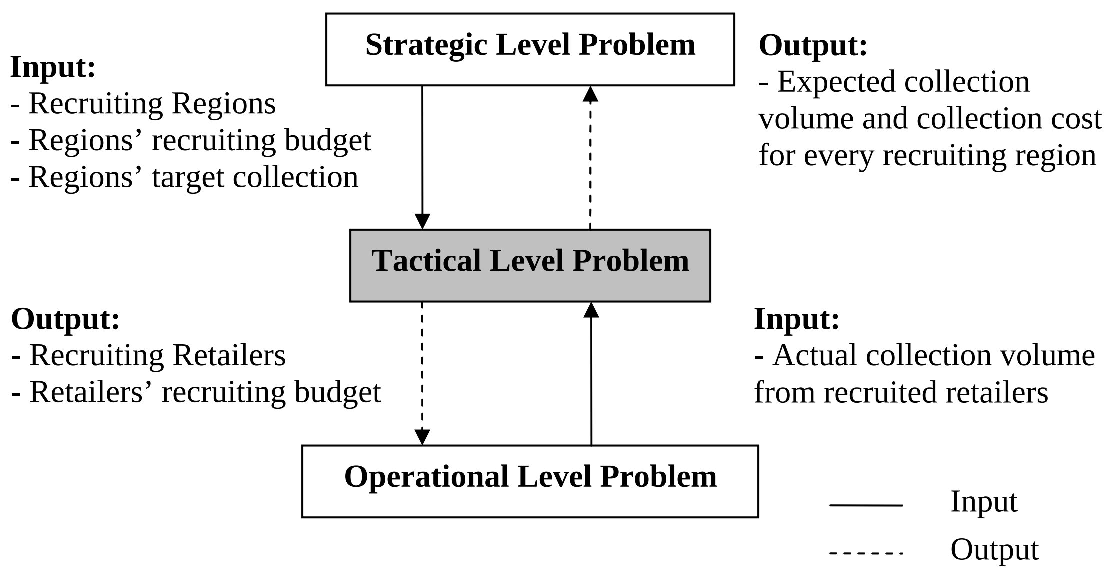

Figure 4.1: Information Flow for Tactical Level Problem

For each region, I assume that the retailers are dispersed around the collector. In some regions, such as rural areas, the retailers may locate sparsely around the collector. While in other regions, such as metropolitan areas, the retailers may cluster tightly near the collector. This dispersion plays a role in the regional collection cost for the strategic level decision making which will be discussed in Chapter 6. I denote the term "recruitment" as the negotiating process that involves two parties: the regional recruiter and the recruitee. The regional collector cannot retrieve material from a source or retailer unless it is agreed by both parties through the recruitment process. The regional collector serves as a recruiter and the retailer acts as a recruitee. The initial objective of the tactical level problem for a region is to recruit the retailers for the sources by using the limited recruiting budget efficiently to reach the target collection volume. For simplicity, throughout this chapter I represent the regional collector's tactical objective as a maximization of the expected collection volume obtained from recruiting the retailers with a limited recruiting budget. The regional tactical problem is solved per period over one macro-period. Given the solutions from strategic level, the tactical for a region decisions are:

1) Which subset of retailers to recruit, and

2) How to allocate the budget to those selected retailers.

Given the details of the tactical level problem, the general framework of the regional recruitment model is described in the next section.

4.2 Multi-period Recruitment Framework

In this section, a general framework for the regional recruitment process is developed. This framework attempts to capture the heterogeneous attitudes of retailers towards participating to join the processor's collection network. In the literature, recruitment models focus on employment recruitment, human resource management, and physiological models in medical research (Darmon 2003, Treven 2006, Hawkins 1992, Georgiou and Tsantas (2002). No research has been found on recruitment in supply chain applications. In this research, I develop a general framework for the regional recruitment model from industrial engineer's perspective using the available methodologies in operations research. The overall problem I consider has M regions and each region has η_m retailers. Consider region m. There are two types of entities in this model: recruiter and agent.

Recruiter

A regional recruiter is an entity who is responsible for recruiting agents to join its network in order to gather required resources from agents. I assume that there is only one recruiter in this region framework and it does not compete with other recruiters for the resource. It is given a resource, called Resource A, which can be used for the agents' recruitment process. This resource typically can be interpreted as money or a discount that can be used as an incentive to recruit the agents. I call this resource the recruiting budget, which is determined for each region by the strategic model. For example, in carpet recycling, the recruiter can represent the regional collector trying to establish a collection network in its region for used carpets from carpet retailers. The recruiting budget can be the amount of initial cost reduction to offer to the retailer for waste collection beyond the cost of actual service.

Agent

An agent (recruitee) represents the owner of the resource that the recruiter desires. In this framework, there is only one type of resource from an agent, called Resource B. The region contains a heterogeneous set of agents. Heterogeneity is in:

a) The quantity of the resource they generate,

b) Their geographical location,

c) Their initial willingness to sell/give the recruiter the resource based on some predefined factors, and

d) Their predisposition towards becoming recruited to the network.

The agents can represent the carpet retailers that have used carpet to be recycled. The regional recruiter interacts with a subset of the agents in order to achieve its objective. The objective of the regional recruiter is described next.

Objective

The objective of the regional recruitment model (recruiter) is to build the collection network that yields the highest collection volume by the end of the planning period, $T-1$, by making decisions in each period with a limited regional recruiting budget. The recruiter is given a recruiting budget that must be allocated over discrete finite time period, $t=(0,1,\ldots,T-1)$, with budget allocation decision required for each period. Figure 4.2 depicts the growth of recruitment network over time.

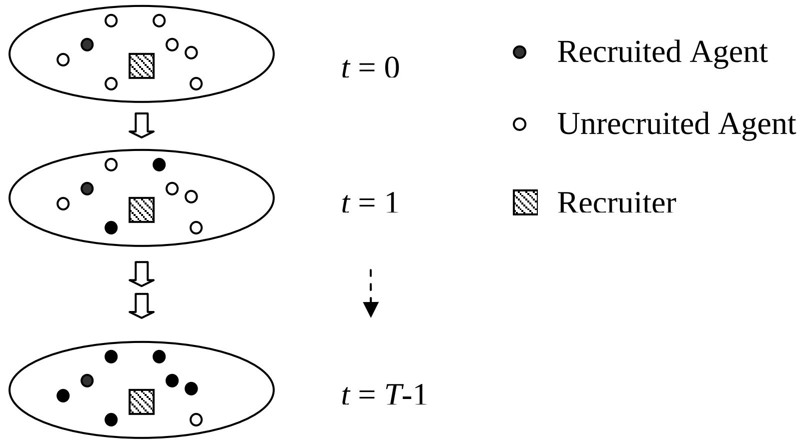

Figure 4.2: Growing Recruitment Network

In order to achieve the objective, the recruiter makes the recruitment budget allocation decision in each decision period. Its decision also affects the decisions of the agents. The decisions of the recruiter and the agents are described in the following paragraphs.

Recruiter's Decisions

In each period, the recruiter decides how to allocate the recruiting budget, Resource A, among a subset of retailers based on the time period, remaining amount of budget and the recruiting willingness state of each agent. I assume that the willingness state of each agent is updated to the recruiter in every period. The total spending budget in all periods must not exceed the recruiting budget limit provided to the recruiter at the beginning of planning period.

Agent's Decisions

At the beginning of the period, after the agent receives the allocation of budget (Resource A) from the recruiter, it decides whether to give/sell its resource, Resource B, to a recruiter. Its decision is based on its willingness to part with the resource. In the case that the agent does not contract to provide the resource to the recruiter, the agent's overall willingness state can change by being influenced by the incentives it receives from the recruiter.

Figure 4.3 summarizes how the decision of the recruiter is related to the decision of the agents in each period.

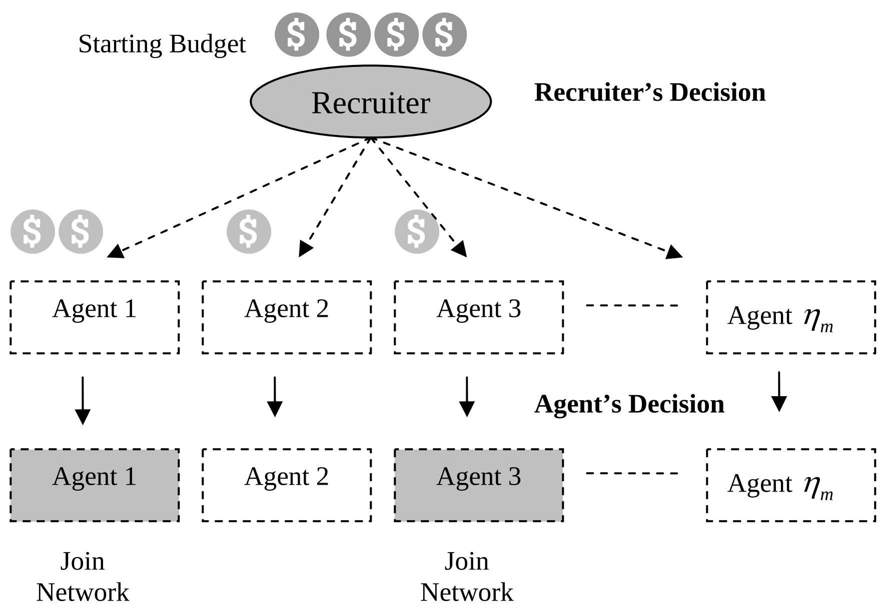

Figure 4.3: Decisions of Recruiter and Agent in One Period

In the recruitment framework, there is information passing back and forth between the recruiter and the agents. The decision timeline for the recruitment model is shown in Figure 4.4, where the solid arrows from recruiter to agents indicate the resource allocation from recruiter to the agents and the dashed arrows show the responses of the agents to the recruiter. The response from the agents in each period provides the feedback the recruiter needs to make better decision in the later periods.

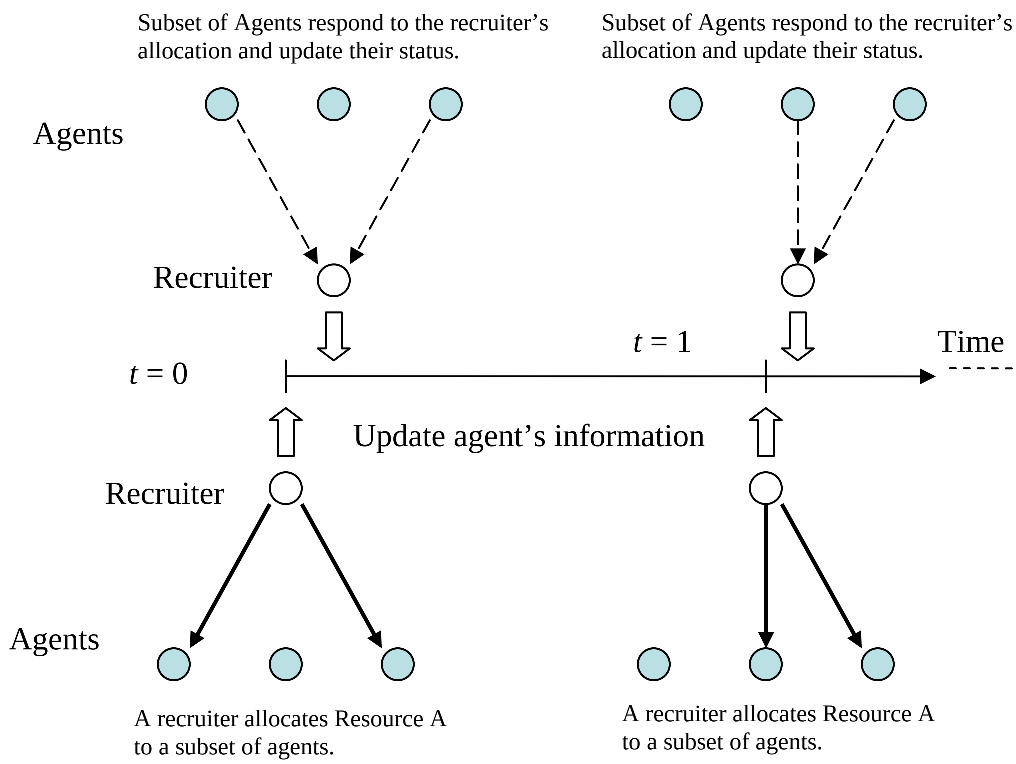

Figure 4.4: The Decision Timeline for a Recruitment Problem

A key element of the recruitment framework is the model of an agent's willingness to participate. This model can be as simple as a one-variable function of the given incentive. However, it is more likely that agents have a more sophisticated "state" relative to their willingness to participate. A Markov model for the agent, which I call the Agent's Resource Willingness Model (ARW), is developed in order to capture a more complicated structure of the agent behavior and yet retain reasonable representational and computational simplicity.

Agent's Resource Willingness Model (ARW)

The key elements of this model are the willingness state and the transition probabilities. I model each agent's resource willingness as a Markov chain with a "recruited" state that is absorbing. This means that the recruited agent never leaves the collection network. In other words, for simplicity, I am assuming that the recruiter is not confronted with the agent retention issue. The model also consists of other states that represent a "distance" from recruitment based on the probability of reaching recruitment and connection to other states. Each agent has its own Markov model. Consider agent i in this model.

<u>Willingness State Definition</u> (Agent's state)

$$s_{it} = \{R, L, M, H\}.$$

This describes what state agent i is in at time period t. There are four possible states for each agent:

<u>1. Recruited (R)</u> - The agent agrees to give the resource to the recruiter.

<u>2. Low (L)</u> - The agent is not recruited by the recruiter. Also, the agent is in the state that is very hard to be recruited.

<u>3. Medium (M)</u> - The agent is not recruited by the recruiter. The agent has no bias against the recruitment.

<u>4. High (H)</u> - The agent is not recruited by the recruiter. Also, the agent is in a state that makes recruitment easy.

I assume that when the agent is recruited, it resides in the R willingness state, an absorbing state. The states L, M, and H represent a "distance" from recruitment based on the probability of reaching the recruitment state and connection to other states. In other words, if the agent is not recruited, it can be only in either the L, M, or H state. Figure 4.5 shows a symbolic representation of the states and possible transitions.

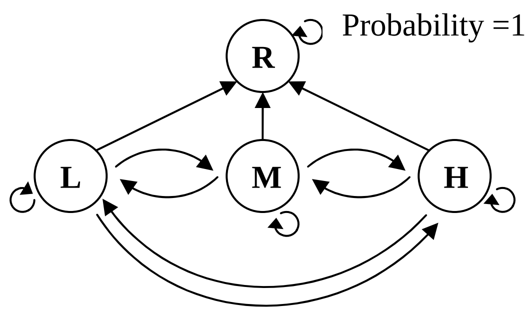

Figure 4.5: Agent's State Diagram

I denote the amount of resource that agent i can generate between each decision epoch (each period) as g_i, which is assumed to be deterministic. According to section 3.2, I assume that the operational level or the pick-up problem can be solved exactly. Thus, this means that I am assuming that the recruiter can collect the full amount of resources from every recruited retailer in each period. In addition, I denote a_{it} as the amount of budget that agent i receives from the recruiter in period t. Given the action a_{it}, the agent transits to the next state with following transition probabilities.

<u>Transition Probabilities</u>

The probability of moving to state $s_{i,t+1}$ from state s_{it} by action a_{it} is denoted by $p(s_{i,t+1} | s_{it}, a_{it})$ or $Pr_{s_{it}s_{i,t+1}}(a_{it})$. There are two types of transition probabilities to consider.

The probability of recruitment is the probability of moving to state R from the L M, or H states ($Pr_{LR}(a_{it}), Pr_{MR}(a_{it}), Pr_{HR}(a_{it})$). The difficulty of recruitment depends on three factors: state of the agent, budget allocation or action, and agent's recruitment budget threshold, μ_i. The recruitment budget threshold has the same units as the budget allocation. In general, it represents a minimum value required to recruit the agent. In the case of carpet retailers, there may be some correlation between the size of the retailer and the budget required. The reason is that the incentives offered may directly scale with the amount of carpet available for pick-up. The threshold may be interpreted as subsidizing the cost of the retailer's disposal fee. In Chapter 5, when I generate the data for numerical study, I assume that if the agent can provide a significant amount of Resource B, it also demands large amount of allocation of Resource A from the recruiter. Thus, the recruitment budget threshold depends on the amount of Resource B collection volume from agent i. A higher g_i implies a higher μ_i. This means that it is more expensive to recruit agents who have higher Resource B generation rates.

In order to capture these three factors together, I apply a Sigmoid function (Seggern 1993) to calculate the probability of recruitment. A sigmoid function is a mathematical function that produces a sigmoid curve — a curve having an "S" shape. Often, a sigmoid function refers to the special case of the logistic function shown in Figure 4.6. In general, a sigmoid function is real-valued and differentiable, having a non-negative or non-positive first derivative, one local minimum, and one local maximum. In addition, I define the recruitment willingness factor, β_s, based on the state of the agent such that $\beta_H > \beta_M > \beta_L > 0$.

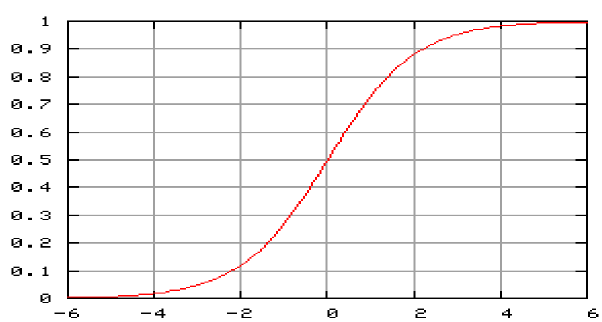

Figure 4.6: A Sigmoid Function

In this research, I assume that the function to calculate probability of recruitment of agent i at time t is expressed as:

$$Pr_{s_{it}R}(a_{it}) = \frac{1}{1+e^{-\beta_{s_{it}}(a_{it}-\mu_i)}}\,. \tag{4.1}$$

Using the probability of recruitment function in (4.1), I can vary the value of the recruitment willingness factor so that each state has different difficulties to recruit as shown in Figure 4.7. I set $\beta_H = 2$, $\beta_M = 1$, and $\beta_L = 0.5$.

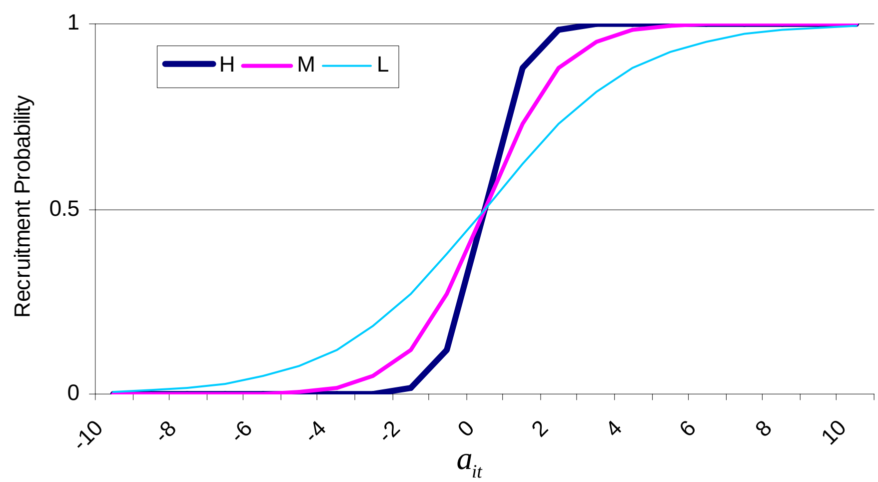

Figure 4.7: Recruitment Probability for Different Recruitment Willingness States

The probability of (unrecruited) state transition can be specified according to how easy the particular agent is moved among the L, M, and H states if it is not recruited. The probability of state transitions can be set up such that it is easy to move to M and H from L. This makes the agent easier to recruit. On the other hand, the probability of state transition can be set up such that it is more difficult to move to M and H from L. This makes the agent more difficult to recruit. Figure 4.8 displays the overall transition probabilities of an agent.

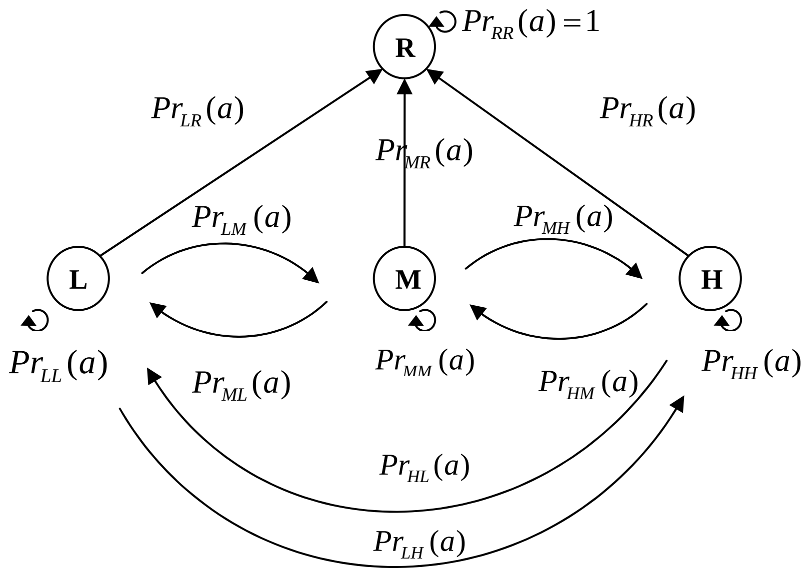

Figure 4.8: Transition Probabilities for An Agent

Using the Agent's Resource Willingness (ARW) Model, the decision for one period that is shown in Figure 4.3 can be modified as shown in Figure 4.9. The ARW model provides a better representation of each agent's participation status for the recruiter.

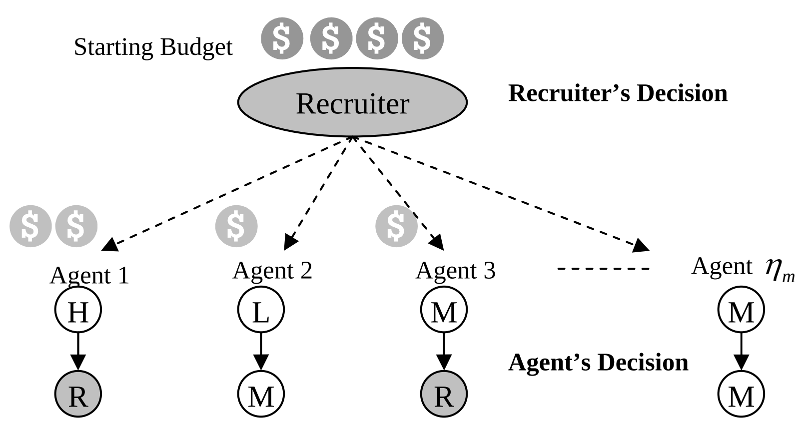

Figure 4.9: Decisions of Recruiter and Agents with the ARW

Given the general framework of the recruitment model, I now introduce a stochastic dynamic programming formulation of the recruitment problem for each regional collector.

4.3 A Stochastic Dynamic Programming Formulation for Recruitment Problem

This section develops a stochastic dynamic programming model for the recruitment problem that capitalizes on the Markov property in the Agent's Resource Willingness model. The formulation of this model consists of the definition of decision epochs, state space, actions, transition probabilities, and rewards. A solution for this model provides the optimal recruiting policy to the regional collector. In this formulation, I assume that the precise information for the parameter values is available. If the values for the parameters are not exactly known, White and Eldeib (1984) discuss how to handle this situation in dynamic programs.[5] The overall problem I consider has M regions and each region has η_m retailers. Each regional collector has its own recruitment problem. In this section, I consider the recruitment problem for region m. The number of agents is η_m, the maximum recruiting budget is $B^{\max}$ and the total number of planning periods is T.

The decision epochs and state spaces are defined as follows.

Decision Epochs

$$t = \{0,1,...,T-1\}$$

State Space

$$Y_t = \{t, w_{1t}, w_{2t}, ..., w_{\eta_m t}, B_t^{Start}\} \text{ for all } Y_t \in \mathbb{S},$$

where the willingness state of retailer i at decision epoch t is $w_{it} \in \{L, M, H, R\}$ and the starting recruitment budget at the beginning of period t is represented by B_t^{Start}.

[5] Special thanks to my dissertation committee member Dr. Chelsea C. White III for providing this insight.

In this model, I define the action set as follows.

Action Sets

$$A_{lt} = \{a_{1lt}, a_{2lt}, ..., a_{\eta_m lt}\},$$

where the amount of resource A allocation to agent i from action set index l at time period t is represented by such that $\sum_{i}^{\eta_m} a_{ilt} \leq B_t^{Start}$ and $0 \leq a_{ilt} \leq B_t^{Start}$ for $l = 1, ..., |A_t|$. At the first period, $B_0^{Start} = B^{\max}$. The size of the action set depends on η_m and B_t^{Start}.

In this model, I define the state transition rules as follows.

State Transition Rules

(a) Initial State

There is more than one possible initial agent state at $t = 1$ depending on the initial value of w_{it}. One example initial state is $Y_0 = \{0, w_{10}, w_{20}, ..., w_{\eta_m 0}, B_0^{Start}\} = \{0, \underbrace{L, ..., L}_{\eta_m}, 10\}$ where all agents begin in the 'L' willingness state and starting budget is 10.

(b) State Transition Probabilities

These probabilities depend on the ARW model for each agent. I assume that each agent's willingness state changes independently, so the state transition probability is the multiplication of the probability of the willingness state transition for each agent given the specific Resource A allocation provided by the recruiter. The representation of the transition probability in the product form can be simplified into a representation based on simple product form through algebra manipulation according to Economou (2005) and Thomas (2005).[6]

[6] Special thanks to my dissertation committee member Dr. Chelsea C. White III for providing this insight.

If the current state is Y_t and action A_{lt} is taken in period t, the probability transition of moving to state Y_{t+1} or $P_t(Y_{t+1} | Y_t, A_{lt})$ can be represented in the following form,

$$\begin{aligned}
P_t(Y_{t+1} | Y_t, A_{lt}) \quad &= P_t\Big((t+1, w_{1(t+1)}, w_{2(t+1)}, ...w_{\eta_m(t+1)}, B_{t+1}^{Start})\Big| \\
&\qquad (t, w_{1t}, w_{2t}, ...w_{\eta_m t}, B_t^{Start}), (a_{1lt}, a_{2lt}, ..., a_{\eta_m lt})\Big), \\
&= Pr_{w_{1t}, w_{1(t+1)}}(a_{1lt}) \cdot Pr_{w_{2t}, w_{2(t+1)}}(a_{2lt}) \cdots Pr_{w_{\eta_m t}, w_{\eta_m(t+1)}}(a_{\eta_m lt}), \\
&= \prod_{i=1}^{\eta_m} Pr_{w_{it}, w_{i(t+1)}}(a_{ilt}). \qquad (4.2)
\end{aligned}$$

where $\sum_{i=1}^{\eta_m} a_{ilt} \leq B_t^{Start}$ and $B_{t+1}^{Start} = B_t^{Start} - \sum_{i=1}^{\eta_m} a_{ilt}$. Here, the probability of moving to willingness state $w_{i(t+1)}$ is $Pr_{w_{it}, w_{i(t+1)}}(a_{ilt})$ if the previous willingness state is w_{it} and allocation a_{ilt} is taken for agent i. These probabilities can be obtained from the ARW model described in section 4.3. If the recruitment willingness factor parameter $\beta_{w_{it}}$ cannot be obtained precisely, the transition probability can not be obtained exactly. White and Eldeib (1994) provide an algorithm to approach this difficulty by assuming that the transition probability mass vector for each state and action is described by a finite number of linear inequalities[7].

For this model, I define the rewards as follows.

<u>Rewards</u>

In order to compute the rewards, I assume that the willingness state of agent i at time t has its own value, $V_{w_{it}}$. Let V_R be the amount of Resource B that agent can provide to the recruiter. Because the reward should represent the increment in collection

---

[7] Special thanks to my dissertation committee member Dr. Chelsea C. White III for providing this insight.

volume, the values for the non-recruited states V_L, V_M, and V_H can be set to zero. However, the reward should be defined such that there is an incentive to move to a higher willingness state. Hence, the value of V_L, V_M, and V_H are assigned a small value such that $V_L < V_M < V_H << V_R$. For example, V_L = 0.1, V_M = 0.2, and V_H = 0.3. Let $r_t\left(Y_t, A_{lt}, Y_{t+1}\right)$ denote the value at time t of the reward received when the state of the system at decision epoch t is Y_t, action A_{lt} is taken, and the system occupies state Y_{t+1} at decision epoch $t+1$. This value represents the total increment in the collective value of all of the agents' state changes. If the recruiter moves many agents to state R, it can obtain a high reward from the cumulative collection volume for the recruited agents. This value can be obtained by:

$$r_t\left(Y_t, A_{lt}, Y_{t+1}\right) = \sum_{i=1}^{\eta_m}\left(V_{w_{i(t+1)}} - V_{w_{it}}\right). \tag{4.3}$$

The regional recruiter's expected reward of state Y_t and for action A_{lt} can be evaluated by computing:

$$\begin{aligned} r_t\left(Y_t, A_{lt}\right) &= \sum_{\forall Y_{t+1}} P(Y_{t+1} \mid Y_t, A_{lt}) r_t\left(Y_t, A_{lt}, Y_{t+1}\right), \\ &= \sum_{\forall Y_{t+1}}\left[\prod_{i=1}^{\eta_m} Pr_{w_{it}, w_{i(t+1)}}(a_{ilt}) \cdot \sum_{i=1}^{\eta_m}\left(V_{w_{i(t+1)}} - V_{w_{it}}\right)\right]. \end{aligned} \tag{4.4}$$

Given the description of the reward function, the objective function of this model can be defined as follows.

<u>Objective Function</u>

The objective is to maximize the expected collection volume using the given budget from the strategic level. In other words, under a fixed budget, the recruiter wants to move a subset of retailers to state R such that those retailers yield the maximum expected collection volume in the final period. Since the overall purpose is to maximize the total collection volume at the end of the time horizon and the volume is not a quantity associated with time, there is no conventional discount factor involved.

Let $\pi = (d_0, d_1, ..., d_{T-1})$ represent the optimal policy for every time period. Hence, $\pi^* = (d_0^*, d_1^*, ..., d_{T-1}^*)$ denotes the optimal policy in each time period. Define the expected total reward obtained at decision epoch $t, t+1, ..., T-1$ by using policy π to be $u_t^{\pi}(Y_t)$ with starting state Y_t in decision epoch t as:

$$u_t^{\pi}(Y_t) = E_{Y_t}^{\pi}\left\{\sum_{t'=t}^{T-1} r_{t'}(Y_{t'}, A_{lt'})\right\}. \tag{4.5}$$

Let $u_t^*(Y_t)$ denote the maximum expected total reward obtained at decision epochs $t, t+1, .., T-1$ with starting state Y_t in decision epoch t. Then the optimality equation for the recruitment problem is:

$$u_t^*(Y_t) = u_t^{\pi^*}(Y_t) = \underbrace{\max}_{A_{lt}}\left\{r_t(Y_t, A_{lt}) + \sum_{\forall Y_{t+1}} P_t(Y_{t+1} \mid Y_t, A_{lt}) u_{t+1}^*(Y_{t+1})\right\}, \tag{4.6}$$

$$A^*(Y_t) = \underbrace{\arg\max}_{A_{lt}}\left\{r_t(Y_t, A_{lt}) + \sum_{\forall Y_{t+1}} P_t(Y_{t+1} \mid Y_t, A_{lt}) u_{t+1}^*(Y_{t+1})\right\}. \tag{4.7}$$

The optimal action in states Y at epoch t is denoted by $A^*(Y_t)$. In other words, the maximum expected total reward at period t, $u_t^*(Y_t)$, is the realization from all possible actions of the immediate reward and expected future reward from a particular action. Essentially, the objective of the recruitment problem is to find $u_0^*(Y_0)$.

Given the stochastic dynamic programming formulation of the recruitment problem, the exact method to solve this problem is developed in the next section.

4.4 Dynamic Programming Algorithm

In this section, the algorithm to solve for the optimal policy of the recruitment problem is proposed. The algorithm takes advantage of the optimality equation developed in section 4.3. Because the stochastic recruitment problem is a finite horizon problem, it can be modeled as a stochastic path problem where the number of paths is exponentially large. In addition, the recruitment problem's reward falls under the total reward problem. For a finite period stochastic path problem with total reward, one could use value iteration based scheme to solve the problem. Hence, for a small sized problem, backward induction or dynamic programming (DP) provides an efficient method to solve the recruitment problem.

The procedure of the DP algorithm is shown as follows.

Backward Induction (DP) Algorithm Procedure

Step 1 Set $t = T$ and $u_T^*(Y_T) = r_T(Y_T) = 0$ for all possible states in t.

Step 2 Substitute $t-1$ for t and compute $u_t^*(Y_t)$ for each $Y_t \in \mathbb{S}$ from

$$u_t^*(Y_t) = \underbrace{\max}_{A_{lt}} \left\{ r_t(Y_t, A_{lt}) + \sum_{\forall Y_{t+1}} P_t(Y_{t+1} \mid Y_t, A_{lt}) u_{t+1}^*(Y_{t+1}) \right\},$$

$$\begin{aligned} u_t^*(Y_t) = \underbrace{\max}_{A_{lt}} \Bigg\{ & \sum_{\forall Y_{t+1}} \left[\prod_{i=1}^{\eta_m} Pr_{w_{it}, w_{i(t+1)}}(a_{ilt}) \cdot \sum_{i=1}^{\eta_m} \left(V_{w_{i(t+1)}} - V_{w_{it}} \right) \right] + \\ & \sum_{\forall Y_{t+1}} \left(\prod_{i=1}^{\eta_m} Pr_{w_{it}, w_{i(t+1)}}(a_{ilt}) \right) u_{t+1}^*(Y_{t+1}) \Bigg\}. \end{aligned} \tag{4.8}$$

Set

$$A^*(Y_t) = \underbrace{\arg\max}_{A_{lt}} \left\{ r_t(Y_t, A_{lt}) + \sum_{\forall Y_{t+1}} P_t(Y_{t+1} \mid Y_t, A_{lt}) u^*_{t+1}(Y_{t+1}) \right\}. \tag{4.9}$$

Step 3 If $t = 0$, stop. Otherwise return to step 2.

Using theorem 4.5.1 from Puterman (1994), it can be shown that the optimal value for all decisions epochs is $u_t^*(Y_t)$ and corresponding to the optimal action (policy) in all states Y_t at epoch t is optimal action $A^*(Y_t)$. An illustration to find $u_1^*(Y_1)$ using the dynamic programming algorithm for a 2-agent problem is displayed in Figure 4.10.

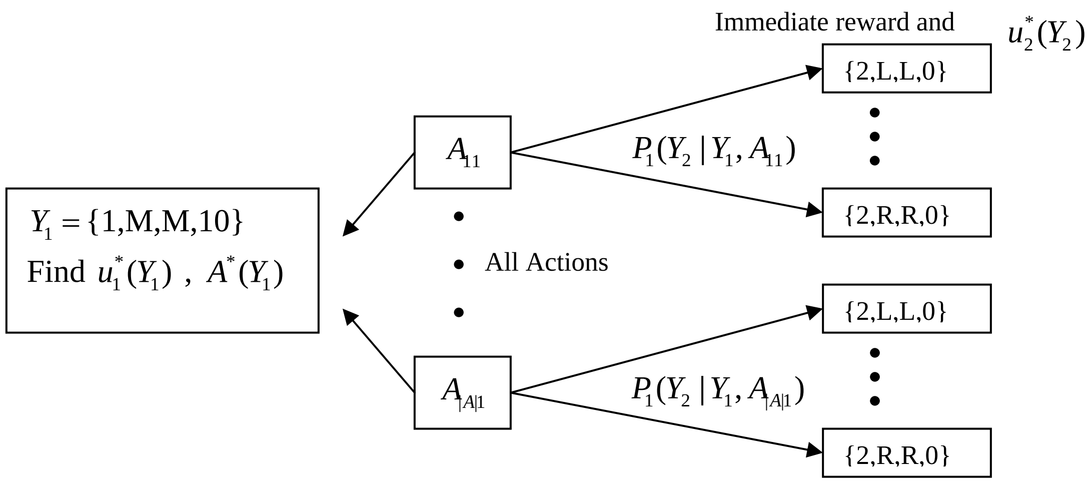

Figure 4.10: Example Steps of the DP Algorithm

For a small sized problem, the DP algorithm gives us an optimal policy for decision making based on the states and the time period. It enables the recruiter to find which agents to recruit and how much of Resource A to allocate to each agent for each period. However, this algorithm suffers from “the curse of dimensionality” as described in Bellman (1957). This means that computational efforts grow exponentially with the number of state variables or with the problem size.

Consider a problem with three periods and a recruiting budget bounded by the largest budget required to recruit an agent multiplied by the number of agents. Table 4.1 displays the number of states, an upper bound on the number of actions, and the total number of state-action pairs. For a large problem, it is difficult to compute an exact number of actions because of the large combinatorial combinations. As shown in Table 4.1, a one-agent problem has a small number of states and actions. As the number of either agents or action sizes increases linearly, the number of states and actions increases at an exponential rate. In addition, the computation of the transition probabilities in (4.2) becomes more troublesome.

Table 4.1: Number of State, Actions, State-Action Pairs

Number of Agents	Action Sizes	Number of States	Upper bound on Number of Actions	Number of state-action pairs
1	5	60	6	360
5	5	15,360	7,776	1.2E+08
10	5	1.5E+07	6E+07	9.5E+14
20	5	1.6E+13	3.6E+15	6 E+28
1	10	120	11	1,320
5	10	30,720	16,1051	4.9E+09
10	10	3.1E+07	2.5E+10	8.1E+17
20	10	3.3E+13	6.7E+20	2.2E+34
1	20	240	21	5,040
5	20	61,440	4E+06	2.5E+11
10	20	6.2E+07	1.6E+13	1E+21
20	20	6.6E+13	2.7E+26	1.8E+40

For large-scale problems, the DP algorithm is difficult to solve in reasonable time because it has to examine every possible action in each state in order to find the optimal solution, even though many states would not be reached by the optimal policy. With the notion that the tactical level problem has to interact efficiently with the strategic level problem to grow the network, the recruitment problem must be solved with a small amount of computational effort. In the next section, I introduce two heuristics as a way to solve the large-scale recruitment problem in reasonable time.

4.5 Q-Learning Based Heuristic

This section develops a heuristic based on the Q-Learning method to obtain a solution policy for the recruitment problem. This heuristic provides an alternative way to solve the large-scale recruitment problem within reasonable effort. Q-learning (Watkins 1989) is an extension to traditional dynamic programming or value iteration. Q-Learning is one of the methods of reinforcement learning (RL) or simulation-based optimization concepts. According to Kaebling (1996), RL is the problem where a solver must learn how to achieve the best action via trial-and-error with interaction in a dynamic environment. Typically, there are two strategies for RL. The first is to do a search of actions in order to find one that performs well in the environment. This approach has been taken by genetic algorithms. The second is to use statistical techniques and DP methods to estimate the utility of taking actions in the states. Q-Learning falls under the second strategy.

With a large scale in the number of actions and the number of states in complex recruitment problems, the exploitation-exploration dilemma is encountered. To achieve high reward, a reinforcement learning solver most likely takes actions that have been tried in the past and found to produce high rewards. In order to discover such actions, it must choose actions that have not been tried before. The solver must *exploit* the on-hand information in order to obtain high rewards, but it also must *explore* to make better action selections in the next periods. The dilemma is that neither exploration nor exploitation can be pursued exclusively without failing at the task. Failing to explore enough state-action pairs results in an inability to choose an action that may produce higher future reward, while failing to exploit makes the current information lead to slow convergence. In general, the solver must choose a variety of actions and progressively favor those that appear to be best. For a stochastic task, each action must be tried many times to gain a reliable estimate of its expected reward. These two characteristics--trial-and-error (exploration) search and delayed reward (exploitation)--are the two most important distinguishing features of reinforcement learning.

The main difference between RL and DP is depicted in Figure 4.11. Both methodologies are similar in the distribution of the random variables that control the system's behaviors. However, things work differently after that. In order to perform the DP algorithm, the transition probabilities and reward functions must be known in advance. In RL, instead of computing these values or estimating them, the system is simulated using the distributions of the random variables. Then, within the simulator, a suitable RL algorithm is executed to obtain a near-optimal solution.

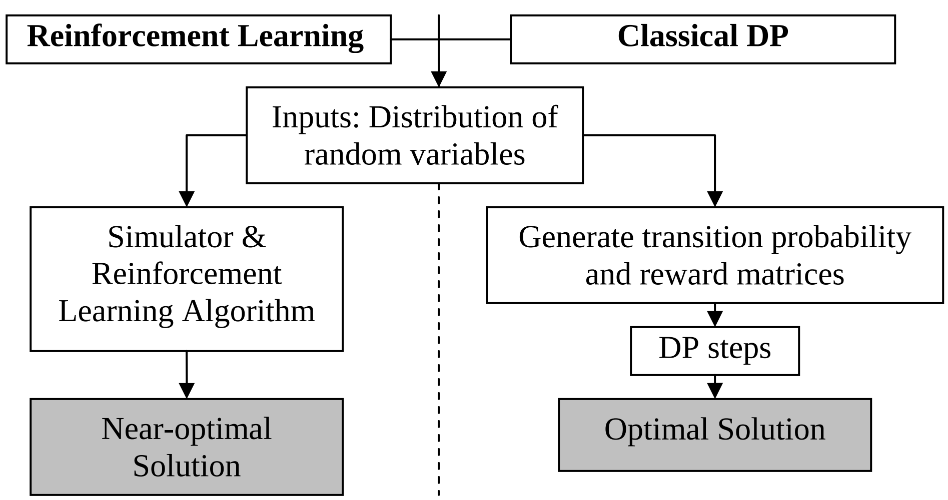

Figure 4.11: A Schematic Difference in RL and DP

Since computational effort is a primary concern for solution of the recruitment problem, I adapt the Q-Learning approach discussed by Gosavi (2003) with a modification to exclude visit factors in order to lessen the computational burden. First, I introduce the Q-value, $Q(state, action)$ or $Q(Y_t, a)$, that corresponds to value of each state-action pair.

The step-by-step procedure of the QBH method is shown as follows.

Q-Learning Based Heuristic (QBH) Procedure

<u>Step 0</u> Set the iteration number to 0.

<u>Step 1</u> Initialize time period, t, to 0 and starting state to Y_t. This represents the initial budget and initial willingness state of each agent.

<u>Step 2</u> Generate an action a using an <u>action selection heuristic</u>, described below.

<u>Step 3</u> Simulate action a to retrieve the next period action, Y_{t+1}. Let $r_t(Y_t, a, Y_{t+1})$ be the immediate reward earned in the transition to state Y_{t+1} from state Y_t under the influence of action a.

<u>Step 4</u> Update $Q(Y_t, a)$ using the following equation:

$$Q(Y_t, a) \leftarrow (1-\alpha)Q(Y_t, a) + \alpha[r_t(Y_t, a, Y_{t+1}) + \max_{b \in A(Y_{t+1})} Q(Y_{t+1}, b)], \quad 0 < \alpha \leq 1, \quad (4.10)$$

where $A(Y_{t+1})$ represent all possible actions in state Y_{t+1} and if $Q(Y_{t+1}, b)$ has no value, set its initial value is set to 0.

<u>Step 5</u> If $t = T$, increase the iteration number by 1 and go to step 1. Otherwise, increase t by 1 and go to step 2. If iteration number exceeds the limit go to step 6.

<u>Step 6</u> For each Y_t, select

$$A^*(Y_t) = \underbrace{\arg\max}_{b \in A(Y_t)} Q(Y_t, b). \quad (4.11)$$

The learning rate is represented by α in (4.10). Its value weights how much the previous value of $Q(Y_t, a)$ and the evaluation of immediate reward with future reward should affect the new value of $Q(Y_t, a)$. The Q-value is a prediction of the sum of the reinforcement one receives when performing the associated action and the following given policy. To update the prediction $Q(Y_t, a)$, one must perform the associated action a, causing a transition to the next state Y_{t+1}, and returning a scalar reinforcement $r_t(Y_t, a, Y_{t+1})$. Then one only needs to find the maximum Q-value in the new state, $\max_{b \in A(Y_{t+1})} Q(Y_{t+1}, b)$, to have all necessary information for revising the prediction (Q-Value) associated with the action just performed. Q-learning does not require one to calculate the transition probabilities to successor states. The reason is that a single sample or a successor state for a given action is an unbiased estimate of the expected value of the successor state.

The action selection heuristics in step 2 of the QBH method procedure are described as follows.

Action Selection Heuristics

Budget allocation for each agent represents the action a in step 2 of the QBH. It is very important to select an action wisely as this is the exploration part of the RL. Three heuristics are introduced. In the Q-Learning QBH procedure, one of the heuristics is randomly selected during each execution of action selection.

Heuristic 1: Random Allocation

In this heuristic, the remaining budget is allocated to a random set of agents at a random amount level.

Heuristic 2: High Willingness State Agent First

This heuristic focuses on allocating the remaining budget to those agents who have a higher chance of recruitment success. This may not be the best way to obtain the highest payoff because the agents with a high willingness state may generate smaller amount of Resource B collection volume compared to agents with a low willingness state who generate a higher amount of collection volume.

Heuristic 3: High Collection Volume Agent First

This heuristic focuses on allocating the remaining budget among those agents who generate higher amount of Resource B collection volume. This may not be the best way to obtain the highest payoff because agents with a higher collection volume may be very hard to recruit. In other words, recruiting many willing small agents may result in a higher amount of total collection volume.

The QBH uses the action selection heuristics to explore the action and state spaces. The exploitation applies (4.10) to update the Q-value for a state-action pair. According to Gosavi (2003), the Q-Learning method gives a near-optimal solution when the maximum number of iteration is large enough.

In order to perform a large number of iterations in a reasonable computation time, the computational complexity of the algorithm should be analyzed. In step 4, the number of steps required to update $Q(Y_t, a)$ in (4.10) requires first a search for the initial value of $Q(Y_t, a)$ and second the maximization of $Q(Y_{t+1}, b)$ for every value of $b \in A(Y_{t+1})$. The value look-up for $Q(Y_t, a)$ is performed in $O(\Omega)$ steps, where Ω is the size of a typically large Q-table. Q-table is a look-up table that stores the value of $Q(state, action)$ for every encountered state-action pair. This step takes $O(\Omega)\times$ number of actions at states Y_{t+1}, which is typically large. In summary, every computation of (4.10) in step 4 of the Q-Learning Based Heuristic requires:

$$\text{Time Complexity of (4.10)} = O(\Omega) * |A|. \tag{4.12}$$

Two modifications are introduced to speed up this step. The first is to set the learning rate α equal to 1 and the second is to store the Q-values using a hash table. Each of these modifications is described in the following paragraphs.

Learning Rate Equal to One

With α = 1, equation (4.10) becomes

$$Q(Y_t, a) \leftarrow r_t(Y_t, a, Y_{t+1}) + \max_{b \in A(Y_{t+1})} Q(Y_{t+1}, b). \tag{4.13}$$

Instead of storing $Q(Y_{t+1}, b)$ for every value of $b \in A(Y_{t+1})$ and searching for the maximum of $\max_{b \in A(Y_{t+1})} Q(Y_{t+1}, b)$ in every iteration, it is much simpler to store the maximum of $Q(Y_{t+1}, b)$ into $Q_{\max}(Y_{t+1}, b_{\max})$. Under this modification, the update of $Q(Y_t, a)$ becomes:

$$Q_{\max}(Y_t, a_{\max}) = \max[Q_{\max}(Y_t, a_{\max}), r_t(Y_t, a, Y_{t+1}) + Q_{\max}(Y_{t+1}, b_{\max})]. \tag{4.14}$$

Basically, retrieving $\max_{b \in A(Y_{t+1})} Q(Y_{t+1}, b)$ can be done in complexity of $O(\Omega)$ from looking up $Q_{\max}(Y_{t+1}, b_{\max})$. The update of $Q_{\max}(Y_t, a_{\max})$ is performed if the new value of $r_t(Y_t, a, Y_{t+1}) + Q_{\max}(Y_{t+1}, b_{\max})$ is higher than the previous value of $Q_{\max}(Y_t, a_{\max})$. In this step, the best action $a_{\max}$ is also updated accordingly.

In addition to this modification, the Hash Table data structure is applied to the QBH method. It is described as follows.

Hash Table

The QBH method requires a large size of Q-table in order to look up Q-values of corresponding states and actions. Computationally, it is time-consuming to retrieve the selected Q-value using a traditional array for the data-structure. As a better alterative, Hash Tables (Knuth 1973) are used as a Q-value data structure to improve the look-up time. There are two components in this data structure--an array (the hash table) and a mapping (the hash function). The hash function maps keys into hash values. Items stored in a hash table must have keys. In this case, the key corresponds to the state-action pair. The hash function maps the key of an item to a hash value, and that hash value is used as an index into the hash table for that item. This allows the Q-value to be inserted and located quickly. There are many hash function forms in the literature (Fox et. al 1992 and Knuth 1973). I implement one that converts a string into an integer value. The Q-value can be retrieved in complexity of $O(1)$ in the average case and best cases. The worst case search time is $O(\Omega)$; however, the probability of this happening is vanishingly small. This data structure technique does not have an impact on the solution quality of the QBH method. The procedure is the same. The only change is the retrieval time of the Q-value of any state-action pair in (4.10).

Employing the hash table data structure for Q-values and fixing the learning rate α to one, the computational complexity of (4.10) in step 4 of the QBH is reduced from $O(\Omega)*|A|$ to $O(1)$ in the average and best cases. In the worst case, it is $O(\Omega)$. This improvement reduces computational requirements for exploitations. Completely ignoring the previous value of Q-value by setting the learning rate α to one may affect the resulting quality of the heuristic solution, but the computational effort is significantly reduced to facilitate overall problem solution.

4.6 A Rolling IP with DP Heuristic

The previous section developed the Q-Learning Based Heuristic, or QBH, approach to solve the recruitment problem. As an alternate to the QBH, this section develops a heuristic called Rolling IP with DP to solve the recruitment problem. This heuristic provides a way to solve the large-scale recruitment problem within reasonable computation effort. The intuition behind this heuristic and the step-by-step procedure are presented in this section.

The heuristic is based on an observation of the DP algorithm described in section 4.4. Solving for an optimal recruitment policy for an individual agent using the DP algorithm requires a relatively small computational effort because the number of states and actions are small as shown in Table 4.1. Using this characteristic may be beneficial for solving the overall recruitment problem. The main concept of this heuristic is to shrink a multi-period problem so as to think of it a one-period problem. First, the optimal policy for each individual agent is solved for T periods. Then, all of these individual agent solutions are used to find the best combination of budget allocations among agents. The resulting solution is implemented for the first time period where the selected agents receive their given first period budget allocations. Next, the optimal policy for each individual agent is solved for the remaining $T-1$ periods. Then the procedure repeats itself until the final period is solved. The example of an agent for whom a recruitment budget is allocated in the first and second period is shown in Figure 4.12.

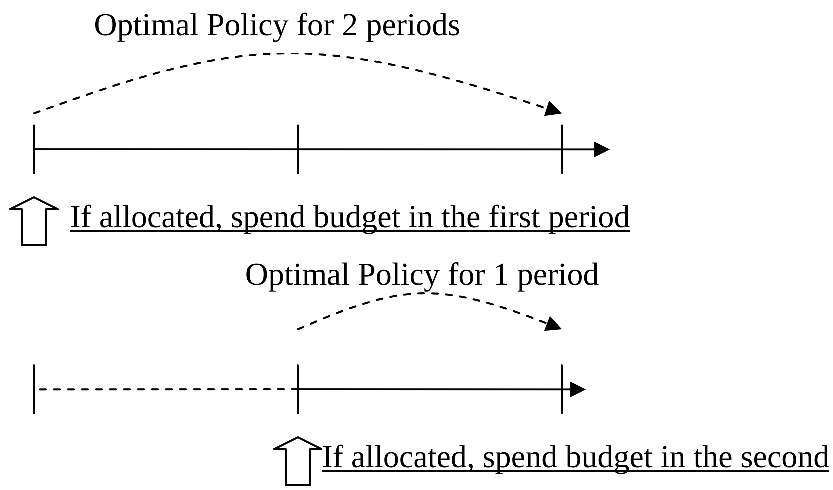

Figure 4.12: The Rolling Horizon Concept

A key step in this approach is an optimization problem that selects the best combination of individual policies to use to maximize the collective recruited agents' collection volume, subject to the overall recruitment budget constraint. This approach is suboptimal because it does not take advantage of the ability to observe and respond to recruitment during the policy execution. To improve performance, a rolling horizon implementation is applied. The remaining unspent funds allocated to those agents who have been recruited and unspent funds allocated to retailers not recruited are added back to the available recruitment budget amount and then the optimization problem is resolved with the updated information.

The reason why this approach is suboptimal is because it does not take action based on the information about how the retailer responds to the expenditures. In other words, it allocates a budget to be spent for the entire period on the retailer and does not account for the ability to reallocate money from among retailers who are recruited early on in the process. It is important to emphasize that the recruitment outcome is a random outcome.

Denote the stochastic recruitment function $SR(i, B^{\max}, t)$ as a function that returns the solution from solving the recruitment problem with the DP algorithm (developed in section 4.4) for retailer i for t periods given the starting total budget $B^{\max}$. The solution yields the optimal budget allocation policy in each period and the expected collection volume from retailer i over t periods. Next, the optimization problem that selects the best combination of individual policies among all retailers is formulated. The index, parameters, and variables are defined as:

Index:

i	Index of retailers	(i = 1, 2, …, η_m)
j	Index of budget amounts	(j = 1, 2, …, J)

Parameters:

$B^{\max}$	Maximum starting total budget over total T periods
B_t^{start}	Maximum starting budget at period t
b_j	Budget allocation the collector choose to spend on the retailer, which is the value of j^{th} entry in $B = (b_1, ..., b_j, ..., b_J)$
$\tilde{v}_{ij}$	Maximum expected increment of capacity volume that can be collected from retailer i if budget amount b_j is allocated to that retailer. The value of $\tilde{v}_{ij}$ can be obtained from solving $SR(i, b_j, t)$ using the DP approach.

Variables:

$$x_{ij} = \begin{cases} 1 & \text{if budget amount } b_j \text{ is allocated to retailer } i \\ 0 & \text{Otherwise} \end{cases}$$

The integer programming problem (IP) called Rolling IP for period t can be formulated as:

Rolling IP (RP_t) for Period t

Maximize $$\sum_i \sum_j \tilde{v}_{ij} x_{ij} \tag{4.15}$$

Subject to: $$\sum_j x_{ij} \le 1 \qquad \forall i \tag{4.16}$$

$$\sum_i \sum_j b_j x_{ij} \le B_t^{start} \tag{4.17}$$

$$x_{ij} = \{0,1\} \qquad \forall i, j . \tag{4.18}$$

The objective function (4.15) is the sum of collection volume. Constraints (4.16) permit only one budget amount to be allocated to retailer i. Constraint (4.17) restricts the overall spending budget to be less than the budget limit. Constraints (4.18) force x_{ij} variables as binary variables.

The procedure for the Rolling IP with DP method is discussed next by combining the Rolling IP formulation together with the rolling horizon concept.

<u>Rolling IP with DP Heuristic (RIDH) Solution Procedure</u>

<u>Step 0</u> Set $t = 1$ and $B_0^{start} = B^{\max}$.

Solve for $\tilde{v}_{ij}$ from $SR(i, b_j, T)$ as defined earlier in this section for all i, j using the DP approach developed in section 4.4. The initial state of $SR(i, b_j, T)$ is [0, initial willingness state of retailer i, b_j].

<u>Step 1</u> Formulate the rolling IP (RP_t) model and solve for x_{ij}.

<u>Step 2</u> For the retailers for where a recruiting budget has been allocated, simulate the action in period t only.

If $t = T$, obtain the total increment in collection volume from period 1 to period T and exit. Otherwise, go to Step 3.

<u>Step 3</u> Set $t = t + 1$.

Update the value of $\tilde{v}_{ij}$ from for all i, j. Note that there is no need to resolve MDP for each retailer. Obtain the $\tilde{v}_{ij}$ by changing the starting initial state to [t+1, new willingness state, remaining budget]. For example, if the initial state is [0,M,30], a budget amount 10 is applied to this period, the next period status change to H, and the overall remaining budget is 10, then $\tilde{v}_{ij}$ can be looked up from state [1,H,10].

Update remaining budget B_t^{start}, ($B_t^{start} = B_{t-1}^{start} -$ actual budget spent in the previous period).

Go to Step 1

These steps can be summarized by the flow chart Figure 4.13.

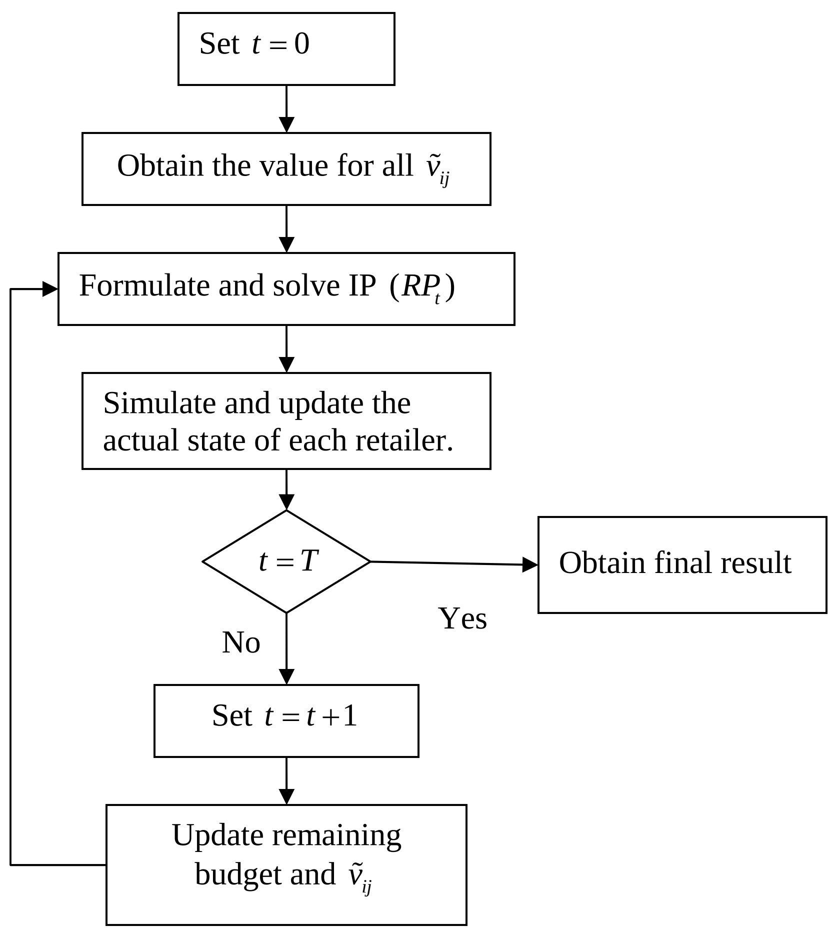

Figure 4.13: Procedure for the RIDH Solution Approach

In this chapter, the general framework for the recruitment problem is developed. Using the developed framework, the recruitment problem is posed as the stochastic dynamic programming. The DP algorithm is proposed to obtain the optimal policy. In addition, two heuristics are introduced in order to solve the recruitment problem for the large-scale problem that the DP method cannot solve with reasonable computational effort. The two heuristics' performances will be tested with small and large scale examples in Chapter 5. The insights and potential algorithm improvements are also developed based upon the computational testing.

CHAPTER 5

NUMERICAL STUDY ON THE

TACTICAL LEVEL PROBLEM

In Chapter 4 I introduced the conceptual model for the tactical level problem and provided an exact method, and two heuristics, to determine the recruiting decisions in each time period. In this chapter, the solution approaches are applied to small and large examples. Since the Dynamic Programming (DP) algorithm can find an optimal solution for the small example in a reasonable computation time, its solution can be used as a benchmark against the solutions obtained by the Rolling IP with DP heuristic (RIDH), and Q-learning based heuristic (QBH). For the large example, the computational requirements are prohibitive for the DP algorithm. Thus, only the results from the two heuristics are compared. In addition, the performance of the QBH is investigated by varying the computational effort and using problem decompositions motivated by the structure of the problem. I also study how incorrect information from the retailers may affect the performance of the recruitment model. All the computation experiments in this chapter and Chapter 7 are solved using a Windows 2000-based Pentium 4 1.80 GHz personal computer with 640MB of RAM with CPLEX version 8.0 (www.ilog.com) for the optimization software.

5.1 Small Example

For the small example, I examine a recruitment problem with five retailers. Each retailer i has a different amount of collection volume, g_i, that it can generate in one period or one month. The recruitment budget threshold, μ_i, is assumed to have a positive correlation with g_i. Hence, a higher amount of collection volume g_i implies higher recruitment budget threshold value μ_i.

Throughout the numerical study in this chapter, and Chapter 7, I use the following numerical parameters. The recruitment willingness factor for the willingness state of each retailer is defined as $\beta_H = 2$, $\beta_M = 1$, and $\beta_L = 0.5$. Equation (4.1) is used to compute the probabilities of recruitment, ($Pr_{LR}(a)$, $Pr_{MR}(a)$, $Pr_{HR}(a)$), for the given willingness state and budget allocation (a). These probabilities also depend on the recruitment budget threshold of the retailer. When the given budget allocation fails to recruit the retailer, two cases are considered.

Case A: If $a \geq \mu_i$, the transition probabilities are depicted in Figure 5.1.

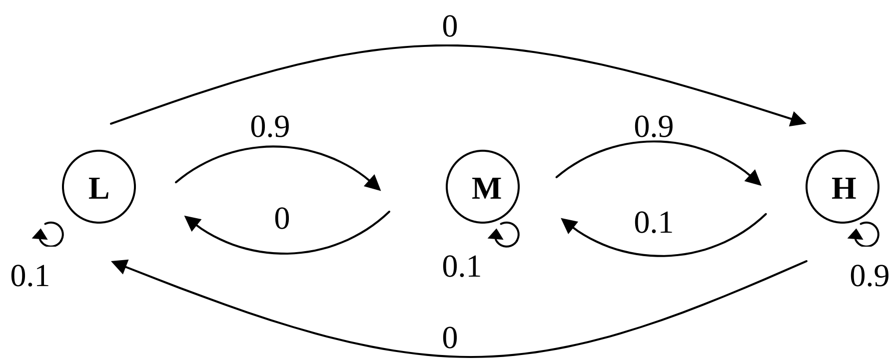

Figure 5.1: Transition Probabilities for Case A

Case B: If $0 \le a < \mu_i$, the transition probabilities are depicted in Figure 5.2.

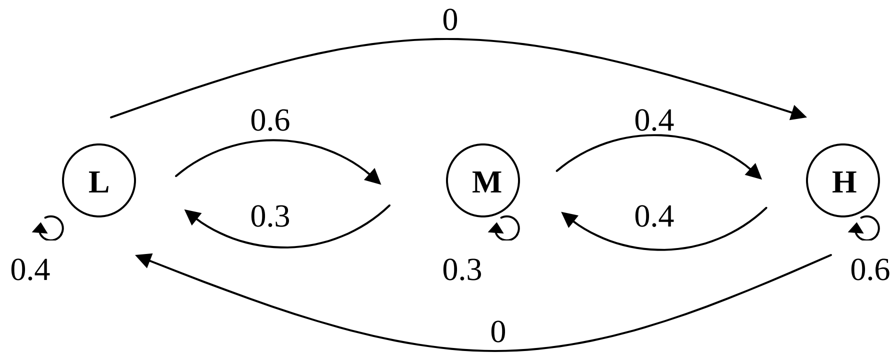

Figure 5.2: Transition Probabilities for Case B

These two cases describe the situation that if the recruiter allocates enough recruitment budget amounts to the retailer, it is more likely to move the retailer to a more favorable state, even though it fails to recruit the retailer. On the other hand, if the recruiter does not allocate a large enough budget amount to the retailer, it is less likely to move the retailer to a more favorable state, even though it fails to recruit the retailer. I assume that all retailers follow these transition probabilities when the given budget allocation fails to recruit the retailer.

With these settings, I generate five test cases that have different retailers' initial willingness states. Table 5.1 shows the amount of collection volume and recruitment budget threshold of each retailer. The alternative budget limitation settings are spaced 10 units apart 10, 20, …, 100 and the budget allocation settings are similarly spaced. The number of time periods is chosen to be three. From the collective retailer collection volumes, the maximum system collection volume in all these five cases is 220 pounds.

Table 5.1: Small Example Data

Retailer	Collection Volume (lb.)	Recruitment Budget Threshold
1	10	5
2	30	15
3	70	49
4	20	20
5	90	81

To test the algorithms, several cases are evaluated where each case differ from the others based on the initial states of the retailers. Of these cases is described in the paragraph below.

<u>**Case 5.1:** All retailers start in state L</u>

In this case, all of the retailers' initial willingness states are set to the low (L) value. The probability of recruitment is computed with $\beta_L = 0.5$. Table 5.2 displays the solution average collection volume, computation time, and optimality gap for solution approaches DP, RIDH, and QBH for different maximum budget settings. The average collection volume is computed from the results obtained by applying the policy resulting from the different solution methods for 100 replications. The optimality gap illustrates the solution quality found by the RIDH and QBH methods compared to the optimal solution obtained by the DP algorithm. For the DP algorithm, the computation time is obtained by examining every possible state and action in every period and selecting the policy that yields the maximum average collection volume. For the QBH solution approach, the maximum number of iteration is set to 100,000.

Table 5.2: Average Volume Collection, Solution Time and Optimality Gap for Case 5.1

	Solution Approaches							
	DP		RIDH			QBH		
Maximum Budget	Average Collection Volume (lb.)	Solution Time (sec.)	Average Collection Volume (lb.)	Solution Time (sec.)	Optimality Gap (%)	Average Collection Volume (lb.)	Solution Time (sec.)	Optimality Gap (%)
10	12.4	16	12.4	5	0.0	11.6	6	6.4
20	27.6	118	28.5	7	0.0	21.5	6	22.1
30	46.3	568	41.4	9	10.5	40.4	7	12.7
40	76.8	2,086	80.4	14	0.0	8.0	8	89.5
50	90.1	8,282	91.0	19	0.0	27.9	9	69.0
60	106.7	61,229	103.0	23	3.4	75.9	9	28.8
70	-	86,400[8]	114.7	32	-	92.8	10	-
80	-	86,400	138.1	41	-	92.1	11	-
90	-	86,400	153.8	59	-	93.0	11	-
100	-	86,400	165.2	65	-	95.6	11	-

The solution averages for the collection volume obtained by the RIDH method are close to the optimal solution for every maximum budget setting. The largest optimality gap is only 10.5%. For maximum budget settings of 20, 40, and 50, the average solution for collection volume found by the RIDH approach happens to be slightly higher than the value found by the DP approach because of the random numerical evaluation found by simulating 100 replications. For this situation, I set the optimality gap to zero.

The computation time requirements for the RIDH approach are much smaller than those from the DP method. The QBH method requires the least amount of solution time for every maximum budget setting, but the optimality gap is larger than that found by the RIDH approach. In fact, its solution is worse than solution obtained by the RIDH approach for every maximum budget setting. For maximum budget settings of 80 to 100, the DP method cannot obtain optimal policy within the stopping time limit of one day.

[8] Algorithm was stopped when the computation time requirement reached 86,400 seconds or one day.

Case 5.2: <u>All retailers start in state M</u>

In this case, all of the retailers' initial willingness states are set to the medium (M) value. The probability of recruitment is computed with $\beta_M = 1$. Table 5.3 displays the solution average collection volume, computation time, and optimality gap for solution approaches DP, RIDH, and QBH for different maximum budget settings.

Table 5.3: Average Volume Collection, Solution Time and Optimality Gap for Case 5.2

	Solution Approaches							
	DP		**RIDH**			**QBH**		
Maximum Budget	Average Collection Volume (lb.)	Solution Time (sec.)	Average Collection Volume (lb.)	Solution Time (sec.)	Optimality Gap (%)	Average Collection Volume (lb.)	Solution Time (sec.)	Optimality Gap (%)
10	12.2	1	12.6	4	0.0	1.5	6	87.7
20	47.2	8	44.4	6	5.9	5.6	6	88.1
30	84.8	40	82.6	8	2.5	16.3	6	80.7
40	100.6	138	99.1	11	1.4	56.7	7	43.6
50	131.4	393	130.7	15	0.5	52.1	8	60.3
60	158.9	962	155.5	16	2.1	45.4	8	71.4
70	175.3	2,097	165.4	23	5.6	82.3	9	53.0
80	187.2	4,419	180.1	26	3.7	84.7	9	54.7
90	201.5	8,417	197.5	38	1.9	117.9	9	41.4
100	210.0	15,309	201.4	41	4.1	181.3	10	13.6

The overall results follow the same trends as Case 5.1. For the same maximum budget setting, the average solution's collection volume in this case is higher than one in Case 5.1 because the retailers in Case 5.2 start in more favorable states than ones in Case 5.1. Furthermore, the computation time requirements are less in this case because the probability of the retailer moving back to state L is small.

Case 5.3: <u>All retailers start in state H</u>

In this case, all of the retailers' initial willingness states are set to the high (H) value. The probability of recruitment is computed with $\beta_H = 2$. Table 5.4 displays the solution average collection volume, computation time, and optimality gap for solution approaches DP, RIDH, and QBH for different maximum budget settings.

The overall results follow the same trends as in Cases 5.1 and 5.2. For the same budget limit, the average solution's collection volume in this case is higher than ones in Cases 5.1 and 5.2 because the retailers in Case 5.3 start with highest favorable states compared to the ones in Cases 5.1 and 5.2. Furthermore, the computation time requirements are less in this case. When the maximum budget equals 70, every retailer can be recruited into the system. It is interesting to see that the QBH approach performs almost as well as the DP algorithm when the budget limit is equal or greater than 60.

Table 5.4: Average Volume Collection, Solution Time and Optimality Gap for Case 5.3

	Solution Approaches							
	DP		**RIDH**			**QBH**		
Maximum Budget	Average Collection Volume (lb.)	Solution Time (sec.)	Average Collection Volume (lb.)	Solution Time (sec.)	Optimality Gap (%)	Average Collection Volume (lb.)	Solution Time (sec.)	Optimality Gap (%)
10	44.1	1	46.9	5	0.0	5.4	6	87.7
20	90.0	1	90.0	3	0.0	54.9	6	39.0
30	137.7	2	139.7	5	0.0	63.1	6	54.1
40	171.1	5	166.9	6	2.4	84.8	7	50.4
50	199.1	11	189.4	6	4.8	100.8	7	49.3
60	211.5	27	210.0	10	0.7	194.8	8	7.9
70	220.0	59	220.0	9	0.0	205.3	8	6.6
80	220.0	116	220.0	16	0.0	220.0	8	0.0
90	220.0	215	220.0	22	0.0	220.0	9	0.0
100	220.0	377	220.0	24	0.0	220.0	9	0.0

Case 5.4: Retailers start in state MHHMH

In this case, retailers' initial willingness states are randomly assigned to the setting 'MHHMH' respectively.

Table 5.5 displays solution average collection volume, computation time, and optimality gap for solution approaches DP, RIDH, and QBH for different maximum budget settings.

Table 5.5: Average Volume Collection, Solution Time and Optimality Gap for Case 5.4

	Solution Approaches							
	DP		RIDH			QBH		
Maximum Budget	Average Collection Volume (lb.)	Solution Time (sec.)	Average Collection Volume (lb.)	Solution Time (sec.)	Optimality Gap (%)	Average Collection Volume (lb.)	Solution Time (sec.)	Optimality Gap (%)
10	46.2	1	47.5	4	0.0	30.0	6	35.0
20	90.0	1	84.8	4	5.7	50.0	6	44.4
30	138.3	4	139.2	5	0.0	60.0	7	56.6
40	177.7	15	173.6	6	2.3	64.6	7	63.6
50	178.2	44	192.5	8	0.0	77.8	7	56.3
60	199.1	103	199.1	10	0.0	90.1	8	54.7
70	212.5	226	213.4	12	0.0	111.0	8	47.7
80	219.9	468	219.8	19	0.1	127.3	8	42.1
90	220.0	863	220.0	24	0.0	168.4	9	23.4
100	220.0	1,615	220.0	26	0.0	150.5	9	31.5

In this case, the quality of the solution obtained by the RIDH approach is almost as good as the solution obtained by the DP algorithm for every maximum budget setting. The QBH approach does not provide very good solutions for this case.

The results from the small example show that the RIDH solution approach performs almost as well as the optimal DP procedure in all cases, with much lower computational effort. The QBH method solves the small example recruitment problem with the least computational effort, but yields the worst average collection volume solutions compared to the DP and RIDH methods. Next, I attempt to solve a larger problem.

5.2 Large Example

To structure the data in the large example, I classify the retailer into different groups according to its collection volume and recruitment difficulty. The recruitment difficulty for the retailer is set according to its initial willingness state: low, medium, or high. For collection volume, I categorize the retailer into three groups: small (0-30 lbs/month), mid-size (40-60 lbs/month), and large (70-90 lbs/month). Table 5.6 displays the solution average volume collected for the one-retailer recruitment that is solved using the DP algorithm with the different collection volumes, initial willingness states and maximum recruitment budget settings

Table 5.6: Average Collection Volume Obtained For Different Collection Volumes, Initial Willingness States and Maximum Recruitment Budget Settings for One Retailer

			Maximum Recruitment Budget Amount									
Collection Volume (lbs.)	Collection Volume Type	Initial State	10	20	30	40	50	60	70	80	90	100
10	Small	L	9.4	10	10	10	10	10	10	10	10	10
10	Small	M	9.9	10	10	10	10	10	10	10	10	10
10	Small	H	10	10	10	10	10	10	10	10	10	10
20	Small	L	10.6	19.8	20	20	20	20	20	20	20	20
20	Small	M	13	20	20	20	20	20	20	20	20	20
20	Small	H	20	20	20	20	20	20	20	20	20	20
30	Small	L	0	27.3	30	30	30	30	30	30	30	30
30	Small	M	0.3	30	30	30	30	30	30	30	30	30
30	Small	H	30	30	30	30	30	30	30	30	30	30
40	Mid-size	L	0	22	40	40	40	40	40	40	40	40
40	Mid-size	M	0	35.2	40	40	40	40	40	40	40	40
40	Mid-size	H	40	40	40	40	40	40	40	40	40	40
50	Mid-size	L	0	0	46	50	50	50	50	50	50	50
50	Mid-size	M	0	42	49.5	50	50	50	50	50	50	50
50	Mid-size	H	50	50	50	50	50	50	50	50	50	50
60	Mid-size	L	0	0	49.8	53.4	60	60	60	60	60	60
60	Mid-size	M	0	52.8	60	60	60	60	60	60	60	60
60	Mid-size	H	60	60	60	60	60	60	60	60	60	60
70	Large	L	0	0	37.1	52.5	57.4	70	70	70	70	70
70	Large	M	0	46.2	65.8	69.3	70	70	70	70	70	70
70	Large	H	42.7	70	70	70	70	70	70	70	70	70

Table 5.6 (contintued)

Collection Volume (lbs.)	Collection Volume Type	Initial State	Maximum Recruitment Budget Amount 10	20	30	40	50	60	70	80	90	100
80	Large	L	0	0	15.2	64	67.2	69.6	76.8	80	80	80
80	Large	M	0	13.6	70.4	80	80	80	80	80	80	80
80	Large	H	12	80	80	80	80	80	80	80	80	80
90	Large	L	0	0	3.6	72	75.6	78.3	81.1	83.5	89	90
90	Large	M	0	4.5	79.2	89.1	90	90	90	90	90	90
90	Large	H	2.7	90	90	90	90	90	90	90	90	90

The results show that if the retailer's collection volume is large and its initial willingness is L, this is the hardest situation for recruitment with a low budget as shown in Figure 5.3. On the other hand, if the collection volume is small, the initial state does not have much effect on the average collection volume as shown in Figure 5.4. For a small budget amount, say 20, it is very likely that a retailer in willingness state H will be recruited.

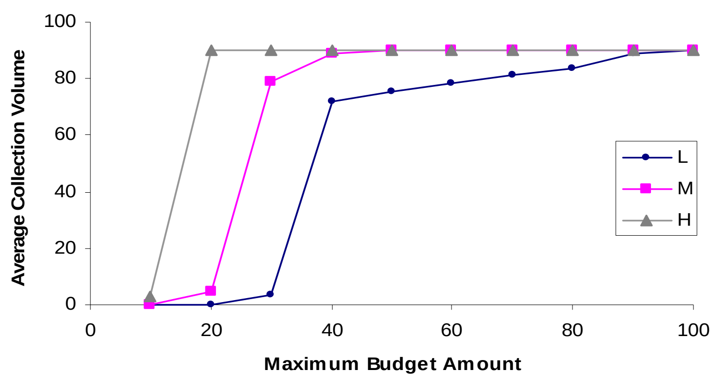

Figure 5.3: Budget Allocation and Average Collection Volume of Different Initial State for Retailer with Collection volume 90

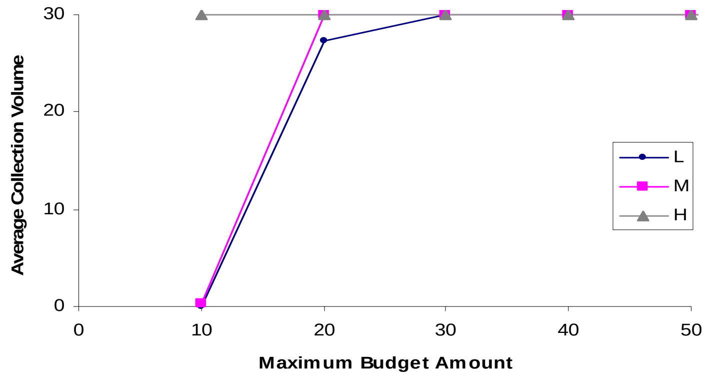

Figure 5.4: Budget Allocation and Average Collection Volume of Different Initial State for Retailer with Collection Volume 30

Using the results in Table 5.6, I construct three cases to examine in the large example. Case 5.5 consists of a different number of retailers (5,10,15,20) with large collection volumes and all retailers starting in willingness state L. The solution results are shown in Table 5.7. Case 5.6 consists of a different number of retailers (5,10,15,20) with small and mid-size collection volumes and all retailers starting in willingness state M. The solution results are shown in Table 5.8. Case 5.7 consists of a different number of retailers (5,10,15,20) with large collection volumes and all retailers starting in willingness state H. The solution results are shown in Table 5.9.

Table 5.7: Results for Case 5.5

		Solution Approaches					
		RIDH			QBH		
Number of Retailers	Maximum Budget	Average Collection Volume (lb.)	Best Collection Volume(lb.)	Solution Time (sec.)	Average Collection Volume(lb.)	Best Collection Volume (lb.)	Solution Time (sec.)
5	50	66.3	90	9	6.1	90	6
	100	146.4	230	16	112.6	150	7
	150	242	320	31	141.6	220	9
	200	325.8	390	52	187.7	240	10
10	50	76.5	90	12	8.5	90	11
	100	162	230	25	85.4	140	16
	150	251.3	340	56	81	170	26
	200	347.5	430	114	193.2	230	32
15	50	74.7	90	16	1.6	80	17
	100	152.5	230	33	93.8	140	30
	150	261.7	340	82	86.7	170	54
	200	362.8	450	165	196.8	230	65
20	50	74.7	90	16	12.8	80	23
	100	147.5	230	41	93.1	140	45
	150	263.1	340	124	114.1	210	74
	200	359	450	191	192.4	230	85

Table 5.8: Results for Case 5.6

		Solution Approaches					
		RIDH			QBH		
Number of Retailers	Maximum Budget	Average Collection Volume (lb.)	Best Collection Volume(lb.)	Solution Time (sec.)	Average Collection Volume(lb.)	Best Collection Volume (lb.)	Solution Time (sec.)
5	50	98.2	110	8	79.8	110	8
	100	144.6	150	9	141.6	150	9
	150	150	150	19	150	150	8
	200	150	150	31	150	150	9
10	50	117.8	130	9	94.6	120	13
	100	207	230	17	134.8	180	20
	150	274.3	300	33	210.8	250	26
	200	302.7	310	62	271.4	310	27
15	50	122.2	140	12	73.6	120	21
	100	235.9	260	23	146	180	35
	150	323.6	350	49	223.8	260	51
	200	399.1	430	96	298	340	56
20	50	127.2	140	12	103.5	110	30
	100	260.1	280	27	148	190	51
	150	360.1	390	63	212.2	270	66
	200	451.7	490	114	327.6	330	75

Table 5.9: Results for Case 5.7

		Solution Approaches					
		RIDH			QBH		
Number of Retailers	Maximum Budget	Average Collection Volume (lb.)	Best Collection Volume(lb.)	Solution Time (sec.)	Average Collection Volume(lb.)	Best Collection Volume (lb.)	Solution Time (sec.)
5	50	218.5	310	6	155.6	320	6
	100	387.9	390	6	317.5	390	7
	150	390	390	15	381.6	390	9
	200	390	390	21	390	390	9
10	50	231.7	300	8	156.6	330	11
	100	462.7	550	11	349.3	530	13
	150	661.3	790	21	528.5	560	16
	200	788.6	790	33	600.4	720	16
15	50	241.5	350	10	122.2	250	15
	100	465.9	550	14	412.3	500	18
	150	688.7	810	27	464.6	570	23
	200	879.7	970	51	736.4	790	26
20	50	248.5	350	10	188.1	320	20
	100	473.4	600	18	425.6	480	25
	150	695.3	860	36	586.8	680	37
	200	908.6	1040	62	667.3	760	43

The results from Cases 5.5, 5.6, and 5.7, which are displayed in Tables Table 5.7, Table 5.8, and Table 5.9, show that the RIDH approach outperforms the QBH method in every case. The best collection volume represents the largest collection volume that the solution method has found so far and has set as a target to achieve. Even though the QBH approach requires less computational effort to obtain the resultant policy, its average solution collection volume is dominated by the one obtained by the RIDH approach, which also provides higher best collection volumes than the one obtained by the QBH method. Next, I examine how the maximum number of iterations and decomposition methods in the QBH approach may affect the solution quality.

5.3 Q-Learning Based Heuristics Performance

In this section, I study how the performance of QBH solution approach can be affected when the maximum number of iteration allowed changes. Also, decomposition ideas are examined in order to explore how dividing the large problem into smaller problems may affect the performance of QBH.

5.3.1 Maximum Number of Iterations

One parameter that may affect the performance of the QBH approach is the maximum number of iterations permitted. Of course, a larger number of iterations require more computational effort. To test the impact, I apply the QBH approach to two different data sets. The first test case uses the data from Case 5.2 with a maximum budget 50. Table 5.10 displays the solution results for different settings of the maximum number of iterations parameter including required computation time, number of state-action pairs, average collection volume obtained by simulating the resultant policy, and the best collection volume so far. The best collection volume represents the largest collection volume that the QBH method has found so far and has set as a target to achieve.

Table 5.10: QBH Results on Case 5.2 with 5 Retailers and Maximum Budget 50

Maximum Number of iterations	Solution Time (sec.)	Number of State-action Pairs	Average Collection Volume (lb.)	Best Collection Volume (lb.)
100	0.1	71	51.1	70.0
1,000	0.1	183	88.5	110.0
10,000	0.6	253	46.2	120.0
100,000	5.9	339	52.1	140.0
1,000,000	59.0	391	42.2	160.0

Figure 5.5 illustrates the relationship between number of state-action pairs and required computation time to solve Case 5.2 using the QBH method. The number of state-action pairs increases quickly as the iteration limit begins to grow. The problem size is small enough that the number of state-action pairs seems to converge at about one million iterations. When the iteration limit increases, there is also an increase in the required computation time, the number of state-action pairs, and the best collection volume.

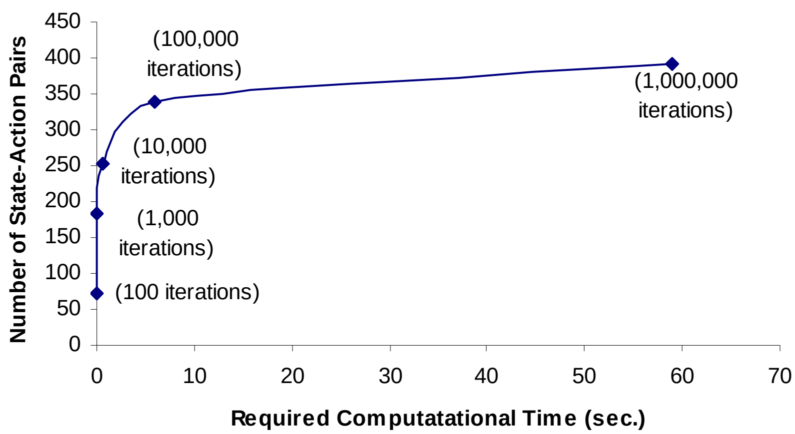

Figure 5.5: Number of State-Action Pairs and Required Computational Time for Solution Obtained by the QBH Method for Case 5.2 with 5 Retailers and Recruitment Budget 50

Figure 5.6 displays two important data sets. The average collection volume is computed as the average result obtained from applying the resultant policy of the QBH procedure 100 times. The best collection volume is the target that the QBH approach is trying to achieve from the policy. As the maximum number of iterations rises, the QBH algorithm explores a larger number of state-action pairs. Consequently, the best collection volume increases because it has more opportunities to find a better collection volume.

Nevertheless, in Figure 5.6, the average solution collection volumes do not approach the best collection volume at the higher number of iterations settings. This may be the result of setting the learning rate to one in QBH's update equation. With this setting, I gain the benefit of reducing the computation time and allowing the QBH method to explore a larger number of state-action pairs. However, I lose the ability to exploit action paths that return good solutions. The QBH approach chooses a riskier path with high return myopically but fails to determine the best move in the next period if the retailer's state does not change to the expected state.

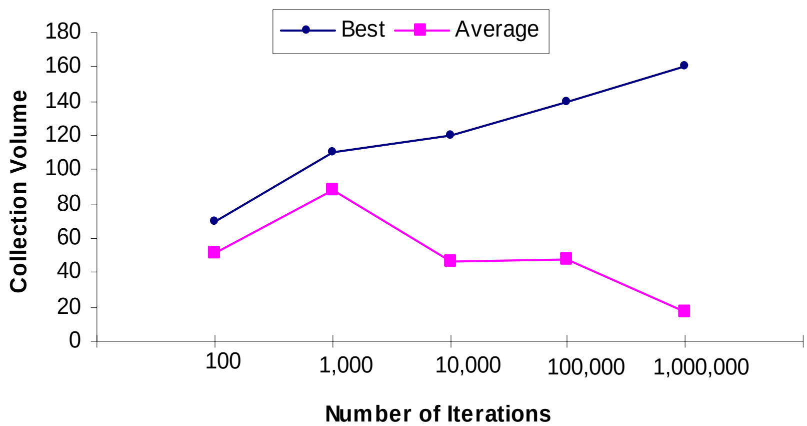

Figure 5.6: Average and Best Collection Volume for Solution Obtained by the QBH Method for Case 5.2 with 5 Retailers and Recruitment Budget 50

Next, I examine the QBH on a larger problem. The second test case replicates each retailer's information in Case 5.2 twice. Hence, there are total of 10 retailers and total budget is 100. Table 5.11 displays the solution results for different settings of the maximum number of iterations parameter.

Table 5.11: QBH Results on Case 5.2 with 10 Retailers and Maximum Budget 100

Maximum Number of iterations	Solution Time (sec.)	Number of State-action Pairs	Average Collection Volume (lb.)	Best Collection Volume (lb.)
100	0.02	160	110.0	110.0
1,000	0.09	819	119.7	120.0
10,000	0.95	3,287	151.0	160.0
100,000	9.70	9,263	105.9	190.0
1,000,000	106.00	19,183	98.7	200.0

Figure 5.7 illustrates the relationship between number of state-action pairs and computation time for the solutions obtained by the QBH method for Case 5.2 with 10 retailers and recruitment budget of 100. The problem size is so large that the number of state-action pairs fails to converge even at one million iterations. The computation time in this case is almost twice the time it requires in the previous study.

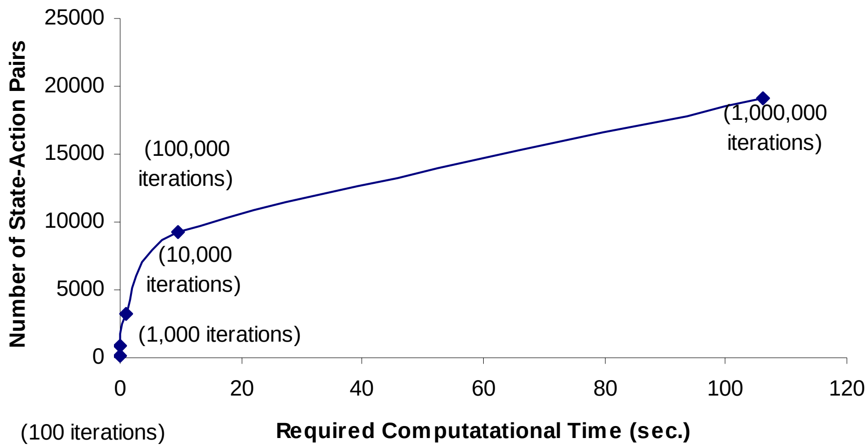

Figure 5.7: Number of State-Action Pairs and Required Computational Time for Solution Obtained by the QBH Method for Case 5.2 with 10 Retailers and Recruitment Budget 100

Figure 5.7 shows that the average solution collection volumes for the solution obtained by the QBH method tend to deviate from the best collection volume at the higher maximum number of iterations. However, at the lower best collection volume, the QBH method achieves the best collection volume. A possible explanation is provided by the fact that the probability of recruitment of a retailer with a lower collection volume is higher than one with a higher collection volume. Aiming to recruit a retailer with a small collection volume indicates two alternative strategies of budget allocation. In one strategy, the budget allocation is concentrated on a small subset of retailers who have high collection volumes. In this strategy, each retailer's budget allocation is large enough to increase the probability of recruitment. Alternatively, in a second strategy, the budget allocation is spread over a large set of retailers who have low collection volumes. In this strategy, each retailer's budget allocation is small but large enough to keep the probability of recruitment high for those retailers. As the QBH method explores a large number of state-action pairs, and aims for a higher collection volume, the recruitment budget is allocated among the retailers who have high collection volumes (retailer 3 and retailer 5 in Table 5.1).

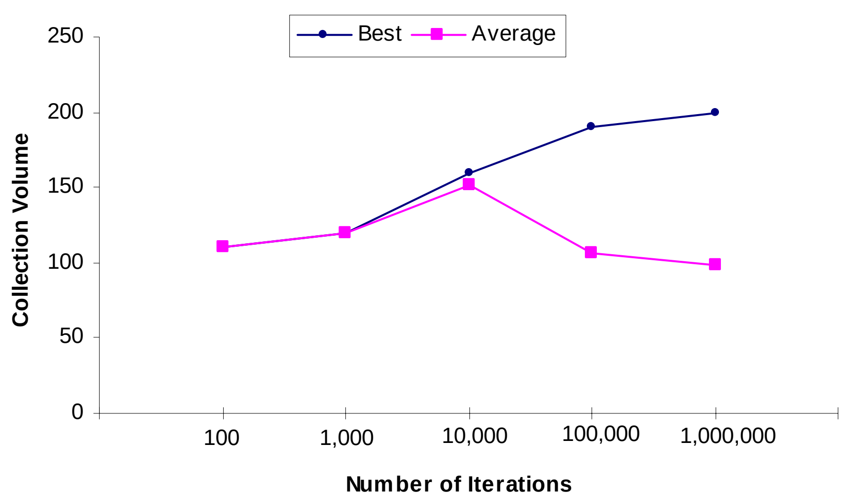

Figure 5.8: Average and Best Collection Volume for Solution Obtained by the QBH Method for Case 5.2 with 10 Retailers and Recruitment Budget 100

Two conclusions can be drawn from this study. First, the QBH approach can find better best collection volume when the maximum number of iterations increases. The reason is that the QBH method explores more number of state-action pairs. However, the average solution collection volume does not always increase when the maximum number of iterations increases. Second, when the problem size is large (10 retailers or more), the number of state-action pairs is too large for the QBH method to explore them adequately and may prohibit the ability to find good solutions. This motivates trying to find ways to keep the problem small by using decomposition.

5.3.2 Decomposition Ideas

For large scale problems, I attempt to improve the solution quality by decomposing the problem into many small subproblems. I divide the budget limit evenly among subproblems and solve each one individually. The overall solution is composed as the union of the solutions for each subproblem. With a smaller problem size for each subproblem, less total computation effort should be required. I define number of retailers as N, the number of retailers in each subproblem as $\tilde{n}$, and the number of sub problems as $\gamma = \left\lceil \frac{N}{\tilde{n}} \right\rceil$. I explore three different decomposition ideas: random, symmetric and sorted decomposition. Three test cases are generated and solved with different decomposition methods, and each subproblem is solved using the QBH method.

1. Random Decomposition

In this decomposition, the retailers are randomly divided into subproblems. Each subproblem has $\tilde{n}$ retailers.

2. Symmetric Decomposition

The retailers are divided into many subproblems such that the sum of all the retailers' collection volumes for each subproblem is about the same. The dividing procedure starts by sorting the retailers in increasing order by the amount of collection volume. Then, the first subproblem includes $(k\gamma+1)^{th}, k=0,1,...,\tilde{n}-1$ retailers on the list. This step is repeated until the γ^{th} subproblem which includes $(k\gamma+\gamma)^{th}, k=0,1,...,\tilde{n}-1$ retailers. Figure 5.9 depicts the symmetric decomposition scheme.

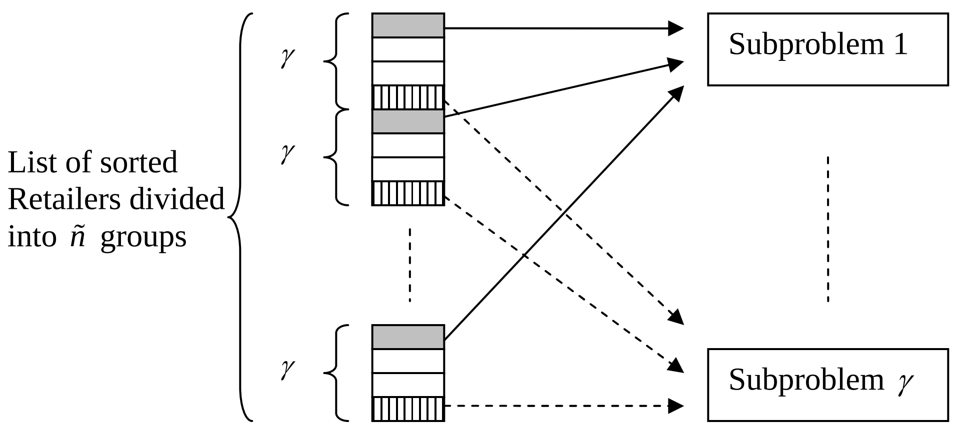

Figure 5.9: Symmetric Decomposition

3. Sorted Decomposition

First, the retailers are sorted in increasing order by the amount of collection volume. Then, the first subproblem consists of the first $\tilde{n}$ retailers. The second subproblem consists of the next $\tilde{n}$ retailers. This procedure is repeated until all retailers are included in a subproblem. The first subproblem groups the retailers with low collection volume together while the last subproblem groups the retailers with high collection volume together.

For the test problem, the collection volume for each retailer is generated according to the lognormal distribution with mean 1.2 and standard deviation 0.5. Limpert *et al.* (2001) show that this distribution has been used for many science applications. Log-normal distributions have proven useful as distributions for rainfall amounts, for the size distributions of aerosol particles or droplets, and for many other cases. The lognormal distribution of for collection volume seems to be reasonable from an empirical standpoint. One feature of the log-normal distribution is that it is positive-definite, so it is often useful for representing quantities that cannot have negative values.

For the test problem, the number of retailers is 40. I consider three cases with different initial willingness states. Each subproblem is solved using the QBH method.

Case 5.8: Decomposition of 40 retailers with low initial willingness states L

In this case, every retailer has a low initial willingness state L. The problem is decomposed into different subproblem size settings. Different decomposition methods are used to generate the subproblem according to subproblem size parameter. Each subproblem is solved using the QBH method. Then, the overall solution is composed as the union of the solutions for each subproblem. Figure 5.10 displays the average solution collection volumes on different number of retailers in the subproblem from all decomposition methods. The benchmark solution is obtained when there is no decomposition at all (subproblem size equals to 40). The decomposition worsens the overall solution quality compared to the solution obtained with no decomposition. All decomposition methods suggest that decreasing the number of retailers per subproblem decreases the average solution collection volume. Since all retailers are difficult to recruit, dividing the available recruitment budget reduces the flexibility of spending the rest of the budget in the later periods. Hence, the overall probability of recruiting the retailers is reduced. The comparative performance of the different decomposition method is indecisive in this case.

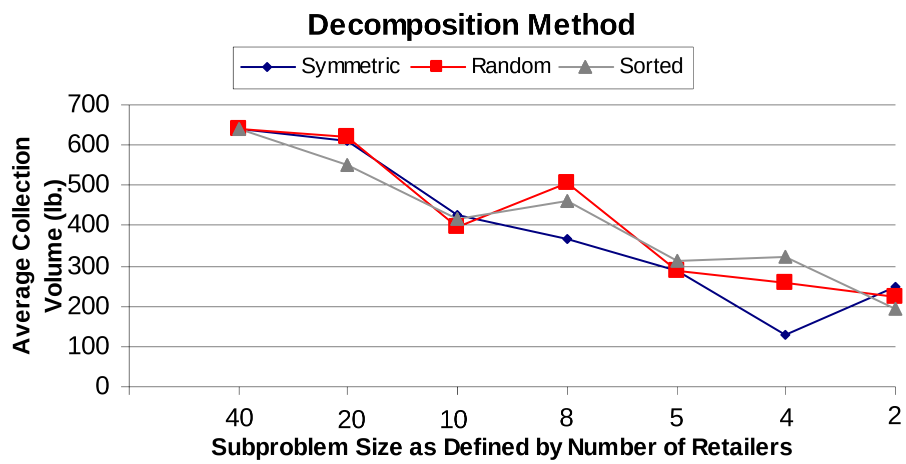

Figure 5.10: Average Solution Collection Volume and Subproblem Size for Case 5.8

<u>**Case 5.9:** Decomposition of 40 retailers with medium initial willingness states M</u>

In this case, every retailer has a medium initial willingness state M. The problem is decomposed into different subproblem size settings. Different decomposition methods are used to generate the subproblem according to subproblem size parameter. Each subproblem is solved using the QBH method. Then, the overall solution is composed of the union of the solutions for each subproblem. Figure 5.11 displays the average solution collection volumes for the case problem, for alternative subproblem sizes, for the three decomposition methods. Problem decomposition seems to worsen the overall solution quality for the test case compared to the solution obtained with no decomposition. For all decomposition methods, lower numbers of retailers per subproblem decreases the average solution collection volume. In this case, random decomposition outperforms the sorted decomposition. For all subproblem sizes, the average solution collection volume obtained by the symmetric decomposition is higher than the one obtained by the random decomposition.

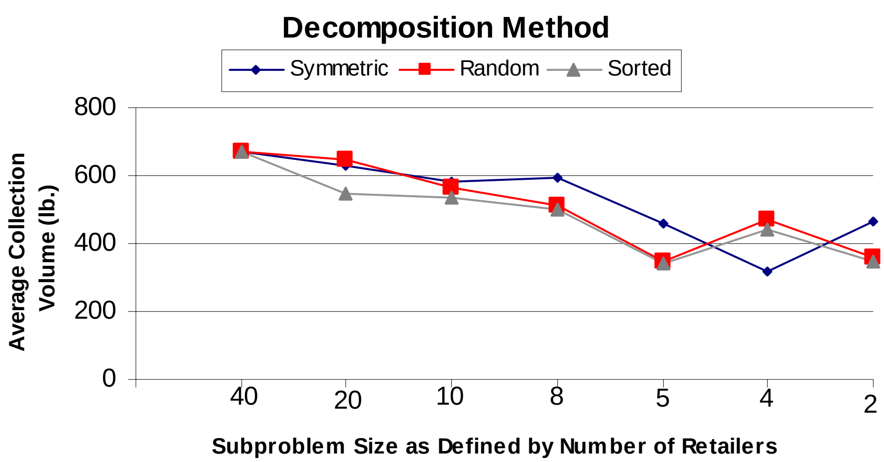

Figure 5.11: Average Solution Collection Volume and Subproblem Size for Case 5.9

<u>**Case 5.10:** Decomposition of 40 retailers with high initial willingness states H</u>

In this case, every retailer has a high initial willingness state H. The problem is decomposed into different subproblem size settings. Different decomposition methods are used to generate the subproblem according to subproblem size parameter. Each subproblem is solved using the QBH method. Then, the overall solution is composed as the union of the solutions for each subproblem. Figure 5.12 displays the average solution collection volumes for the case problem on different sub problem sizes for all decomposition methods. For this test problem, it seems that the decomposition improves the overall solution quality compared to the solution obtained with no decomposition. For all decomposition methods, increasing the number of retailers per subproblem can increase the average collection volume for certain subproblem sizes. For the test problem, symmetric decomposition outperforms the other two decompositions.

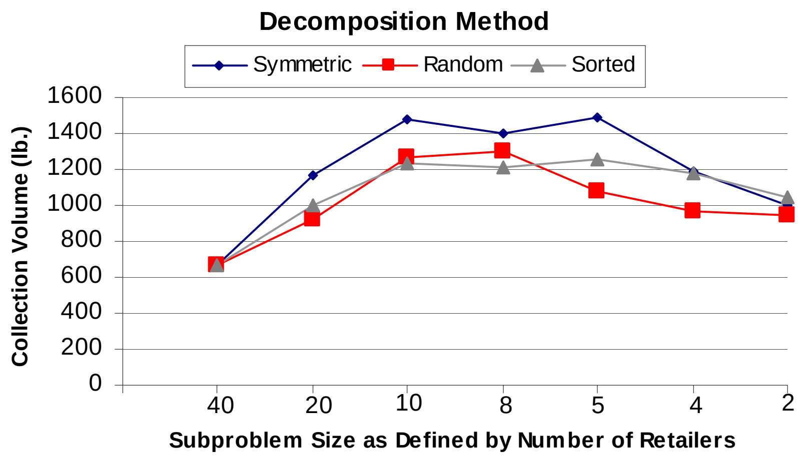

Figure 5.12: Average Solution Collection Volume and Subproblem Size for Case 5.10

From the results in Cases 5.8, 5.9 and 5.10, decomposition with the QBH method yields better results only when the retailers are easy to recruit. Also, the symmetric decomposition approach for the test case performs better than other decomposition methods for most subproblem sizes. The peak in the graph might be explained by the tradeoff between increasing the number of state action pairs explored relative to the total number as the problem size gets smaller, versus decreasing the opportunities to exploit certain combinations of retailers and budget allocations because of the division.

Next, I consider the case when each subproblem is solved using the RIDH approach. For the subproblem size of 5 I evaluate the RIDH approach to solve the case. The results are shown in Table 5.12.

Table 5.12: Applying the RIDH Approach Using Different Decomposition Methods When Subproblem Size is 5 for 40 retailers

Case Data		**Decomposition Methods**			**No Decomposition Solution Obtained by the RIDH Method**
		Symmetric	**Random**	**Sorted**	
Case 5.8	Average Collection Volume (lb.)	538	500	408	489.5
	Best Collection Volume (lb.)	620	630	610	610
	Time (sec.)	50	53	55	252
Case 5.9	Average Collection Volume (lb.)	766	732	572	717.5
	Best Collection Volume (lb.)	1240	1160	760	1000
	Time (sec.)	61	61	56	212
Case 5.10	Average Collection Volume (lb.)	1,685	1,618	1,357	1,625.2
	Best Collection Volume (lb.)	1,820	1,720	1,420	1,770
	Time (sec.)	31	28	31	192

From Table 5.12, symmetric decomposition shows only a slight improvement over solving the original problem without the decomposition for this case, but with a dramatically reduced computation time. However, sorted decomposition worsens the average collection volume in all considered cases for this problem. Thus, overall, it would appear that the symmetric decomposition approach is a useful strategy in this problem for RIDH, the solution quality is not degraded and the computation time is reduced.

Next, I investigate how incorrect information from the retailers may affect the performance of the recruitment model.

5.4 Imperfect Information Study

In the recruitment model, I make the key assumption that the state of the recruitee is known perfectly by the recruiter. In this section, the DP algorithm is used to solve recruitment problem where the transition probabilities may be incorrect. I assume that the initial state of the recruitee is correctly known but the transition probabilities are assumed incorrectly. Hence, the recruiter determines a policy using the DP algorithm with incorrect transition probabilities. The resultant policy is then evaluated using the true values for the transition probabilities. I hypothesize that the approach may be able to approximate the actual recruitee behavior even with incorrect information for the transition probabilities.

The data from the in Case 5.4 problem is used to study the impact of imperfect information. In addition, three sets of transition probabilities are defined. Transition probabilities for set A are provided such that it is the easiest to move the willingness state to a higher state. Transition probabilities for set B are provided such that it is not too easy or too difficult to move the willingness state to a higher state. Transition probabilities for set C are provided such that it is the most difficult to move the willingness state to a higher state.

First, the Case 5.4 problem is solved assuming each of the three sets of transition probabilities as the input data to obtain the optimal policy for each set.

Table 5.13 displays the result when resultant solution policy is evaluated as though the true information for transition probabilities was found in set A. Table 5.14 displays the result when resultant solution policy is evaluated as though set B contained the true transition probabilities and similarly Table 5.15 displays the result for set C.

Table 5.13: Imperfect Information Solution Results with Correct Information from Probability Transactions A

Incorrect	Transition Probabilities A	Transition Probabilities B	Transition Probabilities C
Correct	Transition Probabilities A	Transition Probabilities A	Transition Probabilities A
Maximum Budget	Average Solution Collection Volume (lb.)	Average Solution Collection Volume (lb.)	Average Solution Collection Volume (lb.)
10	0.0	0.0	0.0
20	0.0	0.0	0.0
30	56.2	56.8	58.1
40	74.0	73.7	71.8
50	79.3	78.7	61.9
60	138.0	70.0	70.0

Table 5.14: Imperfect Information Solution Results with Correct Information from Probability Transactions B

Incorrect	Transition Probabilities A	Transition Probabilities B	Transition Probabilities C
Correct	Transition Probabilities B	Transition Probabilities B	Transition Probabilities B
Maximum Budget	Average Solution Collection Volume (lb.)	Average Solution Collection Volume (lb.)	Average Solution Collection Volume (lb.)
10	0.0	0.0	0.0
20	0.0	0.0	0.0
30	35.2	35.1	34.3
40	42.6	42.9	43.8
50	58.7	56.0	61.6
60	85.9	70.0	70.0

Table 5.15: Imperfect Information Solution Results with Correct Information from Probability Transactions C

Incorrect	Transition Probabilities A	Transition Probabilities B	Transition Probabilities C
Correct	Transition Probabilities C	Transition Probabilities C	Transition Probabilities C
Maximum Budget	Average Solution Collection Volume (lb.)	Average Solution Collection Volume (lb.)	Average Solution Collection Volume (lb.)
10	0.0	0.0	0.0
20	0.0	0.0	0.0
30	17.7	18.1	16.4
40	22.1	22.4	23.9
50	31.7	35.1	61.0
60	58.9	70.0	70.0

Overall, even though the incorrect information is used to determine a solution policy, the results do not differ greatly from those with correct initial information. From this test problem, I empirically demonstrate that the accuracy of the transition probability data may not be absolutely critical to the overall solution quality.

In this chapter, the RIDH approach provides the average solution collection volumes almost as good as the ones obtained by the DP approach when the small example is considered. In the large example, the RIDH approach outperforms the QBH approach. In addition, I use the test case to illustrate that the QBH approach can find improved best collection volumes when the maximum number of iterations increases. However, the average solution collection volume does not always increase when the maximum number of iterations increases. Using different decomposition methods, the QBH method yields better results only when the retailers are easy to recruit. Lastly, I empirically illustrate that the correctness of the transition probability data may not be very critical to the overall solution quality.

I have shown in this chapter the RIDH approach performs well for the recruitment problem with a reasonable computation effort. In the next chapter, I propose a strategic model which requires solving many recruitment problems. Using the result in this chapter as a subroutine, the solution method for strategic model is proposed and discussed.

CHAPTER 6

MODEL AND SOLUTION APPROACH FOR THE STRATEGIC LEVEL PROBLEM

Building upon the solutions of the operational level and tactical level problems, I provide a model for the strategic level problem and a methodology to solve it in this chapter. The strategic problem addresses the processor's need to grow a collection network over a time horizon of several macro-periods. An important mechanism of the methodology is to generate and evaluate many network growth trajectories in order to achieve a predetermined target collection volume for the end of final macro-period, while minimizing the total collection costs over time. *The overall problem is to determine the marketing budget allocation over time in different regions in order to follow a selected trajectory while considering uncertainty in collection volumes from the regions of collection networks.*

This chapter is organized as follows. In section 6.1, the description of strategic level problem is discussed. Then I introduce a way to estimate the collection cost for the strategic problem in section 6.2. In section 6.3, the strategic problem is formulated within a stochastic dynamic programming framework. In section 6.4, the strategic problem is viewed as a resource allocation-collection multi-time scale model. The general framework of this model is developed and presented. In section 6.5, the strategic model is adapted to fit the general framework of the multi-time scale model. I then propose a strategic trajectory heuristic to solve the strategic problem in section 6.6. However, one step in the strategic trajectory heuristic requires solving the target recruitment problem. The corresponding methodology is discussed in section 6.7.

6.1 Strategic Level Problem Model

In this section, I provide the details of the strategic level problem from the processor's standpoint. According to the discussion in Chapter 3, the processor's two objectives are minimizing 1) the collection costs and 2) the deviation of the actual collection volume from the target collection volume. In order to accomplish its goals, it must manage collection network growth "smartly" by investing wisely in the recruitment of collection regions.

I attempt to build the strategic model by exploring the benefit of information sharing in a vertical coordination of the supply chain. The processor in the upper tier and the regional collector in the middle tier exchange information back and forth as shown in Figure 6.1. I assume that the processor has information on every retailer in the every region. However, it does not control the recruitment in the regions. Without knowing good estimations of collection volume from all regional collectors, the processor can not grow the collection network to reach the collection target effectively. The significance of information sharing between tiers is discussed by Fiala (2005), Cachon and Fisher (2000), and Lee et al. (1997). With no input from the regions, the processor may wrongly allocate a large amount of marketing budget to a region where it is difficult to recruit and expensive to collect. This mistake may return a lower collection volume than expected and lower success probability for achieving the target collection volume in the final macro-period. Thus, the processor can employ these estimations to help grow collection network wisely by coordinating with the regional collectors.

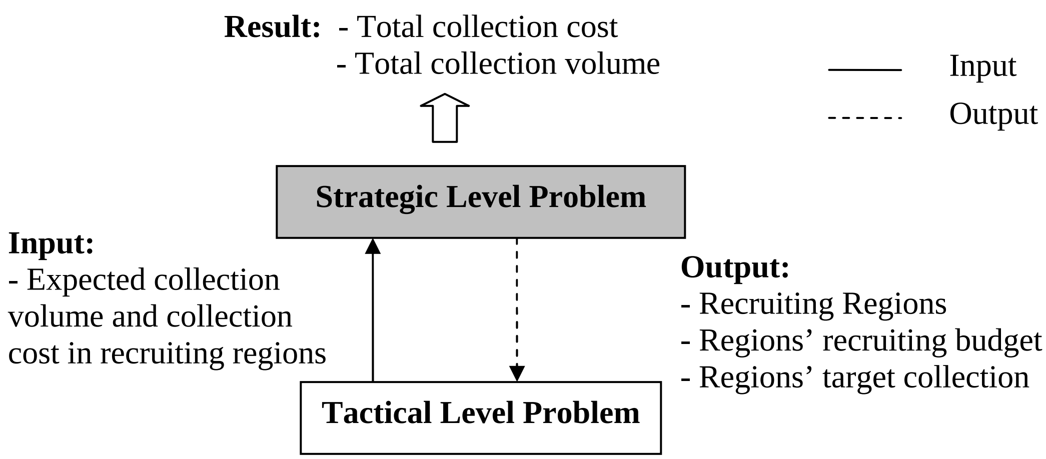

Figure 6.1: Information Flow for Strategic Level Problem

I assume that the regional collector or recruiter for each region can use current information for its retailers to provide an estimate of the collection volume to the processor for a given budget amount and target collection volume. Then the processor makes use of this information in determining the overall marketing budget allocation across regions. When the regional collector receives its budget allocation from the processor at the beginning of any macro-period, it decides how to spend the recruiting budget in each period over one macro-period.

The timeline of how the recruiting budget is spent among regions at the strategic level and how each regional collector allocates the given budget among the retailers in each region in the tactical level is displayed in Figure 6.2. In the figure, macro-period and period present in a year and a month respectively. For each year, the processor determines the regional recruiting budget and the target collection volume. Then the regional collector decides how to allocate its recruiting budget monthly to its retailers.

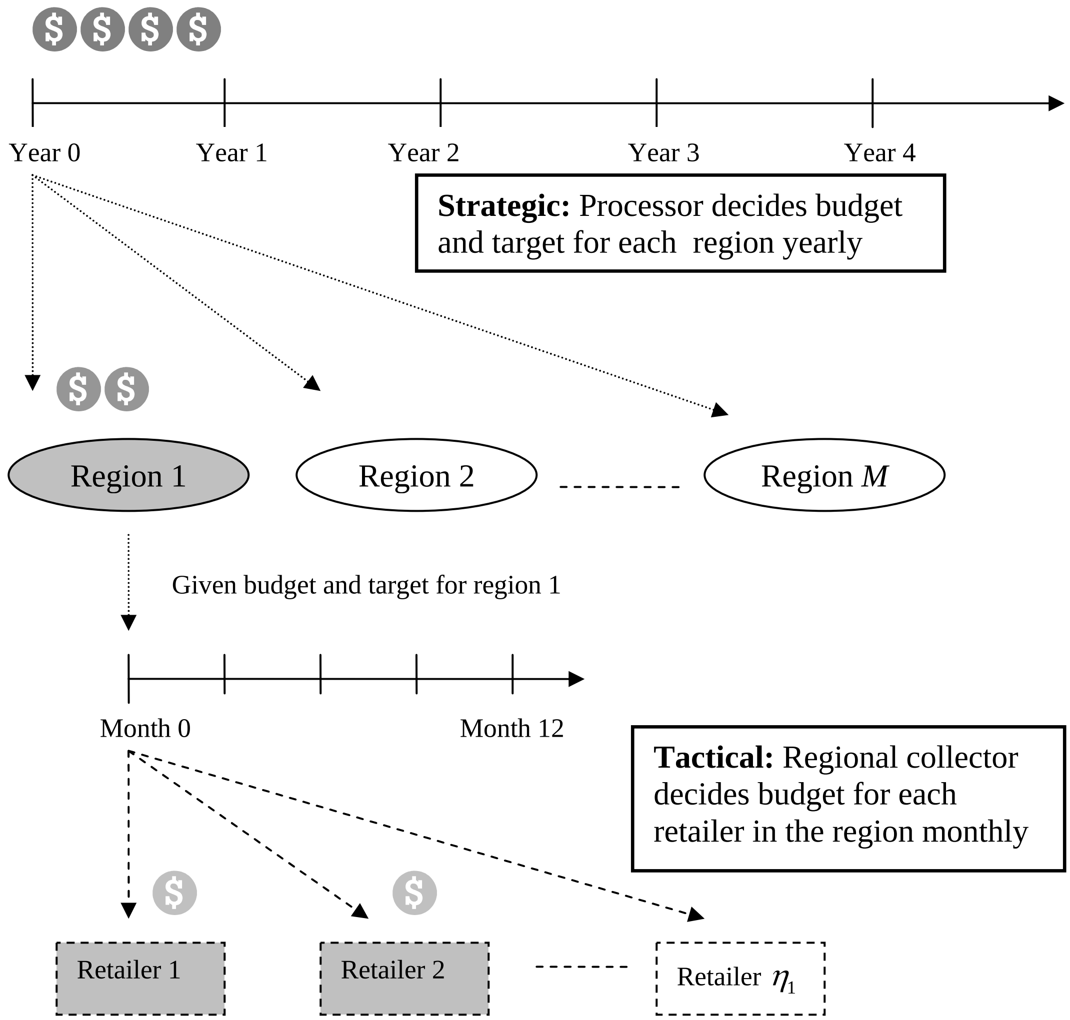

Figure 6.2: Budget Timeline for Strategic and Tactical Level

The processor makes three decisions each year. First, it decides in which region(s) to grow the collection network. Second, it determines how much marketing budget to spend in each region. Third, it sets the regional targets for collection volume for the next year. In order to determine these decisions the processor uses the current information for collection volumes and costs along with the inputs from all the regional collectors.

Before making these decisions, the processor receives an estimate of the collection volume and the collection cost for a given budget and collection target from all regional collectors. The decision timeline of the processor is shown in Figure 6.3.

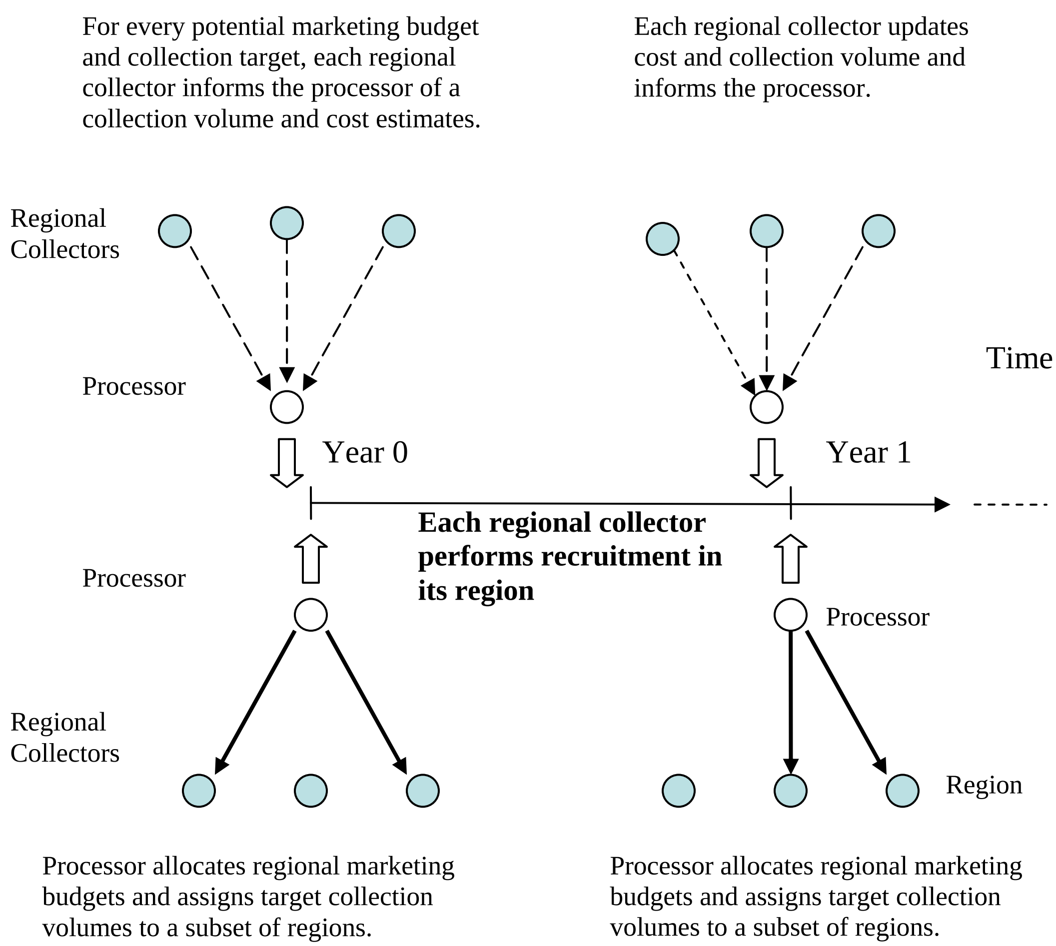

Figure 6.3: The Decision Timeline for a Recruitment Problem

Let M represent the total number of regions. Each region contains η_m, $m = 1, 2, ..., M$, retailers that can be recruited to join the collection network. In the carpet industry, there are thousands of retailers in the southeastern region of the U.S. Even though there is a possibility that some of them can be removed from consideration based on the minimum volume the processor is prepared to collect, there are still hundreds of retailers to be considered for collection recruitment. A large number of retailers imply the high complexity of the collection network growth problem. At this strategic level, I do not aim for a mathematically optimal solution that requires a great deal of computational effort but for a heuristic, not necessarily optimal, good solution that requires a reasonable computational time.

In this section, I provide a description of the strategic level problem. One of the objectives of the strategic level problem is to minimize the total collection cost through the final macro-period. A total collection cost function that is a function of many factors is introduced in the next section.

6.2 Collection Cost Function

This section provides a model to estimate the total collection cost for each regional collector. Each region can have different collection costs. In Chapter 4, I assume that the locations of the retailers were dispersed differently in each region. This has an impact on the regional transportation cost, which is part of the total collection cost. For transportation cost, I use the fact that the collection center in each region can be thought as a depot that transports the collected material to the processor. There are two types of transportation costs to be considered. The first type is the cost to transport the large amount of collected materials to the processing site. This is done by a third party long-haul trucking company. The second type is the cost to collect the materials from retailers in regional collection network. An average distance estimate is employed to reflect the estimated dispersion in that region. The more the retailers cluster together near the center, the lower the average distance becomes. Consequentially, the transportation cost in this situation should be lower. On the other hand, the further the retailers disperse from the center, the higher the average distance becomes. Consequentially, the transportation cost in this situation should be greater. Let $D_m^{(Pr)}$ be the distance between the processor and collector in region m and c_D be the long-haul transportation cost per unit volume and per unit distance.

The fixed collection cost in region m is denoted by $F_m^{(co)}$. Fixed cost is a fixed transportation cost that occurs no matter the level of the collection volume. This could include a truck rental charge, insurance costs, etc. Let $V_m^{(co)}$ be the actual volume collected in region m and $\bar{D}_m$ be the average distance between collector in region m and all retailers in that region. Lastly, I denote $\bar{c}_D$ as transportation cost per unit volume and per unit distance within the region. I estimate for the total collection cost as:

$$c_m^{(total)}(V_m^{(co)}) = F_m^{(co)} + c_D D_m^{(Pr)} V_m^{(co)} + \bar{c}_D \bar{D}_m (V_m^{(co)})^{\rho}, \tag{6.1}$$

where $0 < \rho < 1$. The volume parameter, ρ, controls how much the collection volume affects the total collection cost. The reason ρ is restricted to values between zero and one is to represent collection cost typically as a concave cost function (Konno and Yamamoto 2003, Tishler and Lipovetsky 2000).

To illustrate the collection cost function formula, I construct the following examples. For three different configurations, the distance between the retailers and the collector in a region is given in Table 6.1. For these examples, larger range of distances (1-15) implies large dispersion and a smaller range of distances (1-5) implies small dispersion.

Table 6.1 displays the distances for the different dispersion examples along with the associated calculation for average distance. The volume parameter ρ is set to 0.5 in this example. In addition, $\bar{c}_D$ = 1, c_D = 0.5, $D_m^{(Pr)}$ = 20, and $F_m^{(co)}$ = 20.

Table 6.1: Distance Data for the Example Regions with Alternate Dispersions of Retailers

	Region with Large Distance Dispersion	Region with Medium Distance Dispersion	Region with Small Distance Dispersion
Retailer	Distance 1-15	Distance 1-10	Distance 1-5
1	5	8	3
2	5	9	5
3	7	2	3
4	10	1	5
5	2	7	5
6	1	2	4
7	14	4	3
8	13	8	3
9	9	7	1
10	5	5	1
11	9	3	1
12	13	9	2
13	14	7	2
14	6	6	4
15	3	3	4
16	12	8	1
17	2	6	3
18	5	8	5
19	9	7	5
20	3	6	1
Total	147	116	61
Average Distance	7.35	5.80	3.05

Using the data from Table 6.1, I plot the regional total collection cost versus collection volume as shown in Figure 6.4 and unit cost versus collection volume for the region as shown in Figure 6.5. It can be seen that the total collection cost increases at a higher rate when the regional collection volume is very low than when the volume is high. The figures also show that unit cost approaches a limiting value as the regional collection volume increases.

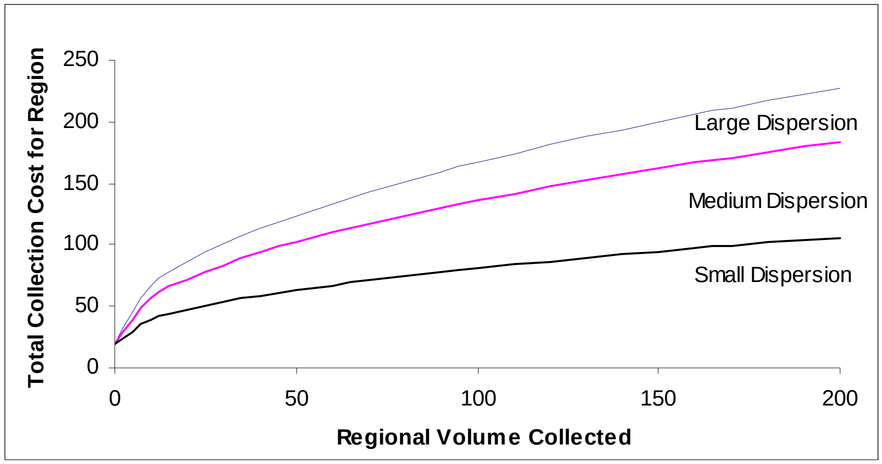

Figure 6.4: Total Collection Cost and Collection Volume

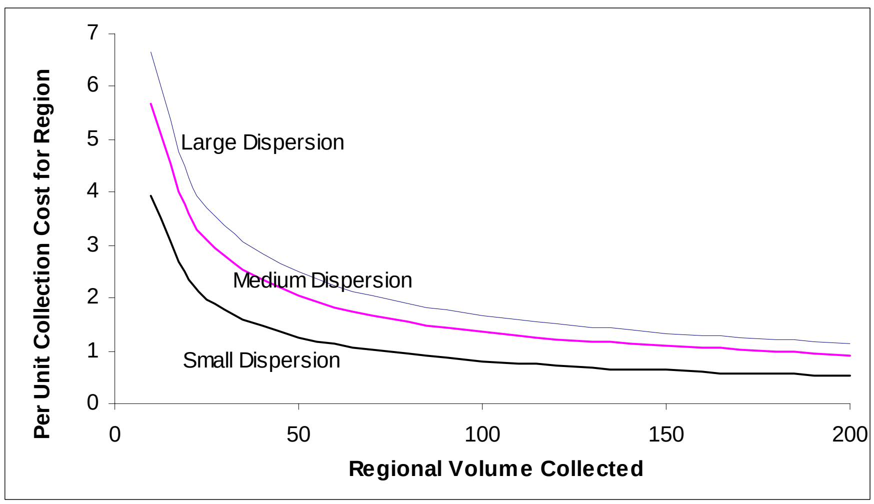

Figure 6.5: Unit Cost and Collection Volume

In sections 6.1 and 6.2, I have provided a general description of the strategic model and showed how to acquire importation information on collection costs. In the next section, I present a stochastic dynamic programming formulation for the strategic problem.

6.3 A Stochastic Dynamic Programming Formulation for the Strategic Problem

As a starting point, I model the strategic problem from the processor's standpoint to provide a heuristic policy. To explain this model, I begin with the definitions of decision epochs, state space, actions, transition probabilities, and rewards. Let the maximum starting recruitment budget be $\theta^{\max}$ and the number of total macro-periods be N to define the decision epochs and state space as follows.

Decision Epochs

$$n = \{0,1,...,N-1\}$$

State Space

$$z_n = \{n, \underbrace{w_{11n}, w_{12n}, ..., w_{1\eta_1 n}}_{\eta_1}, ..., \underbrace{w_{M1n}, w_{M2n}, ..., w_{M\eta_M n}}_{\eta_M}, \theta_n^{Start}\} \text{ for all } z_n \in Z ,$$

where the willingness state of retailer i of region m {from among potentials states (L,M,H,R) as previously described in Chapter 4} at decision epoch n is w_{min} and the maximum starting budget at each macro-period is represented by θ_n^{Start} such that $\sum_n \theta_n^{start} = \theta^{\max}$. Although I assume that w_{min} is known to the processor, I assume that the willingness state can be updated through the recruitment process of that region by the regional collector only. Using the willingness states for every retailer of each region in each macro-period, the processor can convert this information into the estimated collection volume and collection cost of that region. Furthermore, the information also signals the recruitment difficulty of each region to the processor. Hence, this information along with the target collection volume forms the basis for the collector to make decisions on the recruiting budget allocation in each macro-period.

For my model, I define the action set as follows.

Action Set

$$\Lambda_{ln} = \{\lambda_{1ln}, \lambda_{2ln}, ..., \lambda_{Mln}, \bar{\lambda}_{1n}, \bar{\lambda}_{2n}, ..., \bar{\lambda}_{Mn}\},$$

where the amount of recruiting budget provided to region m from action set index l at macro-period n is represented by λ_{mln} and the target collection volume set for region m from macro-period n is represented by $\bar{\lambda}_{mn}$. There are restrictions on λ_{mln} and $\bar{\lambda}_{mn}$ such that $\sum_{m=1}^{M} \lambda_{mln} \le \theta_n^{start}$, $0 \le \lambda_{mln} \le \theta_n^{start}$ for $l = 1, ..., |\Lambda_n|$, and $0 \le \bar{\lambda}_{mn} \le H_{mn}$, where the size of the m^{th} region's target collection volume in macro-period n is H_{mn}. The action set's size depends on the number of regions, the size of budget limit, and the size of target collection volume in each macro-period. Since the amount of recruiting budget allocated to region m, λ_{mln}, limits how much the collector of region m can spend, it constrains the action space of the recruitment problem of that region.

For my model, I define the state transition rules as follows.

State Transition Rules

(a) Initial State

The initial state depends on the initial willingness states of all retailers in every region and the maximum starting recruiting budget. For example, one possible initial state value is $z_0 = \{0, \underbrace{L, ..., L}_{\eta_1}, ..., \underbrace{L, ..., L}_{\eta_M}, 100\}$ where all retailers in each region are in the 'L' willingness state.

(b) Transition Probabilities

The transition probabilities, $P_n(z_{n+1} \mid z_n, \Lambda_{ln})$, are difficult to compute because the processor does not control how each region spends its recruiting budget allocation. If the problem is set up as a centralized system where the processor manages every retailer in every region itself, the transition probability can be computed similarly to what I have shown in Chapter 4.

<u>Rewards</u>

It is hard to express the reward for the state-action pair in earlier macro-periods since I cannot compute the transition probabilities exactly. In addition, the objective of this problem concerns minimizing the total collection cost while reaching the target collection volume in the final macro-period only. I set the reward for each macro-period except the final macro-period as the lowest collection cost. The final macro-period reward includes a penalty for deviating from the target collection volume after the actual collection volume is realized.

<u>Objectives</u>

The objectives of the strategic level problem are to minimize the final collection cost and the deviation between the final collection volume and the collection target. Since these two objectives apply to the end of the planning horizon, it is difficult to represent the optimality equation in a closed form. By its definition, the optimality equation expresses the fact that the value of a state under an optimal policy must equal the expected return for the best action from that state. The expected return on the final collection cost and deviation between the final collection volume and the collection target is difficult to approximate for earlier planning periods.

The strategic network growth problem is a multi stage problem where the first stage decisions generate uncertain results that affect the decisions in the next stage. The problem could be formulated using a multi-stage stochastic knapsack (SKP) model[9] (Birge, J.R. and Louveaux 1997). For the SKP model, previous research results provide a decomposition technique and approximation methods (Vondrak et al. 2004) that simplify solution of the strategic problem. However, there are two reasons that the strategic problem is not formulated here as a stochastic knapsack problem. First, the strategic network growth problem has more than two planning periods, which is a typical horizon used for SKP approaches. Second, the number of possible scenarios in the next decision period becomes too large for the SKP formulation because the state of the strategic problem is the aggregation of each individual state. If the state of the strategic can be represented in a simple form and the probability transition can be computed exactly, a two-stage strategic problem can be formulated as a SKP model. The approximation methods then can be applied to simply solution of the strategic problem.

Next, I attempt to generalize the strategic level problem by examining how the entities in all levels interact. The planning in the tactical level is taken into consideration while making decisions in the strategic level.

6.4 Resource Allocation-Collection Multi-time Scale Model

In this section, I propose a multi-time scale model to illustrate how the strategic and tactical models are connected. The multi-time scale model is motivated by the work of Chang et al. (2003), Panigrahi and Bhatnagar (2004), and Panigrahi and Bhatnagar

[9] Special thanks to my dissertation committee member Dr. Ozlem Ergun for providing this insight.

(2006). Chang et al. (2003) propose a model called Multi-time-scale Markov Decision Process (MMDP) for hierarchically structured sequential decision making processes. The decisions in each hierarchy level are made in different discrete time-scales. The hierarchical decision making in semiconductor fabrications is studied by Panigrahi and Bhatnagar (2004). Panigrahi and Bhatnagar (2006) consider the same problem and develop a simulation based two-timescale actor-critic algorithm in a general framework.

I develop a framework for my hierarchical resource-allocation control problem with a two time-scale Markov Decision Process (MDP) model. The basic framework is similar to Chang et al. (2003) and Panigrahi and Bhatnagar (2006) but some assumptions are different. There are two levels of decisions. The higher level (HL) problem has a slow time-scale. The lower level (LL) problem has fast time-scale. I assume that the problems in both levels are categorized as finite horizon discounted MDPs. The planning period on the slow time-scale is from $n = 0,1,..,N-1$. I denote time in the fast time-scale as $t \in \{t_0, t_1, t_2, ...\}$ and time in the slow time-scale as $t_{nT} = n,$, $n = 0,1,..,N-1$ where T is a fixed finite scale factor between slow and fast time-scales. Figure 6.6 illustrates different time-scale in more details. The LL MDP has T planning periods and the HL MDP has N planning macro-periods. The model requires the LL MDP to be resolved every T periods. I assume that $t_{nT} = n + \varepsilon$ where ε is a positive number arbitrary close to zero. The purpose of this assumption is to allow a small gap between t_{nT} and n such that a LL decision at t_{nT} is made slightly after a HL decision.

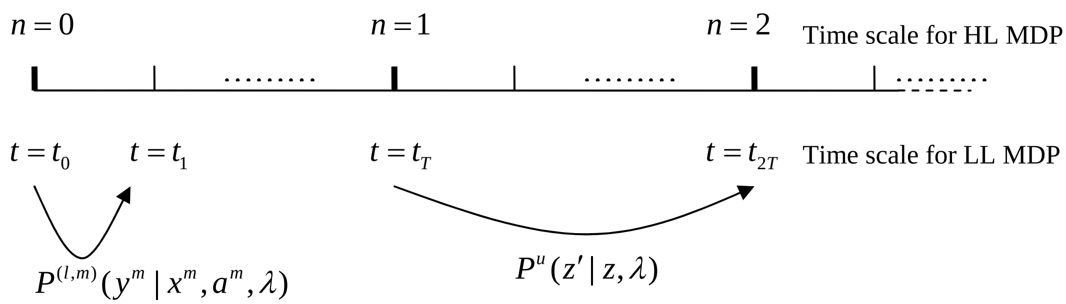

Figure 6.6: Graphical Illustration of Time Evolution in the Two Time-scale Problem

At the LL, there are M agents which correspond to regional collectors. Each agent has different finite number of clients which correspond to retailers. Each client belongs to only one unique agent and follows the Markov model proposed in Chapter 3. I assume the state representation of each agent MDP depicts the status of each client. Each client contains amount of Resource B that the upper level decision maker (corresponding to the processor) desires. The MDP of each agent involves interaction with the clients to retrieve the Resource B. Each agent has a different collection cost function which depends on the amount of Resource B that it can collect from the clients. I assume that the amount of Resource B that the client is willing to offer is a function of its status and the amount of Resource A that it receives from the agent. The allocation of Resource A among all clients represents the action space of each agent in the lower level. In the LL problem, the size of the state and action space of each agent's MDP is different. Figure 6.2 represents how Resource A, recruitment budget, is allocated at different problem level. This is different from the model assumptions by Chang et al. (2003), where it is assumed that every MDP in the LL problem shares the same state and action space.

My model assumes that each agent makes its own decisions based on the action taken by the HL decision, the amount of Resource A provided each agent. The action set in the lower level (LLA) is confined by the action from the higher level (HLA). The transition dynamics of the LL MDP depends on the state at the LL (LLS), the action at the LL and the action at the HL. It does not depend on the state at the HL (HLS). In contrast, Chang et al. (2003) include the HLS as part of the transition dynamics of the LL MDP also.

The objective of the HL is to achieve the target amount of resource B, VN, that is attained at time N with the least collection costs. The allocation of resource A among all agents represents the action space of the upper level decision maker in the upper level. I assume that the maximum starting amount of Resource A, θ_n^{start}, is given and can be used to retrieve Resource B by allocating it to each agent in each slow time-scale period. The defined problem involves Resource A allocation and Resource B collection.

I assume that the performance of overall system is evaluated based on the performance of each agent in the LL obtained from the state transition over the T-horizon of the lower level. Hence, the state at the HL problem at time n is an aggregation of all agents' states at t_{nT}. The transition dynamics for the HL problem do not depend on the LL decisions even though the performance produced by the lower level decisions affect the selection of the higher-level actions.

Both Chang et al. (2003) and Panigrahi and Bhatnagar (2006) offer an initialization function in order to reinitialize the state of the LL MDP. However, the state of the LL MDP in my model is only initialized at t_0. The state at t_{nT} stays the same until the next action is taken.

The higher level MDP has a finite state space Z and a finite action space Λ. At each decision time $n = 0,1,..,N-1$ and at state $z_n \in Z$, an action $\lambda_n \in \Lambda$ is taken and z_n makes transition to a state $z_{n+1} \in Z$ with probability $P^u(z_{n+1} \mid z_n, \lambda_n)$. The superscript u on P represents the transition probability of the HL problem. The action λ basically assigns the amount of Resource A among the sets of agents and set the target collection amount of Resource B from allocated set of agents. After the action has been taken in the HL problem, the solution for the LL MDP over one-step slow time-scale period is determined accordingly. I denote the finite state space and the finite action space for the LL MDP of agent m by X^m and A^m respectively.

Let the initial state in the lower level of agent m be $x^m \in X^m$ and the initial state in the upper level MDP be $z \in Z$ ($z_0 = z = \{x_0^1, x_0^2, ..., x_0^m, ..., x_0^M\}$ at $n = 0$) or $z \in X^1 \times X^2 \times .... \times X^M$. The system follows the lower level MDP evolution from t_0 to t_{T-1}. At the state x^m at t_0, an action $a_0^m \in A_0^m$ is taken and x^m makes transition to next state $y^m \in X^m$ which is a state at time t_1, according to the probability $P^{(l,m)}(y^m \mid x^m, a^m, \lambda_0)$. The superscript (l,m) on P represents the transition probability of agent m in the LL problem. The action $a^m \in A^m$ at t_0 is constrained by λ_0. During this transition, a nonnegative and bounded reward of $R^{(l,m)}(x^m, a^m, \lambda_0)$ for agent m in the LL is incurred and this process is repeated at the state y^m at t_1 until the time t_{T-1}. The superscript (l,m) on R represents the reward of agent m in the LL problem. The expected total reward over the T-horizon represents the expected collection cost when trying to reach the target defined in λ_0. Hence, the state transitions and the reward functions in the LL MDP (over the T-epoch) are induced by the upper level action. This has an impact on the optimal policy of the lower level problem. At time $n=1$, an upper level action λ_1 is taken at z_1. Then the lower level MDP evolves over the next T-epoch.

I define a lower level decision rule for agent m, $d^{(l,m)} = \{\pi_n^{(l,m)}\}$, $n = 0,1,\ldots,N-1$ as a sequence of T-horizon nonstationary policies defined such that for all n, $\pi_n^{(l,m)} = \{\phi_{t_{nT}}^m,\ldots,\phi_{t_{(n+1)T-1}}^m\}$ is a sequence of functions for agent m in the LL problem where $\phi_{t_q}^m : X^m \times \Lambda \to A^m$ for $\forall q \geq 0$.

Given a lower level problem decision rule $d^{(l,m)} \in D^{(l,m)}$ and a nonnegative and bounded immediate reward function $\mathcal{L}^u$ defined over $Z \times \Lambda$ for the upper level problem, I define a function R^u such that for all $n \geq 0$, for $(x^1 \in X^1,\ldots,x^M \in X^M)$ or $z_n \in Z$ and $\lambda_n \in \Lambda$,

$$
\begin{aligned}
R^u(z_n,\lambda_n,(\pi^{(l,1)},\ldots,\pi^{(l,M)})) = \quad & E_{z_n,\lambda_n}^{x^1}\left\{\sum_{t=t_{nT}}^{t_{(n+1)T-1}} \bar{\alpha}^{\sigma(t)} R^{(l,1)}(x_t^1,\phi_t^1(x_t^1,\lambda_n),\lambda_n)\right\} \\
& + E_{z_n,\lambda_n}^{x^2}\left\{\sum_{t=t_{nT}}^{t_{(n+1)T-1}} \bar{\alpha}^{\sigma(t)} R^{(l,2)}(x_t^2,\phi_t^2(x_t^2,\lambda_n),\lambda_n)\right\} \\
& \vdots \\
& + E_{z_n,\lambda_n}^{x^M}\left\{\sum_{t=t_{nT}}^{t_{(n+1)T-1}} \bar{\alpha}^{\sigma(t)} R^{(l,M)}(x_t^M,\phi_t^M(x_t^M,\lambda_n),\lambda_n)\right\} \\
& + \mathcal{L}^u(z_n,\lambda_n), \qquad 0 < \bar{\alpha} \leq 1, \qquad (6.2)
\end{aligned}
$$

where $\sigma(t_{nT+q}) = q$ for all n with $q = \{0,1,\ldots,T-1\}$ and $\bar{\alpha}$ represents the discount factor for each fast time-scale period in the LL problem. The superscript x^m on E represents the initial state of agent m, $x_{t_{nT}}^m = x^m$, and the subscript z_n,λ_n on E represents that z_n

and λ_n for the expectation are fixed. The reward function for upper level R^u is the summation for all m of the T-horizon total expected reward for following the T-horizon nonstationary policy $\pi^{(l,m)}$ given $z_n \in Z$ and $\lambda_n \in \Lambda$ starting with state $x^m \in X^m$ plus an immediate reward of taking an action λ_n at the state z_n at the upper level. In other words, it represents the total expected collection cost to hit the target set by the upper level problem, λ_n. The immediate reward of the upper level problem, $\mathcal{L}^u$, helps guide the upper level decisions to reach the desired target, VN, at the slow time-scale period N.

Hence, the summation for all regions, where region m has a total of total expected reward obtained by the lower level T-horizon nonstationary policy $\pi^{(l,m)}$, acts as a single-step reward for the upper level MDP. I define an upper level stationary decision rule d^u as a function $d^u : Z \to \Lambda$ and I denote D^u as the set of all possible such stationary decision rules. Given the initial states $z \in Z$, the objective is to achieve the decision rules of $d^{(l,m)} \in D^{(l,m)}$ for $\forall m$ and $d^u \in D^u$ that obtains the following functional value defined over Z:

$$
\begin{aligned}
V^*(z) &:= \underbrace{\min}_{d^u \in D^u} \underbrace{\min}_{d^{(l,1)} \in D^{(l,1)},\dots,d^{(l,M)} \in D^{(l,M)}} E^z \left\{ \sum_{n=0}^{N-1} \gamma^n R^u(z_n, d^u(z_n), (\pi^{(l,1)},\dots,\pi^{(l,M)}) \right\}, \\
&= \underbrace{\min}_{d^u \in D^u} \underbrace{\min}_{\forall (d^{(l,m)} \in D^{(l,m)})} E^z \left\{ \sum_{n=0}^{N-1} \gamma^n \left(E_{z_n,\lambda_n}^{x_{t_{nT}}^1} \left[\sum_{t=t_{nT}}^{t_{(n+1)T-1}} \bar{\alpha}^{\sigma(t)} R^{(l,1)}(x_t^1, \phi_t^1(x_t^1, d^u(z_n)), d^u(z_n)) \right] \right.\right. \\
&\qquad + E_{z_n,\lambda_n}^{x_{t_{nT}}^2} \left[\sum_{t=t_{nT}}^{t_{(n+1)T-1}} \bar{\alpha}^{\sigma(t)} R^{(l,2)}(x_t^2, \phi_t^2(x_t^2, d^u(z_n)), d^u(z_n)) \right] \\
&\qquad \vdots \\
&\qquad + E_{z_n,\lambda_n}^{x_{t_{nT}}^M} \left[\sum_{t=t_{nT}}^{t_{(n+1)T-1}} \bar{\alpha}^{\sigma(t)} R^{(l,M)}(x_t^M, \phi_t^M(x_t^M, d^u(z_n)), d^u(z_n)) \right], \\
&\qquad \left.\left. + \mathcal{L}^u(z_n, \lambda_n) \right) \right\}, \qquad 0 < \gamma \le 1, 0 < \bar{\alpha} \le 1, \qquad (6.3)
\end{aligned}
$$

where I define V^* as the two-level optimal finite horizon discounted value function and γ as the discount factor for each slow time-scale period in the HL. In summary, the decisions at the upper level must depend on the aggregation of the lower level state, which is the initial state for the lower level MDP evolution of each agent over the T-horizon in the fast time-scale.

Using the definition of the reward for the upper level and the lower level along with the value function, the optimality equation can be defined as follows.

Optimality Equation

For a given pair of $z \in Z$ and $\lambda \in \Lambda$, I define a set $\Pi^{(l,m)}[z,\lambda]$ of all possible lower level T-horizon nonstationary policies of agent m under the fixed pair of the upper level state z and action λ:

$$
\begin{aligned}
\Pi^{(l,m)}[z,\lambda] := \Big\{ \pi^{(l,m)}[z,\lambda] \,|\, \pi^{(l,m)}[z,\lambda] := \{\phi_{t_0}^{(m,\lambda)},\dots,\phi_{t_{T-1}}^{(m,\lambda)}\}, \\
\phi_{t_q}^{(m,\lambda)} : X^m \times \{\Lambda\} \to A^m, \text{ and } q = 0,\dots,T-1 \Big\}. \qquad (6.4)
\end{aligned}
$$

In addition, I define $P^{(T,m)}_{x^m y^m}(\pi^{(l,m)}[z,\lambda])$ as the probability that a state $y^m \in X^m$ is reached by T-steps starting with x^m by following the T-horizon nonstationary policy $\pi^{(l,m)}[z,\lambda]$. The probability $P^{(T,m)}_{x^m y^m}(\pi^{(l,m)}[z,\lambda])$ can be obtain by $P^{(l,m)}$. Using the definition of $P^{(T,m)}_{x^m y^m}(\pi^{(l,m)}[z,\lambda])$, I can obtain:

$$\begin{aligned} P^u(z_{n+1} \mid z_n, \lambda_n) &= P^u((x^0_{n+1}, x^1_{n+1}, ..., x^m_{n+1}, ..., x^M_{n+1}) \mid (x^0_n, x^1_n, ..., x^m_n, ..., x^M_n), \lambda_n), \\ &= P^u((x^0_{t_{(n+1)T}}, x^1_{t_{(n+1)T}}, ..., x^m_{t_{(n+1)T}}, ..., x^M_{t_{(n+1)T}}) \mid (x^0_{t_{nT}}, x^1_{t_{nT}}, ..., x^m_{t_{nT}}, ..., x^M_{t_{nT}}), \lambda_n), \\ &= \prod_{m=1}^{M} P^{(T,m)}_{x^m_{t_{nT}} x^m_{t_{(n+1)T}}}(\pi^{(l,m)}[z_n, \lambda_n]). \end{aligned} \tag{6.5}$$

Next, I can define how the MDP in upper level operates. The state at time n is an aggregation of the lower level state of each agent, $z_n = (x^0_{t_{nT}}, x^1_{t_{nT}}, ..., x^m_{t_{nT}}, ..., x^M_{t_{nT}})$. An action at state z_n is a composite control of $\lambda_n \in \Lambda$ and $\pi^{(l,m)}[z_n, \lambda_n] \in \Pi^{(l,m)}[z_n, \lambda_n]$ for $\forall m$. Eventually, I can write Bellman's optimality equation for this problem. The upper level sequential decisions are an MDP with a reward function that is defined over all agents' MDP dynamics. From Puterman (1994) and Hernandez-Lerman (1989), I can adapt the standard MDP theory for Bellman's optimality equation, for all $z \in Z$:

$$\begin{aligned} V^*(z) = \underbrace{\min}_{\lambda \in \Lambda} \Bigg(\underbrace{\min}_{\forall(\pi^{(l,m)}[z,\lambda] \in \Pi^{(l,m)}[z,\lambda])} \Big\{ & R^u(z, \lambda, (\pi^{(l,1)}[z,\lambda], ..., \pi^{(l,M)}[z,\lambda]) \\ & + \gamma \sum_{z' \in Z} P^u(z' \mid z, \lambda) V^*(z') \Big\} \Bigg). \end{aligned} \tag{6.6}$$

Using similar arguments as those in Chang et al. (2003), the unique solution to (6.6) is V^*.

In this section, I have developed the resource allocation-collection multi-time scale MDP model as a general framework to represent the overall strategic problem. This model defines the states, transition probabilities, and the value function. In addition, this model provides a general understanding for how the decisions in each hierarchical level are connected. Using this general framework, I interlink the recruitment problem and strategic problem within the framework proposed in the next section.

6.5 Multi-time Strategic Model

In this section, I apply the multi-time scale MDP structure developed in section 6.5 to the strategic level problem. The decision maker for the slow time-scale is the processor and the decision makers for the fast time-scale are the regional collectors (agents) who can recruit the retailers (clients) in their regions. The number of agents or regional collectors is M. The number of retailers for regional collector m is η_m. Resource A represents the recruiting budget and Resource B represents used carpet.

The state of each regional collector m represents the willingness state of its retailers, $x^m = \{w_{m1},..,w_{m\eta_m}\}$. Hence, the state of the processor is the aggregation of all regional collectors' states or $z = \{x^0,..,x^m\} = \{w_{11},..,w_{1\eta_1},...,w_{M1},..,w_{M\eta_M}\}$.

The strategic model must decide how to allocate the available quantities of Resource A at the slow-time scale. I denote $\theta^{\max}$ as a total amount of Resource A with a limit θ_n^{Start} on how much the resource can be spent in each macro-period n. The size of the action in the upper level is bounded by the number of regional collectors and θ_n^{Start} because it is a combinatorial set of resource allocations. Action λ defines the resource allocation and the target collection volume for each regional collector. The resource allocation to regional collector m limits how much the regional collector m can spend over the next T-horizon. Hence, for a regional collector m, the given resource allocation and the number of retailers control its action space, $a^m \in A^m$, and the probability transition, $P^{(l,m)}(y^m \mid x^m, a^m, \lambda)$ of the lower level problem. The probability transition in the lower level can be obtained using (4.2).

The two objectives of the strategic model, minimizing total collection cost and minimizing deviation from the period N collection volume target, are measured in different units. I introduce a penalty cost per collection volume unit, φ, to convert the unit of deviation (pound) from collection volume target to a cost unit (dollar). In the recruitment problem of each collector m, I introduce $R_t^{(l,m)}(x^m, a^m, \lambda)$ as the reward function in period $t \in \{0,1,..,T-1\}$. For $t \in \{0,1,..,T-2\}$, I set the reward $R_t^{(l,m)}(x^m, a^m, \lambda)$ to represent the increment in collection cost only. For $t = T-1$, the reward $R_t^{(l,m)}(x^m, a^m, \lambda)$ includes the penalty cost of deviation from the collection volume target set by λ. This penalty is higher than the collection cost so that that collection volume target is achieved. Consequently, the policy that minimizes the total expected cost over T-horizon, $E_{z,\lambda}^{x^m}\left\{\sum_{t=t_{nT}}^{t_{(n+1)T-1}} \bar{\alpha}^{\sigma(t)} R^{(l,m)}(x_t^m, \phi_t^m(x_t^m, \lambda), \lambda)\right\}$, provides the budget allocation over the T-horizon in order to meet the collection volume target with the least cost.

The immediate reward function $\mathcal{L}^u$ for the strategic level problem can be redefined as $\mathcal{L}_n^u(z, \lambda) = 0$ for $n = 0,1,..., N-2$. For last macro-period reward $\mathcal{L}_{N-1}^u(z, \lambda)$, it can be set to the penalty of failing to reach the overall collection volume target given at the beginning of planning macro-period. Before making the strategic decision, the processor can make better decisions if it has the information of all the retailers in each region. Each regional collector provides its current collection volume, collection cost and possible recruitment success rate in each region. This explains why the status of each retailer in each region is part of the state in the upper level.

With these settings, solving the multi-scale model is the same as obtaining the optimal policy to grow the collection network with the least cost and still meet the collection volume target. However, the computational effort of using an exact method to solve for an optimal policy is high. By relaxing some assumptions and fixing the number of agents to one, the strategic problem can be modeled using the Multi-time scale MDP approach proposed by Chang et al. (2003). One iteration in their proposed value iteration requires $O((|X||Z|)^2 \cdot |\Lambda| \cdot |A|^{T|X|})$ effort. The strategic problem includes the consideration of m different MDPs so the complexity increases by at least the magnitude of m. Since it is very challenging to pose and solve the strategic level problem exactly, I propose an heuristic in the next section to obtain "good" solution.

6.6 Strategic Trajectory Heuristic

I present a strategic trajectory heuristic to solve strategic problem in this chapter by examining the problem in each level. The strategic trajectory heuristic procedure is composed of three main parts. The first part is to formulate a strategic integer programming problem (IP) to make the selection of actions for the current and future macro-periods. The solution from the IP in each slow time-scale period is then simulated over T-periods to obtain the new state. The second part is to update the parameters in each slow time-scale period in order to reformulate the IP in the next slow time-scale period. Determining the parameters for the IP requires solving the lower level MDP (recruitment problem) many times. The last part is to apply the concept of the Q-learning based heuristic to obtain the final policy.

I approach the strategic problem by using reference trajectories that explicitly set targets on the increment amounts that are to be collected in each macro-period. This constrains the upper level problem in every macro-period. If this heuristic searches through enough reference trajectory combinations, a good solution can be obtained. In the remainder of this section, the strategic trajectory heuristic is detailed. First, the strategic IP formulation is developed and the process to updating parameters is specified. These are incorporated into a Q-learning based heuristic.

6.6.1 Strategic IP Formulation

The purpose of creating a trajectory in each slow time-scale period is to set a target on the Resource B and then to try to achieve the target by selecting action $\lambda \in \Lambda$ with the least cost. Achieving the desired target collection volume at time N is what I attempt to accomplish. The strategic IP is formulated in order to select the action $\lambda \in \Lambda$ to achieve this target while minimizing the upper level reward R^u, which is the total expected collection cost when all m regions achieve their lower level targets set by the upper level, λ_n. It is important to emphasize that λ controls the transition dynamics of each of the lower level MDPs. Hence, it is necessary to obtain the estimation of $R^{(l,m)}(x^m, \phi^m(x^m, \lambda), \lambda)$ for all $\lambda \in \Lambda$.

For a given time period, the model's representation of the target collection volume does not include the collection volume of the retailers recruited in the previous period. Alternatively, in the model a time period's target collection volume is actually the desired increment in collection volume, not the cumulative total collection volume attained. For the model, the target collection volume is the range shown by a solid bracket in Figure 6.7.

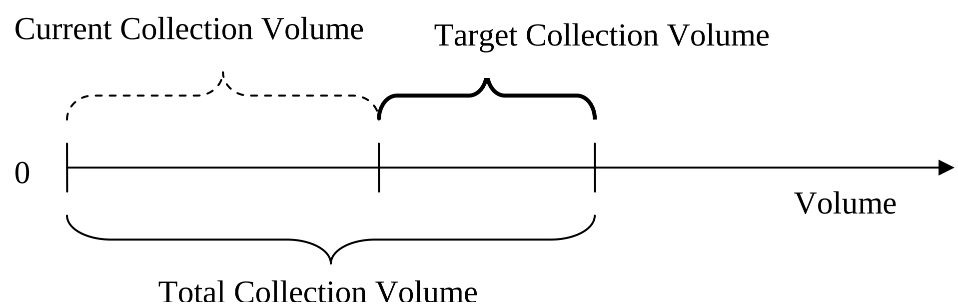

Figure 6.7: Target Collection Used in the IP Model for One Time Period

The strategic IP is used to solve the budget allocation problem for a given macro-period to represent the actual state trajectory as closely as possible based on the state of the macro problem at the beginning of the macro-period. Thus, this is a myopic IP in the sense that it does not try to determine budget allocations beyond the end of the given macro-period. In order to capture different strategic choices by using reference trajectories, the strategic trajectory heuristic uses the strategic IP to try to get the actual trajectory to match the reference as closely as possible.

For the strategic problem, the strategic IP formulation attempts to reach the target collection volume at the end of each macro-period by growing collection network along a specific trajectory. A vector of target collection volumes for each macro-period represents a trajectory over the strategic planning horizon. In each macro-period, the objective is to minimize the collection costs such that the deviation between actual collection volume and the target collection volume from the trajectory in that macro-period is bounded within a certain range. This concept is illustrated in Figure 6.8, where the solid line represents the linear trajectory and the dashed line represents the solution for the strategic IP in each year.

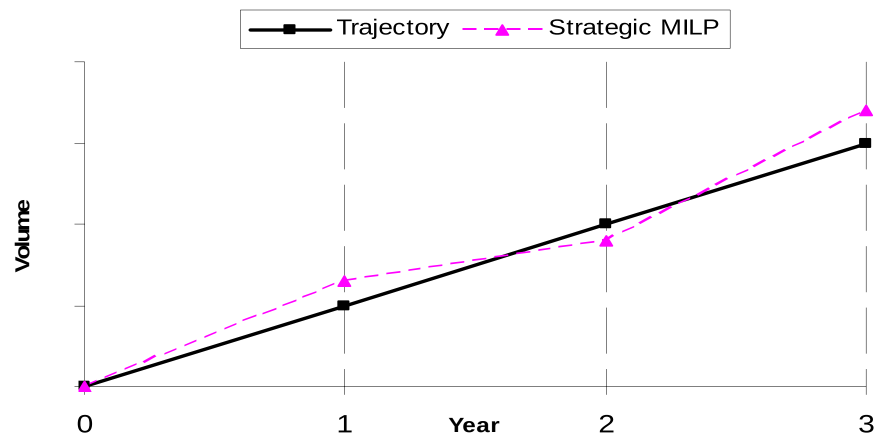

Figure 6.8: Strategic IP and Linear Trajectory

I define the index, parameters, and variables for the strategic IP below.

Index:

m	Index of regions	$(m = 1, 2, \ldots, M)$
j	Index of budget amounts	$(j = 1, 2, \ldots, J)$
k	Index of collection target amounts	$(k = 1, 2, \ldots., K)$
n	Index of macro-periods	$(n = 0,1, \ldots, N-1)$

Using these indices, the parameters for the strategic IP can be defined. Before explicitly defining each term, a modeling device to establish discreet value alternatives for budget and collection volume settings needs to be described. I create a list of alternative budget settings to be used to form a budget array for each region m and for each time period n, $\bar{B}_{mn} = \{\bar{b}_{1mn},..,\bar{b}_{jmn},...,\bar{b}_{Jmn}\}$, where $\bar{b}_{jmn}$ represents the value of the j^{th} entry in $\bar{B}_{mn}$. Also, the list of available target collection volumes is used to create a collection target array for each region m and for each time period n, $\bar{H}_{mn} = \{\bar{h}_{1mn},...,\bar{h}_{kmn},...,\bar{h}_{Kmn}\}$, where $\bar{h}_{kmn}$ represents the value of k^{th} entry in $\bar{H}_{mn}$. Using this notation, the parameters are defined as follows.

Parameters:

$\bar{b}_{jmn}$	Amount of marketing budget that the processor chooses to spend on the region m in period n, which is the value of j^{th} entry in $\bar{B}_{mn}$.
$\bar{h}_{kmn}$	Target collection volume for region m at the end of period n from the target array, which is the value of k^{th} entry in $\bar{H}_{mn}$.
$\tilde{v}_{mjkn}$	Expected increment in collection volume from region m by using budget $\bar{b}_{jmn}$ to achieve target collection volume $\bar{h}_{kmn}$ at the end of macro-period n.
$c_{mn}^{(total)}(\tilde{v}_{mjkn})$	Expected increment in total collection cost of region m at the end of macro-period n to collect volume $\tilde{v}_{mjkn}$ from region m in period n by using budget $\bar{b}_{jmn}$ to achieve target collection volume $\bar{h}_{kmn}$.
VN	Target collection volume at the end of planning horizon $N-1$.
$\bar{V}_n$	Target collection volume from the trajectory that should be achieved at the end of macro-period n.
$\bar{V}_n'$	Adjusted target collection volume that should be achieved at the end of macro-period n.
v_{mn}	Actual collection volume of region m at the start of macro-period n.
vt_{mn}	The summation of actual collection volume for all regions ($m = 1, 2, ..., M$) at the start of macro-period n.
ε	Allowable deviation between expected collection volume and the target collection volume.

The strategic IP embeds a parametric solution of the lower level MDP with the expected increments in collected volume and cost. The solution represents the estimate of the expected collection volume responses for given target and given initial conditions for the regions. The estimate of the cost can be obtained from (6.1). The estimate of the expected collection volume is found by solving the *target recruitment problem* in each region which will be explained in Section 6.7.

In each trajectory, I represent the total target collection volume from all regions at macro-period n as $\overline{V}_n$. There are an infinite numbers of trajectories to consider if I restrict the set of intermediate volume targets to rational numbers. One simple trajectory is a linear trajectory, where $\overline{V}_0 = \frac{VN}{3}$, $\overline{V}_1 = \frac{2VN}{3}$, and $\overline{V}_2 = VN$ for a three-year plan, which is shown as a thick line in Figure 6.9. I also show many other trajectories to reach VN. The algorithm randomly chooses the trajectory by selecting $\overline{V}_n$. The heuristic is thus to use many different, but reasonable, trajectories that capture the different tradeoffs in the problem. The goal is then to further combine the states visited by these trajectories into better solutions through the Q-learning approach.

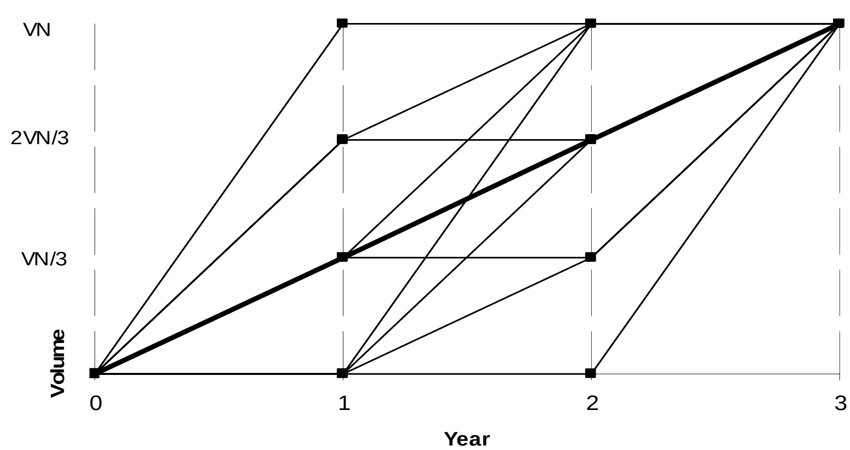

Figure 6.9: Sample Trajectories of Target Collection Volume

Using defined indices and parameters, the decision variables for the strategic IP can be defined as follows.

Variables:

$$x_{mjkn} = \begin{cases} 1 \text{ if budget amount } \bar{b}_{jmn} \text{ is assigned to region } m \text{ in order to achieve the target } \bar{h}_{kmn} \text{ in macro-period } n \\ 0, \text{ otherwise.} \end{cases}$$

Strategic IP for macro-period n, $(SP_{n=0,\ldots,N-1})$, can be formulated as:

Minimize
$$\sum_m \sum_j \sum_k c_{mn}^{(total)}(\tilde{v}_{mjkm})x_{mjkn} \tag{6.7}$$

Subject to:
$$\sum_j \sum_k x_{mjkn} \le 1 \qquad \forall m \tag{6.8}$$
$$\sum_m \sum_j \sum_k \bar{b}_{jmn} x_{mjkn} \le \theta_n^{Start} \tag{6.9}$$
$$\sum_m \sum_j \sum_k \tilde{v}_{mjkn} x_{mjkn} - \bar{V}_n' \le \varepsilon \tag{6.10}$$
$$\sum_m \sum_j \sum_k \tilde{v}_{mjkn} x_{mjkn} - \bar{V}_n' \ge -\varepsilon \tag{6.11}$$
$$x_{mjkn} = \{0,1\} \qquad \forall m, j, k\,. \tag{6.12}$$

The objective function (6.7) is the sum of the total collection cost. Constraints (6.8) require only one budget amount can be allocated in region m for the specific target collection volume. Constraint (6.9) restricts the overall spending budget to be less than the budget limit. Constraints (6.10) and (6.11) bound the deviation between the expected collection volume and the adjusted target collection volume. Constraints (6.12) force x_{mjkn} variables as binary variables.

The strategic IP of period n only looks ahead one macro-period. The solution of strategic IP (SP_n) selects the action λ_n and resultant policy of each collector $\pi^{(l,m)}[z,\lambda_n]$. Then using action λ_n, T-horizon MDP evolution of each collector is performed to retrieve $x^{m}_{t_{(n+1)T}}$ for $\forall m$ or z_{n+1}. This process is called one-period rolling horizon. The strategic IP (SP_{n+1}) is then formulated and the process is repeated until the final macro-period. Hence, by exploring many trajectories, the strategic trajectory heuristics creates possible realizations (z, λ) that meet the objective functions. Some trajectories may be infeasible to reach or feasible but very costly.

I allow the situation where the region starts with some retailers that are already in its network. Hence, each region has an initial seed collection volume. I assume that these recruited retailers are part of the system and so if the processor decides to collect from that region, it gets that initial amount for no recruitment cost. However, that amount is still a part of total collection cost. I then have to keep track of which regions are in the collection network system. If the region is not part of the system, the value of $\tilde{v}_{mjkn}$ of that region becomes the summation of initial seed collection volume and the original value of $\tilde{v}_{mjkn}$. Otherwise, the value of $\tilde{v}_{mjkn}$ remains the same.

In this subsection, I developed the IP formulation for the strategic problem. Its objective is to determine the recruitment budget and the target collection volume to minimize the collection costs such that the deviation between actual collection volume and the target collection volume from the trajectory in that macro-period is bounded within a certain range. The next subsection describes how the parameters of the strategic IP in each macro-period are updated.

6.6.2 Updating Information in Strategic IP Formulation

In order to formulate the strategic IP in each macro-period, three parameters need to be updated. They are the remaining budget (θ_{N-1}^{Start}) in the last period, the expected collection volume ($\tilde{v}_{mjkn}$), and the corresponding collection cost ($c_{mn}^{(total)}(\tilde{v}_{mjkn})$). I describe these updates below.

Remaining Budget (θ_{N-1}^{Start}):

The budget limit, θ_{n}^{Start}, is given for each macro-period. There is an exception in the final macro-period such that the remaining budget from the previous macro-period also can be spent. Thus, the budget limit for the final period decision is the summation of the remaining budget from previous macro-periods and the starting budget limit of that macro-period. The final macro-period maximum budget can be obtained from:

$$\theta_{N-1}^{Start} \leftarrow \theta_{N-1}^{Start} + \theta^{\max} - \sum_{n=0}^{N-2}\sum_{m}\sum_{j}\sum_{k} \overline{b}_{jmn} x_{mjkn} \ . \tag{6.13}$$

Since the willingness state of the retailers in the region may change after the allocation has been assigned, the expected collection volume for the same budget amount and target collection amount in that region may also change. Hence, this value needs to be updated in each macro-period.

Expected collection volume ($\tilde{v}_{mjkn}$):

The algorithm must compute this parameter for every region, every budget amount, every target amount and every macro-period. It can be obtained by solving the recruitment problem which is described in next section (section 6.7). The target recruitment problem takes three parameters: region (m), budget allocation ($\overline{b}_{jmn}$), and target collection volume ($\bar{h}_{kmn}$). I denote the *target recruitment function* $TR(m,\bar{b}_{jmn},\bar{h}_{kmn})$ as a function that returns the expected collection volume from solving target recruitment problem with defined parameters m, $\bar{b}_{jmn}$, and $\bar{h}_{kmn}$. Again, the expected collection volume is the expected increment in collection volume from the amount that can be collected from recruited retailers in region m in macro period n.

Since the expected collection cost $c_{mn}^{(total)}(\tilde{v}_{mjkn})$ depends on the expected collection volume, this value also needs to be updated in each macro-period.

Expected collection cost ($c_{mn}^{(total)}(\tilde{v}_{mjkn})$):

After retrieving the value of $\tilde{v}_{mjkn}$, the algorithm updates the collection cost respectively as shown in Figure 6.10. First, it obtains the actual collection volume of region m at current macro-period n, v_{mn}. Then it computes the increment in collection cost by:

$$c_{mn}^{(total)}(\tilde{v}_{mjkn}) = c_{mn}^{(total)}(v_{mn}+\tilde{v}_{mjkn}) - c_{mn}^{(total)}(v_{mn}) . \tag{6.14}$$

The action space from the high level Λ is (b_{jmn},h_{kmn}) for all j and k. The expected collection cost $c_{mn}^{(total)}(\tilde{v}_{mjkn})$ represents $R^{(l,m)}(x^m,\phi^m(x^m,\lambda),\lambda)$.

These procedures for updating the remaining budget θ_n^{Start} in the final period, the expected collection volume $\tilde{v}_{mjkn}$, and the corresponding collection cost $c_{mn}^{(total)}(\tilde{v}_{mjkn})$ are combined with the IP and the Q-learning based heuristic to obtain the final policy. This approach is described in the next section.

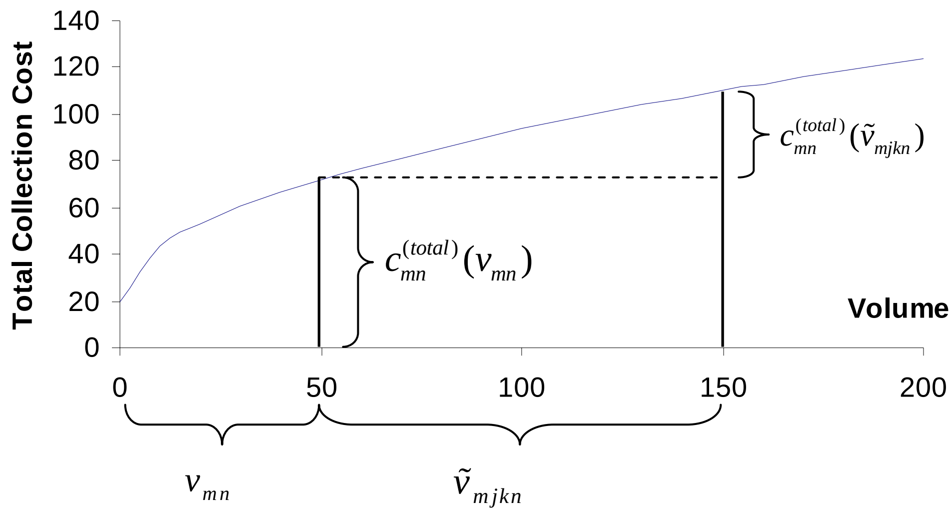

Figure 6.10: Computing Total Collection Cost for Each Region

6.6.3 Strategic Trajectory Heuristics Procedure

This subsection details the complete step-by-step procedure of the Strategic Trajectory Heuristic (STH). The procedure explains how the strategic IP, updating the important parameters, and the Q-Learning based heuristic are combined.

Step 1 First, set the iteration count equal to 1.

Step 2 Set $n = 0$, $vt_n = 0$, $v_{mn} = 0, \forall m$.

Reset all parameters for the first macro-period.

Step 3 **For all** m, j, k **:**

Obtain $\tilde{v}_{mjkn}$ by solving the $TR(m, \overline{b}_{jmn}, \overline{h}_{kmn})$ problem as discussed in section 6.7.

Obtain $c_{mn}^{(total)}(\tilde{v}_{mjkn})$ using (6.14).

End For

<u>Step 4</u> Randomly choose $\overline{V}_n$ such that $vt_n < \overline{V}_n < VN$.

Update $\overline{V}_n' = \overline{V}_n - vt_n$ which is the target in increment of volume target for the end of macro-period n

<u>Step 5</u> Build the (SP_n) model as defined in (6.7)-(6.12) and solve it.

<u>Step 6</u> For all $x_{mjkn} = 1$, simulate the actual collection capacity in region m using budget amount $\overline{b}_{jmn}$ to target $\overline{h}_{kmn}$. This means simulate an instance in the region m using the policy from $TR(m, \overline{b}_{jmn}, \overline{h}_{kmn})$. A different result can be obtained for each simulated instance.

Update $v_{m,n+1}, \forall m$ using the results from the simulation.

Update $vt_{n+1} = \sum_m v_{m,n+1}$.

<u>Step 7</u> Update the Q-Table using the following equation:

$$Q(z_n, \Lambda) \leftarrow r(z_n, \Lambda, z_{n+1}) + \max_{\overline{\Lambda} \in A(z_{n+1})} Q(z_{n+1}, \overline{\Lambda}). \tag{6.15}$$

For $n = 0, 1, ..., N-2$

The initial immediate reward is actually an increment in collection cost. Because the Q-Learning approach is set up as a maximization problem, it expects to be working with the immediate reward as a profit instead so the following adjustment is made. Let ξ be the revenue in each time period. Set this revenue to be very large such that it is always greater than the cost. Then, the immediate reward in period n can be computed from:

$$r_n(z_n, \Lambda, z_{n+1}) = \xi - \left[\sum_m c_{mn}^{(total)}(v_{mn}) - \sum_m c_{mn}^{(total)}(v_{m,n-1}) \right]. \tag{6.16}$$

For $n = N-1$

In the final period, the goal is to minimize the increment in collection cost and also to achieve the collection target, VN. Hence, the immediate reward includes both components. Recall that cost and volume have different measurement units. Let $\bar{\psi}$ be the cost per unit to penalize the collection target deviation. Because the Q-Learning approach is set up as a maximization problem, transform the immediate reward to a profit by subtracting cost from profit. Hence, the immediate reward for period N can be computed from:

$$r_{N-1}(z_{N-1}) = \xi - \left[\sum_m c_{mN}^{(total)}(v_{mN}) - \sum_m c_{mN-1}^{(total)}(v_{m,N-1}) \right] - \bar{\psi} \left| VN - vt_N \right| \tag{6.17}$$

Update the remaining budget for the last period according to (6.13).

Step 8 Increase n by 1. If $n = N$, go to step 9. Otherwise, go to step 3

Step 9 Increase the iteration count by 1.

If iteration count exceeds the limit, exit. Otherwise, go to step 2

Step 10 Use the Q-Table to obtain the final policy.

A summary of this procedure is displayed in the flow chart in Figure 6.11. The iteration count limit represents how many trajectories are examined. In each iteration, the trajectory in each macro-period is randomly generated. Then, the necessary parameters are updated. The strategic IP is solved in order to obtain the collection volume target and recruitment budget for the regional collectors. The solution is simulated to acquire the actual outcome. The state-action pair and the reward in each macro-period are fed into Q-Table for an update. These steps are repeated for N periods in one iteration. After sufficient trajectories are considered, the final policy is obtained.

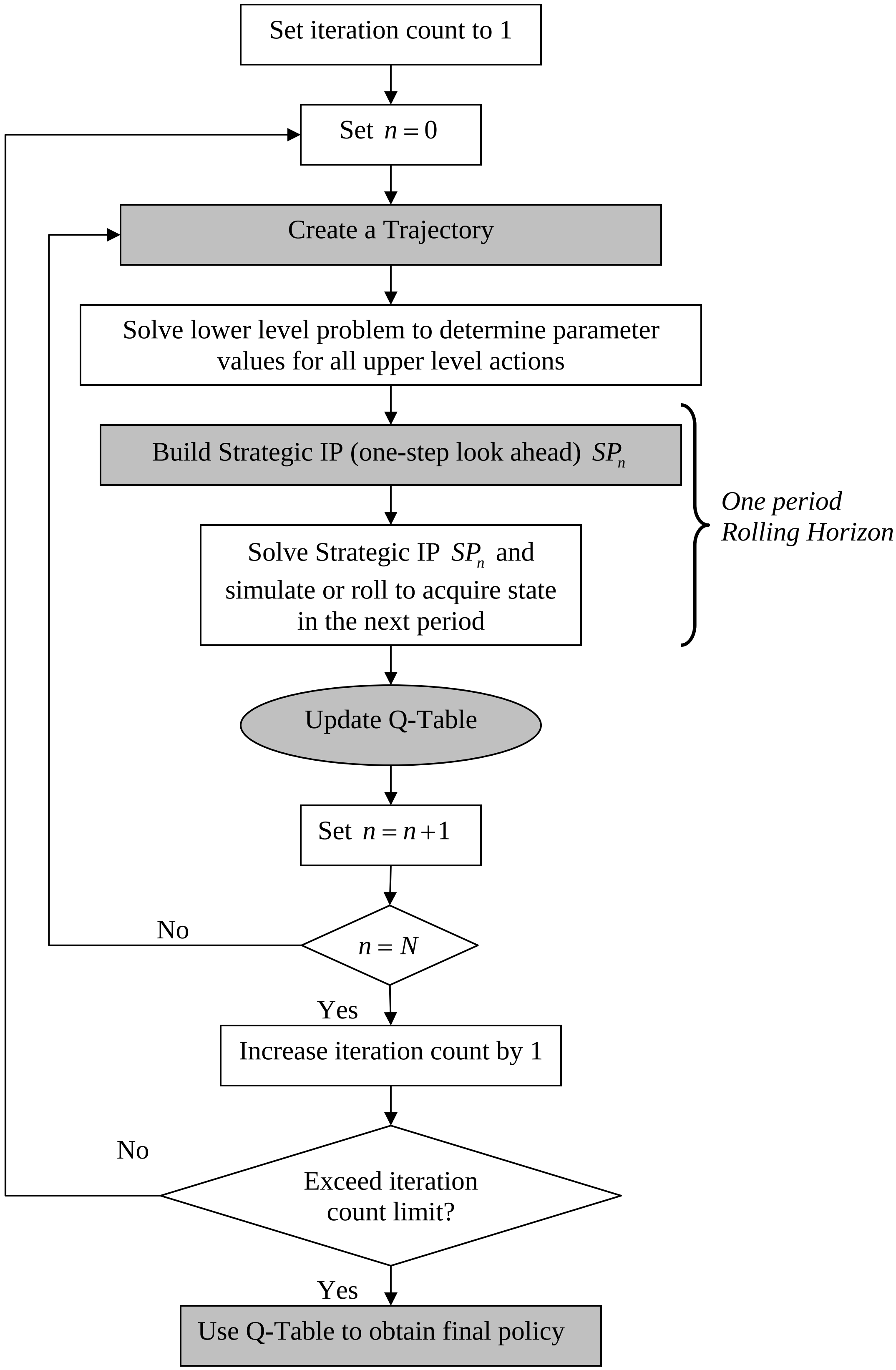

Figure 6.11: Diagram of Strategic Trajectory Heuristics

6.7 Target Recruitment Model

In step 3 of the strategic trajectory procedure described in section 6.6.3, the value of $\tilde{v}_{mjkn}$ is obtained by solving the target recruitment model. Therefore, in this section, I modify the target recruitment model and the methodology based on the discussion in Chapter 4 in order to obtain the value of $\tilde{v}_{mjkn}$. The importance of this model is to provide an estimate of the collection volume from the lower level to the upper level. In section 4.7, I provide a methodology to solve the recruitment problem where the objective is to recruit retailers in order to achieve the maximized collection volume for given recruiting budget. However, the overall strategic goal for each regional collector is to recruit the retailers to minimize the deviation of the expected collection volume from the target collection volume while keeping costs low. In this section, I modify the Chapter 4 model to obtain a target recruitment model and methodology for the strategic problem.

Here the stochastic dynamic programming formulation of the target recruitment problem is similar to the Chapter 4 recruitment model. Decision epochs, states, actions, and transition probabilities remain the same. The only change is for the reward prior the final period, $T-1$. In Chapter 4, the reward represents the summation of the increment in value of the willingness state for every retailer. The value of willingness state R is the collection volume for a retailer. For the strategic level problem, the objective is to maximize the positive state change of all retailers as a proxy for maximizing total collection volume over fixed planning horizon. For the target recruitment model, I modify the reward structure to reflect the new objective. There are η_m retailers to consider. The reward in period $T-1$ of moving to state Y_T from state Y_{T-1} by action $A_{l(T-1)}$ becomes

$$r_{T-1}\left(Y_{T-1}, A_{l(T-1)}, Y_T\right) = \sum_{i=1}^{\eta_m}\left(V_{w_{iT}} - V_{w_{i(T-1)}}\right) - \psi\left|h - \sum_{i=1}^{\eta_m}\left(V_{w_{iT}} - V_{w_{i0}}\right)\right|, \tag{6.18}$$

where h is the given collection target that the processor tries to accomplish according to step 3 in strategic trajectory heuristic procedure, $\overline{h}_{kmn}$, and ψ is the positive penalty of deviating from the target per unit volume. The first term in (6.18), $\sum_{i=1}^{\eta_m}\left(V_{w_{iT}} - V_{w_{i(T-1)}}\right)$, is the summation of increment in value for all retailers. The second term, $\psi\left|h - \sum_{i=1}^{\eta_m}\left(V_{w_{iT}} - V_{w_{i0}}\right)\right|$, is added to penalize the deviation between the target collection volume and the total increment in collection volume for all retailers between the first period and the final period, $\sum_{i=1}^{\eta_m}\left(V_{w_{iT}} - V_{w_{i0}}\right)$. In other words, higher deviation implies smaller reward. The penalty ψ must be high enough to guide the solution toward the target collection volume.

From Chapter 5, because the Rolling IP with DP heuristic provides a solution with better solution quality and solution time than Q-Learning method, I modify the Rolling IP with DP heuristic to solve the target recruitment problem. I denote e as the deviation of actual collection volume from the target. The initial budget allowed B_0^{start} or $B^{\max}$ represents given budget allocation from the strategic level, $\overline{b}_{jmn}$. As a starting point, I can model the target recruitment model for each region as the following quadratic programming problem.

Regional Target Recruitment Problem-Quadratic Programming Formulation (QP_t)

Minimize $$e^2 \tag{6.19}$$

Subject to:

$$\sum_j x_{ij} \le 1 \qquad \forall i \tag{6.20}$$

$$\sum_i \sum_j b_j x_{ij} \le B_t^{start} \tag{6.21}$$

$$e = \sum_i \sum_j \tilde{v}_{ij} x_{ij} - h_t' \tag{6.22}$$

$$x_{ij} = \{0,1\} \qquad \forall i, j \tag{6.23}$$

$$-\infty < e < \infty \,. \tag{6.24}$$

I introduce h_t' as the remaining collection volume target at time t and B_t^{start} as the remaining budget at time t. To avoid the need to solve a challenging integer quadratic programming problem, I change the objective function to minimize the absolute deviation using standard reformulation techniques. I represent e^+ as the deviation of going over the target collection volume $\left(\sum_i \sum_j \tilde{v}_{ij} x_{ij} - h_t' \ge 0 \right)$, and e^- as the deviation of going below the target collection volume $\left(\sum_i \sum_j \tilde{v}_{ij} x_{ij} - h_t' \le 0 \right)$. Then I replace e by the difference of two new nonnegative variables, $e = e^+ - e^-$, where $e^+, e^- \ge 0$. The solution has the property that either $e^+ = 0$ or $e^- = 0$ (or both). Hence, I can use the summation of these two terms to represent the objective function, which is the summation of errors that go over and below the target collection volume. I then reformulate the targeted recruitment problem for each region as a mixed integer linear programming problem (MILP):

Regional Target Recruitment MILP Formulation (TRP_t)

Minimize	$e^+ + e^-$		(6.25)
Subject to:	$\sum_j x_{ij} \le 1$	$\forall i$	(6.26)
	$\sum_i \sum_j b_j x_{ij} \le B_t^{start}$		(6.27)
	$e^+ - e^- = \sum_i \sum_j \tilde{v}_{ij} x_{ij} - h_t'$		(6.28)
	$x_{ij} = \{0,1\}$	$\forall i, j$	(6.29)
	$e^+, e^- \ge 0$.		(6.30)

To solve the target recruitment problem, I follow the recruiting rolling horizon solution procedure as explained in section 4.7 with some modifications. The regional target recruitment rolling horizon solution procedure is described below.

Regional Target Recruitment Rolling Horizon Solution Procedure (RTHS)

Step 0 Set $t = 0$, $h_0' = h$, and $B_0^{start} = B^{\max}$

For all retailers i that are already in the network, set $\tilde{v}_{ij}$ = 0.

Otherwise, solve $\tilde{v}_{ij}$ from $SR(i, b_j, T)$ for all i, j using the DP approach. The initial state of $SR(i, b_j, T)$ is [0, initial willingness state, b_j].

Step 1 Formulate the associated regional TRP_t model and solve for x_{ij} .

Step 2 For the retailers for which marketing budget has been allocated, simulate the action in period t only.

If $t = T$, obtain the total increment in collection volume from period 0 to period T and exit. Otherwise, obtain the total increment in collection volume from prior the period $t-1$ and go to Step 3.

Step 3 Set $t = t + 1$.

For all retailers i that are already in the network, set $\tilde{v}_{ij} = 0$.

Otherwise, look up $\tilde{v}_{ij}$ from current state, [t, current willingness state of retailer i, b_j].

Update the remaining collection target, h'_t ($h'_t = h'_{t-1} -$ the actual increment collection volume from the previous period).

Update remaining budget B_t^{start}, ($B_t^{start} = B_{t-1}^{start} -$ actual budget spent in the previous period).

Go to Step 1.

The RTHS procedure is similar to obtaining a good policy on the fly by adjusting the action in each period corresponding to the uncertainty of the willingness state transition of the retailer. The solution from the target recruitment problem is a value of *target recruitment function*, TR, which is defined in section 6.6. This is a function of region, budget amount, and target collection volume. As shown in Figure 6.12, the algorithm performs this procedure numerous times for given regional target collection volume and budget settings in order to obtain expected collection volume. The strategic model uses the expected collection volume value as the parameter $\tilde{v}_{mjkn}$ in its IP formulation (n represents the macro-period).

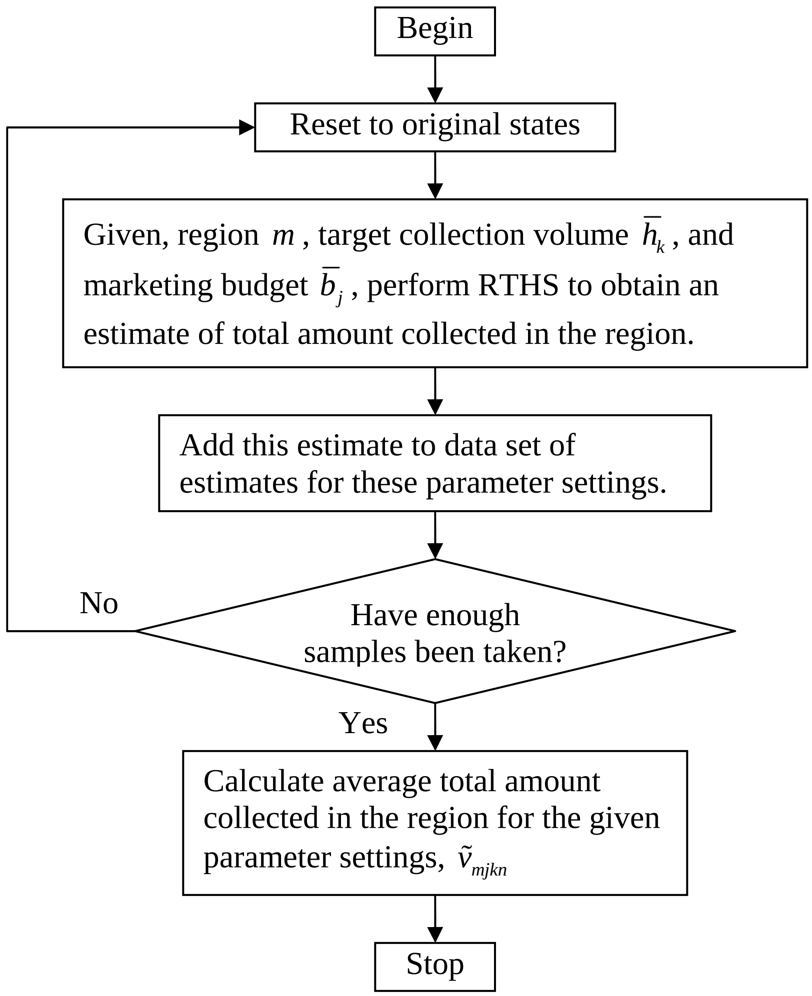

Figure 6.12: Procedure for TR Function

In this chapter, the estimate of the collection cost function is proposed. In addition, the general framework of the strategic problem is examined. I developed the general framework of this problem, called the resource allocation-collection multi-time MDP model. The strategic problem is then adapted for this framework. The complexity of the model makes it prohibitive to obtain the optimal policy in reasonable computation time. Thus, I proposed a heuristics to solve the strategic level problem. This heuristics' performance will be tested with small examples in Chapter 7 and the heuristics will be used in a large-scale example. Insights and possible improvements are also discussed.

CHAPTER 7

NUMERICAL STUDY ON THE

STRATEGIC LEVEL PROBLEM

In Chapter 6 I posed the conceptual model of the strategic problem and provided a heuristic to determine the recruiting budget allocation in each macro-period. I also provided the description and approach for the target recruitment problem. In this chapter, the target recruitment problem is exemplified in section 7.1. Then, the strategic trajectory heuristic approach is applied to small and large examples of the strategic problem in section 7.2 and 7.3. In addition, possible improvements are studied and discussed.

7.1 Target Recruitment Study

In this study, I explore how the region size (number of included retailers), marketing recruitment budget, and target collection volume are related by solving many target recruitment problems by the RTHS method proposed in section 6.7. There are three different regions in this study. Each region has the same proportion of retailer's initial willingness state: low (50%), medium (30%), and high (20%). The number of retailers in regions A, B, and C is 10, 20, and 30 respectively. The collection volume of the retailer in each region is generated according to the lognormal distribution with mean 1.2 and standard deviation 0.5. The complete details of the data are shown in Tables B1, B2, and B3 in Appendix B. Next, I consider fixing two dimensions from 1) recruitment budget, 2) region size, and 3) target collection volume capacity, and investigating the impact of varying the third.

7.1.1 Fixing Target Collection Volume and Recruitment Budget

In this test case, the target collection volume is set to 100 and the recruitment budget is set to 100. The RTHS method is used to simulate 100 replications of the solution for actual collection volume with different region size settings. The solutions for actual collection volumes are then fitted to a normal distribution. The normal distribution curve for each region's solution is shown in Figure 7.1. The narrow curve implies small standard deviation and the center of the curve represents the average value. Figure 7.1 shows how the size of the region has an impact on the solution quality. The solution for region C yields a high probability of obtaining the target collection volume, with a narrow and high curve centered at 100. The solution for region A yields the worst result. Its average value is below the target and the normal curve has a very wide shape.

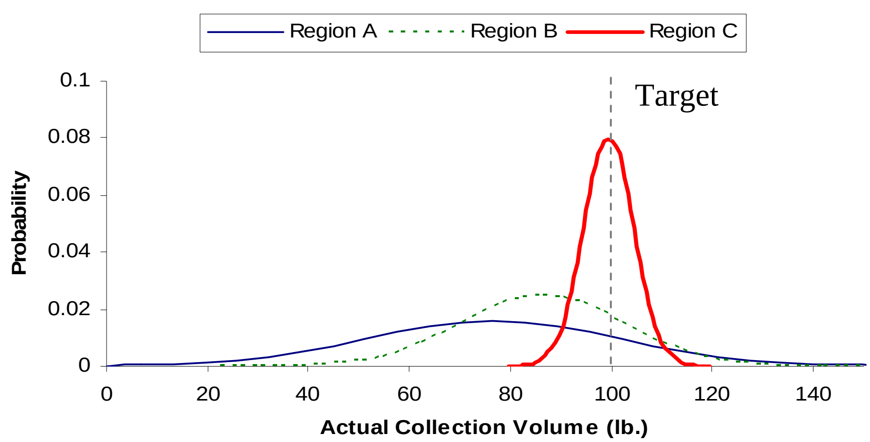

Figure 7.1: Normal Curves of Actual Collection Volume with Target Collection Volume 100 and Budget 40

Figure 7.1 also implies the characteristic of each region. Given that the recruitment budget 40 is considered small, the wide normal curve for region A implies that there are not many retailers who start in the high state. However, as the number of retailers increases in region B and C, the normal curves for region B and C are narrower. The growth in the number of retailers in a region also increases the number of retailers who start in the high state because of the assumption about constant proportions of retailers in each state in each region. Consider the case where the proportion is in favor of the high initial willingness state, i.e. low (30%), medium (30%), and high (40%). The normal curve of each region's solution is shown in Figure 7.3. The normal curve for region A in Figure 7.3 has a higher mean than the one in Figure 7.1. Unlike the normal curve for region B in Figure 7.1, the normal curve for region B in Figure 7.3 is narrow and centered at the target. The normal curve for region C in Figure 7.3 is narrower that the one in Figure 7.1.

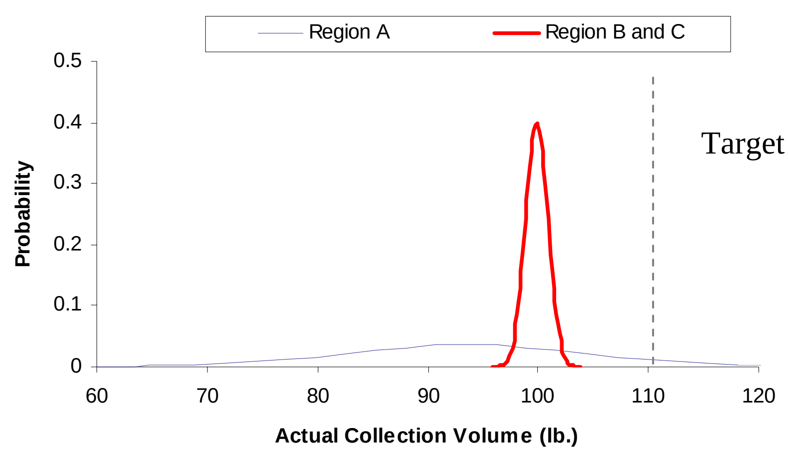

Figure 7.2: Normal Curves of Actual Collection Volume with Target Collection Volume 100 and Budget 40

Next, the target collection volume is set to 200 and the recruitment budget is set to 40. I hypothesize that it is harder to meet the target collection volume in this case compared to the previous case because the target collection volume setting increases while the recruitment budget setting remains the same. The normal curve of each region's solution is shown in Figure 7.3, which shows how the different number of retailers in the region affects the solution quality. The normal curve for region C implies that the largest collection volume it can achieve with high confidence, with the given settings, is only about 130. Thus, it allocates the recruitment budget in order to achieve collection volume of 130 instead of target collection volume of 200. The solution for region B has higher average collection volume than the one for region C, but has a much higher standard deviation. The solution for region A yields the worst result. Its average value is below 100 and its normal curve has a very wide shape.

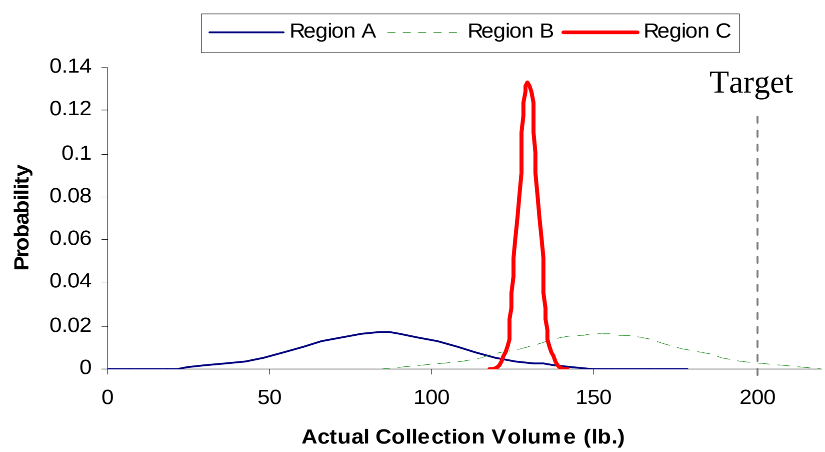

Figure 7.3: Normal Curves of Actual Collection Volume with Target Collection Volume 200 and Budget 40

Next, the target collection volume is set to 200 and the recruitment budget is set to 100. I hypothesize that it is easier to achieve the target collection volume in this case compared to the previous case because of the increase in recruitment budget setting. The normal curve of each region's solution is shown in Figure 7.4, which shows how a different number of retailers in the region affects the solution quality. The normal curves for regions B and C are similar. Both results return the average collection volume close to the target with small standard deviation. The result for region A yields a lower average collection volume than the target with slightly higher standard deviation.

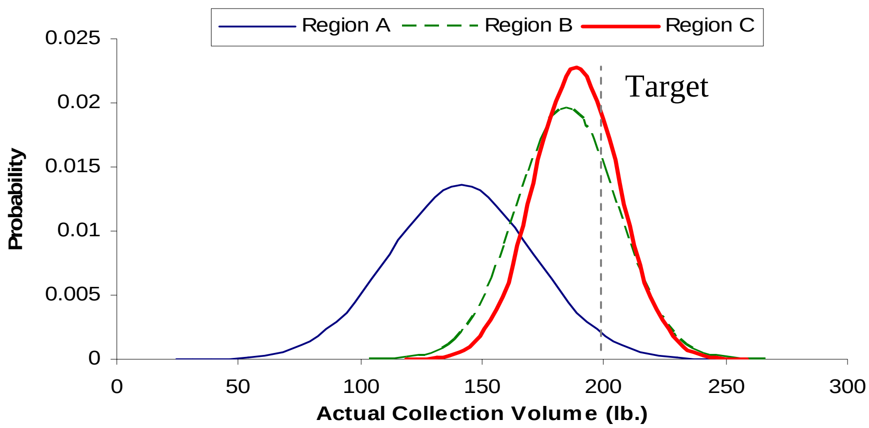

Figure 7.4: Normal Curves of Actual Collection Volume with Target Collection Volume 200 and Budget 100

From Figures 7.1, 7.3, and 7.4, the results show that when the proportion of retailers' initial states is fixed, there is a higher probability to achieve the target collection volume from the region with a larger number of retailers. By fixing the target collection volume and the recruitment budget, it is possible to find the number of retailers in the region or region size so that the target collection volume can be obtained with high confidence. Nevertheless, the budget has to be large enough, and the region contain enough retailers, to meet the volume targets. For example, in Figure 7.3, the recruitment budget of 40 is too small to obtain the target collection volume 200 even in region C, which has the largest number of retailers among the three regions.

Essentially, in this subsection the role of the region's character is highlighted. Each region has a different number of retailers whose initial willingness state are different. Region C has a larger number of retailers who start in the high state than region A. Therefore, region C can recruit more retailers with higher probability than region A for the same amount of budget. Therefore, it is important for the processor to use the information on the characteristic of the region as part of his or her decision making.

The characteristic of the region can be explained by the resultant normal curve for the given target collection volume and recruitment budget. If the normal curve is narrow and tall, it implies that the target collection amount can be achieved with high probability. Therefore, it is possible to achieve a higher target collection volume for the same amount of recruitment budget or to achieve the same target collection volume for the smaller amount of recruitment budget. On the other hand, if the normal curve is wide and flat, it implies that the target collection amount is too high or there is not enough budget amounts to meet the target in that region. Therefore, an increase in the budget amounts may help in meeting the same target, or the reduction of the target collection volume may potentially help achieve the new target with higher probability.

7.1.2 Fixing Target Collection and Number of Retailers

In this test case, the target collection volume is fixed but the recruitment budgets takes different values. The alternative recruitment budget settings are 20, 40, …, 140 and the target collection volume is set to 180. The RTHS approach is used to simulate 100 replications of the solution and the actual collection volumes are found with different recruitment budget settings. Figure 7.5 displays the 90% confidence interval of the actual collection volume solution obtained by the RTHS approach when different recruitment budget settings are used to achieve the target collection volume 180 in region A. The dashed line shows the trend line of the mean collection volume for all recruitment budget settings. The recruitment budget drives the mean collection volume higher but the mean collection volume does not exceed 150. One explanation is because region A is considered hard to achieve a high collection volume target due to the low number of retailers in the favorable state. In addition, the results show that even with a large recruitment budget setting, it is difficult to meet target collection volume with high confidence.

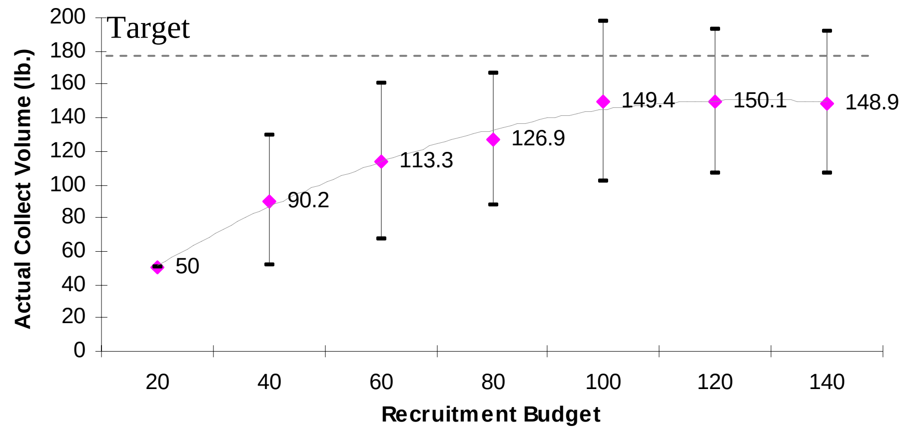

Figure 7.5: 90% Confidence Interval of Actual Collection Volume in Region A with Target Collection Volume 180

Figure 7.6 displays the 90% confidence interval of the actual collection volume solution obtained by the RTHS approach when different recruitment budget settings are used to achieve the target collection volume 180 in region B. The dashed line shows the trend line of the mean collection volume for all recruitment budget settings. The recruitment budget drives the mean collection volume higher until it reaches 178.8. Furthermore, the results show that starting with a recruitment budget setting of 60, the target collection volume can be achieved with high confidence.

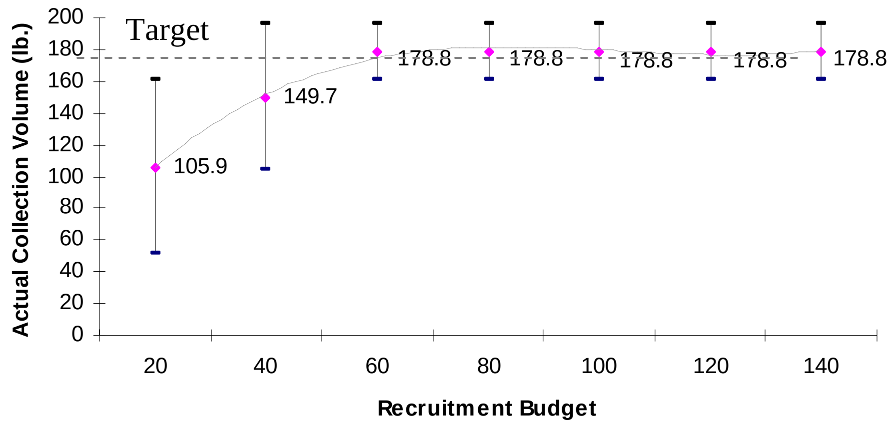

Figure 7.6: 90% Confidence Interval of Actual Collection Volume in Region B with Target Collection Volume 180

Figure 7.7 displays the 90% confidence interval of the actual collection volume solution obtained by the RTHS approach when different recruitment budget settings are used to achieve the target collection volume of 180 in region C. The dashed line shows the trend line of the mean collection volume for all recruitment budget settings. The recruitment budget drives the mean collection volume higher until it reaches 180. In addition, the results show that starting at recruitment budget of 60, the target collection volume can be achieved almost surely.

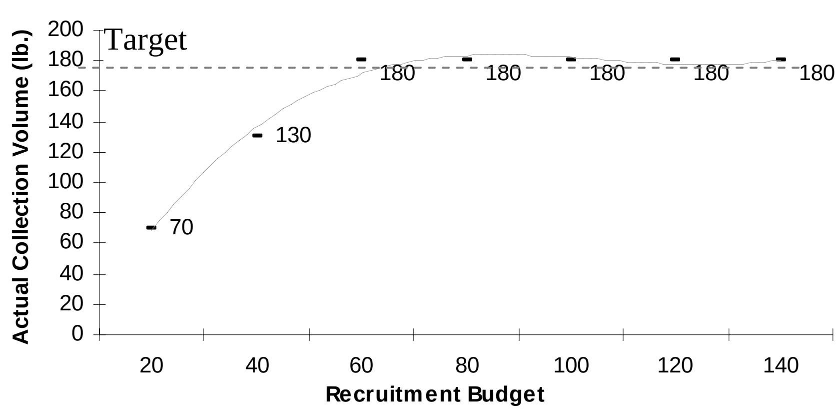

Figure 7.7: 90% Confidence Interval of Actual Collection Volume in Region C with Target Collection Volume 180

Figure 7.8 combines results from the Figures 7.5, 7.6, and 7.7 together. It shows the trend line of the mean collection volume for each region with different recruitment budget settings. The trend lines illustrate the same basic slope. The slope increases when the recruitment budget is small and then converges when the recruitment budget is large.

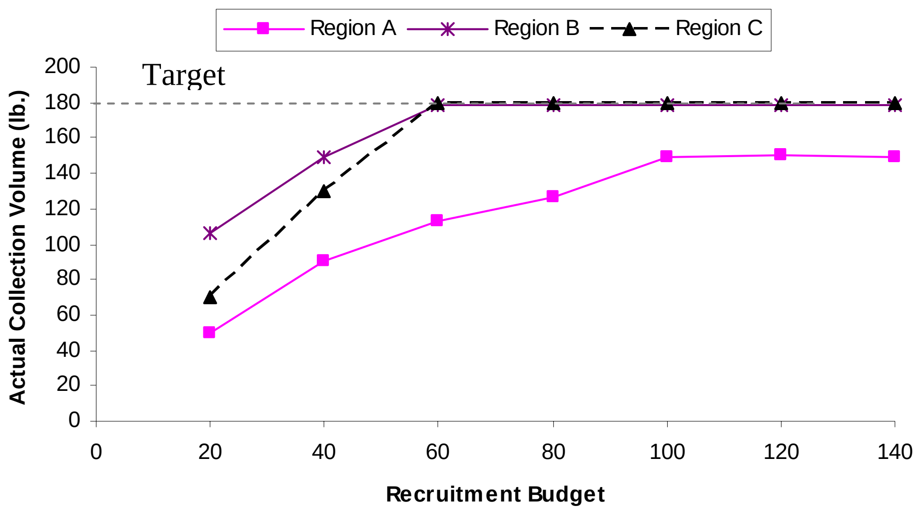

Figure 7.8: Trend line of the Mean Collection Volume in Region A,B, and C with Target Collection Volume 180

From Figures 7.5, 7.6, 7.7, and 7.8, it is possible to find the minimum recruitment budget setting that can achieve the target collection volume with high confidence in different region sizes. However, the characteristic of the region can prevent the recruiter to achieve the target. For example, region A's mean collection volume does not exceed 150 as shown in Figure 7.5.

This subsection also raises the question: what if the poor decisions are made during the recruitment process? Figure 7.7 illustrates that if the recruitment budget of 60 is spent intelligently on Region C, a target amount of 180 can be achieved with high probability by recruiting the retailers who are easily recruited. However, if the recruiter spends the recruitment budget on a retailer who may return a very high collection volume but is difficult to recruit, the recruiter may fail to meet the target. This poor decision may lead to a larger range in the confidence interval of the actual collection volume. Consequently, it becomes harder to find the optimal budget setting that can achieve the target collection volume with high probability. This argument strengthens the significance of the recruitment problem. In addition, achieving the target with high confidence also helps the strategic decision maker with robust planning.

7.1.3 Fixing Recruitment Budget and Number of Retailers

In this test case, the recruitment budget and the region size are fixed but the target collection volume takes different values. The target collection volume settings are 34, 68, …, 340 and the recruitment budget is set to 80. The maximum possible collection volume in region A is 340. The RTHS approach is used to simulate 100 replications of the solution for actual collection volume with different target collection volume settings. Figure 7.9 display the 90% confidence interval of the actual collection volume solution obtained by the RTHS method with different target collection volume settings in region A. The results show that largest target collection volume that it can achieve with high confidence is between 34 and 68. In addition, given the parameter settings, the average solution collection volume does not exceed 147.2.

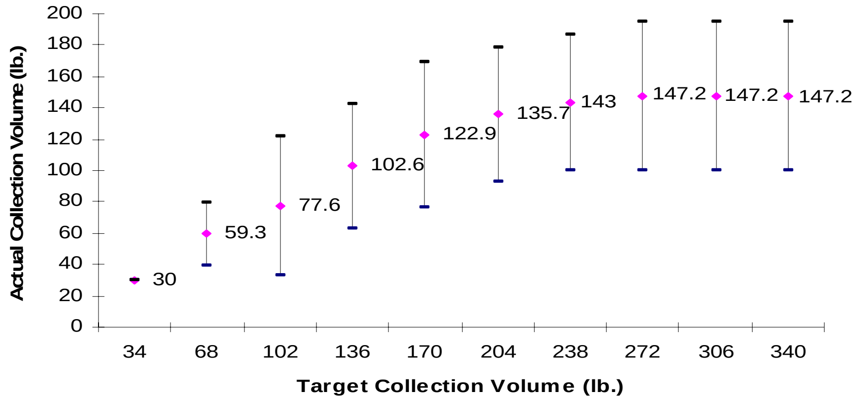

Figure 7.9: 90% Confidence Interval of Expected Collection Volume in Region A with Recruitment Budget 80

Figure 7.10 display the 90% confidence interval of the actual collection volume solution obtained by the RTHS method with different target collection volume settings in region B. The results show that the largest target collection volume that it can achieve with high confidence is between 170 and 204. In addition, given the parameter settings, the average solution collection volume does not exceed 195.6.

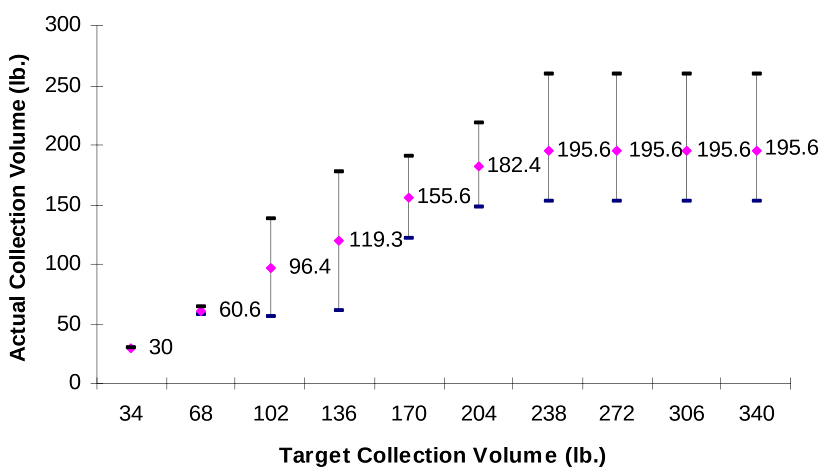

Figure 7.10: 90% Confidence Interval of Expected Collection Volume in Region B with Recruitment Budget 80

Figure 7.11 display the 90% confidence interval when fixed recruiting budget is used to achieve many target collection volumes in region C. The results show that the largest target collection volume that it can achieve with high confidence is between 170 and 204. In addition, given the parameter settings, the average solution collection volume does not exceed 205.8.

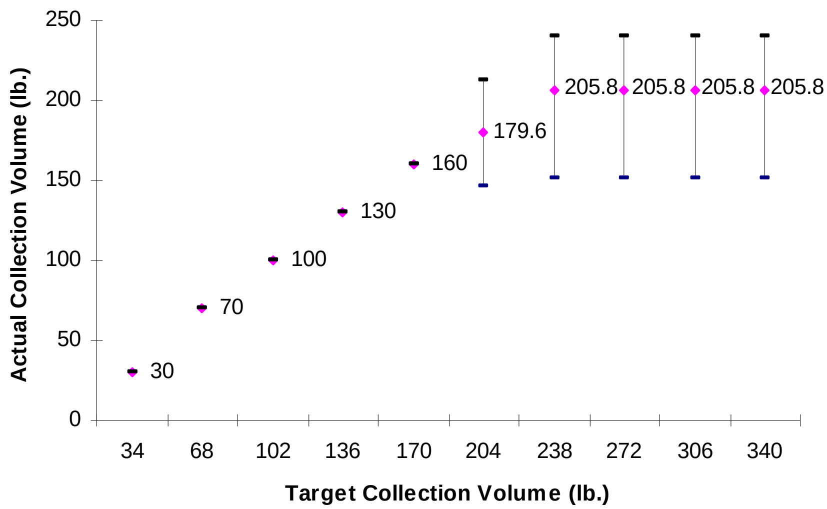

Figure 7.11: 90% Confidence Interval of Expected Collection Volume in Region C with Recruitment Budget 80

From Figures 7.9, 7.10, and 7.11, it is possible to find the minimum target collection volume that the region can achieve with high confidence given a fixed recruitment budget. If the target collection volume is too high, there is a limit on how much the region can collect for a given budget.

In this study, three cases are considered. First, the solutions from the test case in section 7.1.1 show that by fixing the target collection volume and the recruitment budget, it is possible to find the region size that can obtain the target collection volume with high confidence. Second, the solutions from the test case in section 7.1.2 show that it is possible to find the recruitment budget setting that can achieve the target collection volume with high confidence with different region sizes. Lastly, the solutions from the test case in section 7.1.3 show that it is possible to find the target collection volume that the region can achieve with high confidence given a fixed recruitment budget. The conclusions from these cases illustrate a pattern whereby the collection system can fix two dimensions from among 1) recruitment budget, 2) region size, and 3) target collection volume capacity, and then determine the third.

This pattern can enable the processor to perform a pre-selection of the regions in which to attempt recruitment. If the number of regions is large, the processor can choose to concentrate the recruitment from the regions that have a high probability of achieving the predetermined target collection volume. This can be done by fixing the recruitment budget and the region size.

7.2 Small Example

In this section, a data set is generated for each of three test cases. The solutions for these test cases are then obtained by the strategic trajectory heuristic (STH) method. There are two regions and the number of planning years is three. The three-year target collection volume is 50. Region A has 4 retailers and region B has 2 retailers. This small example is used because it is possible to compute the optimal solution to the recruitment problem. The collection cost function is computed by (6.1). The parameter settings in collection cost function are $\overline{c}_D = 1$, $c_D = 0.5$, and $F_m^{(co)} = 20$. The retailers in all regions are centered closely to their regional collector with average distance equal to 1. The difference in transportation cost between the regional collector and the processor is emphasized over the transportation cost within a region. For Cases 7.1 to 7.3, the distance between the processor and the regions are set to $D_A^{(Pr)} = 10$ and $D_B^{(Pr)} = 1$. The total marketing budget is 60 and the maximum budget in each year is 20. The alternative recruitment budget (b_j) settings are 0, 10, 20 and the target collection volume (h_k) settings are 10, 20, 30, 40, 50. The collection volume in each region is generated such that it is not enough to achieve the target collection volume by only collecting from region B, which is closer to the processor. Next, three different test cases are considered where the initial willingness states in each region are different.

Case 7.1: Easy to recruit in both regions

The data in Table 7.1 is given and the heuristic policy is obtained by the STH approach. The iteration number limit is set to 100.

Table 7.1: Data Set for Case 7.1

Retailer	Region	Collection Volume (lb./month)	Recruitment Budget Threshold	Initial Willingness State
1	A	30	15	H
2	A	40	20	M
3	A	20	10	H
4	A	60	36	M
1	B	10	5	M
2	B	20	10	M

The resultant policy obtained by the STH approach is then simulated for 100 replications to find the actual collection volume and the total collection cost at the end of the third year. Figure 7.12 shows the histogram of the final period actual collection volume. In addition, the average cost per pound of the samples in each actual collection volume value (40 and 50 in this case) is computed and displayed because the same actual collection volume does not always produce the same collection cost. In this case, the heuristic policy obtained by the STH method can achieve the target collection volume in almost all replications. Since the data set is small, the minimum collection cost can be calculated. In this case, it is best to recruit retailer 1 and 2 from region B and retailer 3 from region A. Hence, the minimum cost per pound is 2.64.

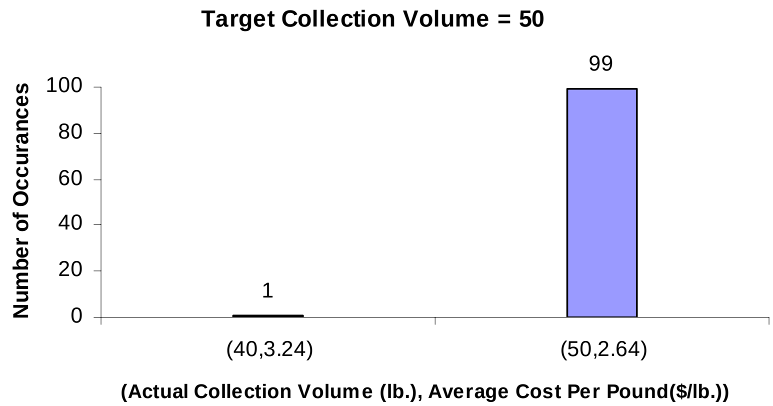

Figure 7.12: Actual Collection Histogram for Case 7.1

For cases 7.1 and 7.2, the actual collection histogram of the optimal policy is shown in Figure 7.13.

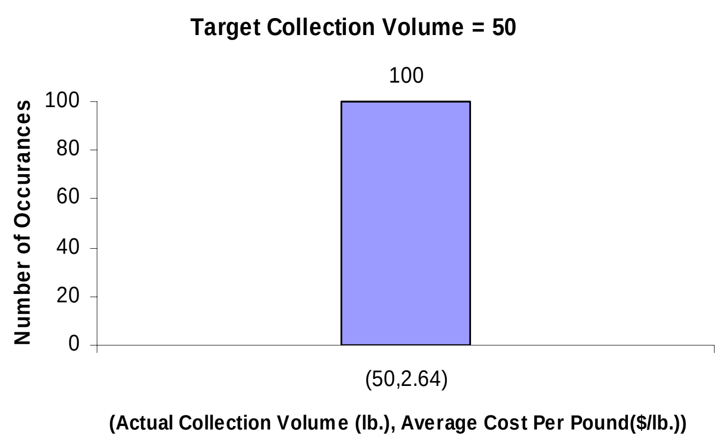

Figure 7.13: Actual Collection Histogram of Optimal Policy for Case 7.1 and 7.2

The total running time for Case 7.1 is 14 minutes. The solution average, minimum, maximum of the collection volume, total cost and, cost per pound are shown in Table 7.2. When the actual collection volume is 50, its average cost per pound is 2.64 which is equal to the minimum value. In this test case, the policy obtained by the STH method is able to achieve the minimum collection cost and meet the target collection volume in 99 replications. The heuristic policy performs almost as well as the optimal policy in this case.

Table 7.2: Result Summary for Case 7.1

	Collection Volume (lb.)	Total Cost ($)	Cost Per Pound ($/lb.)
Average	49.90	132.20	2.65
Minimum	40.00	129.56	2.64
Maximum	50.00	132.23	3.24

Case 7.2: Easy to recruit in region A but difficult to recruit in region B

The data in Table 7.3 is given and the heuristic policy is obtained by the STH approach. The iteration number limit is set to 100.

Table 7.3: Data Set for Case 7.2

Retailer	Region	Collection Volume (lb./month)	Recruitment Budget Threshold	Initial Willingness State
1	A	30	15	H
2	A	40	20	M
3	A	20	10	H
4	A	60	36	M
1	B	10	5	L
2	B	20	10	L

The resultant policy obtained by the STH approach is then simulated for 100 replications to compute the actual collection volume and the total collection cost at the end of the third year. Figure 7.14 shows the histogram of the final period actual collection volume. In this case, the heuristic policy obtained by STH method can achieve the target collection volume in most replications. Since it is more difficult to recruit the retailers in region B, there is a high probability that the allocated budget fails to recruit those retailers in the earlier years.

If the retailers in region B fail to be recruited in the first two years, there are two choices in the heuristic policy in the final year. If only retailer 2 in region B is recruited, the policy chooses to spend the rest of the budget to recruit retailer 3 in region B to achieve the total collection volume of 40 even though it costs more. If only retailer 1 in region B is recruited, the policy chooses to spend the rest of the budget to recruit retailer 2 in region B instead of retailer 3 in region A to achieve the total collection volume of 30 because the total collection volume would be the same but the total cost would be lower. This explains the 30 percent occurrence of the policy obtained by the STH method missing the target collection volume. Even though it is easy to recruit the retailers in region A to meet the target collection volume, its collection cost is higher than the collection cost in region B. Hence, the resultant policy attempts to recruit the retailers from region B first. Compared to the optimal policy in Figure 7.13, the heuristic policy in this case performs well but not as well as in case 7.1.

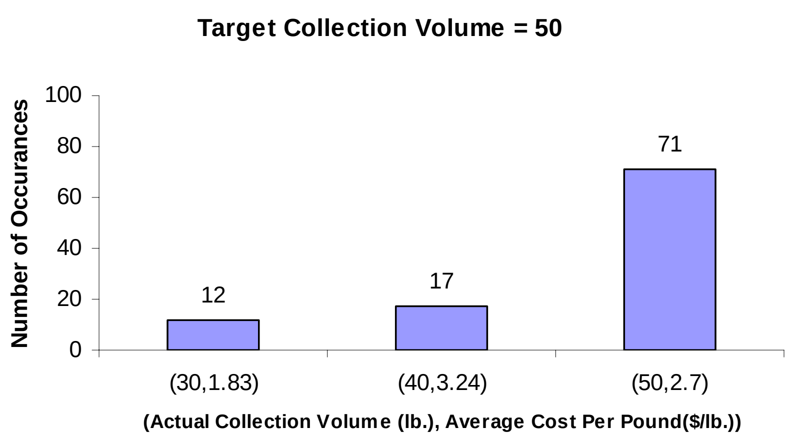

Figure 7.14: Actual Collection Histogram for Case 7.2

The total running time for Case 7.2 is 15 minutes. The solution average, minimum, maximum of the collection volume, total cost and, cost per pound are shown in Table 7.4.

When the actual collection volume is 50, its average cost per pound is 2.70, which is slightly higher than the minimum. When the actual collection volume is 40, its average cost per pound is 3.24. This value is high because the collection volumes are from region A only. On the other hand, when the actual collection volume is 30, its average cost per pound is 1.03. This value is low because the collection volumes are from region 2 only.

Table 7.4: Result Summary for Case 7.2

	Collection Volume (lb.)	Total Cost ($)	Cost Per lb ($/lb.)
Average	45.90	124.64	2.69
Minimum	30.00	30.81	1.03
Maximum	50.00	156.28	4.21

<u>**Case 7.3:** Difficulty in recruiting in both regions</u>

The data in Table 7.5 is given and the heuristic policy is obtained by the STH approach. The iteration limit is set to 100.

Table 7.5: Data Set for Case 7.3

Retailer	Region	Collection Volume (lb./month)	Recruitment Budget Threshold	Initial Willingness State
1	A	30	15	L
2	A	40	20	L
3	A	20	10	L
4	A	60	36	L
1	B	10	5	L
2	B	20	10	L

The resultant policy obtained by the STH approach is then simulated for 100 replications to acquire actual collection volume and the total collection cost at the end of the third year. Figure 7.15 shows the histogram of the final period actual collection volume. In this case, the heuristic policy obtained by STH method can achieve the target collection volume in 26 replications. Since it is more difficult to recruit the retailers in both regions, there is a higher probability that the allocated budget fails to recruit those retailers. The percentage that the policy obtained by the STH method misses the target collection volume is high (74 percent). The actual collection histogram of the optimal policy for case 7.3 is shown in Figure 7.16. Compared to the optimal policy in Figure 7.16, the heuristic policy in this case does not perform well. The optimal policy can meet the target collection volume in 48 more replications.

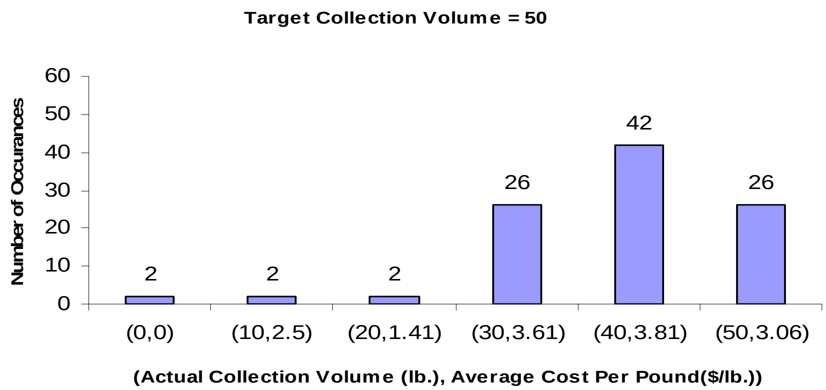

Figure 7.15: Actual Collection Histogram for Case 7.3

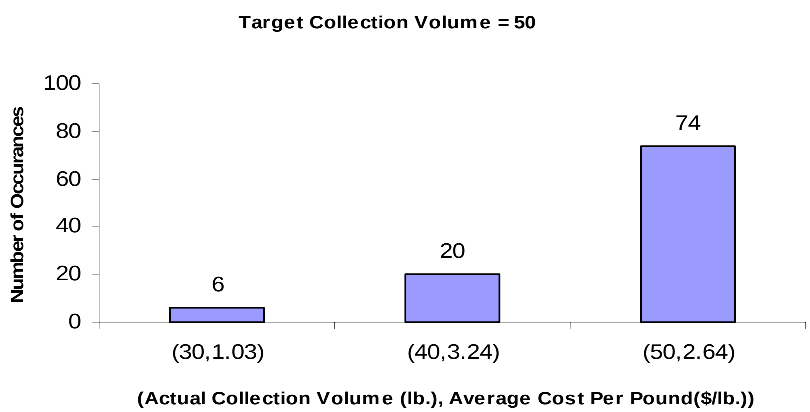

Figure 7.16: Actual Collection Histogram of Optimal Policy for Case 7.3

The total running time for Case 7.3 is 16 minutes. The solution average, minimum, maximum of the collection volume, total cost and, cost per pound are shown in Table 7.6. When the actual collection volume is 50, its average cost per pound is 3.06, which is higher than the minimum. The reason is that since both regions are difficult to recruit, sometimes it is better to concentrate on just region A to achieve the target collection volume because the retailers in region A have higher collection volume. In the worst scenario, there are a few replications where no retailer is successfully joined to the network.

Table 7.6: Result Summary for Case 7.3

	Collection Volume (lb.)	Total Cost ($)	Cost Per lb ($/lb.)
Average	38.20	133.03	3.41
Minimum	0.00	0.00	1.03
Maximum	50.00	174.62	4.27

In order to improve the performance of the heuristic policy, the problem in case 7.3 is resolved with some modifications in the parameter settings. The target collection (h_k) settings are discretized in a finer space {0, 5, 10,…, 50}.The iteration limit is increased to 200. Figure 7.17 shows the histogram of the final period actual collection volume.

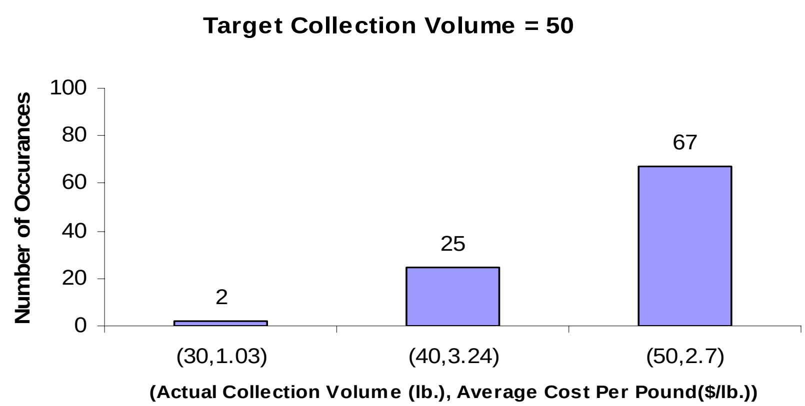

Figure 7.17: Actual Collection Histogram for Case 7.3 with Parameter Modifications

Figure 7.17 shows a tremendous improvement in the solution quality. However, the total running time in this case increases to 41 minutes, compared to 16 minutes in case 7.3.

From the results in Cases 7.1, 7.2, and 7.3, I empirically show that heuristic policy obtained by STH method performs well when the retailers in all regions are easy to recruit. The target collection volume is achieved in most of the replications. However, when the retailers in all regions are difficult to recruit, the performance of the policy acquired by the STH method deteriorates. However, the performance can be improved by adjusting the discretization of the target collection volume, including the iteration limit.

For a small example, the other solution method to solve MMDP such as the approximation method (Chang et al. 2003) and actor-critic algorithm (Panigrahi and Bhatnagar 2006) may provide better solution performance. However, this research aims to develop a methodology to support decision making for a realistic sized problem in order to provide insight to the processor. The STH method is tested on a large sized example in the next section.

7.3 Large Example

In this section, a large data set is generated and explored for four test cases. The computational requirements of the STH approach depend on not only the number of regions and the number of retailers in those regions, but also the size of the target collection volume settings h_k and the size of the recruitment budget settings b_j. In order to solve the large scale problem with reasonable computation effort, the size of the target and budget is chosen carefully through discretization. Next, four different test cases are considered with different parameter settings.

Case 7.4: Easy to recruit

There are two regions (region A and region B) in this case and each region contains 50 retailers. All of the retailers' initial willingness states in both regions are set to the high (H) value. The collection volume of the retailers in region A is smaller than the one in region B. Region A is closer to the processor than region B, so let $D_A^{(Pr)} = 10$ and $D_B^{(Pr)} = 20$. The total marketing budget is set to 600 for three years. The marketing budget limit for each year is 200. The target collection volume at the end of the third year is 2000. The recruitment budget settings are 0, 40, 80, …, 200 and the target collection volume settings are 300, 600,…, 1500.

The resultant policy obtained by the STH approach is then simulated for 100 replications to estimate actual collection volume and the total collection cost at the end of the third year. Figure 7.18 shows the histogram of the final period actual collection volume. In this figure, 2000s represents the collection volume from 2000 to 2100. In this case, the heuristic policy obtained by STH method can achieve the collection volume in the range of 1900-2200 in 94 replications. In this case, the lower bound of the cost per pound is computed from the cost of collecting 2000 from region 1 alone which is 1.04.

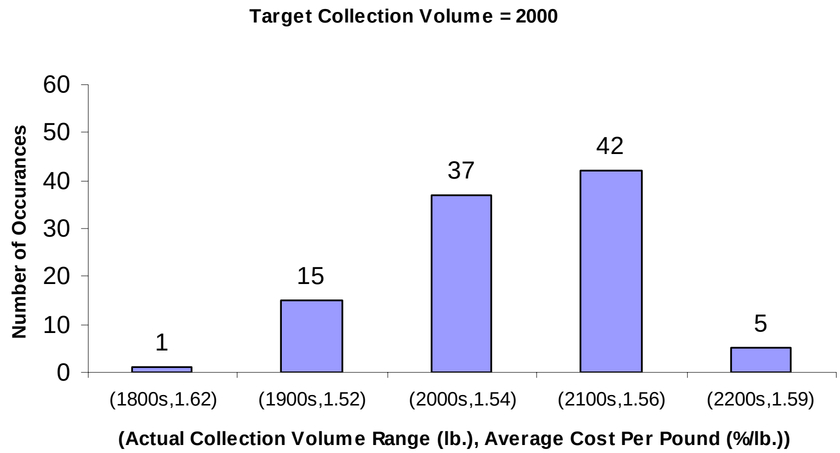

Figure 7.18: Range of Actual Collection Histogram for Case 7.4

The total running time for Case 7.4 is 69 minutes. An increase in number of retailers in the region adds more complexity to the original problem. The solution average, minimum, maximum of the collection volume, total cost and, cost per pound are shown in Table 7.7. The numerical results show that two thirds of the actual collection volume is from region A. This implies that even though it is cheaper to collect from region A, the recruitment budget is not enough to recruit enough retailers to meet the target collection volume. Hence, it has to allocate part of the recruitment budget to recruit the retailers in region B, which contains retailers with higher collection volume, to join the network. As a result, the cost per pound in every replication from the simulation is higher than the proposed lower bound.

Overall, the heuristic policy performs well, as it should, because of the easy recruitment in both regions. Only 16 replications fail to meet the target. In most replications, the actual collection volume is off by only 5%.

Table 7.7: Result Summary for Case 7.4

	Collection Volume (lb.)	Total Cost ($)	Cost Per lb ($/lb.)
Average	2,086.96	3,228.36	1.55
Minimum	1,872.40	2,825.17	1.47
Maximum	2,266.50	3,600.74	1.68

Case 7.5: Three regions with different initial states

There are three regions in this case and the number of retailers in region A, B, and C is 10, 10 and 5 respectively. Region A is the furthest away from the processor but the retailers in this region are easy to recruit (initial willingness state H). Region C is the closest to the processor but the retailers in this region are very hard to recruit (initial willingness state L). Region B locates closer to the processor than region A, but further than region C. All of the retailers' initial willingness states in region B are set to the medium (M) value. The collection volume of the retailers in region C is low compared to the one in region A and B. The complete details of the data for this case are shown in Table B4. Let $D_A^{(Pr)} = 20$, $D_B^{(Pr)} = 10$, and $D_C^{(Pr)} = 1$. The total marketing budget is 150 for three years. The marketing budget limit for each year is 50. The target collection volume at the end of the third year is 300. The recruitment budget settings are 0, 10, 20, …, 50 and the target collection volume settings are 30, 60,…, 300.

The resultant policy obtained by the STH approach is then simulated for 100 replications to acquire actual collection volume and the total collection cost at the end of the third year. Figure 7.19 shows the histogram of the final period actual collection volume. In this figure, 300s represents the collection volume from 300 to 320. In this case, the heuristic policy obtained by STH method can achieve the collection volume in a wide range of 160-360 with the concentration in the middle (260-300).

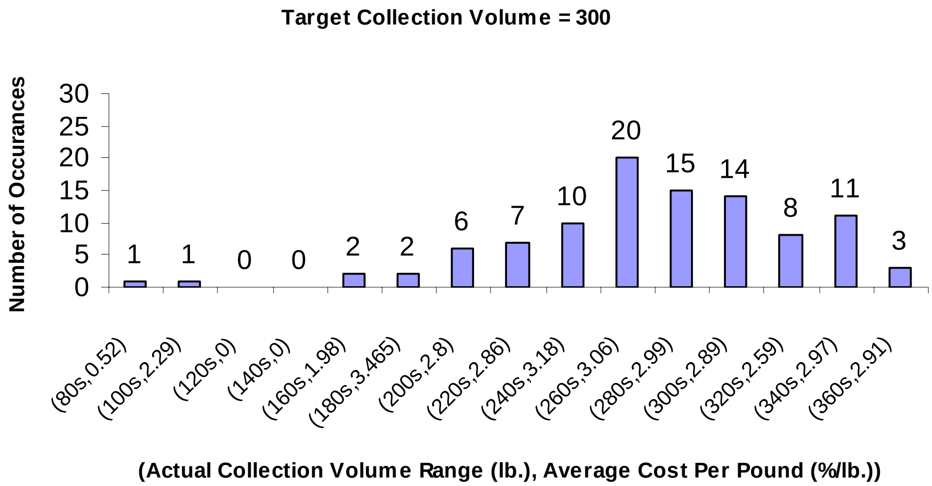

Figure 7.19: Range of Actual Collection Histogram for Case 7.5

The total running time for Case 7.5 is 42 minutes. An increase in the number of regions adds more complexity to the original problem. The solution average, minimum, maximum of the collection volume, total cost and, cost per pound are shown in Table 7.8. The lowest cost per pound of the replication where the actual collection volume meets the target is 1.69. In this replication, it only collects from regions B and C. This is reasonable because the total collection from region C is only 80 and the collection cost in region B is cheaper than the one in region A. By examining the resultant policy, if the regional collector of region A or B fails to meet the target assigned by the processor in the earlier year, the processor is forced to turn to region A in a later year where the retailers are easiest to recruit in order to meet the target collection volume. Consequently, the total collection cost is raised higher.

Overall, the heuristic policy does not perform well. There are 63 replications that fail to meet the target. In the worst case, the actual collection volume is only 80 compared to the target of 300. However, this can be explained from the fact that region C is the hardest to recruit but has the cheapest total collection cost.

Table 7.8: Result Summary for Case 7.5

	Collection Volume (lb.)	Total Cost ($)	Cost Per lb ($/lb.)
Average	264.60	774.95	2.91
Minimum	80.00	41.49	0.52
Maximum	350.00	1,046.65	4.17

<u>Case 7.6: Three regions with random initial states</u>

There are three regions in this case and the number of retailers in region A, B, and C is 20, 20 and 10 respectively. Region C is the furthest away from the processor but the retailers in this region are easy to recruit. Region A is the closest to the processor. Region B locates closer to the processor than region C but further than region A. All of the retailers' initial willingness states in all regions are randomly assigned. The collection volumes of the retailers in all regions are generated from the same lognormal distribution. The complete details of the data for this case are shown in Table B5. Let $D_A^{(Pr)} = 15$, $D_B^{(Pr)} = 20$, and $D_C^{(Pr)} = 25$. The total marketing budget is 180 for three years. The marketing budget limit for each year is 60. The target collection volume at the end of the third year is 500. The recruitment budget settings are 0, 10, 20,…, 60 and the target collection volume settings are 50, 100,…, 500.

The resultant policy obtained by the STH approach is then simulated for 100 replications to acquire actual collection volume and the total collection cost at the end of the third year. Figure 7.20 shows the histogram of the final period actual collection volume. In this figure, 500s represents the collection volume from 500 to 530. In this case, the heuristic policy obtained by STH method can achieve the collection volume in the wide range of 410-590 with the concentration in the middle (500-530). It is not obvious in this case what the lower bound of cost per pound is.

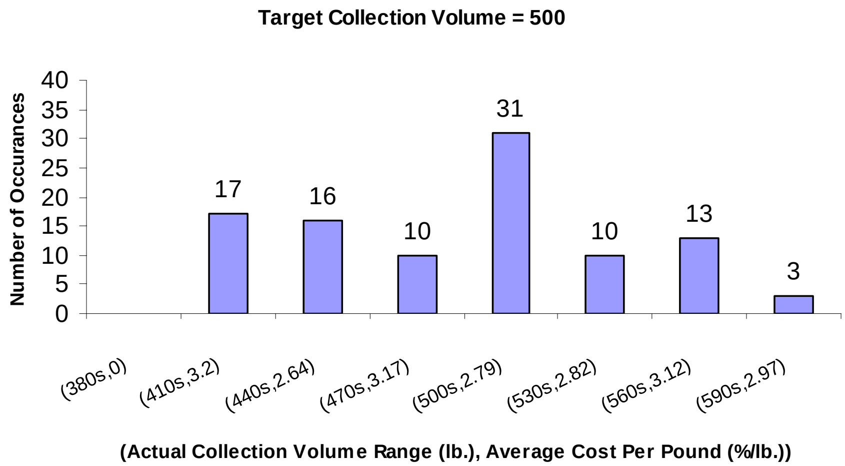

Figure 7.20: Range of Actual Collection Histogram for Case 7.6

The total running time for Case 7.6 is 208 minutes. An increase in running time comes from increase in number of regions and number of retailers per region. The solution average, minimum, maximum of the collection volume, total cost and, cost per pound are shown in Table 7.9. By examining the resultant policy, it attempts to recruit region A and B first. There are some cases where region C is recruited to the network.

Overall, there are 43 replications that fail to meet the target. There is no obvious explanation to determine the performance of the heuristic policy because the initial states of the retailers in all regions are random. If the recruitment budget is large enough, the policy may have done better. However, it is not clear in this case.

Table 7.9: Result Summary for Case 7.6

	Collection Volume (lb.)	Total Cost ($)	Cost Per lb ($/lb.)
Average	473.90	1,380.13	2.92
Minimum	360.00	979.27	2.31
Maximum	590.00	2,020.40	4.11

In the small and large examples, the STH method yields the heuristic policy that performs well when the initial state of the retailers is H. When the initial states of the retailers contain many low (L) and medium (M) values, the STH method finds it difficult to provide a policy that always returns a collection volume that achieves the target. The retailer who starts with a low initial state can move to other states with different probabilities. Hence, there are many more states to consider for the strategic level problem. Still, the STH method can provide a policy that returns the collection volume close to the given target even though the collection cost might not be the minimum.

Considering current computational time requirements, it will take more than 208 minutes to solve the problem with three regions where each region can have hundreds of retailers. Next, I propose how to improve the computation time through a preprocessing scheme.

Preprocessing Phase

I attempt to solve the strategic problem with three regions with a large number of retailers as would be expected in a real case. Region A, B and C have 200, 50, 50 retailers respectively. However, even in case 7.6 where the total number of retailers is only 50 for three regions, it takes 208 minutes to solve, and hence I would expect very significant computation times for regions of this size. I introduce a preprocessing phase that helps improve the computational effort.

In the strategic IP formulation discussed in section 6.6.1, in order to compute for expected collection volume of region m using budget $\overline{b}_j$ to meet target $\overline{h}_k$ in each period, $\tilde{v}_{mjkn}$, the target recruitment problem needs to be resolved for every parameter setting. The majority of the effort comes from solving the recruitment problem for all $\tilde{v}_{mjkn}$. The total number of recruitment problems depends on the number of retailers in region m, the budget size and the target size. Under the RTHS method, in order to solve TRP the expected collection volume for each retailer i from budget b_j, $\tilde{v}_{ij}$, and their action in that period needs to be obtained. This depends also on which willingness state that retailer is in and the period.

Previously, $\tilde{v}_{ij}$ is recomputed every time the value is needed using the exact DP. However, solving for each $\tilde{v}_{ij}$ and its corresponding action for every retailer, starting willingness state, and month period can be computed once at the beginning. The results are stored in a look up table. A hash table is used to store these results in order to improve the retrieval time. The size of this hash table is:

$$\text{The size of hash table} = \text{number of regions} \times \text{number of retailers} \times \text{budget size} \times \text{number of willingness state} \times \text{number of months} \tag{7.1}$$

This preprocessing scheme is then applied to the data set in Case 7.7.

<u>Case 7.7:</u> <u>Large Scale Problem</u>

There are three regions. Region A, B, and C has 200, 50, and 50 retailers respectively. All of the retailers' initial willingness states in all regions are randomly assigned. Moreover, region A, B, and C has initial collection volume of 131.7, 257.1, and 348.8 respectively. Let $D_A^{(Pr)} = 10$, $D_B^{(Pr)} = 20$ and $D_C^{(Pr)} = 30$. The total marketing budget is 300 for three years. The collection volume in retailer for each region A, B and C is drawn from uniform distribution [0,30], [30,60], and [60-90] respectively. The marketing budget limit each year is 100. The target collection volume at the end of third year is 2000. In this case, region A and region B should be chosen to join the network before region C because of the distance between the regional collector and the processor. The recruitment budget settings are 0, 10, 20,…, 100 and the target collection volume settings are 100, 200,…, 1000.

The total running time to generate 100 trajectories is 240 minutes. This is a considerable improvement compared to Case 7.5. The heuristic policy is shown in Table 7.10.

Table 7.10: Heuristic Policy for Case 7.7

	Region	Budget Allocation	Target Collection Volume (lb.)
Year 1	1	60	200
Year 2	1	80	200
Year 2	2	20	100
Year 3	2	100	500
Year 3	3	40	300

The projected collection volume at the end of year 3 is to collect 531.7 from region A, 857.1 from region B and 648.8 from region C. The total projected collection volume is 2,037.6. The resultant policy obtained by the STH approach is then simulated for 100 replications to acquire actual collection volume and the total collection cost at the end of the third year. In all replications the actual collection volume is 2,000.4 and the total collection cost is 5,876.13 (cost per pound 2.94). Since there are numerous retailers in each region that start in state H, the STH method attempts to recruit these retailers first. The STH approach performs very well in identifying these favorable retailers in each region and electing to recruit them to join the network accordingly. It is surprising that the solution policy only recruits the retailers in the state H. However, it is reasonable because the recruiting this type of retailers is easy. Therefore, assigned region can achieve the target collection volume aimed by the processor in each macro-period. Nevertheless, these results do not guarantee the minimum cost per pound.

This numerical study for the strategic problem may be considered a "small" realistic problem. Large scale realistic sized problems would have about 8 regions and each region could have a couple of hundred retailers. The large scale example discussed above is solved using a Windows 2000-based Pentium 4 1.80 GHz personal computer with 640MB of RAM. However, if a parallel processor-based platform is available to solve the strategic problem, the computation time can be reduced tremendously. The advantages of parallel computing can be exploited in three following ways:

1. Solving each retailer in the recruitment problem in parallel to obtain the expected collection volume for every budget setting.

2. Solving each region in parallel to obtain the expected collection volume for every budget and target setting.

3. Solving each trajectory of the strategic problem in parallel.

7.4 A Large Carpet Producer Case Study

To illustrate the approach on a more realistically sized large scale problem that is representative of an actual industry challenge, I have approximated a situation similar to one currently faced by a large carpet producer. In this research case study these data have been estimated by me, and not provided by the company, so in no way do they accurately depict the true financials of the company. I assume that there are five regions to consider. The details for the case study regional information are shown in Table 7.11. The population information for the states under consideration has been retrieved from the

U.S. Census Bureau[10] and the cost per pound has been quoted from American Freight Companies.[11] I assume that the processing site is located at Augusta, Georgia. For purposes of estimating the amount of used carpet available, I assume that one person generates about 17 pounds of carpet per year.[12]

Table 7.11 Case Study Data - Regions Information

Region Name	States included in the Region	Population	Total Population	Approximate Volume (lb.)	Distance to Augusta, GA (miles)	Linear Shipping Cost per Lb ($/lb.)
1. Atlanta	GA	9,072,576	50,321,532	855,466,044	148	0.015
	FL	17,789,864				
	TN	5,962,959				
	AL	4,557,808				
	SC	4,255,083				
	NC	8,683,242				
2. New York	NY	19,254,630	50,311,211	855,290,587	785	0.035
	PA	12,429,616				
	MA	6,398,743				
	NJ	8,717,925				
	CT	3,510,297				
3. Los Angeles	CA	36,132,147	48,127,302	818,164,134	2,363	0.101
	NV	2,414,807				
	OR	3,641,056				
	AZ	5,939,292				
4. Dallas	TX	22,859,968	35,639,018	605,863,306	928	0.041
	NM	1,928,384				
	OK	3,547,884				
	AR	2,779,154				
	LA	4,523,628				
5. Chicago	IL	12,763,371	31,800,951	540,616,167	864	0.039
	WI	5,536,201				
	IA	1,429,096				
	MO	5,800,310				
	IN	6,271,973				

10 http://www.census.gov/popest/datasets.html

11 http://www.freightcenter.com

12 Source: Dr. Matthew Realff, Chemical and Biomolecular Engineering, Georgia Institute of Technology.

In addition, due to the relative size of landfill tipping fees in those areas, I assume that the retailers in the New York and Los Angeles regions mostly have initial states 'H' while the retailers in the Atlanta, Dallas and Chicago regions mostly have initial states 'M' or 'L'. The total population and approximated collection volume in the Atlanta, New York and Los Angeles regions are similar, so I assume that the number of retailers in each of these regions is the same. The data on the amount of carpet generated (square yards) for all available retailers in the Atlanta region are obtained and fitted to a lognormal distribution. Using these data, the data in the Dallas and Chicago regions are approximated by scaling using the total populations. The histogram and the fitted curve of the Atlanta, Dallas and Chicago regions are shown in Figure 7.21. One square yard of used carpet is approximately 4.5 pounds.[13]

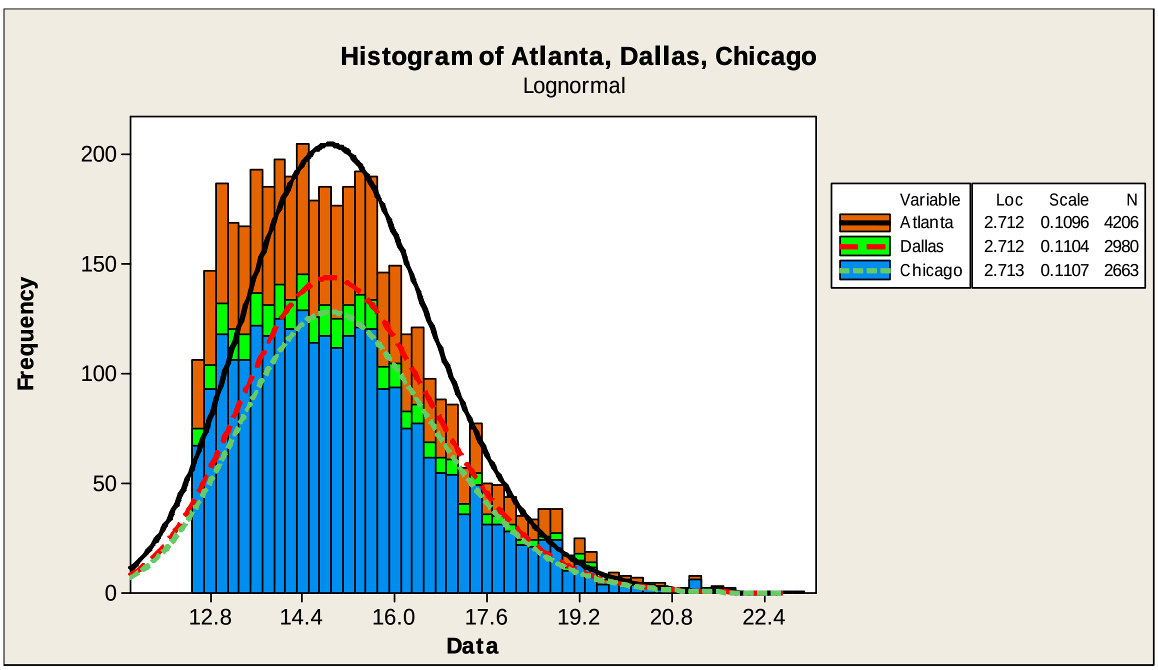

Figure 7.21 Histogram of Retailer Size (Square Yards) in Atlanta, Dallas, and Chicago Regions

13 Source: Carpet and Rug Institute (2003)

For a realistically sized problem, the actual number of retailers is in thousands. Due to computation limits, I consider 50 retailers each in Atlanta, New York and Los Angeles, 35 retailers in Dallas, and 30 retailers in Chicago. As a result, the solution computation time for case study scenario is about 2-3 hours. In the Atlanta region, the top 50 retailers (in term of square yards) are selected, but the total generated volume for these 50 retailers exceeds the approximated collection volume in Table 7.11. Hence, the collection volume in each retailer is scaled by the ratio of the regional approximated collection volume to the total volume of the top 50 retailers. The collection volumes of the retailers for the remaining regions are computed in the same way. The scaling of the volume per retailer is important in the sense that it prevents the problem from being too simple. If the collection volume per retailer is large, it is easy to choose a small number of retailers to achieve the target collection volume. I assume that the target collection volume at the end of the third year is 400 million pounds. Two types of scenarios are considered in this case study, one where no retailers are initially committed to the collection network and the other where the initial network contains recruited retailers in a distant region (Los Angeles). Each scenario is studied first then with an assumption of linear transportation costs and then an assumption of concave transportation costs. Also, the marketing budget is set at small or large settings for alternative problems.

Scenario 1: Small and Large Marketing Budget with Linear and Concave Costs

For this scenario, I assume that there are no retailers initially committed to the collection network. First I study the problem with an assumption of linear transportation costs, using first a small and then a large marketing budget.

Results for the Linear Transportation Cost Assumption

The solution for these assumptions yields an heuristic policy with the marketing budget allocation concentrated on the Atlanta and New York regions. It rarely allocates the marketing budgets to other regions. This is intuitive because Atlanta is the closest region to the processing site (Augusta, GA). Even though the retailers in the Atlanta region are difficult to recruit, it is more cost effective to concentrate the marketing budget in this region. Since the collection costs in this case are assumed linear, there is no financial incentive to recruit additional retailers to obtain larger amounts of collection volume from a region where it is easy to recruit such as the Los Angeles region. The solution results for a large marketing budget assumption show that the number of recruited retailers in the Atlanta region is larger than when recruiting with a small budget. However, due to volume requirements, there is still a need to recruit the retailers from other regions. The New York region is the next closest and the retailers there are easy to recruit.

Results for the Concave Transportation Cost Assumption

Alternatively, I study the large scale industrial case with an assumption of concave transportation costs. Assuming a large marketing budget, the solution produces an heuristic policy that also concentrates the marketing budget allocation in the Atlanta and New York regions. However, for a small marketing budget, the heuristic policy concentrates the allocation in the New York region more than in the Atlanta region. The explanation for this result is that there is a cost incentive (concave cost) to concentrate retailers in a region where it is easier to recruit the retailers.

Scenario 2: Initial Retailers in Los Angeles with Linear and Concave Costs

For this scenario, I assume that there is an initial network containing recruited retailers in the Los Angeles region. As before, I study this scenario first with a linear transportation cost assumption and then with a concave transportation cost assumption. I also assume that the marketing budget is set to be small so that the policy is more likely to recruit in the Los Angeles region in order to meet the target collection volume.

Results for the Linear Transportation Cost Assumption

The solution for this scenario provides an heuristic policy where the marketing budget is allocated to the Los Angeles region in order to obtain the initial collection volume. The later periods' allocations are spent in the Atlanta and New York regions. There appears to be no cost incentive to recruit more retailers from the Los Angeles region.

Results for the Concave Transportation Cost Assumption

For the situation where concave transportation costs are assumed, the solution produces an heuristic policy with a larger portion of the marketing budget allocated to the Los Angeles region compared to the linear cost assumption scenario. An explanation reason is that under a concave cost function, the higher collection volume drives the unit cost from Los Angeles down. The rest of the marketing budget allocation is allocated mostly in the New York region.

Overall, this case study illustrates that if the collection costs are linear, the heuristic policy most likely spends the allocations on the regions that are closest to the processing site. On the other hand, if the collection costs are concave, there is an incentive to recruit more retailers from region where they are easy to recruit even though the region is located further away from the processing site than the region where retailers are more difficult to recruit. Hence, the cost structure has a big impact on the network growth planning.

In this chapter, I provide a range of numerical studies for the target recruitment problem. If two dimensions are fixed among recruitment budget, region size, and target collection volume, the third dimension can be determined. Furthermore, the STH method is applied on a small example where the optimal value can be obtained. The results show that the heuristic policy obtained from the STH method performs well, especially when the initial state is H. For the large scale problem, the initial state of the retailer has a large impact on the performance of the STH approach but the performance can be improved from the parameter setting in the STH approach. A preprocessing scheme is introduced to improve the computational time. This enables the STH method to solve the problem with as many as 300 retailers. Lastly, the large carpet producer case study is performed. The insights from the various scenarios provide useful information to help a company make the marketing budget allocation decisions.

CHAPTER 8

SUMMARY, CONTRIBUTIONS AND FUTURE DIRECTIONS

8.1 Summary

This dissertation addresses a complex set of decisions that surround the growth of reverse supply chain networks over time. The network growth problem is decomposed into strategic, tactical and operational problems. The strategic problem allocates resources to grow the network over regions and within macro-periods. The use of resources over a finer time discretization, termed periods, and within each region is handled by a tactical or recruitment decision problem. Finally, the network obtained from recruitment process must be serviced by trucks over smaller time increments, or micro-periods. In this dissertation, I formulate a higher fidelity representation of the collection system, with an emphasis on the strategic and operational decisions associated with recruiting and retaining a thriving network of collection entities.

In Chapter 1, the problem is motivated through industrial examples that demonstrate the importance of the problem of growing collection networks. An overview of the reverse production system, hierarchical decision making, and control optimization is also provided. The related literature review in this research is discussed in Chapter 2. In

Chapter 3, the research problem is cultivated and decomposed into three levels: strategic, tactical and operational.

In Chapter 4, I develop a general framework for the recruitment of sources in reverse supply chain distribution. The crucial element is the individual agent state behavior, which follows a Markov process. I pose the recruitment problem as a stochastic dynamic programming problem. The DP algorithm is introduced to solve the tactical problem exactly. However, the realistic problems are so large that the DP algorithm cannot solve the problem in reasonable computational time. As a solution alternative, I develop two heuristics: The Q-Learning based heuristic (QBH) and the Rolling IP with DP heuristic (RIDH). The QBH approach is based on a simulation-based optimization technique to avoid computing the large transition probability matrix. The RIDH method utilizes the benefit of a rolling horizon feature and IP capabilities in order to capture the recruitment decisions over time.

In Chapter 5, the QBH and RIDH methods are tested on many examples. The results illustrate the computational efficiency of the heuristics for different types of problems. For small examples, I have shown that the RIDH method can perform almost as well as DP algorithm. In addition, the RIDH method outperforms QBH in the small and large examples, and is therefore the method of choice for this problem.

In Chapter 6, first I develop a function to estimate the collection cost function. Then I pose the strategic problem as a multi-time scale model. Specifically, I generalize the strategic problem as a resource allocation-collection multi-time scale problem where sequential decision making in each hierarchical level impacts the other levels. The decisions and rewards in the upper (strategic) level and the lower (tactical) level are defined in mathematical terms. The optimality equation can be written down, but solving this problem exactly is difficult even for a small problem. I introduce the strategic trajectory heuristic (STH) to solve large scale problems. This approach employs reference trajectories that explicitly put targets on the increment amounts that are to be collected in each macro-period. This constrains the upper level problem in every macro-period. If this heuristic searches enough reference trajectories and combines those trajectories, a good solution can be obtained. I also show how the information from the tactical level can help the strategic level makes better plans to grow the collection network through solving a target recruitment problem.

In Chapter 7, I provide a numerical study for the target recruitment problem. I have shown empirically that if I fix two dimensions from among recruitment budget, region size, and target collection volume, I can determine the third because the heuristic technique is fast enough to run many case studies. Moreover, I test the STH method on a small example where the optimal value can be obtained. The results show that the heuristic policy obtained from the STH approach can perform well, especially when the retailers' initial states are easy to recruit. For the large scale problem, the initial state has a large impact on the performance of the STH approach. However, increasing the number of considered trajectories and adjusting the discretization on the target collection volume in the STH approach may improve the performance of this approach. Also, I propose a preprocessing scheme to improve the computational time. This enables solution for problem containing as many as 300 retailers. Lastly, the large carpet producer case study is performed in order to emulate what may happen in the real situation. The study provides useful information to help the company makes the budget allocation decisions.

8.2 Contributions

Overall, this dissertation presents a framework to grow a recycling collection network from the processor standpoint through layers of decisions. It is a network structure from many sources to few collectors. Unlike most of earlier research for solid waste management problem that concentrates on the operational and tactical levels, the emphasis here on the growing a collection network is on the strategic and tactical level decisions. In addition, this dissertation offers a new perspective for the objective of the collection network. The growing collection network problem focuses on increasing the collection volume to meet a specified target while minimize the cost. These objectives are crucial to the business continuity of a processor like Shaw Industries in the case of carpet recycling. The high set-up cost of the plant makes it essential to have enough low-cost supply that can generate enough revenue to repay capital expenditures. Developing the strategic model to achieve these significant objectives will assist Shaw Industries in understand the collection system better and to make better decisions and have a sustainable recycling business.

At the tactical level, this dissertation is the first to employ a recruitment concept for supply chain applications. I model the behavior of retailers who have different attitudes towards participating in recycling activities as a Markov process. Using this mechanism, the tactical problem is formulated as a stochastic dynamic programming problem. This dissertation provides an exact solution method for small problems and two heuristics for larger problems. The numerical study demonstrates that the RDMH method can solve large tactical problems quickly with good solution quality. The ability to solve an actual size recruitment problem eventually enables the strategic level to solve realistic problem for the processor. In addition, the recruitment model can be used to apply to many types of recruitment problems such as recruiting supermarket stores for collection of plastic bottles in the plastic recycling industry and recruiting the major electronic stores such as Best Buy and Circuit City for the electronics equipment in the electronic scrap recycling industry.

For the strategic level, I develop a new modeling approach to interlink decisions from different hierarchical levels. In addition, the notion of resource allocation and resource collection are incorporated into the multi-time model. The strategic problem is then generalized as to a resource allocation-collection multi time period model. The model enables us to utilize information sharing to aid the decision planning in each level. The proposed strategic trajectory heuristic offers the ability to obtain a policy that can achieve the objectives of the strategic problem. The numerical study shows that it can obtain a good solution for very large problem sizes with reasonable computation time. The ability to solve realistic problem is significant for a processor because of the number of retailers or consumers that are required to meet the demands of large scale recycling plants.

For example, the carpet recycling industry is starting to reach the point where the “easy” carpet recycling sources have been tapped out in some regions. Understanding how retailer attitudes could change the rate of growth the network, because of the need to invest in significant recruitment efforts, could change the level of effort that the industry devotes to this activity. Failure to invest sufficient resources in this arena could lead to a second Evergreen closure for Shaw. This would probably lead to the demise of large scale recycling efforts for carpet for a significant period of time, unless regulatory authorities step in and implement landfill bans. The implications are thus very significant for an industry that forms a major sector of the manufacturing industry located in Georgia.

8.3 Future Research Directions

Results from this dissertation raise new questions and several potential directions of future research. Future extensions can be envisioned in both the modeling and solution methodology areas.

In collection networks, it is important to retain the recruited retailers in the network to reduce the future cost to recruit additional retailers and increase the probability to achieve the collection target volume. In this research, strategic, tactical, and operational decisions affect the retention of collection network entities in the face of their defection opportunities to other processors or markets. Retention may depend on many factors, such as *system service levels* (e.g., allocation of trucks, logistics, and inventory storage capabilities), *profitability*, and *economies of scales* (e.g., volume that justifies a baler).

Currently I have developed a tactical collection model under the assumption that once the agent is recruited to the network, it always stays in the network. However, in actual situations, sometimes a collection agent may opt to leave the network. Future work includes extending the recruitment model to include retention and defection considerations. An important subtask is to define the criteria that determine the *retention* *and defection* actions of the agent after it is recruited. The additional complexity will impact the capability of the current approach to solve large scale collection recruitment problems.

In this research, it is assumed that the regional recruiter will cooperate with the processor and provide the truthful status information. If this assumption is not true, the regional collector may furnish incorrect information that benefits itself in order to obtain more marketing recruiting budget in the next period. Interesting questions focus on *incentives* for the regional recruiter to provide truthful information and what the processor should do to be certain that the regional recruiter will cooperate.[14] Further research on an approach to offer discounts for the correct status information and to set a penalty fee for incorrect status information may be a possible way to incentivize the regional recruiter corporation.

In the real application, the recruitment process and retailer retention may depend not only on the connection between the retailers and the recruiter, but also on the outside market, a competitor. For example, currently the companies in China are buying a large amount of used carpet from U.S. sources to bring it to China for recycling. Hence, there can be a *competition for the desired source*. Retailers both in and out of the collection network may opt to give the source to competitor collectors who provide a better incentive. This also affects the recruitment allocation plan for the carpet recycler in U.S. Adding a competition feature from the game theory perspective to the recruitment model

[14] Special thanks to my dissertation committee member Dr. Ozlem Ergun for providing this insight.

can complicate the model framework but provides a better understanding of how the entities might act in the real situation.

Furthermore, at the operational level, I assume that the retailer's generation rate is deterministic and the pick-up schedule in the operational level can be solved optimally. Hence, a fixed amount of resource is collected from the recruited agents and delivered to the collection center at the end of the period. Future work includes exploring the collection logistics where the *generation rate of collection material varies* among the agents. With this uncertainty, the problem of routing a fixed number of finite capacity trucks to collect the material from the collection agents is more difficult. The performance of the material collection service can be linked to the retention and defection framework where the collection agent might become "unhappy" if no truck shows up to pick up overflowing collection bins. One service level criteria might be the number of days with overflowing bins. In the next period, the "unhappy" collection agents might choose to defect and leave the collection network. The solution from the operational level must be obtained quickly in order to provide information back to the tactical levels of the hierarchical problem.

By extending the model into the operational level, there is an impact on the multi-time model also. In this dissertation, decisions in each level in the two-level hierarchy are made in two different discrete time-scales. The future work includes extending the model into three-level hierarchy with three different discrete time-scales.

APPENDIX A: NOTATION SUMMARY

Table A1: The Summary of Notations

i	: Index of retailers
m	: Index of regional collector
j	: Index of budget type
k	: Index of target
t	: Index of fast-time scale, period or month
n	: Index of slow-time scale, macro-period or year
η_m	: Number of retailers in region m
M	: Number of regions
T	: Number of periods
N	: Number of macro-periods
s_{it}	: Willingness state the agent/retailer i at period t, (L,M,H,R)
g_i	: Amount of resource that agent i can generate in one period
μ_i	: Recruitment budget threshold of agent i
β_s	: Recruitment willingness factor of willingness state s
$Pr_{s_{it}s_{i,t+1}}(a)$	: Transition probability of moving to state $s_{i,t+1}$ from s_{it} by action a
$B^{\max}$	: Maximum budget over T periods for the tactical problem
Y_t	: Aggregate state in tactical model
w_{it}	: Willingness state of retailer i at time t
B_t^{Start}	: Maximum budget at time t for the tactical problem
V_s	: Value of the willingness state s
$Q(s,a)$	: Q-value of given state action pair (s,a)
b_j	: Amount of budget from budget type j in the recruitment model
$\tilde{v}_{ij}$	: Maximum expected increment of capacity volume that can be collected from retailer i if b_j is allocated to that retailer
SR	: Stochastic recruitment function that returns the collection volume
TR	: Target recruitment function
α	: Learning rate
$r_t^{(co)}$	: Reward of the regional collector at time period t
$D_m^{(Pr)}$	: Distance between the processor and collector in region m
c_D	: Long-haul transportation cost per unit volume and per unit distance
$F_m^{(co)}$	: Fixed collection cost in region m

Table A1: (continued)

$V_m^{(co)}$	: Actual volume collected in region m
$\bar{D}_m$	: Average distance between collector in region m and all retailers in that region
$\bar{c}_D$	: Transportation cost per unit volume and per unit distance within the region
ρ	: Volume parameter
$\theta^{\max}$	: Maximum marketing budget for the strategic problem
θ_n^{Start}	: Maximum marking budget at macro-period n for the strategic problem
z_n	: Aggregate state in strategic model
w_{min}	: Willingness state of retailer i in region m at macro-period n
VN	: Target collection volume at the end of planning horizon N
$c_m^{(total)}$	: Collection Cost Function of region m
ψ	: Penalty cost the deviation between the target collection volume and the actual collection volume per unit volume (Recruitment Problem)
$\bar{\psi}$	: Penalty cost the deviation between the target collection volume and the actual collection volume per unit volume (Strategic Problem)
$\bar{\alpha}$	: Discount factor (monthly)
γ	: Discount factor (yearly)
$\bar{h}_{kmn}$	: Target collection volume from target array, which is the value of k^{th} entry in $\bar{H}_{mn}$, for region m at time period n
$\bar{b}_{jmn}$	: Amount of marketing budget that the processor chooses to spend on the region m at time period n, which is the value of j^{th} entry in $\bar{B}_{mn}$
$\tilde{v}_{mjkn}$	: Expected increment in collection volume from region m by using budget $\bar{b}_{jmn}$ to achieve target collection volume $\bar{h}_{kmn}$ at macro-period n
$c_{mn}^{(total)}(\tilde{v}_{mjkn})$	: Expected increment in total collection cost of region i in macro-period n to collect volume $\tilde{v}_{ijk\tau}$ or $R^{(l,m)}(x^m, \phi^m(x^m, \lambda), \lambda)$
$\bar{V}_n$	: Target collection volume from the trajectory that should be achieved in macro-period n
$\bar{V}_n'$	: Adjusted target collection volume that should be achieved in macro-period n
v_{mn}	: Actual collection volume of region m at the start of macro-period n
vt_n	: The total actual collection volume for all regions at the start of macro-period n
ε	: Allowable deviation between expected collection volume and the target collection volume.
ξ	: Revenue in strategic model

APPENDIX B: DATA

Table B1: Data for 10 retailers

Retailer	Collection Volume (lb./month)	Recruitment Budget Threshold	Initial Willingness State
1	10	5	L
2	30	15	L
3	20	10	L
4	30	15	L
5	80	64	L
6	60	36	L
7	40	20	L
8	20	10	L
9	20	10	L
10	30	15	L

Table B2: Data for 20 retailers

Retailer	Collection Volume (lb./month)	Recruitment Budget Threshold	Initial Willingness State
1	10	5	L
2	30	15	L
3	20	10	L
4	30	15	L
5	80	64	L
6	60	36	L
7	40	20	L
8	20	10	L
9	20	10	L
10	30	15	L
11	40	20	M
12	40	20	M
13	80	64	M
14	30	15	M
15	30	15	M
16	40	20	M
17	70	49	H
18	30	15	H
19	60	36	H
20	20	10	H

Table B3: Data for 30 retailers

Retailer	Collection Volume (lb./month)	Recruitment Budget Threshold	Initial Willingness State
1	10	5	L
2	30	15	L
3	20	10	L
4	30	15	L
5	80	64	L
6	60	36	L
7	40	20	L
8	20	10	L
9	20	10	L
10	30	15	L
11	40	20	L
12	40	20	L
13	80	64	L
14	30	15	L
15	30	15	L
16	40	20	M
17	70	49	M
18	30	15	M
19	60	36	M
20	20	10	M
21	20	10	M
22	30	15	M
23	30	15	M
24	60	36	M
25	30	15	H
26	30	15	H
27	30	15	H
28	40	20	H
29	20	10	H
30	30	15	H

Table B4: Data for Case 7.5

Retailer	Region	Collection Volume (lb./month)	Distance	Recruitment Budget Threshold	Initial Willingness State
1	A	60	20	36	H
2	A	40	20	20	H
3	A	90	20	81	H
4	A	60	20	36	H
5	A	50	20	25	H
6	A	70	20	49	H
7	A	70	20	49	H
8	A	50	20	25	H
9	A	40	20	20	H
10	A	80	20	64	H
1	B	50	10	25	M
2	B	50	10	25	M
3	B	50	10	25	M
4	B	40	10	20	M
5	B	50	10	25	M
6	B	60	10	36	M
7	B	50	10	25	M
8	B	70	10	49	M
9	B	60	10	36	M
10	B	40	10	20	M
1	C	10	1	5	L
2	C	20	1	10	L
3	C	10	1	5	L
4	C	20	1	10	L
5	C	20	1	10	L

Table B5: Data for Case 7.6

Retailer	Region	Collection Volume (lb./month)	Distance	Recruitment Budget Threshold	Initial Willingness State
1	A	40	15	20	3
2	A	60	15	36	1
3	A	70	15	49	2
4	A	50	15	25	2
5	A	60	15	36	2
6	A	50	15	25	1
7	A	40	15	20	3
8	A	50	15	25	2
9	A	50	15	25	2
10	A	50	15	25	3
11	A	80	15	64	3
12	A	40	15	20	2
13	A	50	15	25	3
14	A	70	15	49	1
15	A	70	15	49	3
16	A	50	15	25	3
17	A	60	15	36	1
18	A	90	15	81	2
19	A	40	15	20	3
20	A	60	15	36	1
1	B	50	20	25	3
2	B	50	20	25	1
3	B	80	20	64	1
4	B	30	20	15	1
5	B	30	20	15	3
6	B	30	20	15	2
7	B	60	20	36	3
8	B	100	20	100	2
9	B	60	20	64	1
10	B	40	20	20	3
11	B	40	20	20	3
12	B	70	20	49	2
13	B	60	20	36	1
14	B	60	20	36	2
15	B	50	20	25	3
16	B	80	20	64	2
17	B	50	20	25	2
18	B	60	20	36	3
19	B	70	20	49	3
20	B	40	20	20	2
1	C	50	25	25	1
2	C	30	25	15	1
3	C	50	25	25	2
4	C	40	25	20	3

Table B5: (continued)

Retailer	Region	Collection Volume (lb./month)	Distance	Recruitment Budget Threshold	Initial Willingness State
5	C	60	25	36	3
6	C	30	25	15	3
7	C	70	25	49	3
8	C	50	25	20	1
9	C	40	25	16	1
10	C	40	25	16	3

REFERENCES

Anderson, L.E. (1968), “A Mathematical Model for the Optimization of Waste Management System,” University of California at Berkeley, Sanitary Engineering Research Laboratory, SERL Report, No. 68-1, USA.

Assavapokee, T. (2004), Semi-continuous Robust Approach for Strategic Infrastructure Planning of Reverse Production Systems, Ph.D. Dissertation, Georgia Institute of technology, Georgia, USA.

Assavapokee, T., Ammons, J., Realff, M., Hong, I.H. (2005), “A Scenario Relaxation Algorithm for Finite Scenario Based Min-Max Regret and Min-Max Relative Regret Robust Optimization,” Under Review by European Journal of Operational Research, 2005.

Atlanta Business Chronicle (2001), “Augusta Evergreen Nylon Recycling Plant Shuts Down,” http://www.bizjournals.com/atlanta/stories/2001/08/27/daily34.html, viewed 03/16/2006.

Baetz, B.W. (1990), “Optimization/Simulation Modeling for Waste Management Capacity Planning,” *Journal of Urban Planning and Development, ASCE*, 116(2), 59-79.

Barros, A.I., Dekker, R., Scholoten, V. (1998), “A Two-level Network for Recycling Sand: A Case Study,” *European Journal of Operational Research*, 110, 199-214.

Barto, A.G., Bradtke, S.J., Singh, S.P. (1995), “Learning to Act Using Real-Time Dynamic Programming,” *Artificial Intelligence*, 72.

Bass, F. (1969), “A New Product Growth Model for Consumer Durables,” *Management Science*, 15(1), 215-227

Batta, R., Chiu, S.S. (1988), “Optimal Obnoxious Paths on a Network: Transportation of Hazardous Materials,” *Operations Research*, 36, 84-92.

Bellman, R.E. (1957), *Dynamic Programming,* Princeton University Press, Princeton, New Jersey, USA.

Berman O., Drezner, Z., Wesolowsky, G.O. (2001), “The Collection Depots Location Problem on Networks,” *Naval Research Logistics*, 49.

Berman O., Huang, R. (2004), “Minisum Collection Depots Location Problem with Multiple Facilities on a Network,” *Journal of Operational Research Society*, 55, 769-779.

Bertsekas, D.P. (1995), *Dynamic Programming and Optimal Control*, Athena Scientific, Belmont, Massachusetts, USA.

Bertsekas, D.P., Tsisiklis, J. (1987), *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, New Jersey, USA.

Bertsekas, D.P., Tsisiklis, J. (1995), *Neuro-Dynamic Programming*, Athena Scientific, Belmont, Massachusetts, USA.

Beihl, M., Prater, E., Realff, M.J. (2005), “Assessing Performance and Uncertainty in Developing Carpet Reverse Logistics Systems,” *Computers & Operations Research*, In Press, Corrected Proof, Available online 18 April 2005.

Bhatnagar, S., Panigrahi, J. R. (2004), “Hierarchical Decision Making in Semiconductor Fabs Using Multi-timescale Markov Decision Processes,” *In Proceedings of IEEE Conference on Decision and Control*, Paradise Island, Nassau, Bahamas.

Bhatnagar, S., Panigrahi, J. R. (2006), “Actor-critic Algorithms for Hierarchical Markov Decision Processes,” *Automatica*, 42, 637-644.

Birge, J.R., Louveaux, F.V. (1997), *Introduction to Stochastic Programming*, Springer Verlag, New York, USA

Bitran, G., Hass, E.A., Matsuo, K. (1986), “Production Planning of Style Goods with High Set-up Costs and Forecast Revisions,” *Operations Research*, 34, 226-236.

Borkar, V.S. (2005), “An Actor-critic Algorithm for Constrained Markov Decision Processes,” *Systems and Control Letters*, 54, 207-213.

Buwalda, T. (2000), “What’s Under the Rug,” Resource Recycling, January.

Cachon, G., Fisher. (2000), “Supply Chain Inventory Management and the Value of Shared Information,” *Management Science*, 46, 1032–1048.

Canada's Waste Recycling Marketplace (2006), http://www.recyclexchange.com, viewed 03/16/2006.

Carpet America Recovery Effort (2004), “Carpet America Recovery Effort’s Annual Report in 2004,” http://www.carpetrecovery.org/annual_report/04_CARE-annual-rpt.pdf, viewed 03/16/2006.

Carpet and Rug Institute (2003), "The Carpet Industry's Sustainability Report 2003," http://www.carpet-rug.org/pdf_word_docs/03_CRI-Sustainability-Report.pdf, viewed 03/01/2005.

Caruso, C., Colorni, A., Paruccini, M. (1993), "The Regional Urban Solid Waste Management System: A Modeling Approach," *European Journal of Operational Research*, 70, 16-30.

Chang, H.S., Fard, P., Marcus, S., Shayman, M. (2003), "Multi-time Scale Markov Decision Processes," *IEEE Transactions on Automatic Control*, 48, 976-987.

Chang, N.B., Schuler, R.E., Shoemaker, C.A. (1993), "Environmental and Economic Optimization of an Integrated Solid Waste Management System," *Journal of Resource Management Technology*, 21(2), 87-100.

Chang, N.B., Wang, S.F. (1995), "A Fuzzy Goal Programming Approach for the Optimal Planning of Metropolitan Solid Waste Management Systems," *European Journal of Operational Research*, 99, 303-321.

Chang, N.B., Wang, S.F. (1996), "Managerial Fuzzy Optimal Planning for Solid Waste Management Systems," *Journal of Environmental Engineering ASCE*, 122(7), 756-781.

Chang, N.B., Wei, Y.L. (1999), "Strategic Planning of Recycling Drop-off Stations and Collection Network by Multiobjective Programming," *Environmental Management*, 24(2), 247-263.

Choi, J., Realff, M., Lee, L.H. (2004), "Dynamic Programming in a Heuristically Confined State Space: A Stochastic Resource-constrained Project Scheduling Application," *Computers & Chemical Engineering*, 28(6-7), 1039-1058.

Choi, J., Realff, M., Lee, L.H. (2005), "A Q-Learning Method Applied to Resource Constrained Project Scheduling with New Project Arrivals," Working Paper, School of Chemical and Biomolecular Engineering, Georgia Institute of Technology.

Container Recycling Institute (2005), "Container and Packaging Recycling Update: Winter 2005", http://www.container-recycling.org/assets/pdfs/newsletters/CRI-NL-2005Winter.pdf

Coughlan, A.T., Grayson, K. (1998), "Network Marketing Organizations: Compensation Plans, Retail Network Growth, and Profitability," *International Journal of Research in Marketing*, 15, 401-426.

Darmon, R. Y. (2003), "Controlling Sales Force Turnover Costs though Optimal Recruiting and Training Policies," *European Journal of Operation Research*, 154, 291-303.

Del Castillo, E., Cochran, J.K. (1996), "Optimal Short Horizon Distribution Operations in Reusable Container Systems," *Journal of the Operational Research Society*, 47, 48-60.

Economou, A. (2005), "Generalized Product-form Stationary Distributions for Markov Chains in Random Environments with Queuing Applications," *Advanced in Applied Probability*, 185-221

Erkut, E. (1996), "The Road not Taken," *OR/MS Today*, Dec, 22-28.

Feichtinger, G., Mehlmann, A. (1976), "The Recruitment Trajectory Corresponding to Particular Stock Sequences in Markovian Person-flow Models," *Mathematics of Operations Research*, 1, 175-184.

Ferdergruen, A., Prastocos, G., Zipkin, P. (1986), "An Allocation and Distribution Model for Perishable Products," *Operations Research*, 34(1), 75-82.

Fiala, P. (2005), "Information Sharing in Supply Chains," *Omega*, 33(5), 419-423.

Flapper, S.D.P. (1995), "On the Operational Aspects of Reuse," *Proceedings of the 2nd International Symposium on Logistics*, Nottingham, UK, 109-118.

Flapper, S.D.P. (1996), "Logistic Aspects of Reuse: An Overview," *Proceedings of the 1st International Working Seminar on Reuse*, Eindhoven, The Netherlands, 109-118.

Fleischmann, M., Bloemhof-Ruwaard, J.M., Dekker, R., Lann, E.V.D., Nunen, J.A.E.E.V., Van Wassenhove, L. (1997), "Quantitative Models for Reverse Logistics: A Review:," *European Journal of Operational Research*, 1997, 1-17.

Fleischmann, M., Krikke, H.R., Dekker, R., Flapper S.D.P. (2000), "A Characterization of Logistics Networks for Product Recovery," *Omega*, 28, 653-666.

Florida Department of Environmental Protection (2000), "Carpet Recycling Project in Sarasota County," http://www.dep.state.fl.us/waste/quick_topics/publications/shw/recycling/InnovativeGrants/IGyear1/finalreports/sarasota.pdf, viewed 03/16/2006.

Fox, E., Heath, L., Chen, Q., Daoud, A. (1992), "Practical Minimal Perfect Hash Functions for Large Databases," *Communications of the ACM*, 35,105-121.

Fu, M.C. (2001), "Simulation Optimization," *Proceeding of the 2001 Winter Simulation Conference*.

Fu, M.C. (2002), "Optimization for Simulation: Theory vs. Practice," *Journal on computing*, 14 (3), 192-215.

Fuellhart, K. (1999), “Localization and the Use of Information Sources: the Case of the Carpet Industry,” *European Urban & Regional Studies*, 6(1), 39-59.

Georgiou, A.C., Tsantas, N. (2002), “Modeling Recruiting Training in Mathematical Human Resource Planning,” A*pplied Stochastic Models in Business and Industry*, 18, 53-74.

Ginter, P.M., Starling, J.M. (1978), “Reverse Distribution Channels for Recycling,” *California Management Review*, 20(3), 73-82.

Glover, F. (1976), “Chance-constrained Techniques for Integer Programming,” in: M.Dempster (ed.), *Stochastic Programming: Proceedings of the 1974 Oxford International Conference*, Academic Press, New York, USA.

Golbasi, H. (2001), Supply Chain Coordination: Optimization and Game-theoretic Models, Ph.D. Dissertation, Lehigh University, Pennsylvania, USA.

Gosavi, A. (2003), *Simulation-based Optimization: Parametric Optimization Techniques and Reinforcement Learning*, Kluwer Academic Publishers.

Guide V.D.R., Van Wasssenhove L.N. (2001), “Managing product returns for remanufacturing”, *Production and Operations Management*, 10(2), 142-155.

Guiltinan, J.P., Nwokoye, N.G. (1975), “Developing Distribution Channels and Systems in the Emerging Recycling Industries,” *International al Journal of Physical Distribution*, 6(1), 28-38.

Hauskrecht, M., Menleau, N., Kaelbling, L.P., Dean, T., Boutilier, C. (1998), “Hierarchical solution of Markov decision processes using macro-actions,” *Uncertainty in Artificial Intelligence Proceedings of the Fourteenth Conference*, 220-229.

Hawkins, D.A. (1992), “An Activation-recruitment Scheme for Use in Muscle Modeling,” *Journal of Biomechanics*, 25, 1467-1476.

Hernandez-Lerman, O. (1989), *Adaptive Markov Control Processes*, Springer-Verlag, New York, USA..

Hinderer, K. (1970), *Foundation of Non-Stationary Dynamic Programming with Discrete Time Parameter*, Springer-Verlag, New York, USA.

Holland, J.H. (1975), *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Michigan, USA.

Howard, R. (1960), *Dynamic Programming and Markov Processes*, MIT Press, Cambridge, Massachusetts, USA.

Hsieh, H.N, Ho, K.H. (1993), "Optimization of Solid Waste Disposal System by Linear Programming Technique," *Journal of Resource Management Technology*, 21(4), 194-201.

Huang, G.H., Baetz, B.W., Patry, G.G. (1992), "A Grey Linear Programming Approach for Municipal Solid Waste Management Planning under Uncertainty," *Civil Engineering Systems*, 9, 319-335.

Huang, G.H., Baetz, B.W., Patry, G.G. (1993), "A Grey Fuzzy Programming Approach for Municipal Solid Waste Management Planning under Uncertainty," *Civil Engineering Systems*, 10, 123-146.

Huang, G.H., Baetz, B.W., Patry, G.G. (1994, "Grey Dynamic Programming Approach for Municipal Solid Waste Management Planning under Uncertainty," *Civil Engineering Systems*, 11, 43-73.

Huang, G.H., Baetz, B.W., Patry, G.G. (1995), "A Grey Integer Programming for Solid Waste Management Planning under Uncertainty," *European Journal of Operations Research*, 83, 594-620.

IDES The Plastics Web (2006), http://www.ides.com/resinprice/resinpricingreport.asp, viewed 03/16/2006.

International Fiber Journal (1999), "AlliedSignal, DSM, Say They've Cracked Carpet Recycling Code," http://fiberjournal.com/issue/June99/allied.html, viewed 03/16/2006.

Jayaraman, V. (1996), "A Reverse Logistics and Supply Chain Management Model within a Remanufacturing Environment," INFORMS, Atlanta, USA

Kaelbling, L.P., Littman, M.L., Moore, A.W. (1996), "Reinforcement Learning: A Survey," *Journal of Artificial Intelligence Research*, 4, 237-285.

Karlin, S. (1955), "The Structure of Dynamic Programming Models," *Naval Research Logistics Quarterly*, 2, 285-294.

Knuth, D. (1973), *The Art of Computer Programming*, *Volume 3: Sorting and Searching*, Addison Wesley.

Koo, H.J., Shin, H.S., Yoo, H.C. (1991), "Multiobjective Sitting Planning for a Regional Hazardous Waste Treatment Center," *Waste Management Research*, 9, 205-218.

Krikee, H.R. (1998), Recovery Strategies and Reverse Logistic Network Design, Ph.D. Dissertation, University of Twente, Enchede, The Netherlands.

Kroon, L., Vrijens, G. (1995), “Returnable Containers; An Example of Reverse Logistics,” *International Journal of Physical Distribution and Logistics Management*, 25(2), 56-68.

Kumar, P.R. (1985), “A Survey of Some Results in Stochastic Adaptive Control,” *SIAM Journal of Control and Optimization*, 23, 329-380.

Kumar, P.R., Variya, P. (1986), *Stochastic Systems: Estimation, Identification, and Adaptive Control*, Prentice Hall, Englewood Cliffs, NJ, USA.

Kunsch, P. (1990), “Stochastic Programming: Numerical Solution Techniques by Semi-Stochastic Approximation Methods,” in R. Slowinski, and J. Teghem (eds.),” *Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty*, Kluwer Academic Publishers, Dordrecht, 23-44.

Law, A.M., M.G. McComas (2002), “Simulation-based Optimization,” *Proceeding of the 2002 Winter Simulation Conference*.

Lee, H., Padmanabhan, P., Whang, S. (1997), “The Bullwhip Effect in Supply Chains,” *Sloan management Review*, 38, 93-102.

Limpert, E., Stahel, W., Abbt M. (2001), "Log-normal Distributions across the Sciences: Keys and Clues," *BioScience*, 51, 341–352.

Louwers, D., Bert, J.K., Peters, E., Souren F., Flapper, S.D.P. (1999), “A Facility Location Allocation Model for Reusing Carpet Materials”, *Computer and Industrial Engineering*, 36, 855-869.

Lund, J.R., Tchobanoglous, G. (1994), “Linear Programming for Analysis of Material Recovery Facilities,” *Journal of Environmental Engineering, ASCE*, 120(5), 1095-1093.

Marks, D.H., Revelle, C.S., Liebman, J.C. (1970), “Mathematical Models of Location: A Review,” *Journal of Urban Planning Model Development, ACSE*, 81-93.

McDonnel, P., Joshi, S., Qiu, R.G. (2004), “Set-up Games: A Heuristic Game-Theoretic Approach to Set-up Decisions for Heterarchical Manufacturing Systems, *International Journal of Production Research*, 42(19), 4195-4210.

Mehlmann, A. (1980), “An approach to optimal recruitment and transition strategies for manpower systems using dynamic programming,” *Journal of the Operational Research Society*, 31, 1009-1015.

Mihut C., Captain, Captain D.K., Gadala-Maria, F., Amiridis, M.D. (2001), “Review: Recycling of Nylon from Carpet Waste,” *Polymer Engineering and Science*, 2001, 41(9), 1457-1470.

Mirchandani, P.B., Francis, R.L. (1989), *Discrete Location Theory*, Wiley, New York, USA.

MSNBC U.S. News Environment (2005), “Plastic Bottles Pile Up as Mountains of Waste”, http://www.msnbc.msn.com/id/5279230, viewed 03/16/2006.

Newton, D.J. (2000), A Robust Approach for Planning the Strategic Infrastructure of Reverse Production System, Ph.D. Dissertation, Georgia Institute of technology, USA.

North Carolina Division of Pollution Prevention and Environmental Assistance (1998), “Textiles: Carpet & Carpet Pad Commodity Profile”, http://www.resourcesaver.org/file/toolmanager/O16F29842.pdf, viewed 03/16/2006.

Pham, D.T., Karaboga, D. (1998), *Intelligent Optimization Techniques: Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks*, Springer-Verlag, New York, USA.

Plastics Technology (2002), “Big German Plant May Relieve U.S. Bottleneck in Recycling Carpet Nylon,” http://plasticstechnology.com/articles/2005cu2.html, viewed 03/16/2006.

Pochampally K.K., Gupta, S.M (2003), “A Multi-Phase Mathematical Programming Approach to Strategic Planning of an Efficient Reverse Supply Chain Network,” *Proceedings of the 2003 IEEE International Symposium on Electronics and the Environment*, 72-78.

Powell, W.B., George, A.P, Kulkarni, S.R. (2004), “Value Function Approximation using Hierarchical Aggregation for Multiattribute Resource Management,” working paper.

Puterman, M.L. (1994), *Markov Decision Processes*, Wiley Interscience, New York, USA.

Qiu, R., Joshi, S. (1999), “A Structured Adaptive Supervisory Control Methodology for Modeling the Control of A Discrete Event Manufacturing System,” *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 29, 573-586.

Realff, M.J., Ammons, J.C., Newton, D. (1999), “Carpet Recycling: Determining the Reverse Production System Design,” The journal of Polymer-Plastics Technology and Engineering, 38(3), 547-567.

Realff, M.J., Ammons, J.C., Newton, D. (2004), “Robust Reverse Production System Design for Carpet Recycling,” *IIE Transactions*, 36, 767-776.

Rinott, Y. (1978), “On Two-stage Selection Procedures and Related Probability-Inequalities,” *Communications in Statistics: Theory and Methods,* A7, 799-811.

Rockafellar, R.T., Wets, R.J.B. (1991), “Scenarios and Policy Aggregation in Optimization under Uncertainty”, *Mathematics of Operations Research*, 16, 119-147.

Ross, S. (1983), *Introduction to Stochastic Dynamic Programming*, Academic Press, New York, USA.

Rummery, G.A., Niranjan, M. (1994), “On-line Q-learning Using Connectionist Systems,” *Technical Report CUED/F-INFENG/TR 166*, Engineering Department, Cambridge University, England.

Schneeweiss C., Zimmer K. (2004), “Hierarchical coordination mechanisms within the supply chain,” *European Journal of Operational Research*, 153(3), 687-703.

Seggern, V.D. (1993), *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 124, 1993.

Sethi, S.P., Taksar, M., Zhang, Q. (1995), “Hierarchical Capacity Expansion and Production Planning Decisions in Stochastic Manufacturing Systems,” *Journal of Operations Management*, 12, 331-352.

Sethi, S.P., Thompson, G.L. (2000), *Optimal Control Theory: Applications to Management Science and Economics, Second Edition,* Kluwer Academic Publishers, Boston, USA.

Sethi, S.P., Zhang, Q. (1994), *Hierarchical Decision Making in Stochastic Manufacturing Systems,* in series System and Control: Foundations and Applications, Birkhauster, Boston, Cambridge, MA.

Shapiro, Alexander. (1996), “Simulation Based Optimization,” *Proceedings of the 1996 Winter Simulation Conference*.

Shaw Floors: Honeywell (2005), “Shaw Industries Agrees to Purchase U.S. Nylon Carpet Fibers Business from Honeywell”, http://www.shawfloors.com/News/PressRelease/Honeywell.asp, viewed 03/16/2006.

Solomon M., Bloemhof-Ruwarrd, J.M., Van Wassenhove, L. (1994), “On the coordination of Product and By-Product Flows in Two-Level Distribution Networks: Model Formulations and Solution Procedures,” *European Journal of Operational Research*, 79, 325-339.

Solomon M., Bloemhof-Ruwarrd, J.M., Van Wassenhove, L. (1996), “The Capacitated Distribution and Waste Disposal Problem,” *European Journal of Operational Research*, 88, 490-503.

Spengler, T., Puchert H., Penkuhn T., Rentz O. (1997), "Environmental Integrated Production and Recycling Management," *European Journal of Operational Research*, 97, 547-567.

Sutton, R. (1992), *Reinforcement Learning Machine Learning (Special Issue),* 8(3).

Thomas, N. (2005), "Product Form over Components in Markovian Process Algebra," *UKSIM 2003. Sixth National Conference of the United Kingdom Simulation Society*, 25-31

Tishler, A., Lipovetsky S. (2000), "A Globally Concave, Monotone and Flexible Cost Function: Derivation and Application," *Applied Stochastic Models in Business and Industry*, 16, 279-296.

Treven, S. (2006), "Human Resources Management in the Global Environment," *Journal of American Academy of Business*, 8, 120-125.

Tuan T., Park, K. (1999), "Multiple Time Scale Congestion Control for Self-similar Network Traffic," *Performance Evaluation*, 36-37, 359-386.

U.S. Carpet Industry (2002), Memorandum of Understanding for Carpet Stewardship.

U.S. Census Bureau (2003), "Statistic of U.S. Businesses: 2003: Carpet & Upholstery Cleaning Services United States," http://www.census.gov/epcd/susb/2003/us/US56174.HTM, viewed 03/16/2006.

Vondrak, J., Dean, B., Goemans, M. (2004), "Approximating the Stochastic Knapsack Problem: The Benefit of Adaptivity," Symposium on Foundations of Computer Science, 208-217

Watkins, C.J. (1989), "Learning from Delayed Rewards," Ph.D. Dissertation, Kings College, Cambridge, England, May 1989.

White, C.C., III, Eldeib, H.K. (1994), "Markov Decision Processes with Imprecise Transition Probabilities," Operation Research, 42(4), 739-749.

White, C.C., III, Eldeib, H.K. (1986), "Parameter Imprecision in Finite State, Finite Action Dynamic Programs," Operation Research, 34(1), 120-128.

Yudin, D.B., Tsoy, E. (1974), "Integer Valued Stochastic Programming," *Engineering Cybernetics*, 12, 1-8.

VITA

Wuthichai Wongthatsanekorn was born in Bangkok, Thailand, on April 12, 1978. He received his undergraduate degree in Industrial Engineering (with minor area in economics and computer science), from Columbia University, NY, in 2000. He then pursued his master of science in Operations Research from University of Michigan Ann Arbor, MI, and graduated in the summer of 2001. In August 2001, he was enlightened and decided to join Georgia Institute of Technology for continuing his Ph.D. program in Industrial and Systems Engineering. He received his second M.S. degree in Industrial Engineering from Georgia Institute of Technology in 2003. His primary area of interest is the modeling and analysis of practical supply chain and logistics problems. In May 2006, after spending almost 11 years in the U.S., he received his Ph.D. in Industrial and Systems Engineering from Georgia Institute of Technology. His career plans are to seek an academic job in his hometown, Thailand. Apart from his academic interests, he likes to play soccer, tennis, basketball, and squash.