Development of a method to identify problems of a bilevel nature within different engineering applications

Ahmad Khairyanto

2008

Ahmad, K. (2008). Development of a method to identify problems of a bilevel nature within different engineering applications. Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/13507>

<https://doi.org/10.32657/10356/13507>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**DEVELOPMENT OF A METHOD TO IDENTIFY
PROBLEMS OF A BILEVEL NATURE WITHIN DIFFERENT
ENGINEERING APPLICATIONS**

AHMAD KHAIRYANTO

**SCHOOL OF MECHANICAL AND AEROSPACE
ENGINEERING**

2008

**DEVELOPMENT OF A METHOD TO IDENTIFY
PROBLEMS OF A BILEVEL NATURE WITHIN DIFFERENT
ENGINEERING APPLICATIONS**

AHMAD KHAIRYANTO

School of Mechanical and Aerospace Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Master of Engineering

2008

Abstract

Problems of bilevel nature are inextricably linked to decentralized problems involving multiple decision-making typically with multiple objectives, as are those characterized by Nash equilibrium and variations of the Pareto optimization. These problems are to be found within a wide array of seemingly disparate engineering topics ranging from multi-disciplinary complex system design to multi-echelon decentralized inventory planning. Due to the potential for different forms of decision-making (i.e. cooperative, non-cooperative or sequential) to manifest within problems of decentralized settings, there exists regardless of the field of interest a need to determine the exact nature of such problems with respect to the type of decision-making present. This thesis enquires into the issues associated with identifying the form of decision-making that is in line with the bilevel nature and subsequently describes a method to potentially identify such problems within the decentralized setting.

To formally define the identification process of a bilevel nature unfettered to any particular field of interest, a general approach based on complex systems was adopted. Four topics in engineering where bilevel problems have been clearly established, as evident from literature review, namely: i) the multi-disciplinary complex system design, ii) the transport network design, iii) the decentralized capacitated plant selection and, iv) the multi-echelon decentralized inventory planning problems were scrutinized to form the bulk of the area of study. In order to consolidate different events of bilevel problems across the different areas of interest, characteristics defined within the framework of complex systems that are unique to the decision-making process (i.e. Pareto, Nash and Stackelberg/bilevel) were identified. These then form the basis for the method developed to identify the bilevel problem.

The method developed within this thesis identifies bilevel nature by categorizing it under two common headings: i) the pure decentralized system and ii) the hierarchical control complex system. These two headings classify the different problems within a decentralized setting in an exhaustive manner. The method employs an elimination approach in which other variants of the multiple objectives, multiple decision-making problem are ruled out through identifying characteristics inherent to the bilevel nature within a step-by-step manner. The method developed was then evaluated through its

application in a series of case studies based on a set of established problems within different areas of interest in engineering.

Acknowledgments

Verily, all praise is for Allah, we praise Him and we seek His assistance and we ask for His forgiveness. And we seek refuge in Allah from the evils of our selves and from the evils of our actions. Whoever Allah guides, there is no one that can lead him astray, and whoever is led astray there is no guide for him. I bear witness that there is no deity that has the right to be worshipped except Allah and I bear witness that Muhammad is His slave and messenger.

The author would also like to express his sincere gratitude to Associate Professor Sim Siang Kok and Associate Professor Tay Meng Leong for their extreme patience and trust, kind guidance and pupilage throughout the course of the first year research work all of which have been deeply appreciated.

The author would also like thank those who have in one way or another contributed to the completion of the thesis.

Content Page

	<u>Page</u>
ABSTRACT	i
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	vii
LIST OF TABLES	viii
CHAPTER 1	
INTRODUCTION	1
1.1 Problems within a decentralized decision-making setting and its variants	2
1.1.1 The different variants of a problem within the decentralized decision-making setting	4
1.2 The need to correctly identify bilevel problems within the decentralized setting	7
1.3 Defining the Bilevel Problem within the decentralized setting	9
1.3.1 Purely Decentralized System	10
1.3.2 Complex System with Hierarchical Control	10
1.3.3 Summary of Section	12
1.4 Objectives	13
1.5 Organization	14
CHAPTER 2	
LITERATURE REVIEW	16
2.1 Categorization of present works evaluated to distinguish the different types of decision-makings	17

within a single decentralized setting	
2.1.1 Restricted decision-making	18
2.1.2 Cooperative decision-making	18
2.1.3 Isolated decision-making	19
2.1.4 Sequential decision-making	20
2.2 List reviewed	20
2.2.1 Collaborative multidisciplinary design	20
2.2.2 Multi-echelon inventory policy problem	26
2.2.3 Decentralized capacitated plant selection problem	30
2.2.4 Transport network design	32
2.3 Chapter Summarization	37
 CHAPTER 3	
CHARACTERIZING BILEVEL PROBLEMS AND RESOLVING BILEVEL PROGRAMS	40
3.1 Identifying Problems of a Bilevel Nature	41
3.1.1 Identifying a complex system	43
3.1.2 Ascertain Hierarchical or Pure Decentralized Control	46
3.1.3 Ascertain Bilevel Program	48
3.1.4 Summary of Section	50
3.2 Definition and Mathematical Formulation of Bilevel Programming Problem	50
3.2.1 Comments on the Solvability of Bilevel Programming Problem	53
3.2.2 Theoretical Solution Methodologies	57
3.2.3 Section Summarization	66
3.3 Chapter Summarization	67
 CHAPTER 4	
SOLUTION ALGORITHM	68
4.1 Overall Framework	69

4.2 Description of Approach and Algorithm to Resolve the Bilevel Problem	70
4.3 Chapter Summary	72
CHAPTER 5	
CASE STUDY	73
5.1 Description of the multi-echelon inventory problem	73
5.1.1 Evidence of Bilevel Problem in the Multi- Echelon Inventory Problem	76
5.1.2 Results and Discussions	79
5.2 Description of the Decentralized Capacitated Plant Selection Problem	88
5.2.1 Evidence of Bilevel Problem in the Decentralized Capacitated Plant Selection Problem	90
5.2.2 Results and Discussions	91
5.3 Chapter Summary	95
CHAPTER 6	
CONCLUSION AND FUTURE WORK	97
6.1 Conclusion	97
6.2 Further Work	100
6.2.1 Possible further testing for the approach developed	100
6.2.2 Development of a comprehensive approach or methodology to identify and resolve a problem of a bilevel nature	101
6.3 Concluding Remarks	102
REFERENCES	103
APPENDIX	A1

List of Figures

List of Figures	Pages
Fig 1.1 Graph of optimal design solutions for design of a simple pressure vessel corresponding to Stackelberg, Pareto and Nash games	8
Fig 1.2 Variations in design of simple pressure vessel corresponding to different games	8
Fig 2.1 Flow of information within the design process	25
Fig 2.2 A representation of the over-the-wall design process	25
Fig 2.3 Bilevel formulation of the decentralized capacitated plant selection problem	31
Fig 2.4 Classification of problems within the decentralized decision-making setting	39
Fig 3.1 Illustration of thin-walled pressure vessel	44
Fig 3.2 Subsystems sharing the same resource pool	46
Fig 3.3 Flowchart of Three-Step Approach to Detect Bilevel Problems	49
Fig 3.4 Plot of values of x_i against F	55
Fig 3.5 A polytope with its vertices	60
Fig 4.1 Overall System Architecture	69
Fig 4.2 Flowchart of Overall Approach	71
Fig 5.1 The arborescent system	74

List of Tables

List of Tables	Pages
Table 1.1 List of Bilevel Problems	7
Table 2.1 Summary of review findings	37
Table 5.1 Bilevel and Global Optimal Results with $0 \leq \varphi \leq 1$ and $h_0=1, h_i=1$	83
Table 5.2 Bilevel and Global Optimal Results with $0 \leq \varphi \leq 1$ and $h_0=1, h_i=2$	84
Table 5.3 Auxiliary plant data	92
Table 5.4 Product demand	92
Table 5.5 Production capacity consumption a_{ij}	92
Table 5.6 Unit transportation cost R_{ij}	93
Table 5.7 Optimal production plan for each combination of feasible plant selection	93

Chapter 1

Introduction

Many industrial engineering activities involve unstructured, real-life problems that are hard to model and complex in nature [1]. The increasing complexity and sheer size of today's engineering problems has largely driven the need for problems to be resolved within a collaborative manner. This requires harnessing the expertise of multiple engineering teams each working within a specific segment of the whole problem in order to obtain a complete and workable resolution. It is from within such a setting of decentralized decision-making that problems of a bilevel nature typically arise. These are characterized by an inherent hierarchy among the constituent decision-makers where one decision-maker dominates and asserts its decisions over another (or others). The bilevel nature here relates to the innate sequential form of decision-making which occurs as a consequence of the hierarchical state, as well as the eponymous mathematical formulation derived to model such a problem i.e. bi/multilevel optimization or programming [2, 3]. In the preface to the seminal 'Practical Bilevel Optimization Algorithms and Application' [4], Bard described bilevel optimization as addressing:

"...problem in which two decision makers, each with their individual objectives, act and react in a noncooperative, sequential manner."

Yet, a coming together of independent and disparate decision-makers within a decentralized setting may not always be liable to be within a clearly defined hierarchy where one could assert sufficient authority and clout to enforce its decisions before another. Its absence may lead towards a concurrency in the inherent decision-making, or for it to occur collectively without any clearly defined sequence. Another important

aspect of decision-making within the decentralized setting relates to the question of whether decision-makers are poised to optimize individual goals exclusively or be willing to work around individualistic objectives towards a common overall objective instead. Cooperative decision-making arising from the latter case could imply readiness to share individual information and engage in negotiations with regards to each prospective decision to be made towards a common objective [5]. This would be in stark contrast to the former where decision-makers have chosen to only be committed to personal goals. Decision-maker therein would be inclined to make individualistic decisions independently and without concern for other decision-makers, conceivably assuming a worse-case scenario when needing to account for decisions made externally by its peers [6]. Consequently, aside from the sequential decision-making other forms of unique decision-making could potentially emerge from within a decentralized decision-making setting. In that aspect lies the concern for misidentifying a particular problem within a decentralized decision-making setting as a bilevel problem. In this thesis, it is suggested that the ability to distinguish these variants would be crucial towards the identification of a bilevel problem within the decentralized decision-making setting.

1.1 Problems within a decentralized decision-making setting and its variants

Problems are defined here as a decision-making processes [5] whereby a decision-maker, i.e. an engineer or designer, attempts to maximize (or minimize) some desired criteria (e.g. performance, efficiency, cost) by deciding on the optimal control parameter values while being constrained by factors inhibiting its range of possible choices. In problems involving multiple objectives, decision-making also points

towards how a decision-maker resolves conflicting objectives and executes trade-offs between them when the need arises.

The decentralized decision-making setting describes the presence of multiple decision-makers within a single problem. In contrast to the notion of a single problem having only one centralized decision-maker with absolute control over all decisions made, such problems entail the devolution of control and responsibilities between multiple decision-makers [7]. Having been originally fashioned from within that one problem, a strong codependence could be present between individual decision-makers as manifested in the sharing of common resources or overlapping control of coupled parameters [5, 6]. *Local* decisions made by one decision-maker influence and, in turn are influenced by, other decisions made by its peers. The myopic view of the problem as seen by each decision-maker would also impinge on the way decisions are made locally. On a *global* level, the individual decisions made by the host of decision-makers would collectively dictate the outcome of the problem as a whole. Thus in a decentralized problem, local perspectives of individual decision-makers must be taken into consideration in defining and subsequently resolving the problem within the decentralized setting.

In the context of engineering, the presence of multiple decision-makers is a reality which confronts many of today's problems. Designing of airplanes or automobiles, for instance, constitutes a multidisciplinary exercise of extreme proportions which makes it impractical for a single decision-maker to approach the work as a single monolithic problem. Thus the division or *decomposition* of such problems into smaller and more malleable subsystems (in so doing incurring the need for multiple decision-makings) would have become an unavoidable necessity in order to render the

originally large problem more manageable as was postulated by Krishnamachari et al [8] and Kusiak & Wang [9]. From an alternative perspective, problems may already preclude naturally-defined decision-makers *sans* decomposition such as those involving communication or transportation network optimization [10, 11] (in the multitudes of network users that physically ‘inhabit’ the problem topography). Regardless, the presence of multiple decision-makers along with the possibility for different forms of interactions emerging between them decidedly influences the overall trend of decision-making that could take place within the problem. This in turn gives rise to different variants of the problem within a single decentralized decision-making setting.

1.1.1 The different variants of a problem within the decentralized decision-making setting

In describing problems within the decentralized decision-making setting, the way decisions are being made among constituent decision-makers represents an important aspect in defining and thus resolving the problems. There is a need to critically capture the unique interactions which occur between the different decision-makers leading on to its formal identification and modeling within the problem definition. To that end, Lewis & Mistree [13] had identified the utility of game theory¹ as a means to implicitly describe and model distinctive decision-making processes within the decentralized decision-making setting of collaborative multidisciplinary design problems. In their work entitled ‘Collaborative, sequential, and isolated decisions in design’ [13], they describe three possible games which could arise within such a

¹ The field of game theory was first broached in Von Neumann & Morgenstern’s book entitled ‘Theory of Games and Economic Behavior’ [12]. Here it describes the multiple decision-maker setup where each seeks to optimize individual objectives subject to individual constraints while subjected to dependence on design variables controlled by other decision-makers.

decentralized setting namely the Pareto game, the Nash game and the Stackelberg game or bilevel problem. Each game describes a specific decision-making interaction scenario which may occur within the collaborative multidisciplinary design backdrop. The following sub-sections describe the three possible games and the types of decision-making associated with it:

I. The Pareto Game

The Pareto Game was used to model decision-makers/players who are willing to share local information (i.e. their decisions) and compromise or trade-off individual objectives in deference to a common objective. The resolution involves a Pareto-optimal solution² in which the necessary trade-offs between local objectives is on a level agreed upon by all players. At its most ideal, where information could be disseminated instantaneously and without prejudice, this game mimics a single, central decision-maker working on the overall problem and deciding on the best possible course of action to achieve the desired trade-off.

II. The Nash Game

The Nash Game describes decision-makers/players who act in isolation. Such a system excludes perfect information (i.e. knowledge of what decisions are made by other decision-makers) although the presence of complete information may still be considered as a reasonable assumption. Though each player could know the different objectives, constraints and possible decisions or strategies of other players, it would not be privy as to what the actual decisions when made might be due to the absence of information sharing and dissemination. Players in such a setting are at best indifferent to what other players may decide on. The effect is a resolution which advocates a

² Pareto optimal occurs when it would no longer be possible to improve a single objective without the detriment of other objectives.

Nash equilibrium that could be described as a solution from which none of the players could unilaterally diverge from in order to improve its local goal. This means that should a player diverge from the Nash equilibrium solution while other players do not, it will only result in a reduction to its own level of optimization.

III. The Stackelberg Game (Bilevel problem)

The Stackelberg game is used to model the characteristic sequential decision-making process associated with the presence of a dominant decision-maker/player. In a problem with two players, the Stackelberg game describes a sequential decision-making process in which the dominant player (i.e. the leader) takes action first knowing that the dominated player (i.e. the follower) would react in order to attain an optimal response for itself [14]. It is important to note that this brand of sequential decision-making describes the bilevel problem as stated within the opening paragraph.

Despite the fact that Lewis & Mistree had defined the presence of the three abovementioned games within the limited context of a collaborative multidisciplinary design, the game theory model could generally be ascribed to the type of decision-making in other problems within the decentralized decision-making setting. Evidence of multiple forms of decision-making appearing within a single setting are confronted in well-known areas of research including vision tasks [15], supply chain [16, 17] and flow shop problems [18]. In a similar vein, there exist problems of a hybrid nature in which more than one form of decision-making are extolled within a single problem definition at the same time. True to form, the Stackelberg game and bilevel problem has made its appearance within the two aforementioned settings as listed within the table below:

Table 1.1: List of Bilevel Problems

No.	Name of problem	Examples:
1	Bilevel problem appearing within a setting with other possible decision-making processes.	<ul style="list-style-type: none"> - Collaborative design problem [5,6,13] - Multi-echelon inventory problems [16, 17, 19]
2	Bilevel problem in combination with other form of decision-making	<ul style="list-style-type: none"> - Transportation network problems [10, 20 , 21] - Capacitated plant selection problems [22]

1.2 The need to correctly identify bilevel problems within the decentralized setting

The need to distinguish the different possible types of decision-making for a problem within the decentralized decision-making setting stems directly from the potential for different optimal resolutions to arise from different types of decision-making. Rao et al demonstrated in [6] how failure to recognize and correctly identify inherent system traits (e.g. existence of cooperation and perfect information among decision-makers) within a problem may lead to an erroneous final resolution. In their analysis of the optimal design for a simple pressure vessel design with the two objectives of optimization of weight and volume based on the Stackelberg, Pareto (Cooperative) and Nash games (Non-cooperative) formulations, it was found that for each game achieves a potentially unique solution which may or may not overlap each other as shown in Figure 1.1. Figure 1.2 shows the resulting simple pressure vessel design corresponding to each solution. The danger thus lies with the assumption of an incompatible game strategy (be it Stackelberg, Nash or Pareto) when approaching such a problem.

Figure 1.1: Graph of optimal design solutions for design of a simple pressure vessel corresponding to Stackelberg, Pareto and Nash games (adapted from [13])

Figure 1.2: Variations in design of simple pressure vessel corresponding to different games (adapted from [13])

Currently, efforts to identify and distinguish the different variants of problems (inclusive of the bilevel problem) within a decentralized decision-making setting tend to be focused within each specific area of research such as in design [6, 13] or supply chain problems [16, 17]. In identifying and resolving the cooperative, non-cooperative and sequential forms of the multi-echelon inventory problem, both Wang et.al. [21] and Cachon & Zipkin [22] have listed down the different variants associated with the different types of decision-making possible. In Lewis & Mistree's 'Collaborative, sequential, and isolated decisions in design' [13], the different forms of decision-making processes possible within the collaborative multi-disciplinary problems were

described and compared. Thus there is a need to study the process of distinguishing each variant of the decision-making phenomenon for problems within a decentralized setting from a general standpoint. Consequently, the research task in this thesis is presented as the development towards a method aimed at identifying problems of a bilevel nature within general engineering applications. It is envisaged that with an approach developed whereupon easy identification of practical bilevel problem could be facilitated, a larger and more comprehensive classification process for general problems within the decentralized setting could be build upon its back.

1.3 Defining the Bilevel Problem within the decentralized setting

Arguably, the formulation and subsequent growth in practical applications of the bilevel program has had a significant contribution to the formalization of the bilevel problem. One such application concerns the resolution of complex systems of a bilevel nature. Complex systems are characterized by decentralization of control to multiple decision-makers (instead of a single centralized authority) and could be classified into two main branches of control: i) pure decentralized control and, ii) hierarchical control [23]. In essence, a complex system provides a general framework with which to express and conceptualize very large and complex practical problems; a view which was also expressed by Sage in [24]. The complex system thus provides a convenient framework from which to describe and comprehend problems of such nature (i.e. multiple interactive decision-makers, decentralization of authority) and be used as a backbone in the process to identify the different variants of problems within the decentralized setting. The following sections will introduce the various types of complex systems from the two branches and describe evidence of the bilevel problem within both instances.

1.3.1 Purely Decentralized System

The pure decentralized complex system is typified by decision-makers who are purely independent of each other as well as the conspicuous absence of any system-level coordination [23]. Each decision-maker would be in control of a subset of design variables but susceptible to another's decisions in seeking optimality for its own objectives. The pure decentralized systems therefore focus on the individual decision-makers and the need to formally model the unique interactions which go on between them [25, 26]. As stated before in Section 1.1.1, Lewis & Mistree [13] described the three possible games arising: i) the Pareto game, ii) the Nash game and iii) the Stackelberg game which relate to the cooperative, isolated and sequential form of decision-making that had been explained earlier in Section 1.1.1. The bilevel problem within the pure decentralized system is therefore represented by the Stackelberg game which arises when decision-making between the individual decision-makers takes place in a sequential manner.

1.3.2 Complex System with Hierarchical Control

The hierarchical controlled complex system, first conceived within the works of Mesarovic et. al [27], drew motivation from the salient hierarchical structure found within industrial organizations and certain control systems [28]. The hierarchical control structure, acknowledging the presence of multiple decision-makers within the complex system, dedicates a single decision-maker (the coordinator) to be responsible for optimality at a system level. The main objective of the coordinator is to influence the multiple *lower-level* decision-makers present within the system into making decisions beneficial to the overall problem and to resolve conflicts that may arise amongst them [23, 29, 30]. The focus would therefore be on the possible relationships

that exist between the coordinator and the other decision-makers. In general, there are two possible variants of the hierarchical system: i) the Traditional Multi-level/Hierarchical Approach and, ii) the Bilevel System which is representative of the bilevel problem's emergence within the hierarchical controlled complex system.

I. Traditional Multi-level/Hierarchical Approach³

The main idea of the traditional multi-level/hierarchical approach is to optimize a large system consisting of multiple, interrelated objectives and large number of variables by decomposing the system and optimizing the smaller, more tractable subsystems individually [23]. This approach relies heavily on Dantzig & Wolfe's [31] seminal work on decomposition of linear mathematical programs; to simplify the optimization of the larger mathematical program through that of the smaller, more tractable subprograms [32]. Having assigned the task of optimization for each decomposed subsystem to a lower-level decision-maker, a higher-level decision-maker/coordinator is then defined to iteratively coordinate [28] the individual optimizations of the local subsystem in order to attain system-level optimization. An essential condition needed for the traditional approach is that all the local objectives are defined (during the process of decomposition) to always be in *agreement* with the overall objectives (i.e. greater optimality of local objectives coincides with that of the overall problem). In complex systems with inherent multiple decision-makers, such a condition may not always be feasible, a reality which severely limits the effectiveness of the traditional approach. In the absence of the condition, the Stackelberg game offers an alternative approach resulting in the bilevel system.

³ Detailed surveys of these approaches are available in Mahmoud [33] and Nachane [34].

II. Bilevel System

The bilevel system is representative of the bilevel problem and comes into play when the condition of agreement between the local subsystems' objectives and system-level objective could no longer be assured [23]. It advocates a subtler version of the hierarchic control system in which the coordinator would have the ability to influence, but not completely control, the actions of decision-makers at the lower level [14, 35]. To accommodate the absence of agreement, a Stackelberg game/bilevel problem formulation is utilized. It advocates a sequential decision-making process where the coordinator is mandated to make decisions to optimize the overall system first while considering the reactions of the lower-level decision-maker (which are either known or could be reasonably estimated). The lower-level decision-maker would then optimize its own sub-problem; essentially having its feasible space limited or constrained through the prior decisions made by the coordinator. Overall, it is the opinion of Bialas & Karwan [36, 37] and Bard [14] that such an account is a more realistic depiction of actual problems adhering to the hierarchical/multi-level decision-making structure as it would be logical to give greater difference to the decision-maker in charge of the overall rather than local objectives. Such a view would be further strengthened by the many applications of the bilevel system in practical problems involving multi-level management including in economics [38, 39] and transportation [40, 41].

1.3.3 Summary of Section

This section describes bilevel problem within the context complex systems of pure decentralized and hierarchical complex systems. The utility of the complex system is its ability to provide a general framework with which to express and conceptualize

very large and complex practical problems. Therefore combined together, the two types of bilevel problems describe practical instances where the bilevel nature appears within problems of a decentralized decision-making setting. It could be observed that in order to ascertain the presence of a bilevel problem within the complex system, one is required to know when the Stackelberg game comes into play. This refers directly to recognizing the absence of conditions affiliated to the Pareto and Nash games in pure decentralized systems and the infeasibility of the traditional multi-level/hierarchical approach in hierarchical control systems in addition to the emergence of conditions affiliated to the Stackelberg game. This section thus also affirms the role of the Stackelberg strategy within complex systems of a bilevel nature [42] and also of the strong affinity of the Stackelberg game to bilevel problems.

1.4 Objective

The overall aim of this research is to develop a methodology towards identifying the bilevel problem within a decentralized decision-making setting. In order to achieve this aim, the following objectives were identified:

- To show that the bilevel problem tend to manifests itself within the context of decentralized decision-making problems in which different types of decision-making exist or co-exist.
- To adopt a common platform to study the Stackelberg/bilevel, Pareto optimal and Nash equilibrium variants of problems within the decentralized decision-making setting without being limited to within any specific topic of engineering.

- To identify the particular characteristics associated with the bilevel problem which would exclude the possibility of the Pareto optimal or Nash equilibrium problems appearing.
- To formulate a general protocol with which to identify bilevel problems within the decentralized decision-making setting and to test its effectiveness using appropriate case studies.

1.5 Organization

Chapter 2 reviews different literature involving the optimization of problems within decentralized decision-making setting. It concurrently reviews examples of prior works pertaining to the bilevel problem and highlights a general trend in which the bilevel problem could appear within the framework of complex systems. Chapter 3 then characterizes the bilevel problem based on the traits derived in Chapter 2 with the aim of devising an approach based on the traits to aid identification of a practical bilevel problem. It then provides a description of the different mathematical approaches used to solve a bilevel program. Chapter 4 provides a modular description of the algorithm utilized to resolve the sample bilevel problems. Chapter 5 initially describes the process of identifying the traits of the multi-echelon inventory problem and decentralized capacitated plant selection problem which cast them as bilevel problems. Subsequently, actual case studies based on the two problems are undertaken. Results obtained were then tabled and analyzed and a comparison was made between the results obtained and those in related literature. Chapter 6 discusses the approach taken to achieve the objectives set out in Chapter 1 and reiterates whether they have been achieved. The chapter also provides an analysis of those

objectives which have not been attained. It also offers suggestions for further work which could be undertaken to improve the program as well as future implementations.

Chapter 2

Literature Review

Problems of a bilevel nature involve a sequential decision-making process that forms one of three distinct types of decision-making which could emerge from within a decentralized setting. As noted in Rao et.al [6], Wang et.al [16] and Cachon & Zipkin [17], different types of decision-making may result in differing optimal resolutions. Thus far, the research community has tended to approach formulations derived from each type of decision-making individually coupled with a number of comparative studies that are limited to within specific areas of interest. In Chapter 1, the research problem stating that there is no common approach to distinguish the sequential decision-making intrinsic to the bilevel problem from other types of decision-making within the practical, engineering decentralized decision-making settings has been identified. Henceforth, the purpose of Chapter 2 is to present a critical review of existing works relating to the different forms of optimization approaches currently available pertaining to practical engineering problems within the decentralized decision-making setting. The review will categorically cover the different types of decision-makings found within each problem. Consequently, it shows that problems within the decentralized setting do traverse different possible types of decision-making within which the sequential form of the bilevel problem represents a distinct possibility. It also shows that there is a dearth of methods that identify one from another from a generic standpoint. While a number of research work have covered this issue, they have been focused on specific areas of interest. This suggests the need to develop a consolidated approach to identify the bilevel nature for general problems within the decentralized setting. Such an approach could then be incorporated within a

comprehensive resolution framework for all problems within the decentralized setting. Hence the purpose of this review is important in establishing that need.

2.1 Categorization of present works evaluated to distinguish the different types of decision-makings within a single decentralized setting

Although the presence of multiple decision-makers within a single problem has been identified and worked upon in many early research works, comparatively few have focused on allowing for the local decision-makers to be modeled with complete and unlimited independence with regards to decision-making i.e. for each decision-maker to make independent and selfish decisions. The purpose of a critical review of prior works is twofold. First is to establish the existence of multiple variants for a problem within a single decentralized setting due to the possibility for different types of decision-making emerging therein. Second is to show that the bilevel problem represents one of these variants and thus requires a way to identify and distinguish it from the others.

In order to highlight the presence of different types of decision-making of practical applications within the decentralized setting, it is proposed that the review be categorized under the following categories:

- Restricted decision-making,
- Cooperative decision-making,
- Isolated decision-making,
- Sequential decision-making (Bilevel problem).

The last three categories are similar to the different games described in Section 1.1.1. The rationale for such a classification is discussed in the following subsections:-

2.1.1 Restricted decision-making

Restricted decision-making is commonly associated with older resolution approaches for optimizing decentralized problems from a centralized viewpoint. The main idea behind these approaches is to optimize an initially large system consisting of multiple, interrelated objectives and large number of variables by decomposing the system and optimizing the smaller, more tractable subsystems individually [31]. Restricted decision-making refers to the limitation imposed on individual decision-maker who could only implement decisions which are globally optimal, even though such decisions may be detrimental to its own local objectives. In his survey of decentralized and hierarchical models and theories, Findeisen [23] stated that such an assumption could only be practical when describing: i) decision-makers who are without any personal goals and thus could have one assigned to them by an overall system-level designer in charge of the overall problem and, ii) decision-makers who do have personal goal with the caveat that the overall goal of the problem could be tailored to be consistent with each and every personal goal without conflict.

2.1.2 Cooperative decision-making

Cooperative decision-making imply that each decision-maker is willing to compromise its own local objectives to achieve a common overall objective. Chen & Lewis [43] stated that complete cooperation could be achieved within the decentralized problem only when each decision-maker is fully aware of all others (i.e. their objectives, constraints, etc.) as well as the decisions made by each. Typically, this describes decision-makers who make decisions while sharing local information and negotiating potential decisions within a seamless information dissemination infrastructure. It should be noted that in Wang et.al [16], the cooperative formation

was achieved through the mutual acceptance of ‘cooperative mechanisms’ such as contracts or other carefully designed legal agreements in lieu of engaging in active negotiations and information sharing. This does not define the cooperative decision-making and will be discussed further in the review of the multi-echelon inventory problem in Section 2.2.2.

2.1.3 Isolated decision-making

Isolated decision-making stems from decision-makers who make independent decisions without concern for its peers as well as system level optimization within the decentralized problems. Isolated decision-maker make decisions alone, without knowing what the other decision-makers’ actions are (i.e. each decision-maker does not communicate its actions to others). Ideally, a condition of complete information could be assumed in which each decision-maker would know the objectives, constraints and possible actions of other decision-makers. Isolated decision-making typically occur in situations where it would be unfeasible or impractical for individual decision-makers to communicate with one another. Lewis & Mistree [13] described such a situation as arising due to organizational, information or process barriers with geography or red tape having been cited as examples of such a barrier [7]. Migdalas [10] provided an example for the impracticality of information sharing and cooperation in describing the everyday activity of choosing an optimum route from point of origin to destination within a specific transport network system. Wang et.al [16] & Cachon & Zipkin [17] cited rival businesses as representative of decision-makers who engage in isolated decision-making.

2.1.4 Sequential decision-making

This category of decision-making describes the presence of a dominant decision-maker who would make decisions while knowing that the subsequent decision-maker would make a self-optimal response. Lewis & Mistree [13] stated such sequential form of decision-making among decision-makers typically arise when there is a dominant decision-maker present or out of practicality. Cachon & Zipkin [17] gave Intel or Best Buy as examples of dominant decision-makers within a supply chain while Lewis & Mistree [13] described the traditional over-the-wall design process as an example of the latter. As stated before, sequential decision-making is inherently tied to the bilevel problem and forms a unique characteristic.

2.2 List reviewed

Using the set of categories stated in Section 2.1, the following four practical problems in decentralized decision-making setting were reviewed:-

1. Collaborative multidisciplinary design,
2. Multi-echelon inventory problem,
3. Decentralized capacitated planning problem,
4. Transport network design problem.

The four problems reviewed here are not meant to be exhaustive but rather they represent a large selection of bilevel programming applications within engineering.

2.2.1 Collaborative multidisciplinary design

This problem describes the design process of large and complex engineering systems such as automobiles, aircrafts or satellites. Xiao et.al [31] stated that designing such complex structures would typically require collaboration between multiple engineering teams from different disciplines due to the complex nature of

each system which involves amalgamation of different branches of knowledge such as electronics, structural and mechanical. This fact was collaborated by Chanron et.al [45] stating that engineering systems such as those described in the beginning are the results of the interactions between many distinct entities working on different parts of a single design. Sobiesczanski-Sobieski et.al [45] added that the decomposition of complicated engineering systems fulfills an obvious need for enabling distribution of work over to many people.

In describing the problem of collaborative multidisciplinary design, Chen & Lewis [43] mentioned that it is exactly the multidisciplinary nature of the problem that creates organizational challenges along with the possibility of conflicts of interest between the different disciplines. As earlier established approaches such as single-level optimization [46], concurrent subspace optimization (CSSO) [47] and collaborative optimization (CO) [48, 49] looked to solve these problems by focusing on system-level optimization and enabling concurrency of decisions, they tend to turn a blind eye towards decision-makers (i.e. disciplinary engineering teams) engaging in fully independent decision-making, glossing over the interactions which occur between them. They therefore ascribe such problems with multiple disciplinary engineering teams who engage in a restricted form decision-making (i.e. they will not defy system level decisions). In accounting for the different types of interactions, multiple works have instead defined other types of collaborative design efforts in the cooperative, non-cooperative as well as sequential decision-making forms as described before [6, 13, 28, 30, 44 & 45].

Restricted decision-making

In the introduction to ‘Collaborative multidisciplinary decision making using game theory and design capability indices’, Xiao et.al [31] had listed three established approaches for resolving the collaborative multidisciplinary design problem namely: i) the single-level optimization [46], ii) concurrent sub-space optimization (CSSO) [47] and iii) the collaborative optimization [48]. Balling & Sobieszczanski-Sobieski [48] stated that the objective of these approaches is to facilitate concurrent engineering technology for complex advanced systems by enabling a centralized resolution of coupled design parameters. This common focal point for all three approaches therefore results in each of them describing disciplinary engineering teams who only engage in the restricted form of decision-making, ultimately compliant to decisions made by the overall system level designer. The following will describe the different approaches in greater detail.

In his survey of multidisciplinary optimization, Sobieszczanski-Sobieski & Haftka [46] described the single-level optimization framework as an iterative process which involves, in one iteration, system-level decisions being made on all relevant parameter selection; selection being relayed to each local design team; each team assessing the performance of the system-level selections locally and giving feedback to the system-level decision-maker. Consequently, each local design team does not display any form of independent decision-making and is reduced to merely informing the system level designer of the level of performance within each discipline. On the other hand, the CSSO and CO approaches do define multiple sub-problems, each ascribing to a particular discipline for each local decision-maker to optimize. Ultimately however, the objective of each of the two approaches is still very much focused on system level

optimization and resolution of coupled variables as stated before. Thus both frameworks proscribe decisions made by the system-level decision-maker in coordinating the local decision-makers towards a system level optimization as final and non-negotiable. The coordinating decision usually takes the form of a key parameter which affects all local decision-makers which the local decision-makers are not allowed to disagree with even to the detriment of their local objectives. Examples of this formulation could be found in [48, 49 & 50].

Cooperative decision-making

Rao et. al [6] first demonstrated cooperative decision-makers within the collaborative multidisciplinary design problem as one in which the different engineering teams/designers cooperate towards a Pareto-optimal resolution. Lewis & Mistree [28] further developed on the cooperative decision-making formulation by distinguishing full and approximate cooperation as defined by the physical limitations of the infrastructure in which the design teams work in. They stipulated that in order for full cooperation to occur, all design teams have to be within a fully integrated computer environment having access to all non-local information needed. Thus they commented that the approximate cooperation would have been more realistic in which the different designers would not have full access of non-local information but would instead work with approximations derived from sufficient information obtained from each other. In ‘Collaborative, Sequential, and Isolated Decisions in Design’ [13] the difference between the full and approximate cooperation was further elaborated on resulting in the stipulation that each designer must have the ability to communicate and access its counterpart’s analysis at any time in order for full cooperation to take place. This therefore implies the need for designers to not only be willing to cooperate

but also for the seamless dissemination of information between designers to be facilitated in order for full cooperation to take place.

Isolated decision-making

In ‘A Study of Optimal Design Design Under Conflict Using Models of Multi-Player Games’, Rao et. al [6] introduced the isolated decision-making form of the collaborative decentralized decision-making problem as a non-cooperative decision-making process with the Nash equilibrium as the optimal state. This was further developed by Lewis & Mistree [13] stating that the isolated decision-making takes place when any form of coalition among the engineering design teams would not be possible due to organizational, information or process barriers. They stressed that even though the possibility for complete non-cooperation could still be a possibility among the different engineering teams, the main cause for them to engage in isolated decision-making would still ultimately fall on such barriers as mentioned before. They argued that even although the non-cooperative/isolative decision-making was derived from an economical mindset in which complete non-cooperation occurs often, engineering designers should always strive towards cooperation.

Sequential decision-makers

Lewis & Mistree [13] described sequential decision-making as taking place within the traditional over-the-wall design process. An illustration of this process would be having only one team design working on the design and releasing to the next team once the team is satisfied with it. Therefore, it could be assumed that one disciplinary design team gets to make decisions first and enforce them on the next team. Due to the nature of coupling in which decisions external to a discipline would have an effect on its objective, the leading design team would therefore optimize his objectives

while taking all the following design team's reaction into consideration. From another perspective, the values for the leader's non-local variables could therefore be safely assumed as the follower's rational reaction to any decision the leader makes. Once the leader design team decides on the optimal values for its local design variables, it informs the follower design team of its choice. The follower design team then reacts to the leader's decision by optimizing its own objectives by selecting its optimal local variables and treating its non-local variables (i.e. the leader's local variable) as constants. Figure 2.1 describes the flow of information while Figure 2.2 shows a graphical representation of an over-the-wall design of complex engineering systems.

Figure 2.1: Flow of information within the design process (adapted from [13])

Figure 2.2: A representation of the over-the-wall design process (adapted from [13])

2.2.2 Multi-echelon inventory policy problem

The multi-echelon inventory policy problem, first broached by Clark & Scarf in their seminal paper entitled ‘Optimal policies for a multi-echelon inventory problem’ [11], describes a serial supply chain (e.g. a supplier and retailer) whereby one member supplies the other who then faces the external customer demands. The supplier is tasked with replenishing the inventory of the retailer who in turn satisfies customer demands. The quantity stocked by the supplier would have an effect on the retailer’s ability to provide for customer demands as a smaller, more cost efficient quantity (for the warehouse) may result in stock-outs (no inventory) when receiving a request from the retailer. This, in turn, causes possible losses to the retailer in terms of lost customer patronage or possible *backorder costs* (cost paid to customers/retailers for delay in delivering product). Each member therefore attempts to minimize cost by adopting optimal inventory policies which may include parameters such as stocking quantity and transport modes. Wang et.al [16] stated that most supply chain systems are essentially decentralized in nature. Cachon & Zipkin [51] lend credence to this idea stating that “*all supply chains operate as a collection of independent agents, each responsible for managing a subset of the supply chain, each with its own objectives.*”

Although Clark & Scarf’s work represents the first attempt to study the effects of having multiple inter-connected installations (e.g. the warehouse, retailer) on a single supply chain line as, their work essentially focuses on the optimization of the supply chain *as a whole* and therefore defines decision-makers who always agree with overall optimal policy. This implies that both the supplier and retailer are mere non-entities who are willing to accept inventory policies (which may not be individually optimal) in order to attain overall optimality for the supply chain line. Realistically,

such an outcome might not always hold true as demonstrated in the works of Cachon & Zipkin in [17, 51] and Axasäter in [19]. Similar to the collaborative multidisciplinary design problem, all other three forms of decision-making (cooperative, isolative and sequential) formulations have subsequently been established for the multi-echelon inventory policy problem within different works such as [12, 17 and 51].

Restricted decision-making

Traditionally researchers have not distinguished the supplier and retailer to act independently although both were defined with unique objectives (i.e. optimization of individual cost incurred). In most of established works (Clark & Scarf [11], Ishii et al. [53] and Cohen & Lee [54]), a centralized approach was relied upon (i.e. the problem was not set within a decentralized decision-making setting) where it is conveniently assumed that both supplier and retailer would accept any system optimal policy even though it may lead to the detriment of local cost optimization. Lee & Whang [55] provided an extreme example of the possible outcome of the restricted decision-making in their working paper entitled “Decentralized multi-echelon supply chains: Incentives and Information” in which the system optimal inventory policy would have meant that all backorder penalties would be charged to the lowest stage (i.e. retailer) only.

Isolated decision-making

Cachon & Zipkin’s ‘Competitive and Cooperative Inventory Policies in a Two-Stage Supply Chain’ [17] stated the isolated decision-making forms as being present in both the cooperative and non-cooperative forms of the multi-echelon inventory policy problem. In the non-cooperative or competitive form, Cachon & Zipkin

defined a supply chain composed of two self-interested members/firms. Each firm makes decisions alone and without communicating its potential decision to the other thus achieving the isolated decision-making. The final optimal resolution attained is a Nash equilibrium from which none of the members could deviate without decreasing its own optimality. In the cooperative form, each firm still engages within an isolated decision-making although they have agreed to work towards the optimality of the entire supply chain through legal contracts. In order to overcome the competitive nature of the members (both member's fundamental goals are to optimize their local objectives), Wang et. al [16] stated that such contracts should comply with the following rules:

- i) That it ensure the members' profit is not lower than the profit before supply chain cooperation and properly share the total cooperative revenue and,
- ii) That it should diminish the incentives to deviate from the system optimal solution.

Sequential decision-making

Cachon & Zipkin [17] described the sequential decision-making as evident within a supply chain where either the distributor or retailer is a major player (i.e. WalMart, Intel) within the supply chain line and thus dominates the decision-making process by establishing its policies prior to the other member. Cachon & Zipkin's formulation defined a cost structure that is based on each installation adopting a base-stock policy with periodical review, with customer demand occurring at the retailer installation in a random manner defined by a Gaussian distribution. Similar to Clark & Scarf's earlier work, Cachon & Zipkin's defines a penalty cost within the supplier's cost function which is based on the backorder cost incurred at the retailer. In other words,

the backorder cost is shared among the two installations. Thus should the supplier assumes the dominant role, it would have to make its decisions while simultaneously considering the reactions of the retailer as any decisions which may lead to high backorder cost would be detrimental to its cost minimization objectives. The original aim of Cachon & Zipkin's work was to ascertain the difference between optimal results when inventory policies are defined by *echelon* inventory position and *local* inventory positions¹.

An alternative form of the sequential decision-making supply chain described by Axasäter [19] introduced a cost function based on each installation adopting continuous review while preserving the sequential decision-making order. Consequently, the optimal policy calls for an $(s-l, s)$ policy in which the decision-maker would order for inventory each time inventory position falls one below the replenishment s level. Customer demand is simulated as a Markov process with Poisson distribution characterized by the average customer rate λ (no. of customers/time unit). The work focus on establishing a framework for decentralized multi-echelon inventory control which amounted to determining, for both supplier and retailer, optimal inventory policy (the value of S) and transportation mode (which effects the lead or transportation time) individually. The framework defines a cost structure in which the supplier would have to incur a penalty cost equivalent to the increase in cost that would be incurred by the retailer downstream due to changes to the supplier's initial inventory stock and lead time. Axasäter described such a penalty cost as equivalent to a nonlinear function of the delay to retailer deliveries. Thus the formulation is defined as a Stackelberg game in which the supplier gets to decide his

¹ Local inventory position refers to the amount of inventory available within an installation plus those inventories being transported towards it and subtracted by the number of backorder incurred. Echelon inventory position refers to the same value plus all the inventory position values of the installations downstream.

inventory policy and choice of transportation mode before the retailer, but must be willing to recompense to the retailer for any increase in cost due to it changing its optimal policies.

2.2.3 Decentralized capacitated plant selection problem

The decentralized capacitated plant selection problem describes the task of a *principal* plant in selecting a set of *auxiliary* plants that provides sub-assembly parts in order for the principal plant to achieve a specific production rate while minimizing costs. The capacitated plant selection problem describes a variant of the general facility location problem [56] in which the individual capacities of each facility is limited to a certain number, i.e. there is a limit to the amount of product each facility could supply or produce.

Typically, the problem is modeled as a single monolithic problem such that a single centralized decision-maker decides on both the facilities to be opened *and* the production level of each facility. Examples of the single monolithic formulation could be found in Ghianni et al's branch-and-price [57] and Jaramillo et al's genetic algorithm [58] approaches. Cao & Chen [22] stated that such a monolithic model would imply that the different plants managers (i.e the principal and all auxiliary plants) are in effect willing to accept and impose a centralized solution for 'overall' benefit. However, Etrogal & Wu [59] stated that such assumptions could be deemed as unrealistic when the different plants are not owned by the same companies or may have only restricted knowledge pertaining to the cost functions of other plants. This gives rise to the decentralized capacitated plant selection problem developed by Cao & Chen [22] which defines the principal plant's task within a bilevel problem while the collective decisions of the auxiliary plants was represented by the cooperative

form of decision-making. Figure 2.4 shows the flow of decision information within the bilevel problem as formulated by Cao & Chen.

Figure 2.3: Bilevel formulation of the decentralized capacitated plant selection problem (adapted from [22])

Sequential decision-making

Cao & Chen [52] defined two clear, independent decision-makers who engage in a sequential decision-making within the decentralized Capacitated Plant Selection Problem: i) the principal firm who decides on which plants to open based on an overall business plan and, ii) all the auxiliary plants as a single collective decision-making entity which decides on the individual production level at each auxiliary plant. The objective of the principal firm is to select the plants which would represent minimal cost in terms of opening the new plant as well as opportunity cost which arises (i.e. due to the different plants operating at below optimal capacity) while fulfilling its demand for the different components. The cost function of the auxiliary plants consists of the cost associated with producing the particular quantity of components and the cost of transporting the said component to the principal firm. Although the principal firm is defined as the dominant decision-maker and has the

authority to decide on which plants to open, it does not control the operating level at each of the auxiliary plants. The principal plant therefore has to select the appropriate auxiliary plants which would not only be able to provide for its demand but also result in a minimal cost in terms of the fixed, inventory and production costs. In deciding which of the auxiliary plants it should select, the principal plant would have to consider the auxiliary plants combined reactions in terms of production levels as this would have an effect on its cost minimization objective.

Cooperative decision-making

The cooperative decision-making manifests itself within the collective decisions made by the auxiliary plant managers in selecting the most optimal production level. Cao & Chen defined the each of the selected auxiliary plants as operating would work in a cooperative manner thus achieving Pareto optimality.

2.2.4 Transport network design

The transport network design Problem arose due to the need to accurately predict the flow of traffic when changes are affected on a transport system [60, 61]. It addresses the issue of decreased efficiency in an established transport system due to congestions brought about by increased travel demands over time. The issue is remedied by either building new links (i.e. roads) or increasing the capacity of established links [10, 60] in order to reduce congestion within the system. Ultimately, the aim is to reduce travel time between any two points of origin and destination within the transport system. The travel time of an individual user along an individual link is correlated to the number of vehicles which ply along the said link, i.e. the traffic flow along the link. Understandably, large flows cause congestion thereby increasing the travel time experienced by the user [60, 62]. The total travel time, the

sum total of all the different user travel times between any two points within the system, constitutes a dominant contributing factor to the system travel cost. Migdalas [10] surmised that the objective of the transport network design problem is thus to minimize the total system costs consisting of the system travel cost and the amount of investment poured into improving the transport system.

The main complexity involved in the traffic network design problem lies with the independence of individual users in selecting their choice of route between points of origin and destination. Typically, individual users of the network operate in a non-cooperative manner whereby each user seeks to minimize individual travel times by selecting routes independently [60 & 62]. In general, a network designer (i.e. the coordinator in the traffic network design problem) cannot prescribe specific routes to individual users. However, through actions such as increasing the capacity or adding new link, certain routes could be made more attractive to the individual user [10] due to the promise of decreased individual travel time. Essentially, the designer can only control the investment aspect of the total system costs while the system travel cost aspect is determined independently by the traffic flow pattern resulting from the aggregated choices of routes by individual users [10]. The traffic flow pattern, known as the user-optimal flow, is predicted and modeled by solving a separate problem known as the Traffic Assignment Problem² [60 & 62].

The transport network design problem therefore relates closely to a cause-and-effect structure in which a network designer needs to know how best to improve a particular road network based on how road users would eventually utilize the network. Thus the transport network design problem could be described as a bilevel problem in which the network designer represents the coordinator who is concerned

² Also known as the Network Equilibrium Problem [10].

with the overall performance while the road users represents the local or low-level decision-makers who are looking to improve on their individual travel time. The following section will detail the sequential decision-making process within the transport network problem.

Sequential decision-making

In LeBlanc's 'An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem' [62], an approach was devised that incorporates the Traffic Assignment Problem into the Transport Network Design Problem in the role of an oracle [10] to predict the user-optimal flow. The two decision-makers are the network designer (i.e. leader/coordinator) and the aggregated users' selections which ascribe by Wardrop's first condition (i.e. follower/lower-level decision-maker). The designer seeks to fulfill his original objective of minimizing the total system cost through actions such as adding new links and improving the capacity of existing links. This objective is representative of a system or societal aspiration in which the goal is to improve the existing transport system as a whole by minimizing total system cost. However, the designer will be cognizant of not only the non-cooperative nature of each user of the network but more importantly of the resulting traffic flow based on such behaviors. The aggregated choices or actions of the individual users, on the other hand, is expressed through the resulting traffic flow brought about by their selection of routes in the non-cooperative manner (i.e. each user seeking optimize personal travel times resulting in a traffic flow satisfying Wardrop's first condition). Thus when the designer makes decisions regarding addition of links or capacity increments, he would simultaneously consider the aggregated reactions of all network users predicted through the Traffic Assignment Problem. The traffic flow is modeled by

solving the Traffic Assignment Problem once the appropriate decisions (i.e. adding links and increasing capacity) had been made by the designer and information (in the form of a new network) is passed to it.

Isolated decision-making

The isolated decision-making manifests itself within the Traffic Assignment Problem. The objective of the Traffic Assignment Problem is to predict the user-optimal flow that occurs when each individual user independently selects a route between his point of origin and destination in a particular transport system. The motivation behind each user in selecting a route is invariably individual cost minimization, a factor that is most dependent on travel time [10, 62]. To reiterate, traveling time along individual routes (or, more specifically, along the individual links/roads which make up the routes) is adversely affected by congestion (i.e. the traffic flow along the individual links) [62, 62]. This leads to a predicament in which a route between two points, popular due to its short travel time, would eventually cease to offer such a lucrative proposition as the number of users using it increase and congestion ensues. Migdalas [10] had summarized the traffic assignment problem objective as ‘predict(ing) flow patterns and travel times which are the results of the network users’ choices with regard to route from their origins to their destination’.

In order to predict the traffic flow patterns, it is necessary to ascribe a certain behavioral assumption based on the common objective of each user (i.e. seeking the route with minimum travel time for their personal origin and destination points) known as ‘Wardrop’s first condition’ [63 & 64]. The condition stipulates that i) all utilized routes between a particular pair of origin and destination points would incur equal cost (i.e. cost through travel time) to each user and ii) no alternative routes exist

(between the same points) that would incur less cost. The traffic flow that satisfies this condition is known as *user-optimal* or user-equilibrium flow. Migdalas [10] stated that the resulting traffic flow represents a type of network equilibrium (analogous to the Nash equilibrium) in which no user can improve on his travel time by unilaterally choosing an alternative route. The same view was expressed in LeBlanc [62], as well as Roughgarden & Tardos [65].

It is important to note that the user-optimal flow (or network equilibrium flow) usually does not constitute the most efficient utilization of the available network (i.e. realization of the lowest possible time travel for *all* origin and destination points pairings)³. Such a flow ascribes to a behavioral assumption known as ‘Wardrop’s second condition’ which stipulates that the *average* journey time for all origin-destination pairs must be a minimum under such a flow [64]. Thus the traffic flow resulting from Wardrop’s second condition is known as the *system-optimal flow* [10]. The system-optimal flow, rather than focusing on individual user’s limited perception of most time-efficient route, is beneficial from a societal perspective (i.e. obtaining the most out of the transportation network as a whole). The implication here is that it requires a certain degree of cooperation from the individual users in sacrificing a potentially more optimal choice of route for another which would perpetrate a better overall flow. Migdalas [10] described the system-optimal flow as the societal perception of optimal network utilization in which total travel time in the network is minimized.

Migdalas [10] emphasized the importance of the Traffic Assignment Problem when he stated that the observed traffic flow in real life is *closer* to the user-optimal

³ Migdalas [10] states that the only exception to this rule of norm is when there is no congestion, i.e. the ideal case.

flow than the system-optimal flow. Failure to account for the user-optimal flow within a transport network may have adverse effects on efforts to improve on the existing transport system. Braess [66] described a theoretical situation, the Braess Paradox, whereby adding a new link to an existing transport system causes the total travel time to increase rather than decrease. It is therefore essential that any efforts to improve a transport network system include the modeling of the network equilibrium flow within the system. This gives rise to an approach for the Transport Network Design Problem that utilizes elements of the Stackelberg game, which was first proposed by LeBlanc in [62]. The following section describes the problem in greater detail.

2.3 Chapter Summarization

This section has reviewed several practical examples of problems within the decentralized decision-making setting as categorized under the four headings. The four categories relate to the different types of decision-making possible within a particular problem namely the restricted, cooperative, isolated and sequential forms of decision-making. Table 2.1 presents a summary of the findings:

Table 2.1: Summary of review findings

Name of decentralized problem	Restricted decision-making	Cooperative decision-making	Isolated decision-making	Sequential decision-making
Collaborative multidisciplinary design	Yes	Yes	Yes	Yes
Multi-echelon inventory	Yes	No	Yes	Yes
Decentralized capacitated plant selection	No	Yes	No	Yes
Transport network design	No	No	Yes	Yes

Table 2.1 shows that all four of the practical problems within decentralized decision-making setting reviewed may exhibit more than one type of decision-making

in which sequential decision-making is consistently present within all four problems. It also shows that not all four decision-making types would always be present in the four problems as shown in the absence of cooperative decision-making type for the multi-echelon inventory problem. This may be so due to the nature of the problem in which certain types of decision-making are unlikely to occur such as cooperation among different companies within a supply chain as well as individual road users within a transport network.

In addition, a general trend could be observed for the bilevel problem in which it could appear within the context of pure decentralized (i.e. the collaborative multidisciplinary design and multi-echelon inventory problem) and hierarchical complex systems (i.e. the decentralized capacitated plant selection and transport network design problems). The utility of the complex system is its ability to provide a general framework with which to express and conceptualize very large and complex practical problems. Therefore combined together, the two types of bilevel problems could be used to describe practical instances where the bilevel nature appears. It could be observed that in order to ascertain the presence of a bilevel problem within the complex system, one is required to know when the Stackelberg game comes into play. This refers directly to recognizing the absence of conditions affiliated to the Pareto and Nash games in pure decentralized systems and the infeasibility of the traditional multi-level/hierarchical approach in hierarchical control systems in addition to the emergence of conditions affiliated to the Stackelberg game.

In summary, the following conclusions can be made with regards to the aim of this Chapter. It seeks to establish the presence of different types of decision-making within practical problems within the decentralized decision-making setting as well as

obtain confirmation of the sequential nature intrinsic to the bilevel problem as one of them:

- Problems within the decentralized decision-making setting do have the potential to exhibit different types of decision-making,
- The sequential decision-making intrinsic to the bilevel problem is present as one of the decision-making type in each of the four problems reviewed,
- The sequential decision-making intrinsic to the bilevel problem appears within problems of decentralized decision-making as described in Figure 2.4.

Figure 2.4: Classification of problems within the decentralized decision-making setting

Hence in order to identify the bilevel problem within the decentralized setting, this innate structure of the different decision-making types could be utilized.

Chapter 3

Characterizing Bilevel Problems and Resolving Bilevel Programs

In Chapter 2, two sets of traits were derived; one to identify bilevel problem associated with pure decentralized complex system (i.e. the Stackelberg game formulation) and the other for those associated with hierarchical complex systems (i.e. the Bilevel System). The aim of this chapter is to develop a formal means to identify and characterize problems of such nature by incorporating the different traits in a sound and systematic manner. The previous review of bilevel problems pertaining to complex engineering system and transport network design problems have revealed that a key issue is recognizing what type control structure is present within a problem (i.e. whether hierarchical or pure decentralized control). In a pure decentralized complex system, the focus would be on the interactions between individual decision-makers while the focus of the hierarchical complex system is would be on the intra-level relationship between the coordinator and the lower-level decision-makers. Thus the manner in which the bilevel problem emerges within these two structures differs from each other (i.e. one from the perspective of decision-maker to decision-maker while the other from that of the coordinator to the lower-level decision-makers). Therefore, this implies that in order to effectively identify practical instances of problems with a bilevel nature, one must first be able to recognize a complex system and ascertain whether a system of is of pure decentralized control or hierarchical control as well.

This chapter is divided into two parts. The first half describes the identification process of the problem while the second half describes issues pertinent to the resolution of a bilevel problem, i.e. bilevel program. Section 3.1 describes a possible

framework to identify the bilevel problem. Based on the traits derived from Chapter 2 as well as the review of the two practical instances of a bilevel problem, the framework takes a three step approach consisting of: i) identifying the presence of a complex system, ii) ascertaining the type of control at work and iii) determining the presence of a Stackelberg game or bilevel system. Section 3.1.1 describes the first step which consists of looking for tell-tale signs such as the presence of multiple objectives and/or decision-makers and coupling between the individual (decision-maker's) problems. Section 3.1.2 then look for elements unique to each type of control. The presence of a Stackelberg game or bilevel system is then ascertained in Section 3.1.3. The second half of the chapter covers issues pertinent to the mathematical resolution of the bilevel problem. Section 3.2.1 describes the mathematical formulation of a bilevel program in detail, i.e. the leader and follower programs. Section 3.2.2 describes issues pertaining to the solvability of the bilevel program, i.e. the presence of non-singular follower rational reaction. The final part of Section 3.2 describes the different approaches developed to solve bilevel programs. The Chapter is then summarized in Section 3.3.

3.1 Identifying Problems of a Bilevel Nature

From the previous two chapters, it could be observed that problems of a bilevel nature occur almost exclusively where multiple interacting decision-makers are present. This is logical as the Stackelberg game/bilevel problem is associated with the describing dominance among multiple decision-makers and a sequential decision-making process. To reiterate, problems involving multiple decision-makers could be described within the complex system framework. Within the framework of complex systems, the bilevel problem manifests itself in two guises: i) the Stackelberg game

within the pure decentralized complex system and, ii) the bilevel system within the hierarchic system. The following lists the traits of the Stackelberg game followed by that of the bilevel system as derived in Section 2.1.1 and 2.1.2 respectively:

Traits of the Stackelberg game:

- a) the presence of two or more interactive decision-making units, each responsible for decisions made within a specific subsystem;
- b) each decision-maker controls a subset of the total design variables defined within the system;
- c) each decision-maker independently optimizes its own local objectives, but is affected by the actions of the other;
- d) the decision-makers are non-cooperative;
- e) presence of a dominant decision-maker;
- f) a sequential decision-making process;
- g) condition of complete information on behalf of the dominant decision-maker;
- h) dominant decision-maker communicates actions to the dominated decision-maker who makes decisions based on that information.

Traits of the bilevel system:

- a) the presence of two or more interactive decision-making units, each responsible for decisions made within a specific subsystem;
- b) each decision-maker controls a subset of the total design variables defined within the system;
- c) each decision-maker independently optimizes its own local objectives, but is affected by the actions of the other;

- d) a separate decision-maker, i.e. the coordinator, is defined who is responsible for optimization at the system level;
- e) the lower-level decision-makers are real-life decision-makers, each with a defined local objective which cannot be assigned;
- f) the system level objective cannot be made to coincide with the local objectives.

From the above description, it is clear that the two sets of traits share a common subset in addition to those uniquely-defined to each variant of the bilevel problem. The traits described above could generally be grouped into three: i) those associated with the complex system (i.e. the common traits), ii) those describing the type control present and iii) those unique to the bilevel nature within each control structure. In order to positively identify the bilevel problem, a framework is developed which proposes a step-by-step approach based on the traits consisting of i) identifying whether a problem is of a complex system, ii) if so, ascertaining the type of control structure within the complex system (i.e. whether pure decentralized or hierarchic), iii) determining the suitability of applying the Stackelberg strategy (i.e. the bilevel program). The following sections describe the three steps proposed by the framework in order to systematically identify the presence of such interactions representative of the bilevel problem.

3.1.1 Identifying a complex system

The presence of complex system could be described by the first three common traits. These traits are associated to the concepts of decomposition and coupling between the subsystems. The first trait (i.e. the presence of two or more interacting decision-making units, each responsible for decisions made within a specific

subsystem) refers to evidence of decomposition (break down of the original system) within the problem represented by the presence of multiple decision-makers within a particular problem. The second trait (i.e. decision-makers control a subset of the total design variables defined within the system) refers to devolution of control resulting from the decomposition process. In other words, none of the different decision-makers present should have authority over *all* parameters. The final trait (i.e. each decision-maker independently optimizes its own local objectives, but is affected by the actions of other decision-makers) refers to the issue of coupling. The presence of coupling signifies the influence that a particular subsystem may have with another, e.g. linear aerodynamics analysis on a wing to predict pressure distribution is coupled to a linear structural analysis to predict displacements are coupled together as pressure distributions (on the wings) are also dependent on the displacements [46]. The following example demonstrates coupling which may appear within a design problem:

Example 1 (Common variables):

In [6], Rao et al. described a problem pertaining to the design of a thin-walled pressure vessel in which the objectives are to minimize the weight and maximize the volume of the vessel. Figure 3.1 shows the general design of the pressure vessel.

Figure 3.1: Illustration of thin-walled pressure vessel (adapted from Rao et al. [6])

By decomposing the overall design problem into two separate sub-problems, i.e. problem of minimization of weight and problem of maximization of volume, separate equations were defined for each sub-problem, each representing one aspect of the design problem. Within the sub-problem of minimization of weight, the following equation was derived as the objective function to calculate the weight of the vessel:

$$\rho \left[\frac{4}{3} \pi (R+T)^3 + \pi (R+T)^2 L - \left(\frac{4}{3} \pi R^3 + \pi R^2 L \right) \right] \quad [3.1]$$

Within the sub-problem of maximization of volume, the following equation was derived as the objective function to calculate the volume of the vessel:

$$\left[\frac{4}{3} \pi R^3 + \pi R^2 L \right] \quad [3.2]$$

R – radius, T – thickness, L – length, ρ - density of the cylinder.

By observing the two equations, it is clear that the variables R and L are present in both equations. In other words, variables R and L influence both the problems of weight minimization and volume maximization of the final design. Thus R and L are coupled variables within the two objective functions.

Summary

The three common traits used here refers to the concept of decomposition (includes devolution of control and authority), presence of multiple decision-makers and coupling between subsystems, all three of which are associated with complex systems. The next step is to ascertain the type of control present within the problem/complex system.

3.1.2 Ascertain Hierarchical or Pure Decentralized Control

The trait which distinguishes between complex systems of pure decentralized control and those of hierarchic control is the presence of a higher-level decision-maker or a coordinator. This was exemplified in the Transport Network Design Problem reviewed in Section 2.2.2. The coordinator could be identified as the dedicated decision-maker responsible for the system as a whole. Additional task of the coordinator is to resolve conflicts which may arise between the lower-level decision-makers due to the coupled sub-systems (e.g. presence of the local and non-local variables). The following example describes the presence of a coordinator within a resource allocation problem:

Example 2 (Coordinator in resource allocation):

Findeisen [23] described a decentralized resource allocation problem formulated as a complex system consisting of N decomposed subsystems drawing from the same resource pool (figure 3.2). A coordinator is engaged to i) resolve the possible conflicts which may arise between coupled subsystems due to the need to share resources and, ii) steer the decisions made by all the local decision-making units towards the overall goal of the system by allocating the amount of resource to each subsystem:

Figure 3.2: Subsystems sharing the same resource pool (adapted from Findeisen [23])

The following equations represent the objective functions within the different coordinator and a local decision-making unit's sub-problems:

Local decision-making unit's objective function: –

$$y_i = F_i(c_i, u_i), i = 1, \dots, N. \quad [3.3]$$

where y_i represents the performance of the individual subsystems, c_i the local design variables and u_i the resource allocated to the subsystem i

Coordinator's objective function:-

$$Q(y_1, \dots, y_N, u_1, \dots, u_N) \quad [3.4]$$

Equation [3.3] clearly represents the natural dependence of subsystem performance on the choice of local design variables by the local decision-making unit and the resource allocated to it by the coordinator. On the other hand, equation [3.4] shows that the coordinator's objective is directly dependent on the total amount of resources used as well as all the individual performances of the subsystems. Therefore, the coupled variables here are represented through the common variables (i.e. the resource allocated) of u_i while the variables y_i signify the presence of interconnected problems (i.e. the final output of each individual local sub-problem influences the degree of performance that the coordinator could achieve).

Summary

The presence of a higher-level decision-maker or a coordinator is a distinct trait of the hierarchical control structure while its absence signifies the presence of pure decentralized control structure within the complex system. The coordinator could be identified by its tasks which involves resolving conflicts arising between the lower-level decision-maker as well as optimizing the system-level objectives.

3.1.3 Ascertain Bilevel Problem

The final step of the approach diverges into two separate paths. To ascertain a bilevel problem within the context of a pure decentralized system, the approach requires one to check for traits of the Stackelberg game to rule out the other two possible games (i.e. Nash and Pareto games). To ascertain a bilevel problem within the hierarchical system, there is a need to check for absence of agreement between the system-level and local objectives which is the main point of divergence between the traditional multi-level/hierarchical approach and the bilevel system. The two checks will be described in the following:

i) To ascertain Stackelberg Game in Pure Decentralized Complex System

The presence of non-cooperative decision-makers, a dominant decision-maker and a sequential decision-making process is required. In addition, it is crucial that the dominant decision-maker has knowledge of the dominated decision-maker's objectives and possible actions. The dominant decision-maker would also be required to communicate his actions to the dominated decision-maker. On the hand, it is required for the dominated decision-maker know the actions of the dominant decision-maker once taken and to base his decisions on it. Failure of the last two conditions, even with the presence of sequential decision-making and the dominant decision-maker, would cancel out the Stackelberg game as it would be assumed that the dominated decision-maker does not take action based on the dominant decision-maker's actions.

ii) To ascertain the absence of agreement between system-level and local objectives.

This requires that firstly the lower-level decision-makers are actual decision-makers in charge of a subsystem (e.g. a design team responsible for the structural integrity for an airplane) and not mere constructs or proxies [23]. This tow points means that there is no possible way for a system designer to manipulate the local objectives. Additionally, the system-level objective must not be able to coincide itself with the local objectives. Failure of these two conditions would result in the possible emergence of the traditional multi-level/hierarchical approach as there is a possibility of the system-level and local objectives to be in agreement. The following flowchart summarizes the three-step approach developed:

Figure 3.3: Flowchart of Three-Step Approach to Detect Bilevel Problems

3.1.4 Summary of Section

Section 3.1 describes a framework proposing a step-by-step approach in order to positively identify the emergence of a bilevel problem. Based on the common traits obtained from the two sets derived in Chapter 2, the bilevel problem was essentially characterized as a complex system with variants ascribing to either the pure decentralized or hierarchical control. The framework is based on the presence of the traits identified in Chapter 2, and advocates the following three step approach: i) identifying the presence of complex systems, ii) ascertaining the type of control (i.e. whether pure decentralized or hierarchic), iii) determining the suitability of applying the Stackelberg strategy (i.e. the bilevel program). The second half of the chapter looks into issues pertinent to resolving a bilevel problem, the mathematical representation of the Stackelberg game.

3.2 Definition and Mathematical Formulation of Bilevel Programming Problem

To reiterate, Section 2.1.2 had described the Stackelberg game (i.e. a two-person, nonzero-sum game which adopts the Stackelberg strategy) through the bilevel program. The actions within the Stackelberg game followed a sequential order, the first player (leader) declared his strategy and took action first with the second player (follower) reacting to the strategy in a rational manner. In a mathematical programming perspective, the leader problem gets the initial attention, giving value to the design variables under its control before shifting focus to the follower problem. The following set of equations [3.5] presents a generic formulation of the bilevel programming problem:

$$\begin{array}{ll}
 \min_{x \in X} F(x, y) & \leftarrow \\
 \text{subject to } G(x, y) \leq 0 & \leftarrow \\
 \text{(leader problem)} & \\
 & H(x, y) = 0 \\
 \min_{y \in Y} f(x, y) & \leftarrow \\
 \text{subject to } g(x, y) \leq 0 & \leftarrow \text{(follower} \\
 & h(x, y) = 0 \text{ problem)} \\
 & [3.5]
 \end{array}$$

where $F, f: R^n \times R^m \rightarrow R^p$, $G: R^n \times R^m \rightarrow R^p$, $H: R^n \times R^m \rightarrow R^q$, $g: R^n \times R^m \rightarrow R^r$ and $h: R^n \times R^m \rightarrow R^s$.

The bilevel programming problem is made up of two separate mathematical problems, the leader and the follower (as shown within the arrows) problems. Each problem consists of an objective function, $F(x, y)$ and $f(x, y)$ for the leader and follower problems respectively, and may or may not consist of inequality and/or equality constraint equations which define the individual feasible solution space (functions G and H for the leader problem and functions g and h for the follower's problem).

It is convenient to explicitly present the constraints in terms of feasible regions, reaction sets and inducible regions. The feasible region represents all the feasible values for the whole set of design variables, $(x, y(x))$, while reaction set represents the values which the follower would give to achieve optimality in response to a leader's initial valuation of his design variables, $y(x)$. The inducible region, on the other hand, represents the actual valuations of design variables which the leader could consider while optimizing its objective function, $(x, y(x))$. The formulation is as follows:

Equation [3.6] of feasible region of the bilevel programming problem:

$$S = \{(x, y) : x \in X, y \in Y, G(x, y) \leq 0, H(x, y) = 0, g(x, y) \leq 0, h(x, y) = 0\} \quad [3.6]$$

Equation [3.7] of follower's feasible set for each given $x \in X$:

$$S(x) = \{y \in Y : g(x, y) \leq 0, h(x, y) = 0\} \quad [3.7]$$

Equation [3.8] of follower's reaction set for given $x \in X$:

$$P(x) = \{y \in Y : y \in \arg \min [f(x, y) : y \in S(x)]\} \quad [3.8]$$

Equation [3.9] of inducible region:

$$IR = \{(x, y) : (x, y) \in S, y \in P(x)\} \quad [3.9]$$

A typical perspective of the bilevel programming problem is to view the entire follower program as part of the leader program's constraint equation thus reducing the bilevel programming problem to a single-level one. This is strongly reflected in some of the solution methodologies e.g. the reduction of the follower's program to its Lagrange condition. This effectively transforms the problem into a single-level mathematical program as shown in equation [3.10]:

$$\begin{aligned} & \min_{x \in X} F(x, y^*) \\ & \text{subject to } G(x, y^*) \leq 0 \\ & \quad y^* \in P(x) \end{aligned} \quad [3.10]$$

It is essential to note that even with the assumption of favourable conditions for all functions involved within the bilevel programming problem itself¹, the bilevel programming problem may still prove difficult to solve. The important point here is that such assumptions do not normally carry over directly to the single-level bilevel programming problem². The difficulty with solving such problem is further

¹ i.e. objective functions F and f as well as constraint functions G, g and H, h assumed to be continuous, twice differentiable functions as well as the existence of compact real number sets X and Y in eq [3.5]

² This issue will be described further in the different solution methodologies of Section 3.2.3.

compounded by the possible existence of non-singleton follower reaction set which may result in the leader not being able to achieve its optimality³. The following section describes the issue of solvability of the bilevel programming problem and the role that the non-singleton follower reaction set plays in this issue.

3.2.1 Comments on the Solvability of Bilevel Programming Problem

This section explores the main factor that affects the solvability of the bilevel programming problem; the non-singleton follower reaction set as described briefly in the prior section. It has to be mentioned that matters such as empty feasible region S and follower reaction set $P(x)$ corresponding to equations [3.6] and [3.8] respectively, will result in triviality and thus will not be considered.

The non-singleton follower reaction set

The solvability of a bilevel programming problem comes to question when the follower has a range of optimal choice of reactions to the leader's initial optimal action instead of a single optimal reaction. Typically, the leader problem will be dealt with first and thus the variables under its purview will be given value under typical rationality assumptions i.e. that the follower would react with its own optimal variable values. A typical postulation would be that each assigned set of leader variables' values corresponds to a single optimal set of reaction values for the follower variables. Thus the values for the two sets of variables (leader's and follower's) combined would result in an optimal outcome for the leader. However, such an outcome is not assured and the follower problem may instead have a set of points (each point representing unique values for the variables under the follower's control) which would result in the same optimal value for its objective function. These points

³ This applies to bilevel programs with and without favourable assumptions.

form a hyperplane with equal objective function value, thus making the follower problem indifferent to any of the points. In general, the leader does not share this indifference and requires a specific choice of values within the set of points from the follower. The following numerical example was adapted from [4, Chapter 1] and clearly illustrates the issue of non-singleton follower reaction set:

Example 1 (adapted from [4, Chapter 1])

Solve:

$$\begin{aligned} \min_x \left\{ F = (x_1, x_2) \begin{pmatrix} 2 & 3 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} : \begin{matrix} x_1 \geq 0, x_2 \geq 0 \\ x_1 + x_2 = 1 \end{matrix} \right\} & \text{Leader problem} \\ \min_y \left\{ f = (x_1, x_2) \begin{pmatrix} -1 & -4 \\ -3 & -2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} : \begin{matrix} y_1 \geq 0, y_2 \geq 0 \\ y_1 + y_2 = 1 \end{matrix} \right\} & \text{Follower problem} \end{aligned}$$

The solution y as a function x is

$$y(x) = \begin{matrix} (1, 0) & \text{for } x_1 + 3x_2 > 4x_1 + 2x_2; i.e., x_1 < \frac{1}{4} \\ y_1 + y_2 = 1 & \text{for } x_1 = \frac{1}{4} \\ (0, 1) & \text{for } x_1 > \frac{1}{4} \end{matrix}$$

By substituting these values into the leader programme, the following is obtained:

$$\begin{aligned} \min_x F = \begin{cases} 2x_1 + 4x_2 & ; x_1 < \frac{1}{4} \\ 2y_1 + \frac{3}{2} & (0 \leq y_1 \leq 1); x_1 = \frac{1}{4} \\ 3x_1 + x_2 & ; x_1 > \frac{1}{4} \end{cases} \\ \text{subject to } x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

Figure 3.4: Plot of values of x_I against F

Figure 3.4 shows the plot of the values of x_I against F in Example 1. The minimum of the objective function F for the leader is 1.5. This corresponds to the leader's variable set $x = (0.25, 0.75)$ and the follower's variable set $y = (0, 1)$. However, with $x = (0.25, 0.75)$, the follower could choose any value that satisfies $y_I + y_2 = 1$ as this would result in the same optimal value for its objective function f . From the leader's perspective, a follower's choice other than $y = (0, 1)$ is unacceptable as it will prevent it from attaining its optimal value. Thus a well-thought solution methodology for the bilevel programming problem should include measures to overcome such a situation. The following section describes two possible resolutions.

Resolving the non-singleton follower reaction set

The most apparent means of resolution to the non-singleton reaction set issue is to refer the issue to the designer. Essentially, there are two extreme outcomes specific to the condition; the follower selects the choice which would conform to the leader's initial speculation and allows the culmination of the leader's optimality or it selects the choice that renders the leader's optimality unfeasible. Both outcomes eliminate

any ambiguity and thus provide convenient assumptions for any means of resolution. The two assumptions naturally lead to two possible approaches [67].

Casteli et al [68] presented a benevolent follower attitude which translated to the follower selecting the leader's desired values when multiple choices are available. The argument utilized within their research was that the follower would no longer need to be **in competition** with the leader once its optimality is achieved hence the noncooperative nature could be relaxed. Thus the bilevel programming problem could be illustrated in the following equation set [3.11]:

$$\begin{aligned} & \min_{x \in X} F(x, y^*) \\ & \text{subject to } G(x, y^*) \leq 0 \\ & y^* \in P(x) = \left\{ y \in S(x) \left| \begin{array}{l} \min_{y \in Y} f(x, y) \\ \text{subject to } g(x, y) \leq 0 \end{array} \right. \right\} \end{aligned} \quad [3.11]$$

On the flip side, the second approach tackles the issue under the assumption that the follower would always select a choice of action detrimental to the leader. Thus the approach calls for a conservative strategy whereby the leader would always assume the worst and hence shapes his action by choosing a course of action which would lead to the best possible outcome among the worst possible⁴. The formulation is illustrated in the following equation set [3.12]:

$$\begin{aligned} & \min_{x \in X} \max_{y^*} F(x, y^*) \\ & \text{subject to } y^* \in P(x) = \left\{ y \in S(x) \left| \begin{array}{l} \min_{y \in Y} f(x, y) \\ \text{subject to } g(x, y) \leq 0 \end{array} \right. \right\} \end{aligned} \quad [3.12]$$

⁴ This strategy is similar to the minmax optimization.

To reiterate, the final decision on the follower's attitude when faced with a non-singleton reaction set is essentially up to the problem designer. Thus the final selection should be indicative of the actual problem being model and would be specific to each problem. Once decided, the actual resolution of the bilevel programming problem could take place. The following Section 3.2.3 describes a selection of the different solution methodologies available. It is essential to point out that most of these methodologies assumed a singleton reaction set to avoid insolvability.

3.2.2 Theoretical Solution Methodologies

Solution methodologies for bilevel programming problems could generally be classed under three different headings. The three headings are:

1. Solution methodologies that solve linear bilevel programming problems.
2. Solution methodologies that solve nonlinear bilevel programming problems.
3. Heuristics.

The clear separation of solution methodologies for linear and nonlinear bilevel programming problems parallels the nonlinear programming; in addition to familiar issues of non-differentiability and nonconvexity that are symptomatic of nonlinear bilevel programming problems, the linear piecewise boundary of a linear bilevel problems feasible solution space can be exploited in solution methodologies. This however does not imply that resolving the linear variants are less compelling than its nonlinear counterparts as the linear bilevel programming problem is still classed as strongly NP-hard [4, Chapter 5]. This therefore justifies the usefulness of heuristics as a possible solution. The advantage that heuristics have over the traditional analytical method is that it could be utilized for both mathematical forms, linear and nonlinear. The downside of heuristics is that it is ad-hoc and problem-specific. A more detailed

description of the approaches under the linear, non-linear and heuristics heading is available in Appendix 1. The following section will lay the fundamentals required to appreciate the solution methodologies by describing the characteristics of both linear and nonlinear variations in Sections 1 and 2 respectively.

1. Linear Bilevel Programming Problem Definition and Solutions

The linear bilevel programming problem relates to the linearity trait of the functions involved within the bilevel programming problem i.e. objective functions F and f , inequality and equality constraints G , g and H , h respectively. The following equation set [3.13] is an example of a generic linear bilevel programming problem adopted from [4, Chapter 5]:

$$\begin{aligned}
 \min_{x \in X} F(x, y) &= c_1 x + d_1 y \\
 \text{subject to } A_1 x + B_1 y &\leq b_1 \\
 \min_{y \in Y} f(x, y) &= c_2 x + d_2 y \\
 \text{subject to } A_2 x + B_2 y &\leq b_2
 \end{aligned}
 \tag{3.13}$$

where $c_1, c_2 \in R^n$, $d_1, d_2 \in R^m$, $b_1 \in R^p$, $b_2 \in R^q$, $A_1 \in R^{p \times n}$, $B_1 \in R^{p \times m}$, $A_2 \in R^{q \times n}$ and $B_2 \in R^{q \times m}$.

The sets X and Y may consist of additional constraints such as nonnegativity, integrality as well as upper and lower value limits. To reiterate, the feasible region and the rational reaction set of the follower is defined as non-empty to avoid triviality. By further assuming that the reaction set is a point-to-point mapping of each set value of the leader's design variable x to an appropriate y , the following additional remarks could be made regarding the linear bilevel programming problem:

1. The feasible region of the linear problem is located within a polytopes⁵ defined by the constraint equations.
2. The linear bilevel programming problem is equivalent to minimizing F over a feasible region comprising of a piecewise linear equality constraint (polytopes).
3. The solution to the problem lies on the vertices of the inducible region.

Solution methodologies for the linear bilevel programming problem could be classified under three different approaches. The first approach termed vertex enumeration, exploits Remark 3 by testing values corresponding to the vertices of the inducible region. The second approach is referred to as the Kuhn-Tucker approach and deals directly with the *complimentary slackness condition* that manifests itself within the Kuhn-Tucker formulation of the linear bilevel programming problem. The final penalty approach starts by assuming a free design variable condition without any constraints. It subsequently applies penalty values to the objective function when design variable values do not lie on regions defined by the constraint equations. The following section will describe the three approaches in greater detail using the linear bilevel programming problem defined above as an example.

First Approach: Vertex Enumeration

As stated in the previous section, this approach takes advantage of the uniquely shaped feasible region as defined by the linear objective functions and constraints of a linear bilevel program. The most important link between algebraic and geometric solution methodologies of a linear program lies in the formal relation between basic feasible solutions and *extreme points of polytopes*. It is suffice to state that basic feasible solution represents values to the design variables which satisfy the linear

⁵ A polytope is defined as a closed, bounded single/multi-dimensional figure whose faces are hyperplanes. A two-dimensional example would be a closed-polyhedron with its boundaries defined by the constraint equations.

constrain equation. Hence, a search for the optimal solution within a linear program requires a search through all the basic feasible solution⁶. Figure 3.5 illustrates a polytope.

Figure 3.5: A polytope with its vertices

The polytope represents the augmentation of all linear constraint equations thus forming the feasible region. The extreme point of the polytope represents a point which does not lie strictly within a single line segment; it connects two other extreme points instead. In other words, they represent the vertices of the polytope. Therefore a search of all the vertices amounts to a search through all the basic feasible solutions. Thus an optimal solution could be obtained. Examples of solution methodologies which utilize the first approach include the *Kth*-Best Algorithm in [36] which explores all the vertices of the leader problem's polytope, ranks them and determines whether the selected solution is feasible in the follower's problem.

Candler and Townsley [38] developed a solution methodology that involves solving two linear programs: 1) the leader program defined by a free x^7 value and a y^8 value associated with the optimal basis of the follower program and 2) the follower

⁶ For more details regarding the basic feasible solution and linear programming the reader is referred to [23].

⁷ leader's design variable vector.

⁸ follower's design variable vector.

program defined with a fixed x variable. Starting with a fixed x value, the method involved solving for the follower program by exploring the optimal bases of the follower's problem and obtaining an appropriate y value. Bialas and Karwan [36] utilized a ranking method known as the K th-Best Algorithm. Under the assumption that the inducible region is bounded and the rational reaction set $P(x)$ is single-valued for each value of x , the algorithm initially ranks each of the vertices (x, y) value in order of optimality for the leader's objective function. It then solves the follower's program with the x value fixed by the values within the ranking in order and determines whether the y value matches the same initial vertex check. This process is repeated with all the initial vertex values in ranking order until a match is found.

Second Approach: Kuhn-Tucker Resolution Method

This approach initially converts the linear bilevel program to a single level problem by formulating the Kuhn-Tucker conditions for the follower program. Thus a single level mathematical program with an explicit representation of the inducible region is derived. The following equation set [3.14] illustrates the formulation:

$$\begin{aligned}
 & \min c_1x + d_1y \\
 & \text{subject to } A_1x + B_1y \leq b_1 \\
 & \quad uB_2 - v = -d_2 \quad (\text{Stationarity Conditions}) \\
 & \quad u(b_2 - A_2x - B_2y) + vy = 0 \quad (\text{Complementary} \\
 & \quad \text{Slackness}) \\
 & \quad A_2x + B_2y \leq b_2
 \end{aligned}
 \tag{3.14}$$

where $x \geq 0$, $y \geq 0$, $u \geq 0$ and $v \geq 0$.

The Stationary Condition corresponds to the condition whereby there are no other feasible values that would result in an improved (lower) optimality for the follower

programme. The Complementary Slackness directly relates to the active and inactive constraints i.e. when the constraint is active (equal to zero) the dual variables (u and v) will be free to take any positive value and when the constraint is inactive (less than zero), the dual variables will equate to zero.

Although the above formulation results in a standard single-level mathematical programme, the difficulty in solving such problem lies within the Complementary Slackness terms. Thus the solution methodologies devised using this approach feature a procedure to overcome this issue. The general theme running through all the methodologies is the relaxation of the Complementary Slackness terms which would result in an easier program to solve. Fortuny-Amat & McCarl [69] converted the whole formulation into a mixed-integer linear programme by replacing the Complementary Slackness terms with a set of inequalities. Bard & Falk [70] on the other hand represented the terms as piecewise linear, separable functions. The final steps of both methodologies involve a search tree whereby each node represents the final leader objective value when either the dual variable or the constraint itself is given a nought value. The node with the minimum value represents the solution. Other examples classified under this approach include the Complementary Approach [37, 71] and Variable Elimination Algorithm [72]. Both resolve the bilevel programming problem through its Complementary Slackness terms.

Third Approach: Penalty Function Approach

This approach simplifies the linear bilevel programming problem by reducing the bilevel programme into a single-level formulation. This is achieved through exploiting the dual nature of linear programmes; more specifically the *duality gap* of the linear programme. The duality gap is defined as the difference between feasible

values of the primal and dual objectives' functions for corresponding primal and dual design variables. The duality gap is closed when the optimal value is achieved as the value is the same in both the primal and dual formulations.

In equation set [3.14], the term c_2x and A_2x form the follower's objective function and constraint equation becomes a constant once the leader has selected a value for x . Thus the follower's programme could be reduced to the following equation set [3.15]:

$$\begin{aligned} \min d_2y \\ \text{subject to } B_2y \leq b_2 - A_2x \end{aligned} \quad [3.15]$$

The dual formulation of equation set [3.15] is shown below in equation set [3.16]:

$$\begin{aligned} \max_u u(A_2x - b_2) \\ \text{subject to } uB_2 \geq -d_2 \\ u \geq 0 \end{aligned} \quad [3.16]$$

The duality gap defined as the difference between the two objective functions defined above is a function of design variables x , y and u , $\Delta(x, y, u) = d_2y - u(A_2x - b_2)$.

Anandalingam & White [73] augmented the duality gap function weighted down by a large positive constant K on the leader's objective function thus creating a single-level programme. The formulation is illustrated below in equation set [3.17]:

$$\begin{aligned} P(K) = \min \hat{F} &= c_1x + d_1y + K[d_2y - u(A_2x - b_2)] \\ \text{subject to } A_1x + B_1y &\leq b_1 \\ uB_2 &\geq -d_2 \\ A_2x + B_2y &\leq b_2 \\ x, y, u &\geq 0 \end{aligned} \quad [3.17]$$

Under the assumption that the feasible region for (x, y) and u , S and U , are nonempty bounded polyhedra, the optimal solution to the above problem is located at the extreme points of the two space.

The following section will describe the nonlinear bilevel programming problem.

2 Nonlinear Bilevel Programming Problem Definition and Solutions

Nonlinear bilevel programming problems represent bilevel programmes consisting of all of its functions defined as either nonlinear or a combination of linear and nonlinear. One example is the convex-quadratic bilevel programming problem which has a quadratic objective function for both the leader and follower problems as well as a convex feasible region usually defined by linear equality and inequality constraint equations. The solution methodologies involved with solving nonlinear bilevel programs can be classified into three approaches [74]:

1. Kuhn-Tucker Conditions for follower problem.
2. Double-penalty.
3. Steepest Descent.

The Kuhn-Tucker method reduces the bilevel programming problem into a single level problem by replacing the follower problem with its Kuhn-Tucker conditions. The main difficulty with this methodology lies with the complementary term which makes the constraints of the single-level problem difficult to solve. The double-penalty method on the other hand augments both the follower problem and the leader constraints into a single function with the leader's objective function. Although such a neat methodology is attractive, the slow convergence towards a solution is a major drawback. The steepest descent approach converts the bilevel programming problem into a single level nonlinear programming problem and assumes an iterative process from an initial point within the feasible space towards the optimal point i.e. a point which satisfies the necessary optimal conditions. The algorithm utilizing the bundle method is currently the favoured approach as it tackles the non-differentiability issue for $y(x)$ by taking advantage of its Lipschitz compliancy although global optimality is

not guaranteed. The following section will describe certain heuristics to resolve the bilevel programming problem.

3 Solving Bilevel Programming Problems through Heuristics

Heuristics offer an attractive alternative to the approaches reviewed. Examples of heuristics which had been implemented for solving bilevel programming problems include genetic algorithm (GA) [75] and simulated annealing [76]. The following section will provide an example based on an implementation of GA as a means for resolving bilevel programming. It is essential to note that heuristics-based approaches are generally unique to the different practical problems it seek to resolve hence the work reviewed here seek to demonstrate the possibility of its implementation as a means of resolving bilevel programming problems rather than as a concrete guide.

Genetic Algorithm Based Bilevel programming Algorithm (GABBA) (Adapted from [4, Chapter 9])

The GABBA is used to solve linear bilevel programming problems defined in the following equation set [3.18]:

$$\begin{aligned} \min_{x \geq 0} F(x, y) &= c_1 x + d_1 y \\ \min_{y \geq 0} f(x, y) &= d_2 y \\ \text{subject to } g(x, y) &= Ax + By - b \leq 0 \end{aligned} \quad [3.18]$$

The main difference between GABBA and traditional genetic algorithm based method is that it does not utilize the crossover operator. Instead, it relies on the extensive use of the mutation operator as well as injection of random values to ensure variety and change within its population. The underlying ethos of GABBA is that the GA-associated processes are implemented on the leader's problem; although its chromosomes consist of the leader's variables' values x and follower's variables y , its

fitness function is based on the leader's objective function and the mutation and random generation processes are only executed on the x allele of the chromosome. The follower's problem is relegated to a secondary role; for every x generated randomly or through the mutation process at each generation, the follower's problem is solved through conventional simplex method to obtain its optimality y . Feasible chromosomes are then stored within a database for selection to the next generation. The solution could be obtained typically at the end of a user-defined number of generations or when they satisfy certain non-improving characteristics i.e. $F(x^{k+1}, y^{k+1}) - F(x^k, y^k) \leq \varepsilon$ where k represents the number of generations and ε represents a small value.

3.2.3 Section Summarization

This Section covers the possible issues which might arise during a practical implementation of a bilevel programming problem. The Section 3.2.1 initially highlights the general problem of non-singleton follower reaction set to a leader's action and proposed the two methods of resolving the issue under the different attitudes of the follower. The focus is then shifted to the solution methodologies used to resolve bilevel programming problems in Section 3.2.3. These solution methodologies could generally be divided into two classes namely analytical and heuristics. The analytical method could be further divided into those approaches which solve linear bilevel programs and those which solve nonlinear bilevel programs.

3.3 Chapter Summarization

The first part of Chapter 3 formally characterized problems of a bilevel nature based on the traits established in Chapter 2. It begins by re-stating the essential traits which are the tell-tale signs of the presence of a bilevel problem. Based on these traits, a framework for identifying problems of a bilevel nature was developed. The framework attempts to establish a three step approach to detect problems of a bilevel nature: i) identifying the presence of complex systems, ii) ascertaining the type of control (i.e. whether pure decentralized or hierarchic), iii) determining the suitability of applying the Stackelberg strategy (i.e. the bilevel program).

The second half covers the main issues pertaining to the mathematical resolution of the bilevel program used to model the Stackelberg game and bilevel problem. The first part introduces the mathematical formulation of the bilevel program in detail. This is followed by a description of the non-singleton follower reaction set and its implications on the solvability of the bilevel program as a whole. It then proposes two possible answers to the issue; the first assumes that the follower, given a choice of more than one possible solution, would be ambivalent and select values which would be better for the leader while the alternative is to assume the exact opposite (i.e. follower would, on purpose, select values which results in the greatest detriment for the leader). The final part of the chapter provides a short summary of the different approaches to resolve the different variants of the bilevel programming problem. This provides suitable background for the next chapter in which the solution approach for the bilevel problem is described.

Chapter 4

Solution Algorithm

The focus of this chapter is on the solution methodology adopted within this research work to solve the bilevel program. In summary, Chapters 1 had articulated the need to accurately identify and interpret traits inherent within a large and complex problem and had expressed the intention of the present research work to focus on those related to the bilevel problem. Chapter 2 had identified the common traits present within problems of a bilevel nature while Chapter 3 had attempted to formally characterize the bilevel problem based on the traits. In addition, Chapter 3 had described in greater detail the mathematical structure and formulation of a bilevel program, reviewed a number of contemporary analytical and heuristical solution methodologies available as well as highlighted the major pitfall involved in having the non-singleton follower reaction set. The next step is to check the veracity of the claims made in the earlier chapters by resolving a practical example of a bilevel problem (i.e. the multi-level inventory problem). The current Chapter describes the algorithm developed in order to resolve the multi-echelon inventory problem and decentralized capacitated plant selection problem case studies in the following chapter. The algorithm developed was implemented within the Matlab v.6.5 solver environment to facilitate the process of defining of functions made up of more complex mathematical functions (e.g. exponential, factorial). Section 4.1 describes the main framework followed by a description of the algorithm coupled with the approach to solve the problem in Section 4.2. The final Section 4.3 summarizes the main points of the Chapter.

4.1 Overall Framework

Figure 4.1: Overall System Architecture

Figure 4.1 shows the main components of the overall system architecture and how they interact with each other. As stated in the prior opening section of the chapter, the Matlab software was used to facilitate defining complex (sub-)functions. The first part of the framework consists of the generic user interface provided within the Matlab environment. While execution would take place within the command window, each problem requires an initialization process in which the user needs to pre-define the following and store them within a script or an m-file:

1. The leader/follower problem objective function.
2. The leader/follower problem constraint functions.
3. The leader/follower problem design variables.
4. The limits of each design variable for both leader and follower problems.
5. The search step of each design variable value.

6. Define the values for all constant parameters required during the run.

The second part of the framework, the application module, consists of the pre-defined modules within Matlab software. The task of this part of the framework is: i) to calculate the objective function values for both the leader and follower as defined earlier within the script files, ii) to check whether the input design variable data satisfies the constraint functions, iii) output intermediary data into the external storage files and iv) output the final solution on the command screen for viewing and an external file for storage. The following section describes the general algorithm used to solve the different problems.

4.2 Description of Approach and Algorithm to Resolve the Bilevel Problem

The approach undertaken consists of two main parts. The first part consists of generating the optimal follower objective function values based on the potential values of the leader-defined design variables. This consists of first generating the feasible leader design variable values (utilizing the step-size defined within the initialization period) which are within the feasible space as defined by the constraint functions. This is followed by the generation of the follower objective function values for each feasible leader design variable value. The optimal values for each leader design variable value is obtained for the follower's objective function and stored within the external intermediary file (i.e. both the design variables as well as the objective function values). This process is repeated until all feasible leader design variable values are defined.

The next step consists of calculating the optimal leader objective function value. The necessary follower-defined design variable values are obtained from the external storage created in the earlier step. Each leader-defined design variable values act as a

key to ascertain the appropriate values for the follower design variable values. This process is repeated until the optimal leader objective function value is obtained. The following flowchart describes the two-stepped process.

Figure 4.2: Flowchart of Overall Approach

A more detailed description specific to the multi-echelon inventory problem and decentralized capacitated plant selection problem could be found in the Appendix.

4.3 Chapter Summarization

This chapter highlights the main elements and processes of the approach developed to tackle the multi-echelon inventory bilevel problem. Section 4.1 describes the main components within the framework as defined within the Matlab software while Section 4.2 describes the main approach of to resolve the problem through a search method. The following chapter will describe the case studies undertaken.

Chapter 5

Case Study

This Chapter seeks to verify the approach put forward in Chapter 3, i.e. to identify the problem as bilevel and nature, and to subsequently resolve it through the algorithm described in Chapter 4. To reiterate, the three-stepped approach described in Chapter 3 to identify the bilevel problem is based on the two sets of traits derived in Chapter 2, one for the Stackelberg game within the context of the pure decentralized complex system and the other for the bilevel system within the hierarchical control system. The focus of the case study is on the multi-echelon inventory problem first broached by Clark & Scarf in [11] as well as the decentralized capacitated plant selection problem based on Cao & Chen's work in [22]. For each problem, this Chapter will first attempt to verify the presence of a bilevel nature before resolving them. Section 5.1 describes the multi-echelon problem. Section 5.1.1 will go through the step-by-step framework proposed in Section 3.1 to show evidence of a bilevel problem. Section 5.1.2 will display and discuss the results obtained providing comparative results from other relevant works. Section 5.2 will describe the decentralized capacitated plant selection problem while Section 5.2.1 utilizes the approach developed to show that such a problem is of a bilevel nature. This is followed by a discussion of the results obtained in Section 5.2.2. Section 5.3 will summarize the main points of the chapter.

5.1 Description of the multi-echelon inventory problem

The problem under investigation in this case study is based on the works of Axasäter in [19] and [77]. The problem describes an arborescent system consisting of a single warehouse supplying four separate retailers as shown in Figure 5.1. The

warehouse is effectively considered as a single inventory problem requiring its own optimization. The retailers are assumed to operate under similar specifications and would therefore implement similar optimal policies. In other words, it is only necessary to consider for one retailer's inventory problem; multiplying the results by four to obtain total cost accrued at all four installations. The problem describes the warehouse being able to optimize its cost functions and declare its optimal policy (consisting of transportation and inventory policies) before the retailers. The following describes the customer demand, inventory policy and cost functions.

Figure 5.1: The arborescent system (adapted from [54])

Customer Demand

Each retailer faces a stationary and independent Poisson demand. Unfilled demand is backordered thus incurring backorder costs.

Inventory Policy

The inventory policy adopted for both installations is the one-for-one ($s-1, s$) policy in which stock reviewed continuously and for any value below optimal (i.e. s), order

would be immediately placed to the warehouse. Orders delayed due to shortage within the warehouses are ultimately delivered on a first-come-first-serve basis.

Cost Functions

Both the holding cost (i.e. the cost of storing each unit in inventory) and backorder cost (i.e. cost of failing to deliver to customer when demanded) are linearly dependent on per unit of inventory. Although demand only occurs at the retailer installation, the warehouse is made to assume a portion of the backorder cost. Such a policy was also adopted by Cachon & Zipkin in [17]. The mathematical formulation of the retailer and warehouse cost functions are defined as follows:

Warehouse's cost function:

$$\begin{aligned}
 & \min_{L_0, S_0} c_0(L_0, S_0, L_1, S_1) \\
 & = \alpha_0(L_0) + \lambda_0 h_0 L_0 + \lambda_0 \gamma(L_0, S_0, L_1, S_1) \\
 & + \lambda_i \int_0^{L_0} \frac{\lambda_0^{S_0} (L_0 - t)^{S_0-1}}{(S_0 - 1)!} e^{-\lambda_0(L_0-t)} \left(e^{-\lambda_i(L_i+t)} \frac{(1-\varphi)\beta_i}{\lambda_i} \sum_{k=0}^{S_i-1} \frac{(S_i-k)}{k!} (L_i+t)^k \lambda_i^k + (1-\varphi)\beta_i (L_i+t-S_i/\lambda_i) \right) dt \\
 & + \sum_{k=0}^{S_0-1} \frac{\lambda_0^k L_0^k}{k!} e^{-\lambda_0 L_0} \left[e^{-\lambda_i L_i} \frac{(1-\varphi)\beta_i}{\lambda_i} \sum_{k=0}^{S_i-1} \frac{(S_i-k)}{k!} L_i^k \lambda_i^k + (1-\varphi)\beta_i (L_i - S_i/\lambda_i) \right] \quad [5.1]
 \end{aligned}$$

Where $\gamma(L_0, S_0)$

$$= h_0 \left(e^{-L_0 \lambda_0} \sum_{k=0}^{S_0-1} \frac{(S_0-k)}{k!} (L_0 \lambda_0)^k \right) \quad [5.2]$$

Retailer's cost function:

$$\begin{aligned}
 & \min_{L_i, S_i} c_i(L_0, S_0, L_i, S_i) \\
 & = \alpha_i(L_i) + \lambda_i h_i L_i + \lambda_i \overset{\sim}{\Pi}_i^{L_i S_i}(L_0, S_0) \quad i=1-4 \quad [5.3]
 \end{aligned}$$

where $\tilde{\Pi}_i^{L_i S_i}(L_0, S_0) =$

$$\int_0^{L_0} \frac{\lambda_0^{S_0} (L_0 - t)^{S_0-1}}{(S_0 - 1)!} e^{-\lambda_0(L_0 - t)} \left(e^{-\lambda_i(L_i + t)} \frac{h_i + \varphi \beta_i}{\lambda_i} \sum_{k=0}^{S_i-1} \frac{(S_i - k)}{k!} (L_i + t)^k \lambda_i^k + \varphi \beta_i (L_i + t - S_i / \lambda_i) \right) dt$$

$$+ \sum_{k=0}^{S_0-1} \frac{\lambda_0^k L_0^k}{k!} e^{-\lambda_0 L_0} \left[e^{-\lambda_i L_i} \frac{h_i + \varphi \beta_i}{\lambda_i} \sum_{k=0}^{S_i-1} \frac{(S_i - k)}{k!} L_i^k \lambda_i^k + \varphi \beta_i (L_i - S_i / \lambda_i) \right] \quad [5.4]$$

where $0 \leq \varphi \leq 1$ represents the portion of the backorder cost still borne by the retailer, $\alpha_0(\square)$ and $\alpha_i(\square)$ represent warehouse and retailer transportation costs and, $\lambda_0 h_0 L_0$ and $\lambda_i h_i L_i$ represent the holding costs during transportations.

Other parameters described above include:

S_0 - inventory position of supplier (only discrete values),

S_i - inventory position of retailer i (only discrete values),

N - number of retailers,

L_i - transportation time for an item to arrive at retailer i from supplier,

L_0 - transportation time for an item to arrive at supplier from outside source,

λ_i - the demand intensity at retailer i ,

$\lambda_0 - \lambda_0 = \sum_{i=1}^N \lambda_i$ the demand intensity at supplier,

h_i - holding cost per unit and time unit at retailer i ,

h_0 - holding cost per unit and time unit at supplier,

β_i - shortage cost per unit and unit time at retailer i ,

The following section shows that the multi-echelon inventory problem is of a bilevel nature utilizing the three stepped approach proposed in Chapter 3.

5.1.1 Evidence of Bilevel Problem in the Multi-Echelon Inventory Problem

To reiterate, the proposed approach to identify bilevel problems described in Section 3.1 utilizes three steps consisting of: i) identifying the presence of a complex system, ii) ascertaining the type of control at work and iii) determining the presence

of a Stackelberg game or bilevel system. The three steps to identify the multi-echelon inventory problem as a bilevel problem will be described below:

Step 1: Identifying the presence of a complex system

This entails the positive presence of the following traits:

- i) Each decision-maker controls a subset of the total design variables defined within the system;**

This condition is represented by the fact that both the warehouse and retailer managers/ decision-makers control only their individual inventory and transport policies (i.e. (L_0, S_0) for the warehouse and (L_1, S_1) for the retailer).

- ii) Each decision-maker independently optimizes its local objectives, but is affected by the actions of others;**

The first part of this condition is reflected within the problem by the warehouse and retailer manager's desire to minimize only their individual costs. The second part is shown in the way the warehouse's decision affects the retailer due to possibility of backlog orders at the warehouse due to inefficient policies and vice-versa due to the partial backorder charge passed on to the warehouse by the retailer.

Step 2: Ascertaining the type of control at work

The multi-echelon problem could be described as a pure decentralized complex system due to the absence of a coordinator or any decision-maker concerned with the performance or cost-effectiveness of the *entire* supply chain line.

Step 3: Determining the presence of a Stackelberg game.

This entails the positive presence of the following traits:

i) Non-cooperation between the decision-makers;

The condition for non-cooperation is reflected by the desire of both the warehouse and retail manager to only minimize their individual cost without accepting any compromise. Either decision-maker would optimize its objectives even if such actions would prove detrimental to the other.

ii) Presence of dominant decision-maker;

This is evident in the warehouse manager who optimizes its objectives and decides on its policies prior to the retailer.

iii) Sequential decision-making;

The decision-making process adopted for the multi-echelon problem is a sequential decision-making process with the warehouse manager establishing its optimal policies first.

iv) Condition of complete information on behalf of the dominant decision-maker;

Although not explicitly stated, the original problem in [19] implicitly assumes that the warehouse manager would take into account the retailer's reaction while deciding on his optimal policy. Moreover, an approach was developed in the original work in which the potential cost incurred by the warehouse for changing its optimal policies was proportional to the potential delay which would occur due to the change. Thus, this system allows the warehouse manager to estimate the potential cost he would incur from the retailer's reaction.

viii) Dominant decision-maker communicates actions to dominated decision-maker who makes decisions based on that information.

This condition is reflected by the retailer basing his action on the observed decisions of the warehouse manager.

Summary

Overall, the multi-echelon inventory problem satisfies the requirements of a bilevel problem based on the three steps. The following section will show the results obtain and discuss its significance.

5.1.2 Results and Discussions

Two case studies were undertaken. The first seek to compare the performance of the cost functions proposed in this thesis and the cost structure in Axasäter's original work in [19]. The second case study seeks to find the optimal value for φ , i.e. the ratio of backorder cost to be paid by the retailer. Both case studies utilize the algorithm described in Section 4.2, implemented using Matlab 6.5 platform.

Case Study 1

The first case study compares the result obtained under circumstances that are similar to those found in an example conducted by Axasäter in [19]. The main point of difference is the modified cost structure of the warehouse manager which is defined within this case study to be *directly* affected by any backorder cost incurred by the retailer. Such a cost structure is similar to the one proposed by Cachon & Zipkin in [17]. The aim is to compare the results obtained by Axasäter using the cost structure in [19] with that of the modified version defined earlier in Section 5.1.

i) Parameters and Constants

Transportation cost – the warehouse has a choice between $\alpha_0(1) = 3$ or $\alpha_0(0.8) = 4$.

- the retailer has a choice between $\alpha_i(1) = 3$ or $\alpha_i(1.5) = 2$.

Holding cost – $h_0 = h_i = 1$ (i.e. it costs the same to store the item either in the warehouse or retailer).

Backorder cost – $\beta = 5$.

Demand intensities - $\lambda_i = 1, \lambda_0 = \sum_{i=1}^4 \lambda_i = 4$.

Ratio of backorder - $\varphi = 0.5$ (i.e. the warehouse and retailer shares the backorder cost equally)

ii) Results

Results using cost functions as defined within Section 5.1:

The bilevel optimal results obtained are as follows-

Warehouse optimal policy $(L_0, S_0) = (1, 2)$, Cost incurred $c_0 = 7.7045$.

Retailer optimal policy $(L_i, S_i) = (1.5, 3)$, Cost incurred $c_i = 5.3050$.

Total Cost - $c_0 + \sum_{i=1}^4 c_i = 7.7045 + 4 \times 5.3050 = 28.9246$.

The global/Pareto optimal results obtained are as follows-

Global optimal policy $(L_0, S_0, L_1, S_1, L_2, S_2, L_3, S_3, L_4, S_4) = (1, 2, 1.5, 3, 1.5, 3, 1.5, 3, 1.5, 3)$

Warehouse cost = 7.0183, Retailer cost = 4.9835,

Total cost = 7.0183 + 4 × 4.9835 = 28.9246.

Ratio of total cost under bilevel optimization and total cost under global optimality:

$28.9246/28.9246 = 1$.

Results in [11]:

The bilevel optimal results obtained are as follows-

Warehouse optimal policy $(L_0, S_0) = (0.8, 3)$, Cost incurred $c_0 = 7.7635$.

Retailer optimal policy $(L_i, S_i) = (1.5, 3)$, Cost incurred $c_i = 5.6410$.

Total Cost - $c_0 + \sum_{i=1}^4 c_i = 7.7635 + 4 \times 5.6410 = 30.327$.

The global/Pareto optimal results obtained are as follows-

Global optimal policy $(L_0, S_0, L_1, S_1, L_2, S_2, L_3, S_3, L_4, S_4) = (1, 3, 1.5, 3, 1.5, 3, 1.5, 3, 1.5, 3)$

Total cost=30.301.

Ratio of total cost under bilevel optimization and total cost under global optimality:

$30.327/30.301 = 1.0009$.

iii) Discussion

From a direct comparison, both the global optimal total cost and bilevel optimal total cost obtained here are better than those obtained in [54]. Percentage-wise, the global (i.e. Pareto-optimal) and bilevel optimal total costs were 4.5% and 4.6% lower respectively. Comparing the bilevel optimal results head-to-head, both the warehouse and retailer attained more optimal costs (0.8% and 6% respectively). Different optimal policies were also obtained here and in [19]. The optimal warehouse policy was $(L_0, S_0) = (1, 2)$ here and $(0.8, 3)$ while the optimal retailer policy remained the same. The global optimal policy differed from the one obtained in [19] in the warehouse optimal policy only as well $((1, 2)$ compared to $(1, 3)$).

From the results obtained, it could be argued that from a system perspective, i.e. the entire supply chain line, the modified cost structure proposed within Section 5.1 provides a better option as the results cost incurred within a bilevel formulation is equivalent to the one which is most beneficial for the supply chain line. Thus it could be argued that the modified cost was able to account for the non-cooperative nature of

the two decision-makers resulting in overall optimization within a bilevel decision structure.

Case Study 2

The second case study extends the previous one by attempting to find the optimal value of the ratio of backorder cost borne by the retailer φ . The values utilized are from 0 to 1 in increments of 0.1. Table 5.1 shows the results when $h_o=1$ and $h_i=1$ while Table 5.2 shows the results when $h_o=1$ and $h_i=2$. A higher holding cost at the retailer is more reflective of the situation in reality as normally the costs associated with holding costs (i.e. space, rental, opportunity cost) are higher at the retailer than at the warehouse.

i) Parameters and Constants

Transportation cost – the warehouse has a choice between $\alpha_o(1) = 3$ or $\alpha_o(0.8) = 4$.

- the retailer has a choice between $\alpha_i(1) = 3$ or $\alpha_i(1.5) = 2$.

Holding cost – $h_o = h_i = 1$ and $h_o = 1, h_i = 2$.

Backorder cost – $\beta = 5$.

Demand intensities - $\lambda_i = 1, \lambda_o = \sum_{i=1}^4 \lambda_i = 4$.

Ratio of backorder - $0 \leq \varphi \leq 1$ (i.e. the warehouse and retailer shares the backorder cost equally)

ii) Results

Table 5.1: Bilevel and Global Optimal Results with $0 \leq \phi \leq 1$ and $h_0=1, h_i=1$

Bilevel Optimal Solutions					Global Optimal Solutions						
Warehouse			Retailer		Total Cost $c_0 + 4 \times c_i$	Warehouse		Retailer		Overall	
Optimal Policy (L_0, S_0)	Cost c_0	Optimal Policy (L_i, S_i)	Cost c_i	Optimal Policy (L_0, S_0, L_i, S_i)		Cost c_0	Optimal Policy (L_0, S_0, L_i, S_i)	Cost c_i	Optimal Policy (L_0, S_0, L_i, S_i)	Cost	
ϕ											
0	1.0,4	12.200	1.5,1	3.688	26.952	1.0,1,1.0,10	7.018	1.0,1,1.5,1	3.608	1.0,2,1.5,2	26.119
0.1	1.0,4	11.758	1.5,1	4.130	28.278	1.0,1,1.0,10	7.018	0.8,10,1.5,1	4.085	1.0,3,1.5,2	26.823
0.2	1.0,3	9.185	1.5,2	4.582	27.512	1.0,1,1.0,10	7.018	0.8,10,1.5,1	4.446	1.0,3,1.5,2	27.512
0.3	1.0,3	8.956	1.5,2	4.811	28.201	1.0,1,1.0,10	7.018	0.8,10,1.5,2	4.702	0.8,3,1.5,2	28.154
0.4	1.0,1	7.967	1.5,3	5.194	28.743	1.0,1,1.0,10	7.018	0.8,10,1.5,2	4.843	1.0,2,1.5,3	28.568
0.5	1.0,2	7.705	1.5,3	5.305	28.925	1.0,1,1.0,10	7.018	0.8,10,1.5,2	4.984	1.0,2,1.5,3	28.925
0.6	1.0,2	7.586	1.5,3	5.424	29.281	1.0,1,1.0,10	7.018	0.8,10,1.5,2	5.124	1.0,3,1.5,3	29.226
0.7	1.0,2	7.467	1.5,3	5.543	29.638	1.0,1,1.0,10	7.018	0.8,10,1.5,2	5.265	1.0,3,1.5,3	29.495
0.8	1.0,1	7.335	1.5,3	5.827	30.640	1.0,1,1.0,10	7.018	0.8,10,1.5,2	5.405	1.0,3,1.5,3	29.763
0.9	1.0,1	7.080	1.5,4	5.920	30.761	1.0,1,1.0,10	7.018	0.8,10,1.5,3	5.494	1.0,3,1.5,3	30.032
1	1.0,1	7.018	1.5,4	5.982	30.945	1.0,1,1.0,1	7.018	0.8,10,1.5,3	5.389	1.0,3,1.5,3	30.301

corresponding cost incurred while columns 4 and 5 shows the retailer's. The total cost under the bilevel problem is shown in column 6. Columns 7 and 8 show the global optimal policies and cost for the warehouse while columns 9 and 10 shows the retailer's. The final two columns show the optimal overall policy (where global optimal total cost is incurred) and cost incurred.

The results obtained in Table 5.1 shows that the best bilevel result (i.e. most optimal warehouse cost) obtained is when $\varphi=1$ with the warehouse assuming a policy of $(L_0, S_0) = (1, 1)$ while the retailer's optimal policy would then be $(L_i, S_i) = (1.5, 4)$. It could be seen that the result would coincide with the best possible result that the warehouse could attain as well. This comes as no surprise as when $\varphi=1$, retailer will assume all responsibility for backorder costs, the most advantageous situation to the warehouse. However, the total cost incurred of 30.945 is 2.1% higher than the global optimal overall cost of 30.301. In terms of total cost, the best bilevel and global optimal cost occurs when $\varphi=0$ with a bilevel cost of 26.952 and a global optimal cost of 26.119 (bilevel cost is 3.1% higher). $\varphi=0$ also shows the best retailer results with the bilevel cost attaining 3.688 and the global optimal result showing 3.608. These results (i.e. best total and retailers cost) correspond to $\varphi=0$ or when the retailer absolves itself of all backorder cost and the warehouse takes responsibility for all of it.

Generally, the results in Table 5.2 when $h_i=2$ or twice the holding costs at the warehouse reflect a similar trend with a general increase in cost across the board. The best bilevel result (i.e. most optimal warehouse cost) obtained is when $\varphi=1$ with the warehouse assuming a policy of $(L_0, S_0) = (1, 1)$ while the retailer's optimal policy

would then be $(L_i, S_i) = (1, 2)$. As before, the result would coincide with the best possible result that the warehouse could attain (i.e. retailer will assume all responsibility for backorder costs). The total cost incurred of 40.531 is 3.9% higher than the global optimal overall cost of 39.001, which also represents the highest overall cost attained for the bilevel problem. In terms of total cost, the best bilevel and global optimal cost coincide with $\varphi=0$ with a bilevel cost of 33.705 and a global optimal cost of 32.577 (bilevel cost is 3.4% higher). $\varphi=0$ also shows the best retailer results with the bilevel cost attaining 5.377 and the global optimal result showing 5.216. These results (i.e. best total and retailers cost) correspond to $\varphi=0$ or when the retailer absolves itself of all backorder cost and the warehouse takes responsibility for all of it.

In short, the best possible *overall* costs (in both the bilevel and global optimal problems) correspond to when the warehouse would take responsibility of all the backorder costs. In addition, such an arrangement would result in the warehouse taking responsibility for most of the cost and using the most expensive policy (i.e. most expensive transportation mode and high replenishment point). This is irregardless of the difference in holding costs. Unsurprisingly, the best bilevel result which corresponds to the best results that the warehouse could attain occurs when the retailer takes responsibility for all of the backorder costs. It is worthy to mention that when the holding costs are equal, it is possible to attain equivalent total costs under both bilevel and global optimal conditions when $\varphi=0.2$ and 0.5 . The significance of this result is that under such agreements (i.e. pertaining to the division in responsibility for the backorder cost), the seemingly noncooperative decisions made under the bilevel conditions would not diverge from the global optimal results. Thus,

should an independent expert be called to look into optimizing the distribution system by the retailer and warehouse, it would be wise for him to subscribe to such a policy as $\varphi=0.2$ or 0.5 as these would result in independent decisions by either party to attain the most optimal total costs.

From a direct comparison, both the global optimal total cost and bilevel optimal total cost obtained here are better than those obtained in [19]. Percentage-wise, the global and bilevel optimal total costs were 4.5% and 4.6% lower respectively. Comparing the bilevel optimal results head-to-head, both the warehouse and retailer attained more optimal costs (0.8% and 6% respectively). Different optimal policies were also obtained here and in [19]. The optimal warehouse policy was $(L_0, S_0) = (1, 2)$ here and $(0.8, 3)$ while the optimal retailer policy remained the same. The global optimal policy differed from the one obtained in [19] in the warehouse optimal policy only as well $((1, 2)$ compared to $(1, 3))$.

Summary

The differing results were a direct result of the differing cost functions utilized. While the cost function defined a penalty for the warehouse based on partial responsibility over the backorder cost, Axasäter modeled his as a nonlinear function of delay on retailer's delivery caused by a change in warehouse policy (i.e. transportation and inventory policies). Arguably, this could be interpreted for having a warehouse penalty based on backorder costs rather than on delay although situations in real-life may prove otherwise. Another possible interpretation is that cost incurred in *changing* policies is higher than in implementing a new policy. It is essential to remember that the cost structure proposed by Axasäter assumes that the distribution system has already been established and that the cost function proposed may be

interpreted as calculating the cost of changing of changing from one policy to another. Additionally, the results obtained from both case studies hint at the possibility of implementing certain policies (i.e. the sharing of backorder costs) which could result in global optimality event within a prevalent bilevel decision-making structure. This information could prove useful for a decision-maker looking to optimize the entire supply chain line (i.e. a system-level coordinator).

5.2 Description of the Decentralized Capacitated Plant Selection Problem

The problem defines two independent decision-makers within the formulation of the decentralized Capacitated Plant Selection Problem representing: i) the principal firm who decides on which plants to open based on an overall business plan and, ii) the auxiliary plants which decides on the individual production level at each auxiliary plant. The objective of the principal firm is to select the plants which would represent minimal cost in terms of fixed costs from opening a new plant as well as opportunity cost incurred due to the different plants operating at below optimal capacity while fulfilling the demand for the different components. The cost function of the auxiliary plants consists of the cost associated with producing a particular quantity of components and the cost of transporting the said component to the principal firm. Although the principal firm has the authority to decide on which plants to open, it does not control the operating level of each of the auxiliary plants (which falls under the purview of individual auxiliary plant manager). The principal plant would therefore have to consider the auxiliary plants combined reactions in terms of production levels as this would have an effect on its cost minimization objective.

Cost Functions

The mathematical formulation of the retailer and warehouse cost functions are defined as follows:

Principal firm's cost function:

$$\min_{y \in S_{xy}} z_1 = \sum_{i=1}^m f_i y_i + \sum_{i=1}^m p_i \left(\text{Cap}_i y_i - \sum_{j \in \text{IS}_i} d_j a_{ij} x_{ij} \right) \quad [5.5]$$

Aggregate auxiliary firms' cost function:

$$\min_{x \in S_{xy}} z_2 = \sum_{i=1}^m w_i \left(\sum_{j \in \text{IS}_i} d_j a_{ij} x_{ij} \right) + \sum_{i=1}^m \sum_{j \in \text{IS}_i} d_j a_{ij} x_{ij} \quad [5.6]$$

$$\text{Constraints: } \sum_{i \in \text{JS}_j} x_{ij} = 1, j = 1, \dots, n, \quad [5.7]$$

$$\sum_{j \in \text{IS}_i} d_j a_{ij} x_{ij} \leq \text{Cap}_i y_i, i = 1, \dots, m, \quad [5.8]$$

$$\sum_{j \in \text{IS}_i} x_{ij} \leq n y_i \quad [5.9]$$

$$x_{ij} \geq 0, y_i \in \{0, 1\}, i = 1, \dots, m, j \in \text{IS}_i \quad [5.10]$$

m - number of potential plants

n - number of product types

p_i - opportunity cost for unused production capacity of plant i after opened

d_j - customer demand of product j

a_{ij} - capacity consumption ratio for product j in plant i

w_i - cost to use production capacity in plant i

Cap_i - available production capacity in plant i

IS_i - the group of products that can be produced in plant i

JS_j - set of plants that can produce product j

R_{ij} - transportation cost for transferring product j from the principal firm to plant i

f_i - opening cost for plant i

5.2.1 Evidence of Bilevel Problem in the Decentralized Capacitated Plant Selection Problem

Step 1: Identifying the presence of a complex system

This entails the positive presence of the following traits:

- i) Each decision-maker controls a subset of the total design variables defined within the system;**

This condition is represented by the fact that the principal firm only decides on which factory to be opened while the auxiliary plant manager would decide on the production level.

- ii) Each decision-maker independently optimizes its local objectives, but is affected by the actions of others;**

This first part of this condition is reflected within the problem by the principal firm's and auxiliary plant manager's desire to minimize their individual costs (i.e. the firm wishes to minimize cost incurred in hiring all the plants to achieve the level of production desired while plant managers desire to reduce operational running costs during production). The second part is shown in the way the principal firm's decision affects the auxiliary plant manager (the auxiliary plants would only be in production if chosen by the principal firm) and vice-versa (the level of production decided by the auxiliary plant manager could potentially effect the cost incurred by the principal firm due to opportunity cost when the level of production is below the capacity).

Step 2: Ascertaining the type of control at work

The Decentralized Capacitated Plant Selection Problem could be described as a hierarchical complex system due to the presence of a coordinator in the form of the

principal firm. This is reflected by the principal firm's desire to only choose auxiliary plants that would operate optimally at an overall level (i.e. all the plants combined).

Step 3: Determining the presence of a bilevel system.

This entails the positive presence of the following traits:

- i) The lower-level decision-makers truly exists and are not a part of the solution approach, each with a defined local objective which cannot be assigned;**

The lower-level decision-makers within the problem are the auxiliary plant managers, each looking to minimize the cost of running their individual plants.

- ii) System-level objective cannot be made to coincide with the local objectives.**

While the system-level objective of the principal firm requires that each of the auxiliary plant operate at a level close to its capacity, the local objective of each auxiliary plant manager requires them to reduce production as far as possible.

Summary

Overall, the decentralized capacitated plant selection problem satisfies the requirements of a bilevel problem at all three steps. The problem could be defined as a bilevel system within the hierarchical complex system. The following section will show the results obtain and discuss its significance.

5.2.2 Results and Discussions

A case studies was undertaken in which a sample decentralized capacitated plant selection problem was resolved. The case study was adapted from the problem tackled by Cao & Chen in [22]. The objective is to find the most optimal policy for the principal firm in selecting auxiliary plants from a pool of four possible candidates to

produce a set of three different products. The case study utilizes the algorithm described in Section 4.2, implemented using Matlab 6.5 platform.

i) Parameters and Constants

Table 5.3: Auxiliary plant data

Plant i	1	2	3	4
f_i	265	265	200	100
p_i	0.55	0.55	0.5	0.4
w_i	0.55	0.55	0.5	0.4
Cap_i	250	250	200	100

Table 5.4: Product demand

Product j	1	2	3
D_j	60	60	60

Table 5.5: Production capacity consumption a_{ij}

a_{ij}	$j=1$	$j=2$	$j=3$
$i=1$	1	1	1
$i=2$	1.2	0.6	1
$i=3$	0.5	0.8	0.8
$i=4$	1.2	1	1.2

Table 5.6: Unit transportation cost R_{ij}

R_{ij}	$j=1$	$j=2$	$j=3$
$i=1$	0.1	0.1	0.1
$i=2$	0.08	0.08	0.08
$i=3$	0.09	0.09	0.09
$i=4$	0.05	0.05	0.05

ii) Results**Table 5.7: Optimal production plan for each combination of feasible plant selection**

Plant selected i	1,4	1,3	2,3	2,4
Production plan $[j=1,2,3]$	1- [30, 0,60] 4- [30,60,0]	1- [0,0,6] 3-[60,60,54]	2- [0,60,0] 3-[60,0,60]	1-[0,60,42] 4-[60,0,18]
Principal firm's cost	454.60	638.60	643.70	724.66
Aggregate auxiliary plant cost	101.40	80.16	74.40	92.40

Plant selected i	3,4	1,2,3	1,2,4	1,3,4
Production plan $[j=1,2,3]$	3-[60,0,60] 4-[0,60,0]	1-[0,0,6] 2-[0,60,0] 3-[60,0,54]	1-[0,0,6] 2-[0,60,36] 3-[60,0,18]	1-[0,0,6] 3-[60,0,54] 4-[0,60,0]
Principal firm's cost	377.00	1045.30	864.66	778.60
Aggregate auxiliary plant cost	76.80	75.36	92.52	77.76

Plant selected i	2,3,4	1,2,3,4	1	2
Production plan [$j=1,2,3$]	2-[0,60,0] 3-[60,0,54] 4-[0,0,6]	1-[0,0,6] 2-[0,60,0] 3-[60,0,48] 4-[0,0,6]	1-[60,60,60]	2-[60,60,60]
Principal firm's cost	783.22	1184.82	303.50	310.10
Aggregate auxiliary plant cost	74.64	75.60	117.00	106.80

Plant selected i	3
Production plan [$j=1,2,3$]	3-[60,60,60]
Principal firm's cost	237.00
Aggregate auxiliary plant cost	79.2

Optimal plant selection for the principal firm: Auxiliary plant 3.

Optimal production plan for auxiliary plant: 60 units for each of the three types.

Optimal cost incurred by principal firm: 237.

Optimal cost incurred by auxiliary plant: 79.2

iv) Discussion

From the results obtained, it would be optimal for the principal firm to select only auxiliary plant 3 to produce its set of three products. The likely cause for such a

selection is probably due to the high fix cost of opening a plant in the first place. Thus, it could be surmised that it would be prudent for the firm to focus all its production within one plant rather than spread it over multiple plants should the plant be capable of meeting all of the product's needs.

The decentralized capacitated plant selection problem as defined here is rather unique in that the leader, i.e. the principal firm, could only be affected by the plants which it has selected. In addition, the auxiliary plants are combined into a single decision-maker as it was within the original problem defined by Cao & Chen in [22]. In a real-life situation, such an arrangement would be highly unlikely as each auxiliary plant manager would probably be an independent decision-maker whose main concern would be its local plan objectives.

5.3 Chapter Summary

This Chapter had two possible problems of a bilevel nature namely the multi-echelon inventory problem and the decentralized capacitated plant selection problem based on the work of Axasäter [19] and Cao & Chen [22] respectively. In Sections 5.1.1 and 5.2.1, the approach developed in Section 3.1 was successfully applied to ascertain that the two problems were of a bilevel nature. In Section 5.1.2 and 5.1.3, the case studies pertaining to the multi-echelon inventory problem had alluded to the importance of identifying such problems accurately by demonstrating the difference between costs calculated within the global or Pareto context and the bilevel context. Additionally, the case study also demonstrated the possibility of attaining close to Pareto-optimal solutions within the bilevel context by applying the correct policies (i.e. sharing of backorder costs) within the multi-echelon inventory problem. The next

chapter will summarized the salient points of this thesis and discuss possible further works pertaining to the research undertaken.

Chapter 6

Conclusion and Further Works

This research work examined a class of problems characterized by a bilevel nature ergo a bilevel problem. In general, the bilevel problem could be defined as an attempt to optimize a first move advantage by accounting for the potential reactions from other decision-makers to such a move. Such a problem is synonymous with the well-established Stackelberg game or Stackelberg strategy [78, 79] as defined within game theory. Mathematically, both concepts could be modeled through the bilevel programming problem or bilevel optimization problem that is often attributed to Candler & Norton's 1977 technical paper entitled 'Multilevel Programming' [2]. In practice, the bilevel problem tends to be described through either one of the more recognized Stackelberg game or bilevel program. The main objective of this work is to identify the bilevel problem within the decentralized decision-making setting by distinguishing its intrinsic sequential decision-making. Section 6.1 will reiterate and discuss the objectives of this work. This is followed by a brief discussion of potential future developments in Section 6.2.

6.1 Conclusion

In Section 1.4 the following were listed down as objectives for this research work:

- To show that the bilevel problem tend to manifests itself within the context of decentralized decision-making problems in which different types of decision-making exist or co-exist.
- To adopt a common platform to study the Stackelberg/bilevel, Pareto optimal and Nash equilibrium variants of problems within the decentralized decision-

making setting without being limited to within any specific topic of engineering.

- To identify the particular characteristics associated with the bilevel problem which would exclude the possibility of the Pareto optimal or Nash equilibrium problems appearing.
- To formulate a general protocol with which to identify bilevel problems within the decentralized decision-making setting and to test its effectiveness using appropriate case studies.

The following subs-sections will assess whether the research work has achieved its intended objectives.

To show that the bilevel problem tend to manifests itself within the context of decentralized decision-making problems in which different types of decision-making exist or co-exist.

This objective is achieved in the review of Chapter 2. All four engineering problems within the decentralized decision-making setting reviewed i.e. i) collaborative multidisciplinary design problem, ii) multi-echelon inventory policy problem, iii) decentralized capacitated plant selection problem and iv) transport network optimization problem were shown to manifest different types of decision-making namely the restricted, cooperative, isolated and sequential decision-making inherent to the bilevel problem. While the review of problems i) and ii) showed that the one of the decision-making type would be present at any one time, that of problems iii) and iv) showed the possibility of having two different types of decision-making present within the problem at the same time. It was also shown that the bilevel problem was present in all of the four problems reviewed. Although the list of

problems reviewed is not exhaustive, it represents a diverse segment of research interest ranging from multidisciplinary design to transport network. Therefore, this research work has also shown that problems such of a bilevel nature are not restricted to any particular application and may potentially appear in other areas of interest to engineering.

To adopt a common platform to study the Stackelberg/bilevel, Pareto optimal and Nash equilibrium variants of problems within the decentralized decision-making setting without being limited to within any specific topic of engineering.

As described in Section 1.3, observations derived from reviewing the different problems in Chapter 2 show the feasibility of adopting the complex system to study problems within the decentralized decision-making setting. The chart in figure 2.4 has shown that the problems reviewed could be categorized within the fully decentralized and hierarchical control segments of the complex system where the different types of decision-making could also be defined therein. However, due to the restricted number of problems reviewed, the generality of complex system as a platform to exhaustively define all the different problems within the decentralized setting could not be ascertained without expanding the review further into other areas of interest within engineering.

To identify the particular characteristics associated with the bilevel problem which would exclude the possibility of the Pareto optimal or Nash equilibrium problems appearing.

Based on the review performed in Chapter 2, characteristics intrinsic to the bilevel problem were identified and listed down in Chapter 3. This includes, first and foremost, sequential decision-making and the presence of coupled variables.

To formulate a general protocol with which to identify bilevel problems within the decentralized decision-making setting and to test its effectiveness using appropriate case studies.

In Chapter 3, a general protocol to identify the bilevel problem was defined in figure 3.3. The protocol targets problems within the decentralized decision-making setting in general not limited to any specific area of interest. The utility of the protocol was tested in two case studies dealing with the two disparate problems in multi-echelon inventory policy and the decentralized capacitated plant selection and was able to fulfill its purpose in identifying the bilevel problem within. Similar to the comments on the second objective, the full effectiveness and applicability of the general protocol could not be fully guaranteed without further testing it on problems within different areas of interest.

6.2 Further Work

The direction in which the future work takes could be divided into two headings:

1. Further testing for the approach developed.
2. Development of a comprehensive approach or methodology to identify and resolve a problem of a bilevel nature.

6.2.1 Possible further testing for the approach developed

This would involve a more rigorous testing of the approach developed to ascertain its full effectiveness as well as its potential weakness and failure. Due to the general nature of the approach developed, samples of bilevel problems from different fields such as socio-economics and strategic problems could also be used to provide a more thorough platform for testing. In particular, the approach could be utilized new

applications for the bilevel problem/program. This would most probably entail an in-depth and detailed research work focusing on one specific field. In the course of the research work, two main problems of interest which merits further investigation as possible bilevel programming problems are: i) concurrent engineering and ii) manufacturing/Floor shop. Although the implementation of bilevel problem to model concurrent engineering was explored by Rao et al in [6], the intention of the work was intended more as a proof of concept for game theoretical application as a whole rather than just bilevel optimization. However, it deed demonstrate the usefulness of bilevel programming within such a setting.

6.2.2 Development of a comprehensive approach or methodology to identify and resolve a problem of a bilevel nature

This research work admittedly focused more on the identification approach and adopted an ad-hoc manner of resolving the bilevel problems encountered. Thus a framework or software able to combine the two aspects would undoubtedly be a useful and convenient tool for an engineer or a designer. This would require a harmonious synthesis of the two aspects which should be easy to understand and use. However, due to the natural complexities within the mathematical formulation of the bilevel program, such an algorithm or approach capable of resolving *all* bilevel problems would be difficult to achieve. A potential methodology towards such an outcome could be an aggregation of all the different approaches developed or a collection of solution approaches. Heuristics such as genetic algorithm and simulated annealing [4] could also be a potential way to solve bilevel problems without limiting it them to having specific pre-conditions.

6.3 Concluding remarks

Bilevel problems or bilevel programming has the potential to be an essential tool to most decision-makers. It allows the decision-maker to perceive and optimize first hand advantage. Ignoring these advantages short-changes a decision-maker, be it a product designer or engineer, such that it would result in a myopic view of the problem at hand. In reality, this concern has been recognized at a much earlier stage in the economic field where failures to give due consideration to the competition would most often result in catastrophic failures. Within the engineering field in general, there is a growing recognition of the need to consider not just the effects of multiple objectives but that of multiple decision-makers as well. This is evident in problems such as the multi-disciplinary complex engineering systems, transport network design problem, and many other complex or large-scale problems with multiple objectives and decision-makers. To be aware of and having the ability to identify and handle such problems in an appropriate manner, be it Pareto, Nash or Stackelberg, would undeniably be useful to any present and future engineers.

References:

Chapter One

- [1] Michalewicz, Z., Dasgupta, D., Riche, R.G.L., and Schoenauer, M. (1996). Evolutionary Algorithms for Constrained Engineering Problems. *Computers Industrial Engineering*, 30, 4, 851-870.
- [2] Candler, W. and Norton, R. (1977). Multilevel programming, *Technical Report* 258, Washington D.C., World Bank Research Center.
- [3] Candler, W. and Norton, R. (1977). Multi-Level Programming and Development Policy, *World Bank Staff Working Paper #248*, Washington D.C.
- [4] Bard, J.F. (1998). *Practical Bilevel Optimization Algorithms and Applications*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- [5] Kalsi, M., Hacker, K., and Lewis, K. (2001). A Comprehensive Robust Design Approach for Decision Trade-Offs in Complex Systems Design. *Journal of Mechanical Design*, 123, 1-10.
- [6] Jagannatha Rao, J.R., Badhrinath, J.R., Pakala, K., and Mistree, F.R. (1997). A Study of Optimal Design Design Under Conflict Using Models of Multi-Player Games. *Engineering Optimization*, 28, 1/2, 63-94.
- [7] Sandell, Jr, N.R., Varaiya, P., Athans, M., and Safonov, M.G. (1978). Survey of Decentralized Control Methods for Large Scale Systems. *IEEE Transactions on Automatic Control*, AC-23, 2, 108-128.
- [8] Krishnamachari, R. and Papalambros, P. (1997). Hierarchical Decomposition Synthesis in Optimal Systems Design. *Journal of Mechanical Design*, 119, 4, 448-457.
- [9] Kusiak, A. and Wang, J. (1993). Decomposition of the Design Process. *Journal of Mechanical Design*, 115, 4, 687-695.
- [10] Migdalas, A. (1995). Bilevel Programming in Traffic Planning: Models, Methods and Challenge. *Journal of Global Optimization*, 7, 381-405.
- [11] Clark, A.J. and Scarf, H. (1960). Optimal Policies for a Multi-Echelon Inventory Problem. *Management Science*, 6, 475-490.
- [12] Von Neumann, J. and Morgenstern, O. (1947). *The Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press.
- [13] Lewis, K. & Mistree, F. (1997). Collaborative, sequential, and isolated decisions in design. *ASME Journal of Mechanical Design*, 120, 643-652.

-
- [14] Bard, J.F. (1983). Coordination of a Multidivisional Organization Through Two Levels of Management. *OMEGA, The International Journal of Management Science*, 11, 5, 457-468.
- [15] Bozma, H.I. and Duncan, J.S. (1991). Sequential and Parallel Models in Noncooperative Games for Decentralized Decision Making in Vision Systems. In: *Proceedings of the 12th International Conference on Information Processing in Medical Imaging* 358-372.
- [16] Wang, H., Guo, M. and Efstathiou, J. (2004). A game-theoretical cooperative mechanism design for a two-echelon decentralized supply chain. *European Journal of Operational Research*, 157, 372-388.
- [17] Cachon, G.P. and Zipkin, P.H. (1999). Competitive and Cooperative Inventory Policies in a Two-Stage Supply Chain. *Management Science*, 45, 7, 936-953.
- [18] Karlof, J.K. and Wang, W. (1996). Bilevel Programming Applied to the Flow Shop Scheduling Problem. *Computers Operations Research*, 23, 5, 443-451.
- [19] Axasäter, S. (2001). A framework for decentralized multi-echelon inventory control. *IIE Transactions*, 33, 91-97.
- [20] Bell, M.G.H. (2000). A game theory approach to measuring the performance reliability of transport networks. *Transportation Research Part B*, 34, 533-545.
- [21] Castelli, L., Longo, G., Pesenti, R. and Ukovich, W. (2004). Two-Player Noncooperative Games over a Freight Transportation Network. *Transportation Science*, 38, 2, 149-159.
- [22] Cao, D. and Chen, M. (2006). Capacitated plant selection in a decentralized manufacturing environment: A bilevel optimization approach. *European Journal of Operational Research*, 169, 97-110.
- [23] Findeisen, W. (1982). Decentralized and Hierarchical Control Under Consistency or Disagreement of Interests. *Automatica*, 18, 6, 647-664.
- [24] Sage, A.P. (1977). *Methodology for Large Scale Systems*. New York: McGraw-Hill.
- [25] Haimes, Y. Y. and Li, D. (1988). Hierarchical Multiobjective Analysis for Large-scale Systems: Review and Current Status, *Automatica*. 24, 1, 53-69.
- [26] Papavassilopoulos, G.P. (1979). Nonclassical Control Problems and Stackelberg Games, *IEEE Transactions on Automatic Control*. AC-24, 2, 155-166.
- [27] Mesarovic, M.D., Macko, D., and Takahara, Y. (1970). *Theory of Hierarchical, Multilevel Systems*. New York: Academic Press.

-
- [28] Lewis, K. and Mistree, F. (1997). Modeling Interactions in Multidisciplinary Design: A Game Theoretic Approach. *AIAA Journal*, 35, 8, 1387-1392.
- [29] Kalsi, M., Hacker, K., and Lewis, K. (2001). A Comprehensive Robust Design Approach for Decision Trade-Offs in Complex Systems Design. *Journal of Mechanical Design*, 123, 1-10.
- [30] Xiao, A., Zeng, S., Allen, J., Rosen, D., and Mistree, F. (2005). Collaborative multidisciplinary decision making using game theory and design capability indices. *Research in Engineering Design*, 16, 57-72.
- [31] Dantzig, G. and Wolfe, P. (1960). Decomposition principle for linear programs. *Operations Research*, 8, 101-111.
- [32] Kolstad, C.D. (1985). A review of the literature on bi-level mathematical programming, *Technical Report LA-10284-MS*, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.
- [33] Mahmoud, M.S. (1977). Multilevel systems control and applications: a survey. *IEEE Transactions on Systems, Man, and Cybernetics*, 7, 125-143.
- [34] Nechane, D.M. (1984). Optimization methods in multilevel systems: a methodological survey. *European Journal of Operational Research*, 21, 25-38.
- [35] Shih, H-S, Lai, Y-J., and Stanley Lee, E. (1996). Fuzzy approach for multi-level programming problems. *Computers Operations Research*, 23, 1, 73-91.
- [36] Bialas, W.F. and Karwan, M.H. (1982). On Two-Level Optimization. *IEEE Transactions on Automatic Control*, AC-27, 1, 211-214.
- [37] Bialas, W.F. and Karwan, M.H. (1984). Two-level linear programming. *Management Science*, 30, 8, 1004-1020.
- [38] Candler, W. and Townsley, R. (1982). A Linear Two-Level Programming Problem. *Computers & Operations Research*, AC-27, 1, 211-214.
- [39] Bard, J.F., Plummer, J.C., and Sourie, J.C. (1998). Determining tax credits for converting nonfood crops to biofuels: an application of bilevel programming. In A.Migdalas, P.M. Pardalos, and P. Värband (Eds.), *Multilevel Optimization: Algorithms and Applications* (pp. 23-50). Dordrecht:Kluwer Academic Publishers.
- [40] Kara, B.Y. and Verter, V. (2004). Designing a Road Network for Hazardous Materials Transportation. *Transportation Science*, 38, 2, 188-196.
- [41] Brotcorne, L., Labbe, M., Marcotte, P., and Savard, G.(2001). A Bilevel Model for Toll Optimization on a Multicommodity Transportation Network. *Transportation Science*, 35, 4, 345-358.

- [42] Cruz, J.B. (1978). Leader-follower strategies for multilevel systems. *IEEE Transactions on Automatic Control*, AC-23, 244-255.

Chapter Two

- [43] Chen, W. and Lewis, K. (1999). Robust Design Approach for Achieving Flexibility in Multidisciplinary Design. *American Institute of Aeronautics and Astronautics Journal*, 37, 8, 982-989.
- [44] Chanron, V., Lewis, K., Murase, Y., Izui, K., Nishiwaki, S. and Yoshimura, M. (2005). Handling Multiple Objectives in Decentralized Design. In: *Proceedings of IDETC/CIE ASME 2005 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Long Beach, California, U.S.A.* 1-12.
- [45] Sobieszczanski-Sobieski, J., Agte, J.S., and Sandusky Jr., R.R. (2000). Bilevel Integrated System Synthesis. *American Institute of Aeronautics and Astronautics Journal*, 38, 1, 164-172.
- [46] Sobieszczanski-Sobieski, J. and Haftka, R.T. (1997). Multidisciplinary aerospace design optimization: survey of recent developments. *Structural Optimization*, 14, 1, 1-23.
- [47] Wujek, B.A., Renaud, J.E., Batill, S.M., and Brockman, J.B. (1996). Concurrent subspace optimization using design variable sharing in a distributed computing environment. *Concurrent Engineering-Research and Applications*, 4, 4, 361-377.
- [48] Balling, R.J. and Sobieszczanski-Sobieski, J. (1996). Optimization of Coupled Systems: A Critical Overview of Approaches. *American Institute of Aeronautics and Astronautics Journal*, 34, 1, 6-17.
- [49] Braun, R.D., Moore, A.A. and Kroo, I.M. (1997). Collaborative Approach to Launch Vehicle Design. *Journal of Spacecraft and Rockets*, 34, 4, 478-486.
- [50] Kroo, I.M., Altus, S., Gage, P., Braun, R. and Sobieski, I. (1994). Multidisciplinary optimization methods for aircraft preliminary design. In: *Fifth AIAA/USAF.NASA/OAI symposium, Panama City, Florida, U.S.A.* 697-707.
- [51] Gross, D. and Soland, R.M. (1981). Designing a multi-product, multi-echelon inventory system. *TIMS Studies in the Management Sciences*, 16, 1-49.
- [52] Cachon, G.P. and Zipkin, P.H. (1999). Competitive and Cooperative Inventory Policies in a Two-Stage Supply Chain with Lost Sales. *University of Pennsylvania Working Paper, Philadelphia, PA, U.S.A.*
- [53] Ishii, K., Takahashi, K., and Muramatsu, R. (1988). Integrated production, inventory and distribution systems. *International Journal of Production Research*, 26, 3, 473-482.

- [54] Cohen, M., and Lee, H. (1988). Strategic analysis of integrated production-distributed systems: Models and methods. *Operations Research*, 32, 216-228.
- [55] Lee, H., and Whang, J. (1996). Decentralized multi-echelon supply chains: Incentives and Information. *Working Paper, Stanford University*, Stanford, California, U.S.A.
- [56] Holmberg, K., Rönnqvist, M. and Yuan, D. (1997). An exact algorithm for the capacitated facility location problems with single sourcing. *European Journal of Operational Research*, 113, 544-559.
- [57] Ghianni, G., Guerriero, F., and Musmanno, R. (2002). The capacitated plant location problem with multiple facilities in the same site. *Journal of the Operational Research Society*, 32, 783-792.
- [58] Jaramillo, J., Bhadury, J., and Batta, R. (2002). On the use of genetic algorithms to solve location problems. *Computers & Operations Research*, 29, 761-769.
- [59] Etrogal, K., and Wu, S. (2000). Auction-theoretic coordination of production planning in the supply chain. *IIE Transactions*, 32, 931-940.
- [60] LeBlanc, L.J., Morlok, E.K., and Pierskalla, W.P. (1975). An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem. *Transportation Research*, 9, 5, 309-318.
- [61] Abdulaal, M.S. and LeBlanc, L.J. (1979). Continuous Equilibrium Network Design Models. *Transportation Research*, 13B, 19-32.
- [62] Beckmann, M., McGuire, C.B., and Winstern, C. (1956). *Studies in the Economics of Transportation*. New Haven, Connecticut, U.S.A.: Yale University Press.
- [63] Nguyen, S. (1974). A Unified Approach to Equilibrium Methods for Traffic Assignment. In M.A. Florian (ed.), *Traffic Equilibrium Methods, Proceedings of the International Symposium on Traffic Equilibrium Methods in Montreal, 1976* (pp. 148-182). Netherlands: Springer.
- [64] Wardrop, J.G. (1952). Some theoretical aspects of road traffic research. *Proceedings of Institution of Civil engineers*, Part II, 1, 325-378.
- [65] Roughgarden, T. and Tardos, E. (2000). How Bad is Selfish Routing? In: *41st Symposium on Foundations of Computer Science* 93-102.
- [66] Braess, D., Nagurney, A, and Wakolbinger, T. (2005). On Paradox of Traffic Planning. *Transportation Science*, 39, 4, 446-450.

Chapter Three

- [67] Shimizu, K., Ishizuka, Y. and Bard, J.F. (1997) *Nondifferentiable and Two-Level Mathematical Programming*. Boston:Kluwer Academic Publishers.

- [68] Casteli, L., Longo, G., Pesenti, R., and Ukovich, W. (2004). Two-Player Noncooperative Games over a Freight Transportation Network. *Transportation Science*, 38, 2, 149-159.
- [69] Fortuny-Amat, J., and McCarl, B. (1981). A Representation and Economic Interpretation of a Two-Level Programming Problem. *Journal of the Operational Research Society*, 32, 783-792.
- [70] Bard, J.F., and Falk, J.E. (1982). An Explicit Solution to the Multi-Level Programming Problem. *Computers & Operations Research*, 9, 1, 77-100.
- [71] Júdice, J.J. and Faustion, A.M. (1992). A Sequential LCP Method for Bilevel Linear Programming. *Annals of Operations Research*, 34, 1-4, 89-106.
- [72] Hansen, P. and Savard, G. (1992). New Branch-and-Bound Rules for Linear Bilevel Programming. *SIAM Journal of Scientific and Statistical Computing*, 13, 5, 1194-1217.
- [73] Anandalingam, G. and White, D.J. (1990). A Solution Method for the Linear Stackelberg Problem Using Penalty Function. *IEEE Transactions on Automatic Control*, 35, 10, 1170-1173.
- [74] Falk, J.E. and Liu, J. (1995). On Bilevel Programming, Part 1: General Nonlinear Cases. *Mathematical Programming*, 70, 1, 47-72.
- [75] Yin, Y. (2000). Genetic-Algorithms-Based Approach for Bilevel Programming Models. *Journal of Transportation Engineering*, 126, 2, 115-120.
- [76] Friesz, T.L., Anandalingam, G., Mehta, N.J., Nam, K., Shah, S.J., and Tobin, R.L. (1993). The multiobjective equilibrium network design problem revisited: A simulated annealing approach. *European Journal of Operational Research*, 65, 44-57.

Chapter Five

- [77] Axasäter, S. (1990). Simple Solution Procedures for a Class of Two-Echelon Inventory Problems. *Operations Research*, 38, 1, 64-69.

Chapter Six

- [78] Von Stackelberg, H. (1934). *Marktform und Gleichgewicht*. Vienna:Springer-Verlag.
- [79] Chen, C.I., and Cruz, J.B. (1972). Stackelberg Solution for Two-Person Games with Biased Information Patterns. *IEEE Transactions on Automatic Control*, AC-17, 6, 791-798.

Appendix

A.1 Overall Framework of the Multi-Echelon Inventory Problem

The following lists the different modules and external files used:-

List of modules:

1. retcostB.m – Generates retailer's cost.
2. supcostB.m – Generates supplier's cost.
3. AhmModRetGen.m – Generates values for L_0 , S_0 , L_i , S_i and the corresponding retailer cost value and stores them in external file AhmadAxModRetailer.txt.
4. AhmadModTotalGen.m – Generates values for L_0 , S_0 , L_i , S_i and the corresponding total cost (= supplier cost + 4 x retailer cost value) and stores them in external file AhmadAxModTotal.txt.
5. AhmadTestStack.m – Reads AhmadAxModRetailer.txt external file and obtains bilevel optimal solution for supplier's cost. Stores optimal solution in Results.txt.
6. TestAhmModOptimal.m – Reads AhmadAxModTotal.txt and obtains: i) Global optimal supplier's cost and corresponding L_0 , S_0 , L_i , S_i , ii) Global optimal retailer's cost and corresponding L_0 , S_0 , L_i , S_i , and iii) i) Global optimal total cost and corresponding L_0 , S_0 , L_i , S_i . Stores values in Results.txt.
7. Automate.m – Varies the value of φ (the ratio of backorder cost) and runs all the modules in order.

List of external files:

1. AhmadAxModRetailer.txt – stores all feasible values of L_0 , S_0 , L_i , S_i and the corresponding retailer cost values.
2. AhmadAxModTotal.txt – stores all feasible values of L_0 , S_0 , L_i , S_i and the corresponding suppliers, retailers and total cost values.
3. Results.txt – stores all bilevel optimal and pareto/ global optimal solutions.

The following is a flowchart of the overall framework:-

A1: Flowchart of solution algorithm for the multi-echelon inventory problem

A.2 Overall Framework of the Decentralized Capacitated Plant Selection Problem

The following lists the different modules and external files used:-

List of modules:

1. initialize.m- initializes values of $p_i, w_i, a_{ij}, R_{ij}, \text{Cap}_i, f_i$, and d_j .

2. agentcost.m – Generates auxiliary plant's cost.
3. principalcost.m – Generates principal firm's cost.
4. trial2.m – Generates feasible values for i and x (plant selected and production plan respectively) and the corresponding auxiliary plant cost and stores them in external file ValidCoods.txt.
5. constraint2.m – Checks whether values of x are within the constraint limit capacity of each auxiliary plant.
6. agentopt.m – Reads ValidCoods.txt external file calculates minimum retailer's cost. Stores cost and corresponding x values in Optimal.txt.
7. linecnt.m – Counts the number of alternatives within the Optimal.txt file.
8. bilevel.m – Reads Optimal.txt and obtains bilevel optimal solution for i and x and corresponding principal firm, auxiliary plant's costs. Stores solution in Bilevel.txt.

List of external files:

1. ValidCoods.txt – stores all feasible values of i and x and the corresponding auxiliary plant cost values.
2. Optimal.txt – stores optimal values of x and corresponding minimum auxiliary plant cost.
3. Bilevel.txt – stores all bilevel optimal solutions.

The following is a flowchart of the framework:

A1: Flowchart of solution algorithm for the multi-echelon inventory problem