Jat, Mohsin Nasir (2015) Time-differentiated distribution of service parts and repair materials: an investigation of service time, area partitions, and cost relations. PhD thesis, University of Nottingham.

Time-differentiated distribution of service parts and repair materials: An investigation of service time, area partitions, and cost relations

Mohsin Nasir Jat, MSc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

March 2015

Abstract

Motivation: Manufacturers in sectors like Information Technology (IT), Automotive and Aerospace have increasingly become focused on providing after sales services. One of the forms of after sales services is to provide service parts (or spare parts) to customers within different contracted time windows. Commonly offered by large scale IT sector companies, such services are facilitated by Service Parts Logistics (SPL) systems through a network of parts stocking facilities. The number of stocking facilities in a distribution network affects the service responsiveness and service costs related to inventory, transportation and facility set-up. Higher responsiveness can be attained through increasing the number of facilities in a distribution network, which, in turn, usually increases inventory cost. Generally, studies assume that shorter service time windows result in higher costs, but there is a lack of exploration regarding how reductions in service time limits and changes in the fractions of demand for different time-based service types impact on various service related costs. Service area partitioning (or zoning) is another related issue which is unexplored in general facility location literature when considering multiple service time (or distance) constraints and both inventory and transportation costs.

This study is mainly motivated by SPL systems of IT equipment manufacturers that support the provision of service parts at customer sites under different and short service time commitments in a large geographical area. The study is of a generic nature and generates insights that can be relevant for any case where the service responses are provided within different short time windows and involve the provision (or consumption) of some stocks (e.g. emergency infrastructure repairs).

Aim and methodology: The aim of this work is to investigate relationships between time-based service levels, service costs and service zones/areas under a hierarchical organization and a non-hierarchical organization of service facilities. The hierarchical organization has variable capabilities to meet different time-based requirements, while the non-hierarchical organization has a uniform capability to meet the toughest requirement for the entire customer base. The investigation is mainly done through analytical, simulation and optimization modelling with the view of producing answers that provide a general understanding and practical insights rather than producing

situation specific optimization models. Empirical case studies are also conducted to complement the quantitative modelling work so that the research is not divorced from the reality. The case studies point towards the motivation for the modelling study and its relevance to some of the real-world systems, and provide a broader understanding of the issues being researched. The case studies involve two multinational ICT equipment manufacturers and service provides, and a government agency responsible for providing highway emergency services in England.

Key findings: The results from the modelling experiments show that under the non-hierarchical setup, where all facilities provide the full range of service-times in their respective vicinities, inventory and transportation costs are insensitive to the fractions of demand for different time-based service types. However, with an inventory sharing mechanism under the non-hierarchical setup, the increase in the proportion of demand for the service within the longer time window can increase the service availability level while also increasing the average travelling. On the other hand, under the hierarchical setup, which provides a higher level of centralization when there is demand for the service in the longer time window, inventory and transportation costs react to the proportions of demand for different time-based service types. The hierarchical setup results in higher transportation cost compared to the non-hierarchical setup, and, interestingly, does not necessarily lower the inventory level, especially when the overall demand rate is very low. The simulation of the inventory sharing mechanism under the hierarchical setup shows that, in several cases, inventory sharing can not only increase the service availability level, but can also reduce transportation cost.

The analysis based on the optimization models shows that there can be cases where it is more cost effective to serve all demand, regardless of the required service time, in a similar fashion through a non-hierarchical setup. The results also show that the demand fractions for different time-based service types, and inventory and transportation costs can significantly impact on the optimum organization of service zone. There can be distinct optimum patterns of service zones depending on whether the inventory cost or the transportation cost dominates.

Conferences

Jat, M., Muyldermans, L. and MacCarthy, B. (2011), "Time-based service differentiation in a hierarchical system of facilities: Effect of service time on service costs", Presented at the OR53 Annual Conference, The University of Nottingham.

Jat, M., Muyldermans, L. and MacCarthy, B. (2012), "Effect of service time on service costs: Comparison of hierarchical v/s non-hierarchical setup in service parts logistics", Presented at the OR54 Annual Conference, The University of Edinburgh.

Jat, M. and Muyldermans, L. (2012), "Impact of time-based service differentiation on inventory", Presented at the 2nd East Midlands Simulation Special Interest Group Meeting, School of Computer Science, The University of Nottingham.

Jat, M.N. and Muyldermans, L. (2013), "Time-differentiated service parts distribution: Costs under hierarchical and non-hierarchical setups", Presented at the 2013 IEEE Workshop on Computational Intelligence in Production and Logistics Systems (CIPLS), 2013 IEEE Symposium Series on Computational Intelligence, pp. 17–24. doi:10.1109/CIPLS.2013.6595195

Acknowledgements

I would like to thank my supervisor Dr Luc Muyldermans for the continuous and extensive support of my PhD research. I would also like to thank Professor Bart MacCarthy, my second supervisor, for his guidance, and my assessor for the annual reviews for his valuable comments on my work.

I am thankful to the respondents from the case companies who spared time from their busy schedules to provide the industrial insights for this research.

Last but not the least, I thank my parents, wife, siblings and grandmother who have always supported me unconditionally, prayed for me, and reassured me during tough times.

I dedicate this work to my beloved father Dr Nasir Ahmed.

Table of Contents

List of acronyms

DRP   Distribution Resource Planning

CRN   Common Random Numbers

EOQ   Economic Order Quantity

FNS   Fast-Normal-Slow

GPL   General-purpose Programming Language

ICT   Information and Communication Technology

IT   Information Technology

MINLP   Mixed Integer Non-Linear Programming

MRP   Material Requirements Planning

NLP   Non-Linear Programming

OEM   Original Equipment Manufacturer

OM   Operations Management

OR   Operational/Operations Research

SKU   Stock Keeping Unit

SPL   Service Parts Logistics

VRT   Variance Reduction Technique

Chapter 1: Introduction

1.1. Background

A trend can be seen in multinational manufacturing companies to make a strategic reorientation into becoming service providers (Gebauer et al., 2010). Advantages of doing so, including an extension of the reach to customers and their underlying needs, are gradually being recognized by more and more enterprises (Zhen, 2012). It has been known for long now that buyers of high investment equipment do not want to take their chances with their purchases. They expect installation services, application aids, parts, post purchase repair and maintenance, and enhancements to keep the products effective and up to date for as long as possible (Levitt, 1983). The answer to these expectations is an approach which can be summarized as 'selling a function, not a product' (Fortuin and Martin, 1999). This includes offering a range of after-sales services, like the provision of service parts (also known as spare parts and repair parts) for equipment maintenance and repairs to avoid or minimize potentially costly equipment downtimes (Cohen et al., 1997).

Increasing worldwide competition and shrinking profit margins in primary product sales are also forcing high-technology-product manufacturers to find new ways to differentiate themselves from their competitors (Wagner and Lindemann, 2008; Candas and Kutanoglu, 2007). Providing fast and high-quality after-sales services is an important way to achieve this (Cohen et al., 1997; Candas and Kutanoglu, 2007). The potential benefits of superior service management capabilities are diverse and many companies are beginning to view initial product sales primarily as positioning opportunities for pull through sales and services (Dennis and Kambil, 2003). Some

manufacturers sell their primary products (e.g. machines) for a price that is close to the cost of production, with the aim of stimulating future demand for spare parts which can generate high profit margins (Wagner and Lindemann, 2008). The spare parts business is often considered as the highest profit generating function (Suomala et al., 2002; Wagner and Lindemann, 2008; de Souza et al., 2011). There are several publications highlighting the revenues and profits associated with after-sales services generally and service parts businesses specifically to advocate the importance of this area. Some of the published information is summarised below:

- The aftermarket for spare parts and services accounts for 8% of the annual Gross Domestic Product of USA (Cohen, 2005).

- Support and maintenance services continue to constitute a significant part of the US economy, often generating twice as much profit as sales of original products (Dennis and Kambil, 2003; Sang-Hyun Kim et al., 2007).

- After-sales services and parts contribute only 25% of revenues across all manufacturing companies, but are responsible for 40% – 50% of profits (Sang-Hyun Kim et al., 2007).

- After-sales services and parts can contribute as much as 50% of all profits for a typical manufacturing company (Dennis and Kambil, 2003).

- Less than 25% of revenue opportunities in personal computers and automotive are derived from first-time product sales; most revenue opportunities are actually the results of after-sales services (Dennis and Kambil, 2003).

- A study by Wharton School reveals that gross margins for after-sales services in the computer industry generally exceed 50% of enterprise systems and around 20% for non-enterprise systems (Dennis and Kambil, 2003).

-   40 to 50% of the profits made by manufacturers come from parts, maintenance, and servicing, which makes this a $21 billion industry (Cohen et al., 1997; Poole, 2003; Candas and Kutanoglu, 2007).

However, technical equipment support services involving service parts are also reported to be very competitive and implicate complex issues. As the customer expectations for product reliability have increased, the provision of superior after-sales services at a competitive price has become an important qualifier for survival. In particular, after-sales services in IT, communications, and other high-tech industries are facing escalating pressure to improve both the level of service delivered to the customers and the productivity in providing these services (Cohen et al., 1997).

The provision of maintenance and repair services and service parts to the product's end users are facilitated by Service Parts Logistics (SPL). The factors that make the management of SPL challenging include complexity of the facility network, tight constraints on the time and warehouse capacity, and the high costs of inventory and transportation (Cohn and Barnhart, 2006). The importance of the time aspect in service parts distribution is especially highlighted by several authors (e.g. Cohen et al., 1997; Braglia et al., 2004; Cohn and Barnhart, 2006; Candas and Kutanoglu, 2007; de Souza et al., 2011). After-sales services are provided as contractual obligations between customers and manufacturers, and thus designing and operating the logistics network capable of serving customers in a time-responsive manner is crucial (Candas and Kutanoglu, 2007) and strategically important (de Souza et al., 2011). Many of today's business processes, such as production and financial systems, rely on technically advanced equipment. A failure in this equipment can obstruct important functions and any delay in bring the systems up again can result in big losses.

Response times are more critical for certain kinds of products. For computer systems that support critical business functions, one-day service is often not good enough. Service providers of these products are required to guarantee service within hours of a product failure (Cohen et al., 1997). The service time can also differ from customer to customer or equipment to equipment for the same service part. Typically in IT hardware support contracts, depending on the consequences of the equipment downtime, the customers determine different time windows (e.g. 2 hours, 4 hours and 8 hours) within which the requested service part(s) should reach their sites.

Besides IT SPL systems, short service time limits are also synonymous with public sector emergency services. The response times to reach incident sites can vary depending on incident types, and a service system can be required to maintain inventory of the items/material potentially required at incident sites.

In scenarios where there are multiple levels of services, there can be two logical options to meet the requirements of all customers. First, to setup a system that has a uniform capability of meeting the toughest requirement for the entire customer base. Second, to setup a system that has variable capabilities to meet different requirements determined by different customers. Apparently, apart from being less complex, operating a system with a uniform capability (the first option) is unappealing. A uniform capability, which is tuned to meet the most stringent customer requirement, can result in overspending and does not allow any mechanism to transfer cost benefits to customers requiring relaxed services.

As will be discussed in the literature review (Chapter 2), short service times and responsive distribution systems typically increase the costs. However, there is a lack of focus on explicitly investigating the relationship between service time and service

cost. Especially, the impact of the proportions of demand for different time-based service types on service costs is not explored. Multiple service time limits in effect implicate multiple covering ranges in the spatial context, which is an overlooked aspect in the facility location literature. Investigation in this context can provide rich managerial insights into the service time and service cost relationship and its impact on facility locations and service zones.

1.2. Research objectives and scope

The aim of this research is to investigate some of the associations between service time, service areas (or zones), and service costs in a spatial context where demand is covered by a distribution system providing multiple time-based service types. The research is predominantly done through quantitative (analytical, simulation and optimization) modelling. However, to complement the quantitative modelling study and gain a broader understanding of the related issues, an empirical study involving two major ICT equipment manufacturers and a government agency responsible for emergency highway repair services is performed. The objective of the case studies is to investigate:

(1) The characteristics of real world service operations involved in time-differentiated distribution.

Through quantitative (Operational Research) techniques, the research primarily seeks to answer the following questions under a hierarchical and a non-hierarchical setups:

(2) What is the impact of the service time window lengths associated with different time-based service types on service (inventory and transportation) costs?

(3) What is the impact of the demand fractions associated with different time-based service types on service (inventory and transportation) costs?

(4) What is the impact of inventory sharing with varying demand fractions for different time-based service types on transportation and service availability levels?

(5) What is the impact of transportation and inventory costs and the demand fractions for different time-based service types on the optimum facility locations and service zones?

In this research we use the following terminology:

A *time-based service type* is associated with a particular service time constraint corresponding to a guaranteed maximum duration (time window) to deliver a part (or materials) at the demand location after the moment the service has been requested. We use the expressions '*time-based service type*' and '*service type*' interchangeably.

The *Demand fractions* correspond to the percentages of total demand linked with different service types. For example if 40% of the total demand has to be provided within 2 hours and 60% of the total demand has to be provided within 4 hours, then the fractions of demand for the 2 hours service and the 4 hours service are 0.4 and 0.6 respectively.

The *service availability level* is the percentage of demand met from stock on hand, i.e. fill-rate.

Under the *Non-hierarchical setup*, all service facilities (or warehouses) offer the full range of services and are located to cover the entire area considering the shortest travel

time or distance. The facilities provide the distribution service in their vicinities in effect without differentiating the required service time by a particular service request.

Under the *Hierarchical setup*, facilities are located to cover the entire area considering the shortest travel time, however, not all facilities provide the full range of service types. Though all facilities provide the service type with the shortest service time, only a subset of facilities provide a service type with a longer service time in larger service areas to gain a higher level of centralization in the system.

The study assumes a single echelon system and considers the costs related to (1) the maintenance of inventory at local warehouses considering two different inventory policies, and (2) transportation from warehouses to the customers with uniform geographical distribution and Euclidean distances. Inventory and transportation costs account for a significant proportion of the total costs in SPL (Cohen et al., 1997; Wagner and Lindemann, 2008). The analysis is based on two service types, however the models can be extended to consider more than two service time windows without complications. Extending the analysis to include more than two service types though does not generate any further insights with regards to the relationships being investigated in this research.

The aim of this work is to develop models to provide general understanding of the important trade-offs involved in time-differentiated distribution. The research does not seek to produce complex situation specific models with lack of generalizability. The SPL literature, despite of being sizable, has been criticized for the lack of implementation, mainly due to its complexity (Ashayeri et al., 1996; Wagner and Lindemann, 2008), and limited scope (Huiskonen, 2001). Though the strict assumptions in this work make the study rather diverged from the reality and maybe

do not permit that the work produces readily applicable tools for specific practical instances, the logic and insights of the models are easier to communicate and there can be an enhanced likelihood of an implementation of the resulting guidelines (Cohen et al., 1999).

1.3. Outline of the thesis

This thesis is structured into 8 chapters as follows:

Chapter 1 Introduction

      This chapter discusses the background, objectives and scope of this research, and outlines the organization of the thesis.

Chapter 2 Literature review

      Chapter 2 reviews several streams of literature in relation to the distribution issues investigated in this research. The chapter starts with a general review of the literature on SPL and then narrows down to discuss the studies related to location of service facilities, time constrained distribution and service differentiation in SPL. It also briefly discusses the nature of research related to the location of service facilities in emergency services systems. A significant part of the chapter is associated with a review of the literature on different aspects of facility location problems in general, including a review of the types of facility location problems, service area partitioning, hierarchical setups, and (de)centralization. The chapter concludes with the identification of a number of research gaps in the literature and the discussion on how this research addresses some of the identified gaps.

Chapter 3   Research methodology

Chapter 3 describes the methodology used in conducting this research. The chapter first provides an overview of the methodological approaches employed in this research and discusses the strengths and weaknesses of the approaches. This is followed by an overview and a detailed explanation of the stages in this research.

Chapter 4   Empirical case studies

Chapter 4 presents the findings from the empirical study involving the SPL systems of two ICT companies and the service operations of a public agency providing emergency responses. Findings from both organization types are discussed separately. A synopsis of the analysis and the key findings from both sectors is presented at the end.

Chapter 5   Time-differentiated distribution costs under hierarchical and non-hierarchical setups

Focusing on time-based service differentiation, Chapter 5 presents a model to investigate the effects of different service time limits and their fractions in overall demand on inventory, transportation and distribution network setup costs. The model considers an efficient packing of service areas assuming a continuous geographical distribution of customers in a plane.

Chapter 6   Impact of inventory sharing on service availability and transportation levels

Chapter 6 presents a simulation study to analyse the impact of different inventory sharing configurations on service availability (fill-rate) and transportation levels considering varying demand fractions for different service types. Unlike in Chapter 5, the analysis is based on specific cases considering a bounded plane.

Chapter 7   Impact of inventory and transportation costs on optimum zoning for time-differentiated distribution: A unidimensional analysis

Chapter 7 investigates the impact of inventory cost, transportation cost, and the demand fractions for different service types on the optimum locations of service facilities and their service zones. The investigation is done by developing a Nonlinear and a Mixed Integer Nonlinear Programming models which consider uniformly distributed customers on a line segment (or along one road).

Chapter 8   Conclusions and further research directions

Chapter 8 concludes the thesis by summarizing the main contributions of the research and proposing some areas for further study.

Chapter 2: Literature review

2.1. Introduction

This chapter reviews the literature on various aspects of SPL and facility location problems. The review specifically focuses on time constrained and differentiated supply, service area partitioning, hierarchical setups, and centralization-decentralization trade-offs, which are the distribution issues explicitly investigated in this research.

The chapter is organized as follows. The research on SPL is reviewed in Section 2.2. It covers the unique aspects and challenges of managing SPL that differentiate it from the distribution setups of primary products (Section 2.2.1), management of service parts inventories (Section 2.2.2), location of service/distribution facilities and the key decisions for distribution set-ups (Section 2.2.3), time constrained supply of service parts (Section 2.2.4), and service differentiation (Sections 2.2.5). As another area for which this research is relevant, the literature on emergency services systems is briefly discussed in Section 2.3. Section 2.4 broadly discusses the facility location problems that exist in the literature, followed by the review of area partitioning, hierarchical locations, and (de)centralization in subsections 2.4.1, 2.4.2 and 2.4.3 respectively. Section 2.5 summarises the main highlights from the literature review and identifies some gaps. Finally, Section 2.6 briefly discusses the ways in which the objectives of this research can address some of the gaps and contribute to the knowledge.

2.2. Service Parts Logistics (SPL)

2.2.1. Challenges in managing SPL

The requirements for planning the logistics of service parts differ from those of primary products (Dennis and Kambil, 2003; Huiskonen, 2001). The strict requirements on timely service availability set pressures for streamlining SPL systems as the effects of stock-outs may be financially remarkable (Huiskonen, 2001). A supply chain that is explicitly designed to support superior service management is required since a comprehensive service management can significantly complicate a typical supply chain (Dennis and Kambil, 2003).

The control of service parts is a complex matter. Common statistical models for inventory control can lose their applicability because the demand process is often different from those that these traditional models assume (Fortuin and Martin, 1999). Forecasting of demand requires some historical demand figures which are frequently unavailable or invalid for slow moving parts (Fortuin and Martin, 1999). The demand does not remain stable. It can increase as the installed base increases in size (Jin and Tian, 2012) and vanish as new products are introduced rapidly (Cohen et al., 1997). A major challenge for spare parts management after the end of the production is the high uncertainty of spare parts demand over a long period until the end of the equipment use (Inderfurth and Kleber, 2013). Morris (2013) discusses unique problems that occur when forecasting spare parts demand in terms of the part classification, absence of marketing, the use of a forecasting software, and forecasting demand for the parts of a new product.

The service parts inventory turnover is often very low as their use is based on either a product failure in the field or on the consumption of a `usage' part (Fortuin and Martin, 1999; Wagner and Lindemann, 2008). Another factor related to the low demand can be the increased reliability and quality of products (Fortuin and Martin, 1999). Besides, business environmental changes, such as the rapid introduction of new products and the reduction of product life cycles, have increased product varieties and thereby reduced the installed-base of specific models (Fortuin and Martin, 1999), which in turn results in high obsolesce costs (Cohen et al., 1997). Also, as it is guaranteed that the spare parts would be made available for the installed-base under a service contract, the traditional mechanism for a spare parts acquisition is to place a large final order at the end of the regular production of the parent product, causing major holding costs and a high level of obsolescence risk (Inderfurth and Kleber, 2013). Repairing the service parts and using them can mitigate this risk, but there can be considerable product price erosion while the repair costs stay the same. As a consequence, there might be a point in time at which the unit price of the product drops below the repair costs (Pourakbar et al., 2012). If so, Pourakbar et al. (2012) suggest that offering the customer a new product of the similar type or a discount on a next generation product may be more cost effective. An additional aspect is the cost distribution of parts. There is a high disharmony between different parts that a company has to maintain as there can be parts that cost less than one pound and parts that cost thousands of pounds (Amini et al., 2005). However, the cost is not proportional to the part criticality, e.g. a switch may be very cheap but can be extremely critical for a certain type of machine instead of an expensive accessory.

The challenges associated with SPL can be summarized as follows:

- The installed base of products that must be served is large and geographically dispersed.

- There is an enormous variety in the number of service parts which must be maintained.

- The cost of parts is increasing due to the increasing complexity and modularity.

- Product life cycles are short, which is reflected by the high rate of part obsolescence.

- There is a predominance of slow moving parts. This is due to the reduced product sales volumes (per model), which is caused by an increase in the variety and customization. Design improvements have also increased the reliability of parts, which increases the mean time between failures. These trends also contribute towards reducing the predictability of demand.

- Companies can face problems in integrating with their suppliers, mainly due to technological capability issues. These factors contribute to high external replenishment lead-times for spare parts, which in turn have a direct effect on an inventory investment.

(Cohen et al., 1997; Dennis and Kambil, 2003)

- In many capital goods industries, spare parts inventory requires up to 20 times more SKUs than what is needed for current product manufacturing.

- Service locations are usually more decentralized than manufacturing operations.

- Product failures are non-routine.

- There are varying customer service requirements.

- Inconsistencies in the performance of service personnel or third party service providers can undermine customer perceptions of the brand.

(Dennis and Kambil, 2003)

- Most repairs require several parts, and an absence of even one part can cause a delay.

(Lele and Karmarkar, 1983)

Similar challenges and complexities are identified by Cohen et al. (2006), Fortuin and Martin (1999), Fortuin (1980) and Huiskonen (2001). An additional factor highlighted by Fortuin and Martin (1999) is that service parts are often manufactured at the same facilities where the parts required for the assembly of a technical system are being made. As the demand of service parts is relatively low, often parts needed for the assembly are produced with a comparatively higher priority. Nevertheless, it should be stressed that not all service parts have low demand, e.g. engine service parts like filters are commonly used and have a high demand. Cohen et al. (1999) mention that in contrast to repairable parts, consumable service parts are much higher in demand and their average unit cost is significantly lower than the average unit cost of repairable parts.

Mathematical modelling can be critical in solving difficult problems related to SPL (Cohn and Barnhart, 2006; Huiskonen, 2001), but basic modelling approaches often suffer from complicating factors such as large numbers of constraints and integer variables, non-linearities, and weak linear programming relaxations (Cohn and Barnhart, 2006). Employees within an organization can also be expected to resist any change resulting from complex solutions (Cohen et al., 1999).

2.2.2. Inventory management for service parts: Literature and practice

Inventory can represent a significant proportion of costs in any distribution system (Jayaraman, 1998). Specifically in SPL, where the inventory efficiency is a principal requirement for the overall effectiveness (Cohen and Lee, 1990; Cohen et al., 2000), the inventory investment is the largest single factor in the average total cost structure associated with providing after-sales services (Cohen et al., 1997). Furthermore, because competition has forced industries to provide very short-call service contracts in order to boost sales, companies have to maintain large inventories of service parts (Ashayeri et al., 1996). Nevertheless, even a small improvement in inventory control can significantly lower the inventory investment or enhance service, or both (Cohen et al., 1999).

Cohen and Lee (1990) define several specific inventory control decisions that, taken together, define a company's part distribution policy:

- Inventory positioning and control: Selecting stock locations for each SKU and selecting the class of replenishment policies to be used.

- Sourcing: Determining the assignment of sources for different demand types (e.g. determining the different sources from where normal and emergency demands are satisfied)

- Transportation: Selection of transportation mode.

- Requirements prioritization: Applying different stock-issuing procedures for each class of customers.

- Service allocation: Setting segment-specific service targets for each part at each location in the network.

- Shortage allocation: Creating rules for allocating a stocking location's inventory when demand exceeds on-hand supply.

A way in which service practices are differentiated is whether the repair is done by a replacement or a rework. For a sudden failure of a system during a critical operation, following the 'repair by replacement' policy might be more appropriate, i.e. a failed part or subsystem is immediately replaced by an identical, ready-for-use part or subsystem (Rustenburg et al., 2001). There is a wide range of literature on both replacement and repairable service parts inventory management, dating back to 1960s. Kennedy et al. (2002) discuss unique aspects of service parts inventories and review the related literature. Muckstadt (2005) presents details of service parts inventory systems and supply chain algorithms, and provides an extensive bibliography on the subject. Yet, Ashayeri et al. (1996) observe that "despite the wealth of literature on the subject, in practice, no attention has been paid to proper management and control of service-parts inventory". Although there is a huge body of academic literature on theoretical inventory planning concepts for various spare parts supply chain settings, few companies seem to apply them rigorously (Wagner and Lindemann, 2008). Ashayeri et al. (1996) suggest two reasons for this. First, the assumptions made in many of the models developed do not fit reality. Second, a lack of awareness exists. It is common practice that the eventual selection of which parts to purchase, and in what quantities, takes place on a rather intuitive basis (Rustenburg et al., 2001). Inventory models found in the literature are either too specific or too complicated for adoption in corporate practices (Wagner and Lindemann, 2008). From the logistics point of view, even the most sophisticated models have been limited to optimizing the inventories within often very strict assumptions. When these assumptions are relaxed to increase the realism of the models, the complexity of the models increases even

faster and makes it difficult for the practitioners to understand and apply them (Huiskonen, 2001). Huiskonen (2001) blames the complexity of the inventory models in SPL and highlights that even if the complications of models are hidden in a computer software, most managers do not feel comfortable if they do not understand on what basis the specific results of models are produced. It can be difficult to communicate the logic and insights of a complex model to managers. They might see such a model as a black box (Cohen et al., 1999).

Cohen et al. (1999) suggest that for complex logistic systems, basic models can be very effective for both operational control and strategic analysis. That is, the policies these models recommend can dominate the decision rules used in practice. The simplicity of the basic models and the policies they generate enhance the implementation likelihood. One can communicate effectively both quantitative and qualitative insights based on a basic model to managers (Cohen et al. 1999). Ashayeri et al. (1996) found that their case company in the IT sector use the classical Economic Order Quantity (EOQ) model for consumable service parts inventory management, which had proved to be reliable enough. The survey by Cohen et al. (1997) also indicates that basic, understandable inventory management techniques are used widely for service parts. They also found an extensive use of the EOQ model. Although several companies in this survey used customized inventory management software, about half of them also used Distribution Resource Planning (DRP) or Material Requirements Planning (MRP) systems for service parts management. Huiskonen (2001) also reports that most basic inventory theory and models (such as EOQ, ABC-analysis, MRP) have been widely applied in practice and there is relatively little evidence of the use of more sophisticated applications, such as multi-echelon models.

Various inventory policies have been considered in service parts management studies. Deshpande et al. (2003a, 2003b) base their model on a reorder point and order quantity (R, Q) policy which was used in military logistics system for service parts. Cohen et al. (1989) and Cohen et al. (1990) consider a (s, S) base stock policy in their study. Alvarez et al. (2013), Ashayeri et al. (1996), Candas and Kutanoglu (2007), Cohen et al. (1999), Graves (1985), Gzara et al. (2014), Jeet et al. (2009) and Kukreja et al. (2001) use the one-for-one, i.e. (S-1, S), replenishment policy to study different industrial and theoretical problems. Though fairly rare, some studies using EOQ model in the context of service parts management can also be found (e.g. Cobbaert and Van Oudheusden, 1996; Schrady, 1967). These are the few examples that show the diversity of inventory policies being used in the context of service parts management. However, the (S-1, S) inventory policy is considered to be the most appropriate policy for managing low demand service parts by several authors (e.g. the authors considering the (S-1, S) policy stated above). The appropriateness of the (S-1, S) policy for slow moving expensive service parts is not refuted by the other authors cited here. Nevertheless, giving the example of the IBM's service parts management case, Cohen et al. (1990) state that the (S-1, S) policy does not provide an adequate cost and service performance for a wide range of demand rates.

A main focus of this research is on the relation between service time and inventory levels. In general, there are few studies that explicitly focus on the relationship between the delivery (or customer) lead-time and inventory. In the context of service parts, the study by Schultz (2004), which illustrates how manufacturing firms can meet shorter customer lead times by maintaining service parts locally, can be loosely related to this theme.

2.2.3. Location of service facilities

After-sales support services are supported by a network of repair and stocking facilities that might provide both local and regional responses to customer needs. Such networks can include hundreds of locations where parts are stocked (Cohen and Lee, 1990). Locating inventory stocking facilities, allocating customer demand to these facilities, selecting stock levels (Candas and Kutanoglu, 2007), determining the number of echelons and the linkage between the locations, and defining customer priority classes for service differentiation are the main decisions to be made when designing a SPL system (Cohen and Lee, 1990).

Due to the time-based service level requirements, that are a critical part of any SPL system, there is a stronger interaction between 'tactical' inventory decisions and 'strategic' network design as service requirements are not only a function of the part availability at a facility, but also a coverage issue. Thus, the effects of network decisions on inventory (and vice versa) in an integrated model becomes critical for the optimization of a SPL system (Candas and Kutanoglu, 2007).

In contrast to the importance of the location structure and allocation decisions, this area is comparatively less explored in SPL research. The lack of research is not just limited to the location decisions, but overall the strategic areas are under-researched in the context of SPL. Criticizing the scope of previous SPL research as being limited to inventory modelling, Huiskonen (2001) states that the process of a logistics system design cannot be done in isolation, without taking into account the numerous links with the other processes of a company. Wagner and Lindemann (2008) argue the same issue and highlight that despite the importance of the spare parts business on the firm and macro-economic level, previous literature on spare parts management is quite

limited (Table 2.1). The literature on spare parts management has focused primarily on the planning and operational aspects (e.g. the determination of optimum spare parts inventory levels) and has neglected the strategic and organisational problems companies have to solve in order to manage their spare parts business effectively. The increasing importance of spare parts sales for the performance of companies calls for improved and innovative concepts and strategies in this area (Wagner and Lindemann, 2008). Though the intensive inventory research in the context of SPL might be considered logical as the large part of overall costs in SPL is linked to inventory, other aspects also deserve investigation for a broad understanding.

Table 2.1: Selected literature on supply chain management for primary products (top) and spare parts (bottom) (cited in Wagner and Lindemann, 2008)

	Level	**Issues/supply chain decisions**	**Selected references**
Primary products	Strategy/design	• Outsourcing • Locations • Long-term capacity planning • Channels of distribution • Supply chain type • Models of transportation • Information & communication technologies	Beamon (1998), Chauhan et al. (2004), Chopra (2003), Feitzinger and Lee (1997), Fisher (1997), Lee (2002, 2004), Meixell and Gargeya (2005), Olhager (2003), Reeve and Srinivasan (2005), Rodrigues et al. (2004), Wagner and Friedl (2000)
	Planning and operations	• Demand forecasting • Service levels • Short and medium term capacity planning • Inventory levels	Cachon (1999), Cachon and Fisher (2000), Konijnendijk (1991), Lee and Billington (1993), Musalem and Dekker (2005), Tibben-Lembke and Bassok (2005), Tyagi (2002), Zhao and Xie (2002)
Spare parts	Strategy/design	• Outsourcing • Locations • Channels of distribution • Supply chain type • Information & communication technologies	Dennis and Kambil (2003), Huiskonen (2001)
	Planning and operations	• Demand forecasting • Service levels • Inventory levels	Aronis et al. (2004), Ashayeri et al. (1996), Caglar et al. (2004), Chang et al. (2005), Cohen et al. (1989, 1990, 1999), Dekker et al. (1998), Diaz (2003), Fortuin (1980, 1981), Hollier (1980), Kalchschmidt et al. (2003), Kennedy et al. (2002), Liu and Lee (2007), Schultz (2004), Sleptchenko et al. (2005), Suomala et al. (2002, 2004), Thonemann et al. (2002), Wong et al. (2005)

The delivery performance of service parts suppliers, in particular the reliability of the agreed delivery times, can be critical and problematic (Candas and Kutanoglu, 2007; Fortuin and Martin, 1999). The percentage of demand met on time is a major performance metric of customer service (Cohen et al., 1999). Any uncertainty with respect to the demand process has to be compensated by the flexibility of the delivery process. If this flexibility is insufficient, safety stocks and/or safety lead times have to be introduced (Fortuin and Martin, 1999), which, of course, result in extra cost.

According to Cohen and Lee (1990), the geographical distribution of service facilities significantly affects the operational performance of after-sales services. If a company maintains a complex and dispersed network of facilities, it can satisfy the customer needs for a prompt response. Responsiveness however comes at a significant cost as a typical SPL system may consist of tens or hundreds of warehouses (Gzara et al., 2014). This cost-service trade-off is one of the most important decisions a company has to make for its SPL system design (Cohen and Lee, 1990). Decentralization is the factor that determines the level of service that a company will provide to a customer and the respective costs. A decentralized network is suitable for a customer base requiring immediate response (e.g. for systems used by air traffic controllers) in contrast to the centralized network which is more suitable for a customer base that is not strict on waiting times and for whom the cost efficiency plays a vital role (e.g. domestic computer systems) (Cohen and Lee, 1990). Case studies by Cohen et al. (1990), Wagner and Lindemann (2008) and de Souza et al. (2011) confirm that, predominately, companies perform after-sales service operations from decentralized locations. This is in line with the urgency factor being a key characteristic differentiating SPL from ordinary logistics. The study by Wagner and Lindemann (2008) reveals that by installing and operating a decentralized setup, the lead time for

spare parts can be significantly reduced and customer service levels can be increased. Also, the regional contact personnel can enhance customer satisfaction by offering more specific information about the regional order processing or customer complaints. On the other hand, central distribution of spare parts avoids keeping the same spare parts at multiple warehouses (Wagner and Lindemann, 2008). Cohen et al. (2000) have identified key differences between the attributes of centralized and decentralized strategies for after-sales services (Table 2.2).

Table 2.2: Difference between centralized and distributed service-supply-strategies (Cohen et al., 2000)

Attributes	Centralized	Distributed
Performance targets	Achieving the highest level of inventory turnover at the lowest cost	Ensuring that customers can rapidly obtain any critical part
Network structure	A small number of central warehouses and repair depots	Inventory and repairs available from locations close to customers
Planning process	Visibility of demand at the point of sale	Inventory and transaction visibility at all levels
	Statistical forecasting of local demand and lead times	Forecasting based on estimates of reliability of parts and installed base (customer region)
	Stocking decisions at retail locations made independently of network decisions	Stocking decisions are made based on what products are required and where they are available for all locations
Fulfilment process	Drop-off or mail-in repairs are a viable alternative	Parts are designed to be easily serviced by the service provider (the manufacturer)
	Little fulfilment coordination needed among stocking locations	A high level of coordination exists among all stakeholders in the supply chain
	Both planning of inventory levels and physical fulfilment may be outsourced	Planning of supply chain management is rarely outsourced

To assess the alignment of service-network strategy with the required urgency, managers can place their organizations on the following matrix (Table 2.3) in which the vertical columns indicate the network structure of service centres, and the rows represent the level of criticality, i.e. how crucial it is for the customers to be served

urgently. Making a part available from several places helps in providing quick responses, while a central warehouse or distribution makes more economic sense when demand is not urgent (Cohen et al., 2000). (De)Centralization in a distribution setup is further discussed in Section 2.4.3.

Table 2.3: Connecting service strategy with criticality (Cohen et al., 2000)

		Service Strategy	
		Centralized	**Distributed**
Service Criticality	**Low**	Matched	Mismatched
	High	Mismatched	Matched

Along with the geographical location aspect, Cohen et al. (2006) add the concept of a product hierarchy to relate service levels and the corresponding service costs (Figure 2.1). In the product hierarchy, the higher the spare is ranked, the more expensive it is likely to be. While, in the geographical hierarchy, the higher a location is ranked, the further it is likely to be from the customer. They suggest that a service is most expensive when a spare from the top of the product hierarchy (i.e. the complete product) is supplied from a location in the lowest level of geographical hierarchy. Depending on customer needs, it may well be most appropriate. In contrast, the slowest and the most economical solution can be to provide the spares at the lowest level of the product hierarchy from a central location, i.e. from the top of the geographical hierarchy (Cohen et al., 2006).

THE PRODUCT HIERARCHY

End products (such as computers)

Modules (such as monitors)

Submodules (such as motherboards)

Piece parts (such as semiconductors)

THE GEOGRAPHICAL HIERARCHY

Central repair facility, spare parts warehouse, and distribution center

Regional repair facilities and spare parts distribution centers

Field repair facilities and spare parts distribution centers

Stocks of spare parts on-site with customers

Figure 2.1: After-sales services hierarchy (Cohen et al., 2006)

2.2.4. Time constrained supply

Fill-rate level, i.e. the percentage of demand satisfied from stock on hand, is important, but, as far as SPL is concerned, customers are not just interested about whether the service provider has the part in inventory or not. They mainly care about the time taken by the service provider to provide the service (Kranenburg and van Houtum, 2008; Kutanoglu and Mahajan, 2009), for which the distance to customers plays a major role (Nozick and Turnquist, 2001). Noting the importance of the service time in service parts distribution, Yang et al. (2013) stress that good customer-oriented performance measures are lacking in the literature. The standard service levels, such as fill-rates,

are supplier-oriented, whereas customers only observe deliveries with no delays and deliveries with certain response time.

Though the importance of the service time in SPL is widely recognized, there are very few modelling studies that explicitly consider time, or distance, to reach customers. To the best of our knowledge, the few SPL studies that do consider this factor are by Candas and Kutanoglu (2007), Iyoob and Kutanoglu (2013), Jalil et al. (2011), Jeet et al. (2009), Kutanoglu and Mahajan (2009) and Kutanoglu (2008); summarised in Table 2.4 at the end of this section. The main focus of Kutanoglu (2008), Kutanoglu and Mahajan (2009) and Iyoob and Kutanoglu (2013) is on inventory sharing. Kutanoglu (2008) considers a stylized model and develops a total cost function as the sum of inventory holding cost, transportation cost, and penalties due to emergency and direct (from central warehouse) shipments. He computes a system wide time-based service level, i.e. what percentage of demand is satisfied in a particular time windows in the whole system. By changing the demand rate, stock levels and number of local facilities (2 and 3) the study analyses the impact on costs and the percentage of total demand met in certain time widows. The analysis is based on a randomly generated set of demand and facility points in a plane. Kutanoglu and Mahajan (2009) present a model to determine the minimum–cost stock levels at facilities considering the same costs as considered by Kutanoglu (2008) and considering system wide constraints on the percentage of demand that is met in particular time windows. Their analysis is also based on a randomly generated set of demand and facility points. Note that both Kutanoglu (2008) and Kutanoglu and Mahajan (2009) do not consider covering the customers within certain maximum service time window. They consider system wide service time levels in which certain percentage of demand is met in certain time. Such a level can be met even if a subset of customers is out of the service time range from

warehouses. Similarly, Iyoob and Kutanoglu (2013) consider a pool of two facilities and a set of customers, such that some customers are in the range of one of the facilities, some are in the range of both facilities and the remaining are outside the range of both facilities. They do not specify the location of facilities or customers. The model allocates each customer to one of the two facilities in the pool to minimize the costs, such that certain system wide percentage of demand is met from within the range. Besides considering fixed facility running cost and inventory holding cost (considering (S-1, S) policy and restricting *S*, the base stock level, to be maximum 1), they consider three different transportation cost constants: 1) transportation cost from the allocated facility, 2) transshipment cost (i.e. from the other local facility), and 3) emergency shipment cost from a central warehouse. That is, the transportation cost does not depend on the distance between customers and facilities. They extend the problem by considering multiple such two facility pools in the system such that each customer belongs to only one pool. In their study, a customer does not have to be within the time range from the allocated facility. Only the system wide constraint of 'certain percentage of demand to be within the range' has to be met.

The studies by Candas and Kutanoglu (2007) and Jeet et al. (2009) consider a distance constraint to cover demand for service parts provision. Candas and Kutanoglu (2007) address a stocking and demand allocation problem. Considering a set of facility points, a set of demand points, and a distance constraint, they decide which facility to stock, how much to stock and which facility to assign to which customer in order to minimize the total cost. Their cost function include facilities operating, transportation and inventory holding (considering (S-1, S) policy)) costs. Jeet et al. (2009) develop a more efficient modelling and solution technique for the same model. Jalil et al. (2011), through a case study, analyse the impact of the information about customer locations

and requirements (such as required service times) on the reduction in service costs through a better demand allocation.

Table 2.4: SPL modelling studies considering service time

Study	Inventory sharing	Demand coverage		Complete demand coverage	Time-based service-differentiation	Nature of study
		Discrete locations	Continuous area			
Kutanoglu (2008)	✓	✓	Not considered		Not considered	Analytical: Impact of stock level on costs and overall time-based availability level.
Kutanoglu and Mahajan (2009)	✓	✓				Optimization: Optimum stocking, subject to certain time-based availability constraint.
Candas and Kutanoglu (2007) and Jeet et al. (2009)		✓		✓		Optimization: Optimum stocking and demand allocation, subject to a distance constraint.
Jalil et al. (2011)	✓	✓				Empirical: Impact of customer location information on costs.
Iyoob and Kutanoglu (2013)	✓					Optimization: Optimum demand allocation to facilities such that certain % of demand is met within time.

2.2.5. Service differentiation

Meeting the criticality of customer needs and matching these with the company's supply chain strategy are important ingredients for efficient after-sales services. Companies that match their part-supply strategy to the criticality of the customer needs can dramatically improve the customer satisfaction in after sales interactions (Cohen

et al., 1990, 2000). However, in most cases the desired package of support services differ significantly from one market segment to another. These different expectations focus on several attributes like failure frequency and downtime of the system on the one hand, and maintenance and repair costs on the other (Deshpande et al., 2003b). Customers differ in their perception about which factors actually or should affect the performance based on their individual experience (Deshpande et al., 2003b). To maximize the marketing impact and in order to provide appropriate services, managers should have an accurate idea of customer support expectations (Lele and Karmarkar, 1983). If a company faces demand with different service requirements and it does not have the ability to offer differentiated service levels, then the only feasible option is to provide a service level to all customers that satisfies the highest service requirement. However, in the scenarios where customer perceptions and requirements differ, the aggregate targets can be inadequate for overall satisfaction (Deshpande et al., 2003b; Fortuin and Martin, 1999). Applying a single policy for the entire assortment can result in too many compromises (Fortuin and Martin, 1999).

Service differentiation, also referred to as customer differentiation, is the opportunity provided to customers to choose between different services against different costs. It is a desired characteristic in SPL and should be facilitated by inventory planning and inventory management (Kranenburg and van Houtum, 2008). The delivery of differentiated levels of service to disparate classes of customers is an increasingly important requirement in today's customer centric environment (Deshpande et al., 2003b). The SPL research reports by Aberdeen Group (Vigoroso and Gecker, 2005) and Deloitte (2006) also stressed this factor. The study by Vigoroso and Gecker (2005) highlights six primary areas that should be considered in the assessment of service parts management. One of these is the service contract management and design, which

describes the agreed service levels. The report from Deloitte (2006) stresses that the capabilities for differentiating customer service levels as per customer requirements are sub-standard and should be addressed as underlying problems. The result of an inappropriate segmentation, or no segmentation, can be that some support areas are over-serviced while others are neglected. This in turn results in under/overpricing of the support services (Lele and Karmarkar, 1983). Effective service differentiation, on the other hand, can mean a high overall customer satisfaction without extensive investment. The results of the studies by Cohen et al. (1999) and Deshpande et al. (2003a, 2003b) suggest that differentiated service requirements can be met without a significant increase in inventory costs for reasonable ranges of service differentiation. It allows an effective way of utilizing the inventory investment because a higher service is provided for the more important parts/customers at the expense of accepting lower service levels for the parts/customers with less impact.

There are several ways in which services can be differentiated. Different kinds of service differentiation have been studied in the inventory and logistics research generally and specifically in the context of SPL. Broadly, differentiation can be done on the basis of part categories or customer demand classes. Some parts are more critical for a customer or product than others (Muckstadt, 2005). Similarly, one part can be more critical to one group of customers than another. A company may be supplying a common part to two customers; one with a high service requirement and the other with a low service requirement (Deshpande et al., 2003b). As classification of parts is not the focus of this study, it is not discussed further in the following sections.

The inventory research relating to differentiation dates back to 1960s. The problem of multiple demand classes was introduced by Veinott Jr (1965). Veinott Jr (1965) focused on the question of how much to order and when to replenish within a periodic review system where in each period the requests are satisfied in a sequence which is in accordance to the priority of their classes. After this study, several variants have been studied. The literature on service differentiation based on customer or demand priority classes is diverse and can be broadly classified into two categories: 1) availability (fill-rate) based differentiation, and 2) service time based differentiation.

2.2.5.1.   Availability/fill-rate based differentiation

Fill-rate is a classical measure of customer service, which, as mentioned previously, is the percentage of demand satisfied from the stock on hand (Cohen and Lee, 1990; Kutanoglu and Mahajan, 2009). Tempelmeier (2006) and Zoller (2005) highlight three strategic options available when serving demand classes of different availability needs; **1)** *Participation*: Extend the highest service level for all customer classes. This of course means over-serving low priority customer classes and denying opportunities for discounts. **2)** *Segregation*: Maintaining exclusive supplies/safety-stocks for the high-priority demand class, which is an expensive way to provide the privileged service and neglects economies of scale. **3)** *Differentiation*: serve all demand from consolidated stocks. The high-priority class is privileged only while the stock is critically low.

Apparently option 3, i.e. differentiation, is the best strategic option to provide exclusive in-stock service to distinguished sources of demand with different fill-rates. In-stock differentiation is mostly provided through policies that introduce some form of intervention levels in a consolidated inventory. These are commonly known as

critical levels. Rationing is a related concept in which service to a low-priority demand is stopped once the stocks drops a certain critical level, while the high-priority demand continues to be served. All rationing policies provide guidance on when to hold back inventory from the lower-priority customers (Deshpande et al., 2003a). That is, at some level of inventory, the system may intervene and deliberately decide to deny the access of some lower priority demand arriving in the future (Pourakbar and Dekker, 2012). The intervention levels can either be set statically (i.e. intervention levels remain the same throughout) or dynamically (i.e. intervention levels change according to the time remaining to a replenishment). The static approach is prone to premature or belated interventions. That is, the service to a low-priority demand is ceased even though the total demand of both priority classes could have been met from the same stock (premature intervention), or even when the service to a low-priority demand is ceased at a level, the full demand of the high-priority class still could not be met (late intervention). On the other hand, the dynamic approach complicates the system by increasing the inventory reviews (Zoller, 2005).

There are several inventory studies adopting critical levels for differentiation – the majority of which are in the context of service parts management. The first study which incorporated rationing/critical levels is conducted by Topkis, (1968) who focuses on how inventory should be allocated between demand classes within a single period of a periodic review model. Although, Veinott Jr (1965) had earlier suggested the use of a critical level without modelling it. Topkis's (1968) model characterizes each demand class by a different shortage cost and accordingly determines the optimal rationing policy between successive procurements of new stock. A significant number of contributions are made in this area since the study by Topkis (1968). The later contributors include Benjaafar et al. (2011), Cattani and Souza (2002), Dekker et al.

(1998, 2002), Deshpande et al. (2003a), Evans (1968), Fadıloğlu and Bulut (2010), Frank et al. (2003), Kaplan (1969), Kranenburg and van Houtum (2008, 2007), Nahmias and Demmy (1981), Pourakbar and Dekker (2012), Tempelmeier (2006), Teunter and Klein Haneveld (2008) and Zoller (2005, 2006).

2.2.5.2. Time-based differentiation

A common strategy used by the firms in service environments is to differentiate their products with respect to time-based characteristics, and use a segment pricing to serve different market segments profitably (Boyaci and Ray, 2006). Studies by Whitt (1999) and Kranenburg and van Houtum (2008, 2009) show that it can be advantageous to partition customers into groups based on service times. In SPL, several examples of the companies that offer different time-based service levels to their customers can be found (Cohen et al., 1997, 1999; Huiskonen, 2001; Kranenburg and van Houtum, 2008, 2009; Wagner and Lindemann, 2008).

Many studies that consider time differentiation in a SPL system consider a service requests as emergency or non-emergency, where the emergency service is quicker and more expensive. There are different ways of how (non-)emergency demand is considered and dealt with, such as: giving priority to certain customers over others for the release of parts if the collective demand cannot be satisfied (e.g. Cohen et al., 1990, 1988), and releasing parts immediately for emergency requests while releasing parts within standard time duration for non-emergency requests (e.g. Cohen et al., 1999; Moinzadeh and Schmidt, 1991; Wang et al., 2002). These studies are however concerned with inventory cost optimization assuming that customers instantly get parts on being released by a warehouse.

The (non-)emergency demand based differentiation overlaps with the concept of inventory sharing (or lateral transshipment). In an inventory sharing or lateral transshipment mechanism, there can be up to three modes of supply; 1) normal supply of part from a local warehouse, 2) transshipment from another local warehouse in a case of a stock-out at the primary local warehouse, and 3) an emergency shipment from outside the sharing facility pool (e.g. from the central warehouse or other external supplier) in a case of a stock-out at all potential local warehouses to satisfy the demand. Different transportation (or replenishment) costs are considered for each of these supply options. The studies related to transshipment are reviewed by Paterson et al. (2011). Inventory sharing is a popular research theme in SPL. In the context of SPL, the recent work on inventory sharing and emergency supplies is carried out by Alvarez et al. (2013), Iyoob and Kutanoglu (2013), Satır et al. (2012) and Yang et al. (2013), who also provide an updated review of the research in this area. Yang et al. (2013) note that though the transshipment time for service parts is not negligible, this aspect (spatial consideration) is hardly considered in the existing service logistics literature. Besides, these studies only differentiate the way the replenishment is done at local warehouses while considering a uniform service for customers.

There are few studies which consider differentiating customers based on the different service (supply) time options that they opt. Kranenburg and van Houtum (2007b) consider a single facility setup and multiple customer groups, each having a service level of maximum average waiting time at the warehouse. The facility has a normal and an emergency replenishment option, where the emergency replenishment mode is used in case of a stock-out and is quicker and more expensive. Their model seeks to determine the stock level at the facility that minimizes the cost (sum of inventory holding, normal replenishment, and emergency replenishment costs) while meeting

the average waiting time target for each customer group. The model provides a framework to compare the use of separate stocks per group to the use of shared stocks for all groups. Kranenburg and van Houtum (2008) study a similar system as in Kranenburg and van Houtum (2007b), but they use a critical level policy as a mean to offer a fill-rate differentiation to the customer groups as well. Kranenburg and van Houtum (2009) is also similar to Kranenburg and van Houtum (2007b), however, instead of a single facility system, they consider multiple warehouses.

Table 2.5: SPL modelling studies considering different service times for different customers

| Study | Factors | | | | Nature of Study |
	Shared stocks for customers	Multi-facility/ transshipment	Customer locations	Fill-rate based differentiation	
Kranenburg and van Houtum (2007b)	✓				
Kranenburg and van Houtum (2008)	✓		Not considered	✓	Inventory parameter optimization
Kranenburg and van Houtum (2009)	✓	✓			

These studies (Kranenburg and van Houtum, 2007b; Kranenburg and van Houtum, 2008; Kranenburg and van Houtum, 2009) only consider customer waiting times at warehouses, i.e. not at customer sites (Table 2.5). The travel distance and time to reach customers is not considered, which can be an important factor when the service times are short, e.g. typically few hours in IT equipment support services. Apparently the travel time can make a considerable difference in the overall service time (Nozick and Turnquist, 2001). Service area partitioning can be an important factor to meet customer service time targets. Based on a case study of an IT equipment support service provider, Jalil et al. (2011) highlight the potential economic value of the machine location information for spare parts logistics where different customers have

different service time requirements. They show that planning performance depends on the quality of installed-base data.

2.3. Emergency services systems

Emergency services are another type of service operations where service (or response) time limits play a critical role. Marianov and ReVelle (1995) and Li et al. (2011) survey the methods related to locating facilities for emergency services. The covering models, discussed in the following section, are especially applicable for emergency facilities, and hence have been very attractive for research (Farahani et al., 2012). The covering models emphasize on providing a coverage for emergency calls within a predefined distance standard (Li et al., 2011). The first study in the context of emergency services is reported to be by Toregas et al. (1971), who consider a distance constraint for locating facilities to completely cover a set of demand points. Plenty of models have since been developed, mostly aiming to maximize the demand coverage with a given number of facilities (maximal covering location problem). Besides, most of the studies on emergency services are related to the medical services. This can be noticed in the review by Li et al. (2011). Brotcorne et al. (2003) specifically trace the evolution of ambulance location and relocation models.

Li et al. (2011) note that, in previous research, emergency calls or demand are treated as discrete points. All demand in an area is generated from the weighted centre of this area. This approximation may result in an inaccurate representation of real world situations. Hence, it is noteworthy to investigate the possibility of using a continuous area instead of discrete points for the demand generation. Also, they point out that, in reality, emergency calls may have different priorities that require different types and/or numbers of emergency services. It can be of interest to integrate the concept of

quality levels as well as priorities in models. Similar observations are made by Farahani et al. (2012) for the covering models in general. Another factor overlooked by the covering models is inventory. As we show through a case study, there can be scenarios where emergency responses involve providing materials at incident locations within a certain time limit.

2.4. Facility location problems

There are several streams of location problems depending on the objective of the facility set-up and optimization criteria. The work in this area is so vast that several review articles and books have been published. Generic reviews on facility location research are provided by Owen and Daskin (1998), Drezner and Hamacher (2002) and Klose and Drexl (2005). Location problems can vary in terms of the way demand and candidate facility locations are considered. The customers can either be considered as discrete points, or as being uniformly distributed over a region. Similarly, for a facility location allocation, either certain candidate points can be considered or a continuous area can be considered in which facilities can be located anywhere. The problems also vary according to the number of facilities that have to be located and the objective function considered. With a given number of facilities, the objective can be to locate facilities to minimize the average distance between customers and facilities (mini-sum/p-median problem), minimize the maximum distance between customers and facilities (mini-max/ p-centre problem), provide a complete coverage of demand or maximize the coverage of demand when there is a maximum distance that can be covered from a facility (complete and maximal covering problems) and so on. These are some broad classes of location problems. There are several variations to these classes and some problems are related to each other. Besides, the same problem have

been referred in the literature by using different names, which can create confusion. For example, Drezner et al. (2002) states fifteen different names used for a mini-sum problem.

In this research the analysis is based on considering uniformly distributed customers and Euclidean distances. The location problems considering continuous geographical spread of customers have been reported to be challenging while managing a single clear objective, and there are no procedures in the literature that provide optimal solutions (Plastria, 2002). Continuous location models always have some geometrical flavour (Drezner and Hamacher, 2002). Normally heuristics are applied to solve these problems with a substantial computation effort involving several movements of facility points resulting in Voronoi diagrams (discussed in the next section) being computed many times (Iri et al., 1984; Plastria, 2002). One way to deal with uniformly distributed customers is to simplify the problem by aggregating demand into discrete points. However, a discrete representation of continuous customer locations can result in coverage errors or uncertainties and imprecise distance measurements (Current and Schilling, 1990; Daskin et al., 1989; Drezner and Drezner, 1997; Murray et al., 2008).

Locating facilities to reach customer sites within certain time limits has traditionally been tackled as covering problems where the objective can be to locate the minimum number of facilities to cover demand, or to cover the maximum demand with a given number of facilities. As with other types of locations problems, there is a vast body of research in covering problems and several authors have reviewed the work in this area. Covering problems are comprehensively reviewed by Farahani et al. (2012), Li et al. (2011), Plastria (1995), ReVelle et al. (2002), Schilling et al. (1993). All these reviews report a lack of research in the case where uniformly distributed customers are

considered. The relatively recent review by Farahani et al. (2012) specifically highlights the need for research in the continuous location domain and on the covering scenarios where different facilities have different covering radii.

Covering location models make a use of an action radius, a threshold distance within which a demand point is considered to be covered (Plastria, 2002). The main multi-facility version of the full covering problem is traditionally called the p-centre problem. This problem is much harder than its single facility version and is NP-hard (Fowler et al., 1981; Masuyama et al., 1981; Megiddo and Supowit, 1984). For continuous location, such questions are known to be 'notorious' in recreational mathematics (Plastria, 2002). The difficulty of this problem is probably the reason why multi-facility location models in a continuous space are much less popular (Plastria, 2002). A full covering Voronoi diagram method for uniformly distributed customers is presented by Suzuki and Okabe (1995) and Suzuki and Drezner (1996). Based on the p-centre problem logic, it seeks to determine facility locations that minimize the maximum distance in a rectangular plane. Since the *p*-centre problem is a non-convex optimization model, a local minimum is identified by this heuristic upon termination. The number of facilities are increased one by one until the distance constraint is satisfied for the entire plane. Murray et al. (2008) evaluate the effectiveness of alternative complete coverage modelling approaches, including the above and the ones where uniformly distributed customers are represented as discrete points, by focusing on an application (siting emergency warning sirens) in an urban area.

Besides minimizing the number of facilities, i.e. the set-up cost, another important objective in facility location decisions is to minimize the average distance to reach

customers (p-median problem) in turn to minimize the transportation costs. A bibliography on this area is provided by Reese (2006). This kind of problem is also underexplored when considering continuous location and Euclidean distances. Suzuki and Okabe (1995) show that the objective function for the continuous p-median problem is non-convex and has non-differentiable points. Considering squared distances between users and facilities, they use a numerical iterative method that gives approximate solution (also in Okabe et al., 2000) and show that in each iteration for the solution, each facility moves towards the centroid of its Voronoi region. As a result, at the end of the procedure, each facility is near the centroid (or centre of gravity) of its Voronoi region. Note that the centre of a regular polygon is its centroid. For example, the centre of a rectangle is its centroid (Anton, 2013).

In spite of a large body of knowledge on facility location problems, a method that tackles the minimization of the required number of facilities along with minimizing the average distance to serve uniformly distributed customers could not be found in the literature. In other words, a method could not be found that, considering a covering distance constraint, decides and locates the minimum number of facilities in a bounded area such that the average distance to reach continuously spread customers is minimized.

2.4.1. Area partitioning (districting)

District design involves the partitioning (according to some criteria) of a large geographical area into smaller subareas or districts for organizational and/or administrative purposes (Muyldermans et al., 2003). The applications of district design include political districting (Bozkaya et al., 2003; Mehrotra et al., 1998), the design of territories for salesmen (Drexl and Haase, 1999; Fleischmann and Paraschis,

1988; Skiera and Albers, 1998), healthcare districting (Blais et al., 2003; Pezzella et al., 1981), school district design (Ferland and Guénette, 1990), police districting (D'Amico et al., 2002), design of response areas for medical emergency services (Iannoni et al., 2009), districting for node routing activities such as product delivery services (Simchi-Levi, 1992; Wong and Beasley, 1984), and district design for the organization of arc-routing activities such as winter gritting, road maintenance (Lotan et al., 1996; Muyldermans et al., 2002, 2003; Perrier et al., 2006, 2008) and refuse collection (Hanafi et al., 1999; Male and Liebman, 1978).

Area partitions in the Euclidean plane are commonly represented in form of Voronoi diagrams. Voronoi diagrams have a long history and are the subject of study and application in various fields (Ash and Bolker, 1985; Aurenhammer, 1991; Hartvigsen, 1992), such as computer science, computational geography, physics, economics and biology (Hartvigsen, 1992). Okabe et al. (2000) provide a thorough account of various forms, concepts and applications of Voronoi diagrams. The basic concept of Voronoi diagrams is as follows:

Assume a set of given points (more than one and finite) in the Euclidean plane. Let all these points be distinct, i.e. consider that these points do not coincide in the plane. Having such point set, every location in the plane is assigned to the closest member of the point set. Hence, the set of locations assigned to a member in the point set forms the point's own region (normally represented as a polygon). If a location is equally close to more than one member of the point set (a location on a boundary between regions), the location is assigned to all those points which are closest and equal in distance. The set of locations assigned to more than one member in the point set forms the boundaries between the regions. The regions are collectively exhaustive in the

plane because every location is assigned to at least one member in the point set. And, other than the boundaries, the set of regions are mutually exclusive (Okabe et al., 2000). Figure 2.2 is an example of Voronoi diagrams.
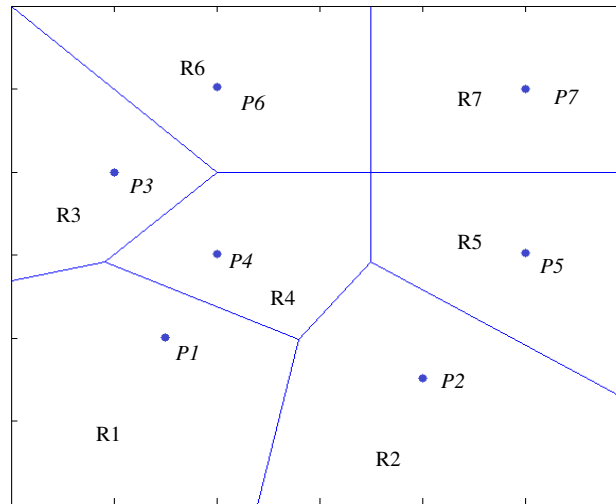
Figure 2.2: Example of a Voronoi diagram partitioning Euclidean plane (Where R_i is the region of point P_i, i = 1,2,3, ..., 7)

Strictly speaking, because the users in real world use streets, the regions in a Voronoi diagram should take into account the road networks, as the street distance is different from the Euclidean distance. However, it is suggested that this difference is not as large as expected and can be justified for an analysis (Cooper, 1983). Catchment areas can be approximated by ordinary Voronoi diagrams (Okabe et al., 2000). Several studies can be found in OR literature that, for analysis, assume Euclidean distances as the travelling distances.

2.4.1.1.   Partitioning of a plane into perfectly packed identical regions

Partitioning an area into identical regular hexagons is considered as an efficient approach of locating facilities in a plane (where each hexagon represents the service catchment area of a facility), assuming Euclidean travel distances. "Of all systems of

regular market areas that will cover a plane completely, the hexagonal one is most efficient in the sense of minimising the distance to be covered between supplier and demander per unit area" (Beckmann, 1968).

Many authors have supported the effectiveness of identical and regular hexagonal partitioning (Bollobás and Stern, 1972; Gusein-Zade, 1992; Haimovich and Magnanti, 1988; Lösch, 1954; Morgan and Bolton, 2002; Stern, 1972). Morgan and Bolton (2002) through their mathematical analysis prove in a certain mathematical sense that for partitioning, the regular hexagons are better than any other collection of congruent or non-congruent shapes of equal or unequal areas, in finite or infinite domains. They consider the average distance from uniformly distributed consumers to the facility to prove that for a unit area and N edges, the regular N-gon has the smallest possible average distance $P_o(N)$ to the centre . As N increases and regular N-gons approach the circle, $P_o(N)$ decreases. Hence for example, regular octagons are better than regular hexagon, but it is not possible to tile a plane with octagons (i.e. perfect packing or a complete coverage is not possible). In fact, only regular hexagons, squares and triangles can be arranged to tile a plane with perfect packing. Hence hexagonal patterns are the best option to achieve perfect packing/tiling with minimum $P_o(N)$.

The average distance ($P_o(N)$) values in Table 2.6 are computed from the following general formula for a N-gon (Morgan and Bolton, 2002):

$$P_o(N) = \frac{2N \int_0^{\frac{\pi}{N}} \int_0^{\frac{Cos(\pi/N)}{cos\theta}} r^2 \, dr \, d\theta}{\left(2N \int_0^{\frac{\pi}{N}} \int_0^{\frac{Cos(\pi/N)}{cos\theta}} r \, dr \, d\theta \right)^{3/2}}$$

Table 2.6: The average distance (approx.) to the centre of regular N-gon of unit area

N	$P_o(N)$
3 (triangle)	0.4036
4 (square)	0.3825
5	0.3784
6 (hexagon)	0.3772
7	0.3766
8	0.3764
20	0.3761
∞ (circle)	0.3761

Besides the study by Morgan and Bolton (2002), which is comparatively recent in this subject, starting from 1930's, several authors have investigated the superiority of the hexagonal packing in context of regional and urban analysis (e.g. Beckmann and Thisse, 1987; Beckmann, 1968; Berry, 1967; Bollobás and Stern, 1972; Hanson, 1997). These studies extend the pioneering work by Christaller (1933) (English translation in Christaller, 1966) and Lösch (1944) (English translation in Lösch, 1954) on Central Place theory which inspired significant literature. These studies show that as long as the social benefit is a decreasing function of distance, congruent hexagons remain optimal. However, Lösch (1954) has noted that the superiority of a hexagon over a square is small and of no practical importance in many instances. Square "is not much inferior to the hexagon" and is the second best region with an added advantage of simply drawn boundaries (Lösch, 1954).

Work by Simchi-Levi (1992), Bertsimas and Simchi-Levi (1996) and Drèze et al. (2008) are examples of the studies related to service systems that consider hexagonal pattern of service catchment areas for their analysis. Simchi-Levi (1992) presents an analytical model to assist the design and control of probabilistic distribution systems that provide services such as delivery, customer pickup, repair and maintenance, and assumed that service stations serve customers inside hexagonal areas. Bertsimas and Simchi-Levi (1996) consider regular hexagonal patterns to analyse some vehicle

routing problems. Drèze et al. (2008) study a problem of locating public facilities (e.g. libraries) considering the hexagonal partitioning of an area.

2.4.2. Hierarchical locations

Hierarchical systems are the ones in which there are functionally coordinated multiple levels; each having some common properties (Dökmeci, 1973). Giving the examples of library services, public health services, emergency services, schools, and marketing structures, Hodgson (1986) expresses that many real-world location problems involve facility systems that are hierarchical in nature and provide several levels of service. The facilities in such service systems are hierarchical in terms of the types or levels of service they offer (Jayaraman et al., 2003). Moore and ReVelle, (1982) describe hierarchical service location systems as follows: "Consider a system with N types of facilities providing N levels of service. Each type of facility is conceived of as the means (building, equipment, staff) required to perform its functions. Each level of service is the set of functions or services provided by that type of facility but not at lower-function facilities. A given level of service is assuredly available from the corresponding type of facility, and at higher-function facilities, but not from a lower-function facility." This explanation though corresponds to the nested hierarchical setup only (Table 2.7). The nested hierarchy is one of the different classes of hierarchical facility setups (discussed latter in this section).

Table 2.7: An example of two level nested hierarchy of services facilities

Facility Types	Service Types	
	Type 1 service	Type 2 service
Type 1 (higher level) service facilities provide: →	✓	✓
Type 2 (lower level) service facilities provide: →		✓

Şahin and Süral (2007) and recently Farahani et al. (2014) have surveyed and classified the literature on hierarchical facility location problems including hierarchical coverage models. According to Dökmeci (1973), Hodgson (1986), Moore and ReVelle (1982), Narula, (1984), O'Kelly and Storbeck (1984), and Şahin and Süral (2007), multilevel systems are not given that much attention in the literature compared to the study of location-allocation problems for single level systems. There are several studies that recognize a hierarchy context, but do not consider hierarchical features explicitly (Şahin and Süral, 2007). Expanding a location-allocation problem to consider more than one level of facility increases the complexity of the problem (Hodgson, 1986; Narula, 1984; Şahin and Süral, 2007). Many traditional location allocation studies have been directed towards the location of systems in which facilities are considered to be homogeneous with respect to size, attractiveness, and the level of service provided (Hodgson, 1986). However, there is an increasing rate of published articles related to the hierarchical facility location problems, which is more noticeable from the early last decade (Farahani et al., 2014). But, Farahani et al. (2014) also highlight that despite the extensive efforts in hierarchical facility location modelling, there are still research gaps. For hierarchical covering they find only one study that considers complete demand satisfaction. This study (Marianov and Serra, 2001) considers complete coverage on nodes through locating two types of servers.

Hierarchical facility systems may be based upon a variety of organizational structures. There may be institutional ties between levels, whereby lower levels are administratively subordinate to higher ones. In many cases, hierarchies have no such inter level linkages; different levels being distinguished solely by the range of goods and/or services they provide (Hodgson, 1986). Narula (1984) describes two types of facility hierarchies: (1) successively inclusive facility hierarchy and (2) successively

exclusive facility hierarchy. In successively inclusive hierarchy, a type k facility ($k =$ *1, 2, … s*) serves demand of types *1, 2, … k*. In successively exclusive hierarchy, a type k facility only serves type k service demand. The healthcare delivery systems, banking systems (drive-ins, branch offices, main offices), and postal systems (branch offices, main post offices) can be considered as examples of the systems with successively inclusive facilities. On the other hand, production distribution systems and electricity distribution systems are two examples of the systems having successively exclusive hierarchical facilities (Narula, 1984). Şahin and Süral (2007) give the same kind of classification using the terminologies of nested and non-nested facilities (refer to Figures 2.3 and 2.4). In a nested hierarchy, a higher-level facility provides all the services provided by a lower-level facility and at least one additional service. While in a non-nested hierarchy, facilities at each level offer different services. Mirchandani (1987) presents an extended classification and classifies hierarchical facilities into three types; successively inclusive, successively exclusive, and locally inclusive and successively exclusive facilities. In the locally inclusive and successively exclusive setting, a type k facility serves demand of type *1* through to k locally (i.e., at the node of its location), but serves only the type k demand from outside its locality (Mirchandani, 1987). Another kind of classification for hierarchical facilities is by their spatial configuration. This classification refers to coherency. In a coherent system, all demand sites that are assigned to a particular lower-level facility are assigned to one and the same higher-level facility. Whereas non-coherent systems are less constrained on the spatial level configuration (Şahin and Süral, 2007). The flow pattern in hierarchical facility systems can also be classified. Customers and/or goods can have either a single-flow or multi-flow pattern. Single-flow starts from level *0*, passes through all the levels, and ends at the highest level (or it starts from the

highest level and ends at level *0*). Multi-flow can be from any lower (higher) level *m*

to any higher (lower) level *n* where *n, m* $\in$ *{0,1,2,...,k}*. In addition to this, multi or

single-flow can also be either referral or non-referral. In a referral system, a proportion

of customers served at each level are referred to the higher levels (Şahin and Süral,

2007).

A hierarchical system with three levels is shown. Sites are marked with letters and numbers denoting the level. Service areas of different-level facilities are circled with different patterns. A customer to be served by the highest-level facility goes to a lowest-level facility first and then passes through all levels until the top. As the facilities are non-nested, different-level facilities are marked with different shapes, denoting that different services are provided at different levels: a level 1 facility is marked with a white circle, a level 2 facility is marked with a triangle, a level 3 facility is marked with a square and dark circles represent the customer sites. In a non-coherent structure, the customers assigned to the same facility at the lowest level may be assigned to different facilities at a higher level: B0 is serviced by B1 for its level 1 demand. Although B1 works with B2 for its level 2 demand, B0 is serviced by A2 for its level 2 demand (Şahin and Süral, 2007).

Figure 2.3: A single-flow, non-nested and non-coherent structure

A three-level system is presented. It is a multi-flow system as goods are shipped from a higher-level facility to any lower-level facility (see the flows from A3 to D0 and B2 to C0). Since the facilities are nested (i.e. all services at lower levels are available at higher levels), a higher-level facility is denoted by all shapes of lower-level facilities (for example, a level 3 facility is marked with a square, a triangle and a circle). As the structure is coherent, a lower-level service area is a subset of a higher-level service area (Şahin and Süral, 2007).

Figure 2.4: A multi-flow, nested and coherent structure

Chistaller's hierarchical central places is a classical nested hierarchical system having a hierarchical hexagonal pattern in a plane with continuous customer spread. In this system (Figure 2.5), 'central places of higher order' are defined as those that serve in a bigger region, in which other central places exist. In a higher level, not only the services of the higher order are offered, but those of the lower orders are also offered. The system comprises different circular ranges of central places depending on their types. To serve the entire land, a perfect and uniform net of central places is created, resulting in a hierarchical hexagonal pattern.
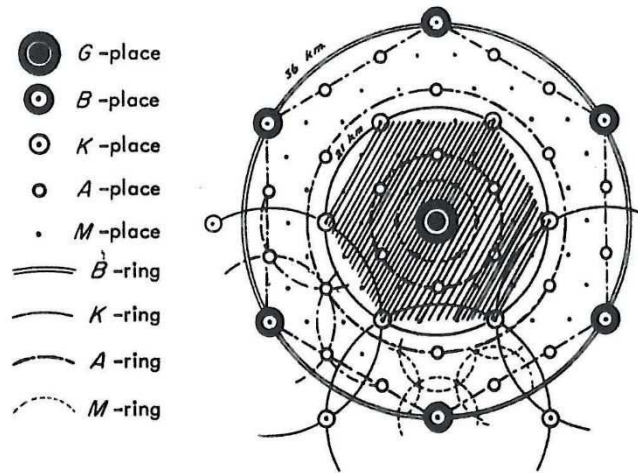
Figure 2.5: Circular range limits in central places system (source: Okabe et al.,

2000)

2.4.3. (De)Centralization

Location decisions require careful attention due to the trade-offs between facility costs, transportation costs, inventory costs and customer responsiveness (Nozick and Turnquist, 2001). One of the basic strategic decisions in the design of any distribution network is whether to set the system as centralized or distributed (Wagner and Lindemann, 2008). Pros and cons for both these options are well argued in literature. Earlier texts discussing various advantages and disadvantages of these strategies include Brown (1967), Heskett (1973), Patton (1986), and Starr and Miller (1962). Generally the advantages associated with these strategies can be identified as follows (Das and Tyagi, 1997):

Centralization: reduced factory-to-warehouse transport costs, improved inventory management, reduced safety stocks, better opportunity for negotiating transportation services, lower stock carrying costs, and easier planning, management and control.

Decentralization: rapid filling of customer orders, reduction in warehouse-to-customer transport costs, better local availability of stocks, and lower delivery time.

However, there is a lack of guidance as to how much (de)centralization is ideal (Das and Tyagi, 1997). Bendoly et al. (2007) stress that the preference of centralization or decentralization should depend on different scenarios of market/business.

Heskett (1973), Maister (1976), and Smykay and Bowersox (1973) discuss centralization in relation to the square root law. According to the square root law, inventory levels increase as the number of warehouses in the system increases. It states that savings from centralization are proportional to the square root of the ratio of the new number of stocking locations over the original number of stocking locations. For example, if the inventory is decentralized from one to two stocking locations, the stock will increase by a factor of $\sqrt{2}$, assuming that demand is equal at both inventory locations. Evers and Beier (1993), Evers (1995), Ronen (1990), Tallon (1993), and Zinn et al. (1989) study the effects of centralization and decentralization on aggregate inventory by modifying the square root law considering the correlation and variability of demand at all locations.

Several authors have used the newsboy problem to study (de)centralization effects (e.g. Chang and Lin, 1991; Chen and Lin, 1989, 1990; Cherikh, 2000; Eppen, 1979). Eppen (1979) show that the expected cost of centralized inventory is lower than that of decentralized inventory. Chang and Lin extend Eppen's results by approaching the concept of centralization with the inventory sharing perspective. They define a centralized inventory system to be the one that allows transfer of stocks between locations. There are numerous studies (e.g. Cherikh, 2000; Granot and Sosic, 2003; Kukreja et al., 2001), both in the context of general inventory management and service

parts inventory management, that consider centralization from this perspective, and not surprising, advocate the benefits of this practice with regards to inventory levels and stock availability. As concluded by Cherikh (2000), "centralisation is also preferable when a portion or all of the excess demand at a location may be reallocated among other locations with remaining inventory… [. This] is not surprising. Due to the aggregation effect, pooling the stocks together reduces the risks from the uncertain demands which results in lower costs and higher profits". As mentioned previously, Paterson et al. (2011) present a review on inventory modelling research related to inventory sharing (lateral transshipment) and provide a classification of these studies.

Meller (1995) investigate the impact of multiple stocking points on system profitability by examining the increased profit needed to offset the inventory cost increases. Sargent and Kay (1995), focusing the storage within one facility (e.g. a factory), examine the trade-off between the saving in material handling costs due to more decentralization of the storage and the additional costs to set up and to run the decentralized storage. Das and Tyagi (1997) analyse the inventory centralization decision by considering the trade-off between inventory and transportation costs. They base their analysis on different roles of facilities and conclude that if each facility is responsible for the costs of distributing and maintaining stocks for all its customers, then a partial centralization of inventories results as a trade-off between inventory and transportation costs. The higher the transportation cost in relation to inventory cost, the greater should be the level of decentralization. On the other hand, a higher service availability level can be achieved through greater centralization. In this case, the actual degree of centralization requires added analysis. Das and Tyagi (1997), like Bendoly et al. (2007), stress that the decisions regarding the degree of centralization should consider the nature of demand and other factors appropriate for a business situation.

2.5. Summary of the literature review and the research gaps

The importance of after-sales support services has been highlighted by several authors in terms of customer satisfaction, revenue and profit. For some industrial sectors, like IT, after-sales support services are of particular importance. Many of the business processes in today's world run on sophisticated equipment which can require support in a form of service parts provision, which is facilitated by SPL. However, there are many complexities and challenges inherent in SPL that distinguish it from regular supply chain logistics. These include special inventory characteristics and urgent and varying service level requirements.

The special characteristics that make service parts inventory management distinct from general inventory management are mainly reported as the low turnover and high demand uncertainty. However, some service parts can have a higher turnover and some degree of certain demand, e.g. in cases of scheduled maintenance. The research on service parts management is predominately based on mathematical programming and OR techniques. But despite the significant research in this area, there is a gap between the literature and practice which is identified by several authors. This is blamed on the complexity and limited scope of the models present in the literature, mostly seeking inventory policy optimization of some kind. Authors who have studied SPL industrial practices indicate that many of the companies use simple inventory policies such as the EOQ, which, in some instances, has proved to be reliable enough for service parts management. Most of the recent inventory research in SPL is based on the one-for-one replenishment policy, which is considered to be more suitable for service parts due to normally low turnovers and high unit costs.

A popular area of investigation in service parts inventory research is inventory sharing. There are many studies that focus on inventory sharing through transshipments within different warehouses in a SPL system. However, apart from few, the studies normally do not take into account the time or distance as a factor in transshipments when determining the availability level. Where time (or distance) is included as a factor, it is considered that parts are made available in time at warehouses, not at customer sites.

Decisions on facility locations are important for designing SPL systems. However, the location aspect of SPL and its relation with inventory is an under-researched area. Most of the SPL research is related to inventory modelling in isolation from the other aspects that can influence the overall SPL design (such as service times). It is normally considered that supplying in a shorter time and meeting urgent customer requirements are costly. However, studies have fallen short to explore how and to what extent the supply time affects costs. A shorter service/supply time can impact on the number of facilities and hence increase the level of decentralization. This in turn can increase set-up and inventory costs due to maintaining stocks at multiple sites. However, decentralization reduces transportation costs which might lower the overall operational cost. These trade-offs have not been investigated in the presence of multiple time constrained services restricting the level of (de)centralization. Considering that a SPL system is set up to cover an area, of course considering the shortest service time option, questions like, 'how do the proportions of demand for different service times affect the costs?' need to be explored.

Service differentiation is well studied in the SPL literature but more attention has been given on the fill-rate based differentiation for distinct customer groups. There are fewer studies that consider the time-based differentiation. Even the studies that do

consider the time-based differentiation assume different average waiting time targets at warehouses, not at customer sites. In real-world systems, equipment support requires the provision of service parts at customer locations. The time to reach customers can play a vital role in terms of meeting service time requirements and this may require considering service facility locations and designing of service areas/zones (area partitioning) in order to ensure that service time targets are met. This broader system view is not taken by the studies in this area. Area partitioning is overlooked in the context of time-differentiated services, i.e. for the systems with multiple service distance constraints. Besides, the only two SPL studies that do consider a distance constraint to allocate customers to facilities, assume discrete demand points. In reality, demand locations can vary over time. Continuous covering considerations can be important so that any changes in the location of existing customers or additions of a new customers do not result in redesigning of the system.

In general, continuous customer location has not received much attention in facility location models either, nor has the area partitioning where there are more than one service time options. There are few methods dealing with continuous geographical distribution of customers in a specific area. These few methods tackle a single clear objective, either to maximize the coverage, or to minimize the average distance. Along with SPL, many emergency service systems also deal with strict and multiple response time targets. The location problems for emergency service facilities are traditionally dealt as covering problems considering a maximum distance constraint. Besides the lack of a focus on continuous customer location, multiple ranges for covering different types of demand and the consideration of inventory cost have not gained attention of the researchers studying the covering problems.

The options in situations where there are different service levels are either to respond differently to different customer requirements, or to provide the most stringent level of service to all customer groups in order to retain the overall satisfaction. A way to respond differently to different service time requirements by customers is to set-up a hierarchy of facilities so that the supply under the relaxed service time constraint is done in a more centralized way by exploiting the longer time allowed to meet the demand. Many real world service systems provide different types of services through hierarchical setups of facilities where different facility types offer different sets of services. Several hierarchical location models have been studied. However, hierarchical systems for time-differentiated distribution, in which different facility types provide different time-based service types, have not been investigated and compared to non-hierarchical (completely decentralized) systems where all facilities provide the complete range of time-based service types.

Table 2.8 provides an overall snapshot of the key studies related to the focus of this research. The table shows that service differentiation and the location aspect are not addressed together, as are continuous area coverage and inventory management.

Table 2.8: Summary table – key studies

ST: Service time, **STD**: Service time differentiation (customer grouping with respect to service times), **CL**: Customer location, **FL**: Facility location, **D**: Discrete, **C**: Continuous, **CC**: Complete coverage (all customers within a service time range), **IC**: Inventory cost, **IS**: Inventory sharing, **TC**: Transportation cost (to reach customer sites)

Study	Factors								
	ST	STD	CL		FL		CC	IC	IS	TC
			D	C	D	C				
Kutanoglu (2008)	✓		✓		✓			✓	✓	✓
Kutanoglu and Mahajan (2009)	✓		✓		✓			✓	✓	✓
Candas and Kutanoglu (2007) and Jeet et al. (2009)	✓		✓		✓		✓	✓		✓
Iyoob and Kutanoglu (2013)	✓							✓	✓	✓
Kranenburg and van Houtum (2007b)	✓	✓						✓		
Kranenburg and van Houtum (2008)	✓	✓						✓		
Kranenburg and van Houtum (2009)	✓	✓						✓	✓	
Suzuki and Okabe (1995) and Suzuki and Drezner (1996)	✓			✓		✓	✓			
Murray et al. (2008)	✓		✓	✓		✓	✓			
Okabe (1995) and Okabe et al. (2000)				✓		✓				✓

2.6. Research contribution

By addressing the research questions stated in Section 1.2 (Chapter 1), the research in a way deals with several of the gaps highlighted above. The research takes several factors into consideration to broadly investigate the distribution with time-differentiated service commitments. The different factors included in the research objectives have not been investigated collectively in the literature. The research generates novel managerial insights into the cost behaviour in relation to service times, and, while considering multiple service time (or distance) constraints, it contributes to the knowledge on hierarchical location problems, (de)centralization, inventory sharing, and continuous covering and area partitioning. The research does so by developing generic quantitative models for broader understanding of the problem.

The research is distinct from the studies in the literature that consider a service distance constraint as it considers a spatial context with continuous geographical distribution of customers together with the inventory aspect. In SPL literature, which is predominately focused on inventory management, studies normally do not consider the time and distance to reach demand as factors in service availability. As pointed out in the previous section, only a couple of studies take the distance factor into account while considering a single distance constraint, but, these studies only consider pre-specified demand and facility location points. On the other hand, the covering problems, considered to be relevant for public emergency service systems, do not consider the inventory factor and also largely consider discrete locations. Moreover, this research incorporates service differentiation, in which different customer groups have different service time requirements, and hence takes multiple service distance constraints into account.

Chapter 3: Research methodology

3.1. Introduction

This research adopts a multi-method approach to address its aim and objectives. Primarily, the research can be related to the Operational Research (OR) discipline. As far as the alignment of a research with a philosophical worldview is concerned, it is an unclear issue when it comes to OR studies. As highlighted by Mingers (2000), there are many ambiguities with regards to the philosophical nature of OR and it is a tangled issue: "Is it [(OR)] science or technology? Is it natural or social science? Can it be realist as well as being interpretivist?"

Besides developing and analyzing quantitative models to investigate the research problem, empirical case studies are performed to gain an understanding of the context of the research problem. This chapter discusses the overall methodology of this research in two main sections. Section 3.2 generally discusses the adopted research approach and highlights its strengths and weaknesses. Section 3.3 discusses the research process and provides details of the specific research stages covered to accomplish this work.

3.2. Overview of research methodology

The investigation in this research is primarily based on axiomatic quantitative modelling. Generally, quantitative model-based research generates rational knowledge. It is based on the assumption that we can build objective models that can explain (part of) the behavior of real life operational processes or that can capture (part of) decision making problems that are faced by managers in real life operational

processes. Models of causal relationships between control variables and performance variables are developed, analyzed or tested. Performance variables can be either physical variables such as inventory position or utilization rate, or economic variables such as profits, costs or revenues (Bertrand and Fransoo, 2002). Meredith et al. (1989) and Bertrand and Fransoo (2002) discuss different classifications of quantitative modelling in Operations Management (OM) research. Bertrand and Fransoo (2002) define normative research and descriptive research as two classifications of axiomatic research (Table 3.1). The quantitative model-based research in this thesis cannot be exclusively labelled as normative or as descriptive. Some parts of the research can be classified as normative (e.g. the optimization models), while some can be classified as descriptive (e.g. the analytical models and simulation study).

Table 3.1: Axiomatic quantitative model-based OM research (Bertrand and Fransoo, 2002)

	Normative research	Descriptive research
Axiomatic research	Normative research is primarily interested in developing policies, strategies and actions, to improve over the results available in the existing literature, to find an optimal solution for a newly defined problem, or to compare various strategies for addressing a specific problem	Descriptive axiomatic research is primarily interested in analyzing a model which leads to understanding and explanation of the characteristics of the model

According to the framework presented by Meredith et al. (1989), axiomatic research is a rational and artificial approach. The research with a rationalist approach is based on the belief that the phenomenon being studied exists out there independent of the research context or the beliefs and assumptions of the researcher (Klein and Lyytinen, 1985; Guba, 1990). Similarly, the artificial approach is based on abstracted and simplified models and is characterized by a significant separation of the phenomenon from the researcher (Meredith et al., 1989). One of the concerns here is that

relationships and observations can be manipulated at will and controlled as needed by the researcher. McKay et al. (1988) highlights this concern as the reason for the gap between OM research's prescriptive advice and workable answers for practitioners.

Näslund (2002) stresses that multiple research methods are much needed since all research questions cannot be answered with the same approach. He states that this, however, is not the correct description of the logistics research in published articles. Dunn et al. (1993) raise the similar concern that "… a given field may be underachieving if all of its research is being conducted within a narrow methodological domain" (cited in Mangan et al., 2004). According to Bolumole (2001), it should be ensured that the exploration of research issues is in-depth and that the research is contextually rich and industrially relevant.

To support the quantitative model-based research and in part overcome the concerns mentioned above, case studies are performed in this research as a supplement. Case study research is an empirical approach to understand real-world phenomena. According to the framework presented by Meredith et al. (1989) to describe the research paradigms in OM, case study research is an interpretive and natural method. The case study method and the axiomatic research method lie on the opposite ends in his framework. Case studies can use multiple methods and tools for data collection in natural settings that consider the temporal and contextual aspects of the phenomena under study (Benbasat et al., 1987; Yin, 1994).

The key strengths and weaknesses of the quantitative model-based research method and the case study research method can be summarised as below.

3.2.1. Strengths and weaknesses of quantitative model-based research

In today's competitive world, a demand for a greater technical competence in managerial decision making has been created by the necessity to efficiently solve the complex problems which arise in modern organizations. At the same time, advancement in technology has also made it possible to process and solve complex business problems (Hillier, 2005). This requirement and opportunity can be catered well by a quantitative model-based approach. The scientific management approach and methods, and techniques developed through OR have been making a serious impact on the design and control of operational processes (Thompson, 1967; Bertrand and Fransoo, 2002; Hillier, 2005). This is especially valid for highly automated operational processes and operations decision problems where the impact of human factor is negligible (Bertrand and Fransoo, 2002). Axiomatic models make trade-offs very explicit and have provided valuable insights in basic tradeoffs at managerial level (Bertrand and Fransoo, 2002). Finally, quantitative modeling tends to yield conclusions with a high reliability and internal consistency (Meredith et al., 1989).

The much debated drawback of quantitative approach is that practitioners seem to view the abstraction of a quantified material as very remote from everyday practice and therefore of little use (Näslund, 2002). Model based research is sometimes not considered very useful to operations managers and practitioners because it fails to recognize the applied nature of production and operations management (Flynn et al., 1990). As the results of mathematical programming are as valid as the assumptions on which the model is based (Flynn et al., 1990), there can be a risk that mathematical models might be based on convenient but unrealistic assumptions. Another critique for model based research is that operational processes can be very complex systems

that are difficult to model from a performance point of view. Performance is generally measured in terms of product quality, production efficiency, and delivering speed and flexibility. These can be affected by many different elements in the process, such as human factors, which are often neglected. As a result, implementation of such problem solutions often turn out to be a tedious process (Bertrand and Fransoo, 2002).

3.2.2. Strengths and weaknesses of case study research

Benbasat et al. (1987) and Meredith et al. (1989) identify relevance, understanding and exploratory depth as three main strengths of the case study approach. They state that the case study approach allows a phenomenon to be studied in its natural setting, which in turn allows to generate a meaningful and relevant theory from the understanding gained through observing an actual practice. Secondly, the case study method allows the much more meaningful question of *why*, rather than just *what* and *how*, to be answered with a relatively full understanding of the nature and complexity of the complete phenomenon. Thirdly, it allows exploratory investigations where variables are still unknown and a phenomenon not understood. Besides this, the case study approach includes a richness of explanation and a potential of investigation in well-described specific situations (Yin, 1994; McCutcheon and Meredith, 1993).

Access and time, triangulation requirements, lack of controls and unfamiliarity of procedures are mentioned by Meredith et al. (1989) as four major disadvantages of the case study approach. Doing case research requires a direct observation in an actual contemporary situation, for which, it can be difficult to gain the required access. Besides, cost and time can also be issues. The data analysis can be difficult and require multiple methods, tools, and entities for triangulation. A researcher may face a lack of control and complications of context and temporal dynamics. Another serious

disadvantage of the case study method is the lack of familiarity of its procedures and rigor by a researcher. A research employing this method can be prone to construct errors, a poor validation, and a questionable generalizability.

3.3. Research process and methods
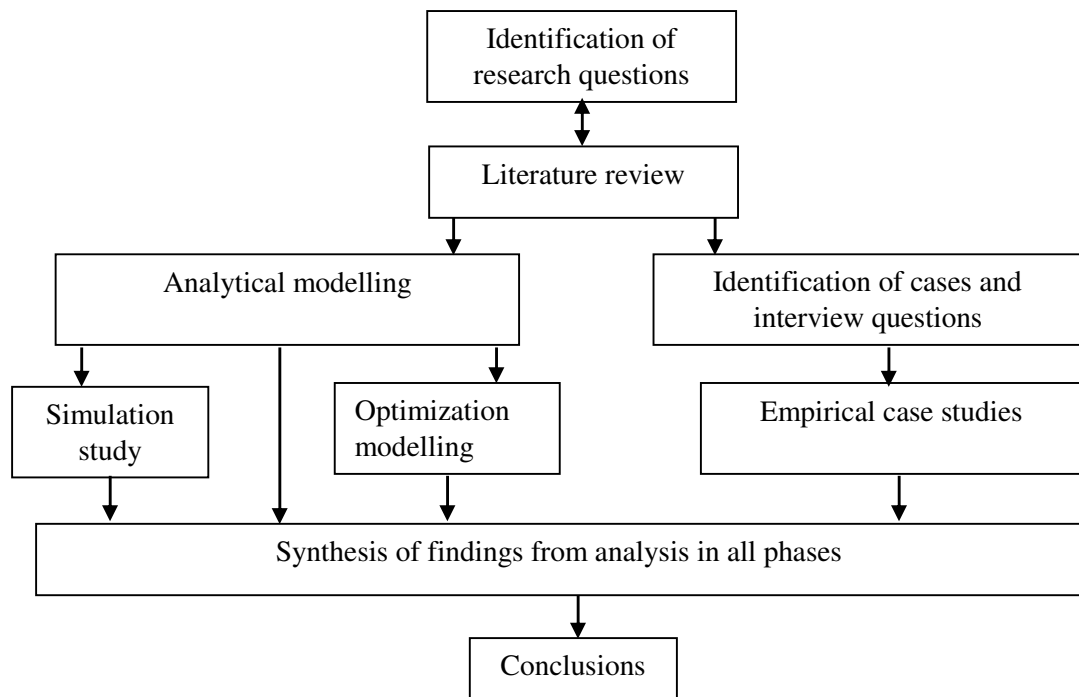
Figure 3.1: Research process

3.3.1. Quantitative modelling study

Quantitative models are developed to generate insights about the impact of service times and the fractions of demand for different service types, as independent variables, on inventory, transportation and setup costs as dependent variables. The costs are compared under the hierarchical and non-hierarchical (completely decentralized) setups. Insights are generated through numerical experiments based on the cost functions using synthetic data and the demand information available from the case studies.

Specifically modelling real-life systems, Like SPL systems, can require substantial time, resources and higher degree of access to organization information. Studying multiple cases in this way may not have been realistic to accomplish this PhD research. Most importantly, specifically basing the research on a particular industrial instance will lack generalization. A more suitable approach to address the research questions can be to formulate the problem independently of any particular instance of the problem in industry. The experimental settings in this research are based on the Euclidean plane. Analytic geometry is used to model and analyse service area partitions according to the service time windows under the hierarchical and non-hierarchical setups. Considering the Euclidean plane and distance to study a problem is not rare in related literature (e.g. Simchi-Levi, 1992; Kutanoglu, 2008; Kutanoglu and Mahajan, 2009).

To begin with, the work does not consider a specific shape or boundary of the area in the Euclidean plane that has to be partitioned into service areas around a number of service facilities. The number of facilities are computed considering the size of service areas, which in turn depend on the service time and the maximum distance that can be travelled from a service facility within that time window. The inventory in the system is computed by realizing the demand handled by each service facility and considering EOQ and (S-1, S) inventory policies and safety stock formula. Similarly, travelling is estimated by realizing the demand handled by each service facility and considering average Euclidean distances within the catchment areas from the facilities.

The analytical model assumes that the Voronoi diagram resulting from partitioning the Euclidean plane has a regular hexagonal pattern. To determine the number of facilities the total area is divided by the area of a hexagonal region, which gives a

lower bound on the number of facilities (Suzuki and Drezner, 1996). These assumptions allow closed form solutions and a flexible and insightful investigation through analytical treatment while capturing all the essential factors in the research scope. However, packing an overall area is not likely to allow full hexagonal regions to be fitted within the boundaries. This can result in a fractional number of facilities and hence underestimation of the number of facilities.

In the next stage, specific cases of facility placements and service area partitions (in form of regular square lattice) are considered inside a rectangular plane for a computational and simulation study. The computational output confirms the insights from the analytical model. The simulation study investigates the impact of distance constrained inventory sharing configurations on transportation and service availability levels (fill-rates) considering different fractions of demand for the service types. Estimating transportation and inventory availability levels numerically can be very challenging due to overlapping coverage ranges of facilities resulting in several sharing zones of different sizes and forms. The simulation models are programmed in C++.

In the analytical model and the computational and simulation study, the setup of facilities and their catchment areas is considered in view of the average distance and demand coverage. Service costs are then analysed with respect to the changes in the fractions of demand for the different time-based service types. For in-depth investigation, finally the problem of locating service facilities and determining their service zones that minimize the service cost is studied through a Non-Linear Programming (NLP) model and a Mixed Integer Non-Linear Programing (MINLP) models. The models consider both inventory (based on EOQ) and transportation costs

to determine the optimum setup of facilities and service zones. This allows a better understanding of the trade-offs of inventory and transportation costs in setting up a distribution system with service-time constraints. The MINLP model also considers the demand fractions for the different service types to generate optimum hierarchical setups. The problem being multifaceted and especially complex as demand and possible facility locations are considered continuous, the problem is explored in one dimension (i.e. customers' geographical distribution is considered to be over a line segment (or a route)).

3.3.2. Empirical case studies

The empirical case studies are carried out to complement the quantitative analysis so that this research is aware of the reality. The case studies indicate the motivation for the modelling study and its relevance to some of the real-world systems. The case studies also offer general insights into the real world service operations providing time-differentiated services requiring parts or material to be delivered at demand sites. The research in logistics is predominantly quantitative (Mentzer and Kahn, 1995; Ellram, 1996; Näslund, 2002; Mangan et al., 2004; Frankel et al., 2005; Mello and Flint, 2009). As is the case with general logistics literature, the specific research on SPL is also predominantly quantitative, with intensive focus on planning and operational issues as compared to long term strategic and design issues (Wagner and Lindemann, 2008). Most of the research related to SPL employs mathematical/OR techniques. However, a number of studies employing OR techniques also conduct case studies initially to understand the context of the problem (e.g. Cohen et al., 1990; Cohen et al., 1999; Rustenburg et al., 2001; Deshpande et al., 2003a; Khawam et al., 2007). In contrast to this approach, some researchers develop mathematical models

and then gather data (mostly quantitative) through case studies to validate the models and/or demonstrate how the modelled policies can behave in real world (e.g. Braglia et al., 2004; Kranenburg and van Houtum, 2007b, 2009). There are also some pure case studies and industrial reviews associated with SPL (e.g. Lele and Karmarkar, 1983; Levitt, 1983; Cohen and Lee, 1990; Cohen et al., 2000; Dennis and Kambil, 2003; Wagner and Lindemann, 2008). These provide an understanding of SPL practices, SPL related issues being faced by the companies, and future challenges. The insights in these case studies have been used as motivational factors in some mathematical studies. As far as public sector emergency service systems are concerned, apart from few studies on medical services, studies investigating different strategic aspects and characteristics of emergency service operations could not be found.

Besides allowing some fine tuning of the analysis based on the quantitative models, and making the discussion contextually rich and industrial relevant, the empirical case studies in this research also reveal new insights into the state of the art SPL systems and emergency service systems in practice. The case studies are conducted in stages similar to those suggested by (Stuart et al., 2002): 1) defining research questions, 2) developing instruments for data collection, 3) data gathering, 4) data analysis, and 5) dissemination.

3.3.2.1.   Interview questions

A set of open-ended and semi-structured questions were identified around the topics which are relevant to the overall research theme and the subject of the presented quantitative modelling study. The framework for inquiry from the ICT case companies is adapted from Cohen et al. (1997) and Wagner and Lindemann (2008). To gather

additional information, a special focus is given on the range of service contracts, demand characteristics, procedures for satisfying demand, network structure and supply capabilities, and inventory policies and rules. The questions asked during the interviews in the ICT case companies (Appendix 1) cover the following topics:

- Range of services/service contracts

- Network structure and capabilities of supplying service parts

- Procedure for satisfying service requests

- Service parts/materials characteristics

- Demand/customer characteristics

- Inventory policies and stocking rules

- Sourcing of service parts/material

- Service cost characteristics/cost structure

- Issues in managing service parts logistics

- Management trends

For the Highways Agency, being a public sector organization as opposed to the commercial ICT case companies, the interview questions are altered, although the overall line of inquiry is kept the same (Appendix 1).

3.3.2.2. Selection of case sectors and companies

After-sales services are more critical for certain kinds of products compared to others. Service providers for ICT products are normally required to guarantee service within short time windows. Companies spend a huge amount of capital on their IT equipment and many of their critical operations are dependent on such equipment. Malfunction of just one computer component can halt many central operations. As service times are comparatively more sensitive in IT sector, an investigation here provides useful

insights to understand the advanced SPL management and is compatible to the theme of this research. The service time options in automotive and aerospace sector range from 24 hour to a week and 12 hours to a week respectively (de Souza et al., 2011), while in the IT sector, the service time range can start from 2 hours.

Relevant position holders in well-known ICT hardware companies and service providers were approached with the information requests. A positive response was received from two organizations, which led to the face-to-face interviews and correspondence through e-mails for the purpose data gathering. These two case companies are major players in the sector on the national as well as global level. The organizations are referred as '*Company A*' and '*Company B*' in this document (Table 3.2). Company B is a frequently cited case in the academic literature as an example of a state of the art company in SPL operations.

Table 3.2: ICT case companies

Company	Presence	Country of origin	Employees
Company A	Global	Japan	170,000 +
Company B	Global	US	430,000 +

Short response times are also commonly associated with public sector emergency services such as ambulance, fire and rescue, and police services. A road/traffic incident can require different forms of emergency responses including infrastructure repairs. In England, the emergency repairs on the highways in response to traffic incidents is the responsibility of the Highways Agency. The repairs require different materials and have to be done within certain time limits. Being a potential beneficiary of this research, the case of the Highways Agency (Table 3.3) is studied to understand its service delivery operations with a focus on the repair services.

Table 3.3: Public sector emergency service provider case

Agency	Parent government department	Presence/ Jurisdiction	Employees
The Highways Agency	Department of Transport	England	3,400 (in 2012-13)

3.3.2.3. Data collection

Face-to-face interviews were conducted in the offices of the participating ICT companies between June 2011 and August 2011. The interviewee from Company A holds the position of '*Business and Systems Manager (Service Business)*'. He is responsible for supply and lifecycle services, spares management, service delivery, and systems management in the company's UK service business. The interview took place at company A's head office and lasted for around 2 hours. The interviewee from Company B holds the position of '*Service Logistics Business Manager (Service Parts Operations)*' and is responsible for the overall management of the company's SPL in the UK and Republic of Ireland. The Company B's representative was interviewed for around three hours in total over two sittings. The Highways Agency forwarded the interview request to the Operations Manager (service delivery) of the agency's Midlands region as the relevant person. The Operations Manager was interviewed on telephone in March 2014 for around forty minutes.

Semi-structured questionnaires were developed to guide the interviews. Prior to the interviews, the interviewees were informed about the areas that would be discussed. After the interviews, the interview scripts were prepared and sent to the interviewees via e-mails to confirm our understanding. Follow-up questions were also e-mailed to the participants. These follow-up questions included clarification of some of the earlier responses and the questions that could not be answered during the interview due to the

information not being readily available. To ensure confidentiality, participants from the ICT companies were also requested to highlight any information in the scripts that can potentially be of a sensitive nature. Demand data of an average service part was requested from both ICT companies in April 2012. Company B accepted this request and provided this information (demand over six months) through an e-mail. This information includes the location of warehouses and the total demand in the locality of each warehouse. This data allows us to assess the hierarchical and non-hierarchical setups partly considering the real world settings.

Chapter 4: Empirical case studies

4.1. Introduction

This chapter presents the findings from the empirical study involving SPL systems in two ICT companies (Section 4.2) and the service operations of a public organization providing emergency responses (Section 4.3). Section 4.4 provides a synopsis of the analysis and highlights key findings from the empirical study.

4.2. SPL for computer hardware support services: The cases of two multinational companies

The case companies are well known global IT brands providing a wide range of products and services. These companies provide IT solutions to clients in various industries, including, but not limited to, automotive, financial services (banks), healthcare, energy and utilities, manufacturing, transportation, and retail. One of the case companies also serve central and local governments. Their hardware products include servers, personal computers, and storage and networking equipment. Both companies offer a range of business and IT services which include maintenance and technical support for IT equipment. The case companies operate large scale SPL systems in the UK to support computer hardware infrastructures of corporate customers by providing service parts within short time windows. Table 4.1 provides a brief overview of the scale and scope of the IT hardware service operations of these companies in the UK.

Table 4.1: An overview of IT hardware service operations of the case companies in the UK

Case companies	Scope of service parts distribution services	Scale of service parts distribution operations
Company A	- offers 2 hours, 4 hours, 8 hours and next day service response time commitments (warranty contracts offer two business days response) - services products including networking equipment, printers, electronic point of sale devices, personal computers, printers, circuit boards, and chip and pin devices - maintains around 23,000 Stock Keeping Units (SKUs) in service parts inventory - has the capability to repair 45% of the parts	- has 18 warehouses and 300 secure/lock boxes in the service parts distribution network with the capability to reach any customers within 90 minutes - dispatches around 20,000 service parts per month
Company B	- offers 2 hours (response or fix*), 4 hours (response or fix*), 8 hours (response or fix*), next business day and 2nd business day service options - services products including servers, storage devices, networking equipment, printers, electronic point of sale devices, ATMs, and pumping stations - maintains around 17,000 SKUs in service parts inventory - has the capability to repair 90% of the parts * 'fix' option means that the service time commitment corresponds to the fixing of parts at customer sites rather than just reaching at customer sites.	- has 17 warehouses, 4 island stocking points and 119 pickup drop off points in service parts distribution network with the capability to reach any customers within 60 minutes. - dispatches approximately 4,200 service parts per month

The analysis of the gathered information in consolidation with the findings from the literature is arranged in the following subsections. Section 4.2.1 examines the range of services offered by the case companies and discusses the nature of demand. Section 4.2.2 discusses the characteristics of service parts. Sections 4.2.3 to 4.2.5 cover several important aspects of service parts distribution. These sections provide an analysis on strategic and operational decisions that can define a company's parts distribution policy. These decisions, as highlighted by Cohen and Lee (1990), relate to stock locations, replenishment policies, sourcing of stocks, and transportation. Section 4.2.6 discusses the inventory and transportation costs, which are two major components of

SPL operating costs (Cohen et al., 1997). Section 4.2.7 discuss the issues and trends in the management of SPL in the case companies.

4.2.1.   Range of services and demand characteristics

The delivery of differentiated levels of service to disparate classes of customers is an increasingly important requirement in today's customer centric environment (Deshpande et al., 2003b). The case companies provide a wide range of service response times ranging from 2 hours to second business day. The same day services by both companies include 2 hours, 4 hours, and 8 hours responses. These service times are offered by the companies as part of the service contracts with a clients and there can be different service time commitments for different equipment with one customer. According to Company A, most contracts have specific terms regarding which type of equipment requires what service time level. At Company A, the service time commitments are normally associated with the time to supply requested part(s) and to get service engineers at customer sites. Company B provides more flexibility in setting contracts by allowing customers to choose whether 2 hour, 4 hours, and 8 hours windows associate with the arrivals of requested parts and engineers at customer sites, or the fixing of parts. In a case of 'fix' service, part(s) and a service engineer typically reach a customer site an hour before the end of the service time window. For example, for '2 hours fix' service, the service part(s) and an engineer reach a customer site within one hour. Besides the above stated service options, Company B can also tailor service contract terms for customers to include other service times.

Figures 4.1 (a) and (b) show approximate proportions of service calls for different service time options excluding warranty service at Company A. Within the same day services, the strictest service time option, i.e. 2 hours service, is the least common. The

proportion of calls for the other two same day service time options, i.e. 4 hour and 8 hour services, do not differ significantly at both companies. There is also a high proportion of demand for the next day service. Specifically at Company A, a high majority of service calls are not for the same day services. This can probably allow a higher level of stock centralization to meet customer demand. Allowing to meet the entire demand for the next day service from single stocking facility might result in lower inventory levels and hence lower inventory costs.
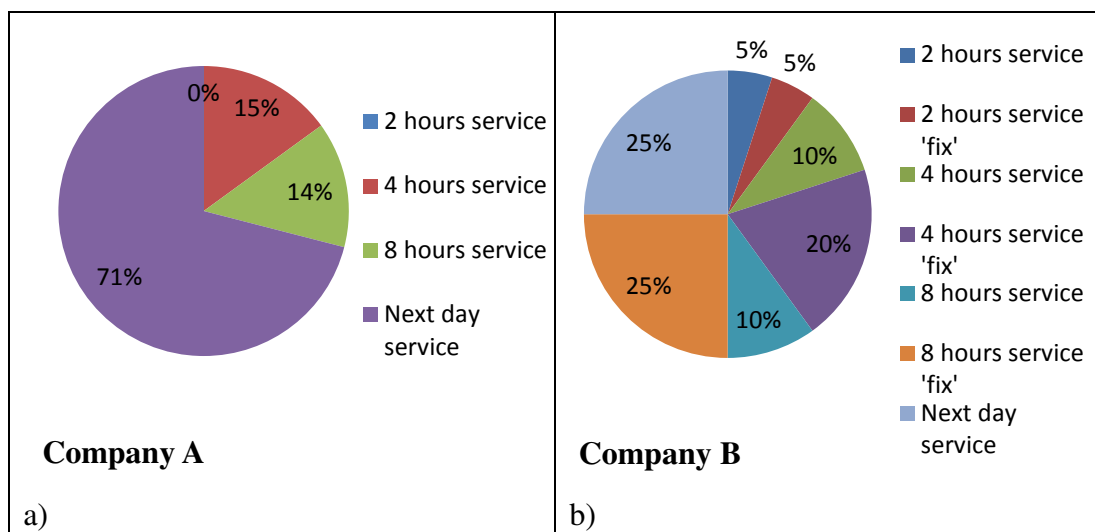
Figures 4.1 (a & b): Approximate proportion of calls for different service time options

Company A provides a standard availability level for all parts, customers, and equipment. Company B however offers different part availability levels that can be set for a certain part type, customer, or equipment.

There is a considerable difference between the numbers of service calls at both companies. Company A and B approximately receive 20,000 and 3,500 service calls per month on average respectively. According to Lele and Karmarkar (1983), most repairs require several parts. Contrary to this, we found that supplying only one part per service call is common in both companies (around 1.2 parts per service call at

Company B with daily part consumption worth around £28,000). Modularity of components might be a factor in this.

The customers' geographical distribution is not uniform. Customer locations are clustered and demand at certain service facilities is reported to be significantly higher than others. However, according to Company B, though clustered, there is a country wide spread of their customer locations. Jalil et al. (2011) show that installed-base information can be useful for an efficient SPL management. However, both companies could not provide information regarding the number of units in the installed base. According to Company B, the installed base for the UK is not precisely known because of the huge number of products under different ranges. Warranty and ad hoc services extended to the customers also affect the ability to forecast the number of units under service. Besides, there are a lot of umbrella customers (customers having several further customers, e.g. local governments) which adds to the complexity in this regard. Company B's customers include around 35 retailers and 516 local governments (which include multiple bodies e.g. police, and fire and rescue services). Another relevant issue is the existence of global contracts which are sometimes not clearly known in the UK.

4.2.2. Service parts characteristics

Cohen et al. (1997) and Dennis and Kambil (2003) report that computer hardware service providers have to maintain an extensive range of service parts. This study also found that the case companies maintain a great variety of service parts in inventory. Company A and B maintain 23,000 and 17,000 SKUs respectively.

Parts delivered to customers are not necessarily new. Company B highlights different factors on which the supply of new or used parts depends. It can depend on the stock

availability at the time of service call and the speed of delivery required. Besides, where legislation mandates, all warranty parts supplied to the customers are new. Large percentage of warranty parts are sent back to Original Equipment Manufacturers (OEMs) for credit or replacement. Parts out of a warranty are more likely to be repaired or reutilized. Providing new or old parts also depends on the lifecycle stage of equipment. For newly announced products the service parts are more likely to be unused. Both companies report that a high percentage of service parts that they maintain are repairable, however, these are not necessarily repaired. Sometimes the parts with minor faults are just replaced and not repaired. Approximately 45% of the service parts at Company A can be repaired. Company B informs that approximately 90% of their parts can be repaired, however, only 60% are repaired.

The lifecycle length of the products requiring service parts are considerably short. The equipment served by Company A have lifecycles ranging from three (e.g. PCs and servers) to seven years (e.g. ATMs). On average, the lifecycle of an equipment served by Company B is six to seven years (equipment with a longer lifecycle can be in use for eight years, whereas, equipment with a shorter lifecycle period can be in use for five years). After obsolescence, Company A normally scraps the service parts with a revenue sharing system in place with their suppliers. That is, the burden of loss through obsolescence is shared between Company A and their suppliers. The unwanted parts at Company B are dealt with in different ways. These parts can be scrapped straightaway or harvested before scrapping or resold.

Fortuin and Martin (1999) and Wagner and Lindemann (2008) report that service parts inventory turnover is often very low as the use of service parts is based on either a failure of a product in the field or on a consumption of a `usage' part. Other factors

contributing to low demand can be the increased reliability and quality of products, and the increased product variety and thereby the reduced installed base of specific models (Fortuin and Martin, 1999). Our findings are in line with the supposition that service parts are predominantly slow moving (Cohen et al., 1997; Dennis and Kambil, 2003). Fast-Normal-Slow (FNS) analysis is a popular SKU classification approach where SKUs are labelled as fast, normal, or slow movers based on their demand volume in a specific period (van Kampen et al., 2012). For example, SKUs with more than 10 orders during the replenishment lead time can be classified as fast moving (Silver et al., 1998), however, there are no standard demand rates for the classification in the literature. Figures 4.2 and 4.3 present the FNS analysis at the case companies. We are though unaware of the specific definition of the classes used by the companies. Figure 4.2 shows the percentages of fast, normal and slow moving parts at Company A. A very high percentage of parts are slow moving (around 85%), whereas the percentage of fast moving parts is considerably low. 13 % of the parts have normal demand rate. Figure 4.3 shows the percentages of demand corresponding to fast, normal and slow moving parts in Company B. The information is based on past demand data over 12 months. An extremely low proportion of overall demand is accounted for the fast moving parts. Similarly, the proportion of demand that corresponds to the normal moving parts is also quite low. This picture however is clouded by the use of '*boot stocks*' with service engineers and local purchases of items such as print-heads for retail printers.
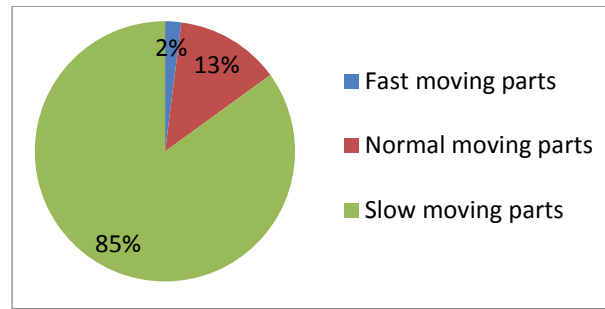
Figure 4.2: Percentages of parts classified as fast, normal and slow moving at Company A
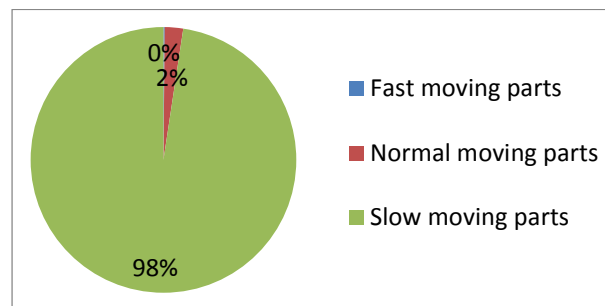
Figure 4.3: Percentages of demand corresponding to fast, normal and slow moving parts at Company B

4.2.3. Network structure and distribution capabilities

The case companies distribute parts through three echelon networks with dual role central warehouses in contrast to traditional arborescent structure. However stocks are not typically maintained at the facilities in the third echelons. These facilities mostly act as transit points near customer sites where the requested parts are stored temporarily before service engineers collect and carry these parts to customer sites. Island stocking points (called Island Kits) in Company B's distribution network are facilities in the third echelon where stocks are maintained. Totalling four in number, extra stocks are kept in island stocking points due to the logistical complexities in shipping parts to the islands. Figure 4.4 and Table 4.2 describe the service parts distribution structures of the case companies. Both companies have very similar distribution structures. All stocking facilities, including the central warehouse, provide

parts to the nearby customers. That is, besides replenishing the local warehouses (or field stores), the central warehouses also serve customers directly via transit points. None of the SPL studies that we have reviewed consider this type of distribution structure. Table 4.2 provides information on the facility numbers at each level in the distribution networks.
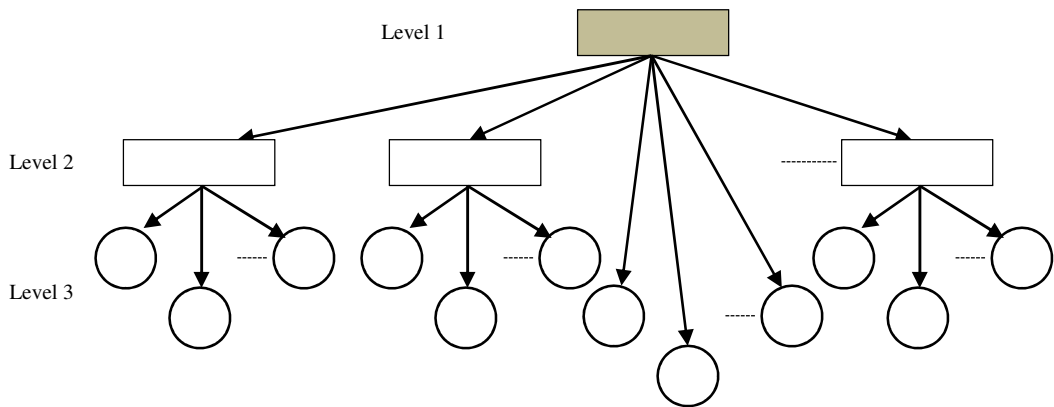
Figure 4.4: Service parts distribution structure – Company A and Company B

Table 4.2: Number of facilities at different levels of service parts distribution networks of the case companies

Levels	Company A	Company B
Level 1	1 Central Warehouse	1 Central Warehouse
Level 2	17 Local Warehouses	16 Field Store Locations
Level 3	300 lock/secure boxes	117 Pickup drop-off points, 2 lock boxes and 4 island stocking points

Dedicated facilities for each customer niche are common in several industries (Boyaci and Ray, 2006). Service systems often include hierarchical facilities in terms of the types/levels of the service they offer (Jayaraman et al., 2003). However, with the exception to next day service responses by Company B, all stocking locations (central and local) in both case companies provide the full range of service time options. In other words, all warehouses are similar in terms of service options they offer.

As discussed in the literature review, the delivery performance of service parts suppliers, in particular the reliability of agreed delivery times, can be critical and problematic (Fortuin and Martin, 1999; Candas and Kutanoglu, 2007). Any uncertainty in the demand process has to be compensated by the flexibility of the delivery process (Fortuin and Martin, 1999). Both companies regard the number of stocking facilities as sufficient for meeting customer requirements. Company A went further to state that there might be an opportunity to reduce the number of warehouses and still meet the requirements. Company A and Company B have the capabilities to reach all their customers in the UK within 90 minutes and 60 minutes respectively. Company A has the capability to reach customers 30 minutes earlier than the strictest response time window it offers (2 hours response), while Company B has the capability to exactly meet the strictest response time window it offers (2 hours 'fix' requiring one hour response). However, as seen earlier, the proportion of demand for the strictest service time window is low in both companies. That is, the decentralization levels of these systems are forced by the small proportions of demand. It is interesting to observe that though Company A has a higher number of facilities, Company B has better capability to reach customers (in terms of response time) bearing in mind that Company B's customer spread is country wide.

Company A's facilities are located near major metropolitan/business hubs, where more customers are clustered. The facility locations are strategically selected to reach the clients within the required time limit. Company B also considers customer base and geographical coverage to decide the location of storage facilities besides the costs.

4.2.4. Procedure for satisfying service requests

All service calls are received centrally. Mostly customers perform diagnosis themselves to determine which part to order. Company B uses multiple diagnostic tools to determine the parts required for service, including '*human*' - (call screening/front office) and '*remote*' – (machine driven dial home or diagnosis). At Company A, the nearest stocking location where the requested part is available is selected to supply the part in response to a service call. Parts are delivered to '*lock boxes*' or '*secure boxes*' near customer sites from where these parts are collected by service engineers for installation at customer sites.
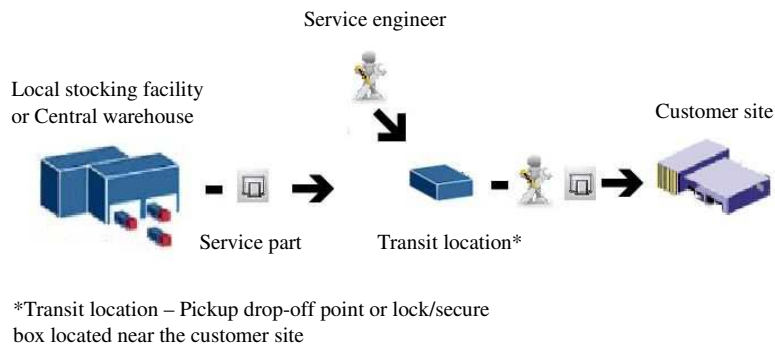
Figure 4.5: Procedure for meeting customer service requests involving service part supply (Company A and Company B)

At Company B, the procedure to deliver parts for same day service calls is similar to that in Company A (see Figure 4.5). However, if a request is for the next day service, the request is met from the central warehouse in the UK or from the Netherlands, so that the stocks at a Field Store Locations are not sacrificed for low priority demand. Inventory is also managed at customer sites for critical products and sometimes engineers are posted at customer sites for '*mission critical*' issues. If a part has to be recovered, a service engineer collects the part from the customer site and returns it to any of the Company B sites, from where it reaches the central warehouse in maximum

three days. Figure 4.6 depicts Company B's reverse logistics process involved in service parts recovery. Company B recovers a faulty part from a customer if the part is 1) valuable, 2) sensitive, e.g. the ones used in weapon systems, or 3) under warranty. Parts can be repaired in-house at Company B, by a major repair vendor, or by an OEM depending on the situation. For example the faulty parts under warranty are sent to OEMs for repair.
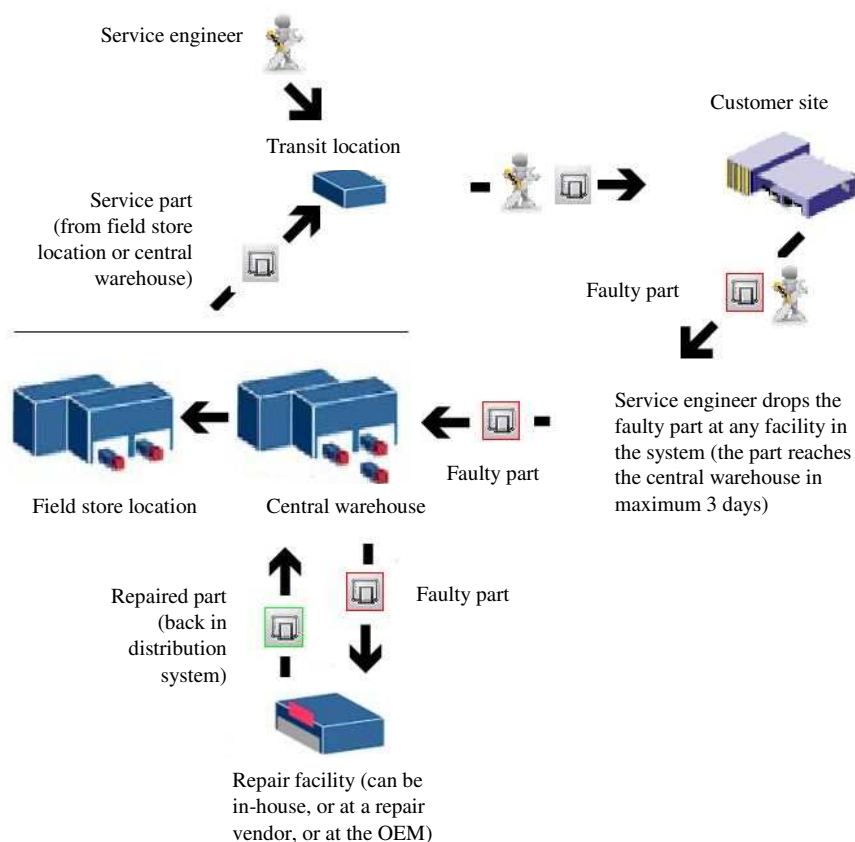
Figure 4.6: Service part recovery (Company B)

Both companies partially outsource servicing at customer sites to third party service engineers. Company A employs 400 service engineers as their own staff and around 30% of the services are carried out by third party engineers. Company B has 350 service engineers in their workforce whereas approximately half of their service

requests are met by third party service engineers. The transportation function for supplying parts is also outsourced in both companies.

4.2.5. Inventory policies, stocking rules and sourcing of service parts

Various inventory policies have been considered in service parts management studies. However, the one-for-one (S-1, S) inventory replenishment policy is considered to be the most appropriate policy for managing low demand service parts. Service parts inventory at Company A is controlled according to both 'minimum/maximum' inventory level and one-for-one inventory replenishment policies. Company B mostly uses the one-for-one replenishment policy for the inventory control. The appropriateness of the (S-1, S) policy for slow moving expensive service parts is not refuted by most authors. However, Cohen et al. (1990) states that the (S-1, S) policy does not provide an adequate cost and service performance for a wide range of demand rates (as some service parts can be inexpensive and fast moving). Nonetheless, the proportion of fast moving parts is very low in the case companies (Figures 4.2 and 4.3), which suggests that the use of (S-1, S) policy is appropriate.

Both companies use forecast driven software packages to manage inventories. Company A uses a commercially available software system known as 'add*ONE'. Company B uses a real-time company developed software system, providing an overall visibility of the stocks in the system. Company B considers mean-time-between-failure for demand forecast. Company A provides a standard availability level of 95% for all parts, customers and equipment. Whereas, Company B can offer availability levels ranging from 85% to 99% for different parts, customers and equipment.

There is a central visibility of stock levels at all stocking locations in both companies. The procedure for stock replenishments at local warehouses (or field stores) from the central warehouse is also similar in both companies. Local warehouses (or field stores) are replenished overnight from the central warehouse. A high percentage of stocks are deployed at the central warehouse (approximately 75% of the total stocks in both companies are maintained centrally). Company A nearly maintains the full range of parts in each stocking facility. However, exceptionally expensive parts are only kept centrally. Company B does not necessarily maintain the same range of parts at all locations. The parts that are maintained depend on the customer requirements for the region where a particular field store is located in. Parts are also maintained only centrally by Company B if they are expensive or bulky or if they have a very low demand.

For most parts, deliveries at Company A's central warehouse are received from suppliers within 2 to 3 days after orders. Parts are not only sourced from OEMs but also from brokers and small suppliers. Replenishment lead time at Company B's central warehouse vary significantly. Normally the replenishment time for buffer stocks is around three days, whereas it takes around one day for emergency replenishments. Company B also sources stocks from brokers and open market along with OEMs. However a preference is given to manufacturers for sourcing the parts. Many parts are bought as single lots from OEMs. The lead time for a new buy can be significantly higher (0-90 days) than the lead time for parts sourced from a secondary market (couple of days).

4.2.6. Service cost characteristics

There is a huge disparity in the cost of service parts maintained by the case companies. The cost of parts in Company A ranges from £0.01 to £28,000, with the average cost of around £144. The cost of parts in Company B's inventory ranges from approximately £10 to £0.25 million, with the average cost of a part at around £1000 (cheap parts, such as cables and screws, are kept by service engineers as boot stock with a fixed allowance). Inventory can represent a significant proportion of costs in any distribution system (Jayaraman, 1998). Specifically in SPL, inventory investment can be the largest single factor in an average total cost structure (Cohen et al., 1997). The average part cost, especially at Company B, is very high, which indicates high inventory holding costs in terms of the capital tied up.

Company B's current inventory reduction target is around £1 million. The cost of delivering a part to a customer is also very high. For same day delivery, the transportation cost incurred by Company A is 60 to 75 pence per mile, while for Company B it is 100 to 125 pence per mile. A factor due to which the transportation costs can be even higher is that service parts and service engineers separately complete a part of their journeys to customer sites.

There are several publications highlighting the revenues and profits associated with service parts and after-sales services (e.g. Cohen et al., 1997; Dennis and Kambil, 2003; Poole, 2003; Cohen, 2005; Sang-Hyun Kim et al., 2007; Candas and Kutanoglu, 2007). The spare parts business is often considered as the highest profit generating function (Suomala et al., 2002; Wagner and Lindemann, 2008). The study by Wagner and Lindemann (2008) reported that because of the high profit margins that can be generated in the service parts business, companies are willing to increase, and in some

cases rely on increasing their service parts sales. However, some studies also indicate that service businesses are highly competitive and face cost and performance pressures from customers (Cohen et al., 1990; Cohen et al., 1997; Wagner and Lindemann, 2008). A contrast is found in how both case companies perceive their service businesses. At Company A, the service business is not considered as a high revenue and profit generating function of the organization. It is rather perceived as a necessary support function. Company B realizes that servicing certain families of products is highly profitable, while for others it is not so profitable. That is, the profitability depends on the market segment or the product category. However, overall, the service function is considered to be of a high strategic importance and is considered profitable at Company B.

4.2.7. Issues in managing SPL and management trends

Cohen et al. (1997) mention that after-sales services in computer, communications, and other high-tech industries are facing an escalating pressure to improve both the level of service delivered to the customers and the productivity in providing these services. Company A confirms that there is a considerable cost pressure from their customers to achieve low maintenance costs. Besides, obsolescence of parts can be a significant issue depending on the equipment. Benefit from reducing the existing stocks can be insignificant due to the very low resale value of parts. Nevertheless, the burden of the loss because of obsolescence is shared between Company A and their suppliers under a revenue sharing system. Company A, which had merged with another major company in the sector, also faces a stocks consolidation issue. Due to this recent merger, some parts in the inventory have the same description but a different identity, resulting in duplication of stocks.

Company B also considers obsolescence as a big issue and states lifecycle management as a significant challenge. Parts worth millions are scraped. On the other hand, some products are in operation years after the manufacturing is discontinued. Supporting such products is expensive. Supporting and procuring parts for low capability old products can be expensive than supporting and procuring parts for similar high capability new products. For example, supplying a hard disk with a higher capacity is more convenient than finding and supplying a 20GB hard disk which is not easily available in the market. Managing availability versus cost is also identified as an issue. The cost of parts is going up while the cost of missing service commitments can be very high. For instance, banks apply heavy penalties on hourly basis for an unavailability of a service for the products that are critical. Flexibility in service contracts also sometimes leads to complexities in service operations. Besides, the software system in use is not totally automated as Company B's staff has to manually check some of the information for each service request (e.g. whether the request is covered under the contract or whether the customer needs to pay extra). The software system can be fully automated, and hence streamline the service operation if contract terms are fixed and rigid. But this can then compromise the flexibility for the customers. Customers can also change the service levels during the contract tenure. This again provides a flexibility but considerably increases the complexity. Another issue related to service contracts is that sometimes there is a lack of information or clarity regarding their nature. Some customers wrongly assume that their service contracts provide a global coverage, whereas on some occasions, customers have a global coverage but the operations in other countries are not made aware of it.

In short, Company A highlights the cost pressure from customers, stock duplication, obsolescence and low resale value of parts, and Company B highlights obsolescence

(life cycle management), service availability v/s service cost trade-off, maintaining old equipment, flexible contract terms, and lack of knowledge or clarity regarding contract nature as issues in managing their SPL systems. In terms of service time commitments, both companies indicated that there is no requirement from the customers to provide service in shorter time windows than what are already on offer.

Table 4.3: Issues in managing SPL

	Company A	Company B
Highlighted issues in managing SPL	- Cost pressure from customers - Obsolescence - Stock duplication - Low resale value of parts	- Cheaper and better options available to customers in certain market segment - Obsolescence - Service availability v/s service cost trade-off - Maintenance of old equipment - Flexible contract terms - Lack of knowledge regarding contract nature

Both companies indicate an increasing trend to outsource services at customer sites to third party engineers and technical couriers. Company B also points out that commoditization is increasing in the industry and products are becoming smaller and more modular.

The percentage of demand met on time is a major performance metric of customer service (Cohen et al., 1999). Company A constantly meets Logistics Service Level Agreements but sees an opportunity for improvement in service costs and the integration of the logistics and the call management system. It is recognized that the cost per call can be reduced, leading to improved future bids to customers. A fully integrated logistics and call management system can help to drive further cost reductions. Company B's on time response is also quite high (99% of the customers are reached on time). Company B realizes that there is a significant competition in the service business in IT sector and considers itself very competitive in serving the top

end IT products. On the other hand, for the service of low end products, there are several cheaper options available to its customers. Also, the customers in this market segment (low end products) demand comparatively more substantiated service than what is offered by Company B.

4.3. Time constrained provision of materials as a part of an emergency service: The case of the Highways Agency

The Highways Agency is an executive agency of the Department for Transport (England), and is responsible for operating, maintaining and improving the strategic road network in England. The strategic road network in England is around 4,300 miles long and comprises motorways and trunk roads (the significant 'A' roads). While the Agency's network represents only 2% of all roads in England by length, it carries a third of all traffic by mileage. Two thirds of all heavy goods vehicle mileage in England is undertaken on the strategic road network, making it the economic backbone of the country. In what follows, the structure of the Highways Agency's service system is discussed in Section 4.3.1, the response procedure and the types of incident responses are discussed in Section 4.3.2, the inventory and demand characteristics are discussed in Section 4.3.3, and the prospect for improvements is stated in Section 4.3.4.

4.3.1. Structure

The service operations of the Highways Agency are divided into seven main regions and there are area offices at Bristol, Bedford, Manchester, Birmingham, Dorking, Exeter and Leeds. This brief case study is focused on the Midlands region and is mainly concerned with the repair works carried out in response of an incident on a

highway. The Midlands region has three out-stations (depots) and one control room. Besides, there are seven stand points, from where the responses are provided. All facilities operate 24 hours a day. The facility locations of the agency are historical. It cannot be said what the considerations were in locating these facilities. The agency tries to utilize these facilities in an effective manner to make the best use.

There are several contractors (known as Managing Agent Contractors) performing different jobs required in the incident responses. From now onwards, we refer to a Management Agent Contractor that performs repairs after an incident as a '*repair contractor*'. Most regions, like the Midlands, have only one repair contractor. The repair contractor has twelve facilities (warehouses) in the Midlands. Some of the facilities are not fully operational during the late hours (i.e. some facilities are toned down at night). The stocks at the repair contractor's facilities are directly sourced from the manufacturers of the items required for repairs. There is no central warehouse, i.e. the system has just one-echelon. Some of the warehouses only have specific roles, e.g. there are two warehouses which are dedicated for lighting repairs only.

4.3.2. Incident response

An incident can be graded as '*immediate*', '*prompt*', '*routine*', or '*no-response*'. An incident of the immediate grade requires immediate response as there is a risk to a life or there is an effect on a live lane. A prompt grade incident is the one that does not affect a live lane but still requires a response. The other two grades, i.e. routine and no-response, do not require a response, although an incident of a routine grade requires checking during a patrol. Service calls for the incidents requiring a response are logged centrally at the regional control centre, and, without any delay, the calls are forwarded to the repair contractor through the Highways Agency's contact centre. The contractor

is given as much details as possible, and the contractor sends the equipment they know will be required from the information they have. The contractor also sends a bronze commander who makes an assessment of what further equipment are required.

A response is normally provided by the nearest stand point of the agency and the nearest contractor facility. For the agency, there is one standard response time window of 20 minutes between 06:00 and 22:00 hours. This response time limit is however not binding during the remaining hours (22:00-06:00). The on-time response target is 90%, which only applies to the service between 06:00 and 22:00 hours.

The repair contractor however has '*resolution time*' targets and not response time targets. The target resolution time starts after the agency, on completion of its own tasks, hands over the incident scene to the repair contractor. Between 06:00 and 22:00 hours, the target resolution time for the contractor is one and a half hours normally, but can vary depending on the incident scale and type. The target resolution time for the contractor between 22:00 and 06:00 hours is 2 hours. Hence, effectively, there is a time-based differentiation based on the time of day as the repair service between 22:00 to 06:00 hours can be provided in a longer time-window than the standard one and a half hours.

4.3.3. Inventory and demand characteristics

There is a vast range of materials used in the repairs (e.g. different sizes and types of barriers, sign boards etc.). The repair contractor is responsible for maintaining inventories completely, however, the inventory is owned by the agency. The contractor keeps a nominal inventory level. The agency does not define any inventory levels for the contractor. Also, the inventory levels are not known by the agency at all

times, but the agency does occasionally audit the inventory kept by the contractor. The agency is mainly concerned about the timely service by the contractor.

Overall the inventory cost is not considered as a big cost by the agency. Obsolescence happens but it is not a big problem as well. Inventory can become obsolete for a particular route, but there are certainly other routes where the stocks can be used. As far as demand is concerned, it is not high as well. There are four to five pothole repairs and one to two collision damage repairs per day on an average. However, the demand is sporadic and the unpredictability of the demand is a challenge for the Agency. The demand cannot be defined as clustered or more concentrated at few specific locations. It cannot be predicted where incidents are going to happen, although the agency maintains more equipment at busier links. Hence the system can be described as a network with continuous spread of demand locations requiring every location to be covered within the strictest service time.

4.3.4. Opportunities for improvements

The interviewee suggests that the main improvements can come from: 1) better communications between all partners to ensure that the right information gets passed, and 2) by not delaying the deployment of the correct resources and equipment to resolve incidents as quick as possible. Another opportunity for improvement might be in relocating the historical facility locations to cover demand efficiently.

4.4. Summary and conclusion

The key insights generated by the ICT case studies covering different aspects of SPL systems are as follows:

- A wide range of service time options are offered to the customers. Previous SPL studies indicate that major IT companies offer time-differentiated services involving service parts supply in short time windows, e.g. 4 and 8 hours. It is however not indicated in these studies whether a contracted time is associated with delivering a part at a customer site or fixing it. Presumably, customers should be concerned with the time it takes to fix a part and recover the system. However, in one of the case companies the service times normally only correspond to the time taken to deliver a part. The other case company does provide the option to fix a service part in the contracted time, however a considerable percentage of service calls only require service parts to reach at customer locations in contracted time windows.

- A high percentage of service calls that the case companies receive is not for same day responses. This can allow a high level of centralization in distribution as most customer requests can be met from a longer distance.

- The findings confirm that there is a great variety of service parts that service providers have to maintain with a significant disparity in their costs. A significant percentage of service parts are repairable. The service parts are predominately slow movers and the products that the case companies serve have comparatively short life cycle. These factors can contribute to a high obsolescence rate, which is an issue highlighted by both case companies. Fortuin and Martin (1999) suggest that collaboration with other parties using the same service parts may provide an opportunity to aggregate the demand and therefore reduce the overall investment in service parts. Both companies believe that it is likely that other major companies in this sector hold similar service parts in their inventories. However there is no mechanism of inventory

sharing with other companies. One of the case companies does however occasionally source parts from middlemen holding a pool of inventory for the case company's competitors.

- The distribution networks of the case companies are decentralized with capabilities to reach customers in short times. Service requests, which commonly require just one part to satisfy, are received centrally. Both companies use (S-1, S) inventory control policy and use forecast driven software systems to manage their inventories. Services are mostly provided in a decentralized fashion. One of the case companies serves all customer requests from the nearest part stocking facility with available stocks, while the other serves all same day service calls from the nearest part stocking facility. However, both companies maintain very expensive parts only centrally.

- There is a high degree of outsourcing. The transportation of parts to pick-up locations near customer sites is completely outsourced. Also, servicing at customer sites is partly outsourced to third party engineers.

- The distribution and recovery network structures are complex and there are features in the systems that the current SPL literature does not consider. For example the studies do not consider that central warehouses, along with replenishing forward/field warehouses, also serve the customers directly. Besides, the part recovery system at one of the case companies has different transportation and storage stages for a part before it is incorporated back in the useable inventory. The studies related to repairable service parts do not take such considerations into account.

- The location decisions are based on demand coverage and being near to customer sites. Inventory is not indicated as a factor in deciding facility locations.

- The companies report high inventory and transportation costs. The average cost of a service part is quite high, hinting high inventory investment. Transportation costs for the same day services are also very high. Besides, requested parts and service engineers cover a part of their journeys separately to reach customer sites.

- Many parts are bought as single lots by one of the case companies. Hence the main opportunity in reducing the service costs remains in the efficient management of the transportation for supplying these parts.

- As opposed to several publications that pronounce service parts business and after sales services as high profit generators, one of the case companies does not see its service business as a profit generating function of the organization. Rather, it considers it as a necessary support function. The other company in the study also reports that servicing certain families of products is not so profitable.

- Common issues faced by the case companies in managing their SPL are obsolescence of parts in inventory and the cost pressure from customers. Other issues highlighted by the case companies include low resale values of parts, expensive maintenance of old equipment, managing the service availability v/s service cost trade-off, complexities due to the flexibility in contract terms, and the lack of awareness regarding contract natures and terms. The case companies see an opportunity for improvement in their call management systems through automation and integration.

97

- There is a trend towards more outsourcing of servicing at customer sites to third party engineers.

The study of the Highways Agency's service operations shows that transporting materials at demand sites within short and multiple time windows is not limited to SPL systems. The following table (Table 4.4) summaries some of the aspects of the Highways Agency's service operations in comparison to that of the ICT SPL systems studied.

Table 4.4: Comparison of service operations of the Highways Agency and ICT cases

Aspects	The Highways Agency	ICT Cases
Service times	• There are only two service time options which are dependent on the time of day. That is, unlike the ICT cases, at a particular time of day, service is provided within a single time constraint • The time constraint for repair services is associated with the 'resolution' (or completion) of the task, i.e. not just reaching the site • The responses are always provided within short time windows (the difference between the service times is not large)	• The companies provide four service time options, with one of the company offering three of its service time options as either 'reach' (just delivery) of 'fix' (deliver and install). In fix services, parts typically reach one hour earlier than the total service time window • Significant proportion of demand is for the next day response (i.e. not for the short/same day responses)
Variety of parts/material	• There is a vast variety of materials/parts used for repairs	
Facility setup	• One-echelon • Service calls are received centrally and the nearest facility having the required stocks is chosen to meet demand • There is no hierarchy of facilities in terms of service time options they provide	• Two-echelon inventory maintenance • Service calls are received centrally and the nearest facility having the required stocks is chosen to meet demand • For the same day service options, there is no hierarchy of facilities in terms of service time options the facilities provide
Demand locations	• Demand locations are not considered to be clustered. There is a 'continuous' demand base	• Customer locations are 'discrete' and can be clustered, but the spread is throughout the country
Lifecycle durations	• Though inventory is slow moving, obsolescence is not an issue as materials can be used in repairs for a long time	• Parts in inventory are prone to become obsolete as lifecycle durations of products being serviced are short and the inventory is generally slow moving
Inventory cost	• Low	• High
Replenishment	• Stocks are directly replenished from manufacturers	• Stocks are replenished from multiple sources (OEM's and brokers)
Outsourcing	• There is a high level of outsourcing	

Chapter 5: Time-differentiated distribution costs under hierarchical and non-hierarchical setups

5.1. Introduction

The aim of the work presented in this chapter is to provide insights into the relationship between service times and some important service cost components in a system where distribution is done within different service time windows. We analyse the problem by developing a cost model of a stylized system under two distinct organizations of stocking facilities, namely '*hierarchical*' organization and '*non-hierarchical*' organization. The modelled hierarchical system can be classified as a successively inclusive (or nested) hierarchical system. The classification of hierarchical systems is outlined in Section 2.4.2 (Chapter 2). We investigate the impact of service time limits on inventory, travelling, and distribution network setup costs. Assuming a reorder point and order quantity (R, Q) and the one-for-one (S-1, S) inventory policies, and Euclidean travelling distances, the analysis shows how a non-hierarchical setup and a hierarchical setup can result in different service costs. The sensitivity analysis of the cost model is performed by altering the demand fractions, the service time window lengths, and the ratio between the service time window lengths of two time-based service types. A part of the work in this chapter is presented in Jat and Muyldermans (2013).

The rest of the chapter is structured as follows. The problem description and assumptions are stated in Section 5.2. The formulation of the cost functions is presented in Section 5.3. Section 5.4 presents the model analysis including the analysis based on a (R, Q) inventory policy (Section 5.4.1) and on the (S-1, S) inventory policy

(Section 5.4.2). Section 5.4.2 also includes an analysis based on historical demand of case Company B. Finally, the conclusion and a brief discussion are presented in Section 5.5.

5.2. Problem description and assumptions

Let us consider customers with identical demand uniformly spread over a large geographical area. Customers have to be supplied with parts within different contracted service time commitments. We refer to the type of service that ensures a supply of parts within the short time window as the '*strict*' service, and the type of service that ensures a supply of parts within the longer time window as the '*relaxed*' service. In order to meet a service time commitment in the entire service area, every client location should be within the maximum distance, which can be covered within the committed service time, from at least one service parts storage facility. From now onwards, we refer to storage facilities, where parts are stored and from where they are dispatched to customers, as '*service facilities*' or just '*facilities*', and we use the term '*setup*' to refer to an organization of service facilities. The aim is to determine the impact of setting different service time constraints and the demand fractions for the relaxed and strict services on inventory, transportation and distribution network setup costs under the hierarchical and non-hierarchical (completely decentralized) setups.

We make the following assumptions to study the problem:

1) A (R, Q) inventory policy is considered, where R (a reorder point) is determined by accounting a safety stock level based on certain probability of not stocking out during the lead time assuming a Poisson demand process, and Q (an order quantity) is determined based on the EOQ model. Although the use of the EOQ model cannot be

considered as the most appropriate policy for managing slow moving items and items with unstable demand, which are two common characteristics of service parts, some service parts can be fast moving with a predictable demand (such as those used in scheduled maintenance). The EOQ model, which incorporates the trade-offs of the inventory order and holding costs, is a commonly used model for inventories and allows a tractable formulation to convey the main insights. A cross industry exploratory investigation by Rumyantsev and Netessine (2007) shows that many of the predictions from classical inventory models, such as the EOQ model, extend beyond individual products to the aggregate firm level. Hence, these models can help with high-level strategic choices in addition to tactical decisions. In case of SPL systems, as highlighted earlier, it has been reported that not many companies apply complex concepts that exist in the literature. Although the service parts are generally characterized as slow moving (Fortuin and Martin, 1999; Wagner and Lindemann, 2008), some previous studies related to SPL (e.g. Ashayeri et al., 1996; Cohen et al., 1997; Huiskonen, 2001) point out the use of basic inventory management techniques for service parts, including the EOQ model.

2) The problem is also formulated and analysed under the (S-1, S) inventory policy, which is considered appropriate for slow moving items and is a widely considered inventory policy by SPL studies. The following commonly used assumptions are considered for the (S-1, S) policy (Gzara et al., 2014):

- Single item
- Demand arrives one at a time according to the Poisson process
- Backorders allowed
- No capacity constraints on the supply (replenishment)

3) Customers' geographical distribution is assumed to be uniform over the plane and it is assumed that travelling distances are Euclidean. These assumptions are commonly made by studies for simplification.

4) It is assumed that the geographical area to be covered to provide services is large and that services have to be provided within short time commitments. Typically, large IT companies cover vast geographical areas for service parts distribution through several service facilities.

5) A single-echelon distribution system is assumed. A regular hexagonal packing of service catchment areas is considered assuming that facility points are located efficiently to cover the entire area for the strictest service time commitment and that a service request is fulfilled by the nearest service facility offering the required service type. The hexagonal partitioning is considered as an efficient partitioning when assuming Euclidian distances. Examples of service system studies that consider a hexagonal pattern of service catchment areas for their analysis include the ones by Simchi-Levi (1992) and Drèze et al. (2008).

6) For an analytical treatment, boundary effects and rounding off errors are ignored in determining the number of facilities. Hence the analysis is an approximation. To calculate the number of facilities we divide the total area by a facility catchment area (a full hexagonal area determined according to the maximum distance that can be travelled within a committed service time), which can result in a fractional number. Secondly, packing an area in a plane with full identical hexagons may not provide a complete coverage, and typically, areas on boundaries of the region may need to be covered by partial hexagons and need additional facilities. Simply dividing an entire service area with a facility catchment area can underestimate the number of facilities

required for the coverage. Nevertheless, the boundary effect becomes less significant in case of a large overall area covered by a high number of regular hexagons (Bollobás and Stern, 1972; Morgan and Bolton, 2002). Considering a hierarchical organization of facilities, we show in Section 5.3.3 that the boundary effect diminishes when the overall area becomes large, and hence, the estimations from the cost formulae improve.

5.3. The model formulation

The model represents a non-hierarchical setup and a hierarchical setup of facilities providing two different time-based service types. The relaxed service has the longer maximum time window (e.g. the service to deliver parts within 4 hours), and the strict service has the shorter time window (e.g. the service to deliver parts within 2 hours). The formulation can be extended to more than two service time windows as shown later. Below is the list of the general notations used in the formulation:

$A =$ total area to be served (a large geographical area)

$\lambda =$ total demand in area A, i.e. the total number of service calls per unit time

$f =$ fraction of total demand corresponding to the strict service

$1\text{-}f =$ fraction of total demand corresponding to the relaxed service

$s_s =$ service distance constraint for the strict service

$s_r =$ service distance constraint for the relaxed service

$C_o =$ cost per inventory replenishment order (for the EOQ policy)

$C_h =$ holding cost per unit per unit time (for the EOQ policy)

$L =$ lead time to receive service parts at service facilities

$C_t =$ transportation cost per unit distance

$S =$ base stock level at a facility under (S-1, S) inventory management policy

$F =$ fixed facility cost per unit time

Cyclic inventory cost (EOQ model): Considering the EOQ policy in a centralized system (having only one facility), the inventory order quantity Q equals to $\sqrt{\frac{2C_o\lambda}{C_h}}$, and the average inventory level equals to $0.5*Q = \frac{1}{2}\sqrt{\frac{2C_o\lambda}{C_h}}$. This gives,

$$Cyclic\ Inventory\ holding\ cost\ per\ unit\ time = C_h\frac{1}{2}\sqrt{\frac{2C_o\lambda}{C_h}} =$$

$$\sqrt{\frac{C_hC_o\lambda}{2}} \tag{5.1}$$

Note that with the EOQ model, the inventory holding cost per unit time (5.1) and the inventory ordering cost per unit time are equal. As Q is the replenishment order quantity and λ is the demand per unit time, inventory orders per unit time equal to λ/Q or $\sqrt{\frac{C_h\lambda}{2C_o}}$, and the inventory order cost per unit time then equals to $C_o\sqrt{\frac{C_h\lambda}{2C_o}}$ or $\sqrt{\frac{C_oC_h\lambda}{2}}$, which is same as the inventory holding cost per unit time (5.1). Hence, the total inventory cost per unit time is twice the inventory holding cost per unit time (5.1) (or twice the inventory order cost per unit time).

$$Total\ Cyclic\ inventory\ cost\ at\ a\ facility\ per\ unit\ time = 2\sqrt{\frac{C_hC_o\lambda}{2}} =$$

$$\sqrt{2C_hC_o\lambda}$$

Within a decentralized system, if λ_i is the demand served by service facility i ($i=1, \dots n$) and n is the total number of facilities, then assuming that each facility applies the EOQ policy, per unit time,

$Total\ cyclic\ inventory\ cost\ in\ a\ decentralized\ system =$

$$\sum_{i=1}^{n} \sqrt{2C_h C_o \lambda_i} \tag{5.2}$$

Safety stock cost: Considering a certain probability of not stocking-out over the lead time,

$Total\ safety\ stock\ holding\ cost\ in\ a\ decentralized\ system =$

$$C_h \sum_{i=1}^{n} z\sqrt{L\lambda_i} \tag{5.3}$$

where z is the safety factor from the Poisson distribution and $\sqrt{L\lambda_i}$ is the standard deviation of a Poisson demand over the lead time L.

Inventory level (one-for-one (S-1, S) inventory policy): The value of S, i.e. the base stock level, under the (S-1, S) inventory policy is determined considering the steady state probability of the quantity of units in resupply. Note that there is a well-known relationship between stock-out probability and fill-rate under the (S-1, S) inventory policy (Muckstadt, 2005, 2010; Zipkin, 2000). The following expression is used in an iterative procedure to determine S under a set fill-rate level (Muckstadt, 2010).

Fill-rate = F(S) = Probability that demand is less than S over the lead time, P(λ<S)

$$F(S) = \sum_{x<S} \frac{e^{-\lambda L}(\lambda L)^x}{x!}$$

Where, $\frac{e^{-\lambda L}(\lambda L)^x}{x!}$ is the unconditional probability that x units remain in the resupply, λ is the rate with which demand is generated by the Poisson process, and L is the mean of resupply time. The iterative procedure to determine S is as follows:

Let $F(S)_r$ be the required fill-rate, i.e. the minimum percentage of demand met from the stocks on hand,

1. If demand λ is positive, set $S = 1$, else set $S = 0$ and go to step 7;
2. Calculate $F(S)$;
3. If $F(S) \geq F(S)_r$, go to step 7, else continue;
4. Increment S by one;
5. Calculate $F(S)$;
6. If $F(S) \geq F(S)_r$, continue, else go to step 4;
7. Set S as the 'Base stock level';

The above procedure is coded in MS Visual Basic for the use in excel sheets for the analysis.

Average service distance: Considering Euclidean travelling distances, a uniform geographical distribution of customers, and hexagonal service areas,

$$Average\ distance\ to\ reach\ a\ customer\ in\ a\ catchment\ area =$$
$$0.60799(s) \tag{5.4}$$

where s is an edge length of the hexagonal catchment area (Figure 5.1), and it is considered that the service facility is located in the middle of the catchment area (details can be found in Appendix 2 and Stone (1991)).

5.3.1. Time-differentiated distribution under non-hierarchical organization of service facilities

Service facilities under the non-hierarchical (completely decentralized) setup provide the full range of service time responses, i.e. both the relaxed and the strict services (Figure 5.2). Catchment areas of service facilities are according to the maximum distance that can be covered from a service facility to provide the strict service. As s_s

is the maximum distance that can be covered from a service facility to provide the strict service, it is also equal to the length of an edge of the hexagonal service catchment area of a facility (Figure 5.1)).

's' is the maximum distance that can be travelled from a facility (located at the centre) within the hexagonal catchment area. In our case $s = s_s$.
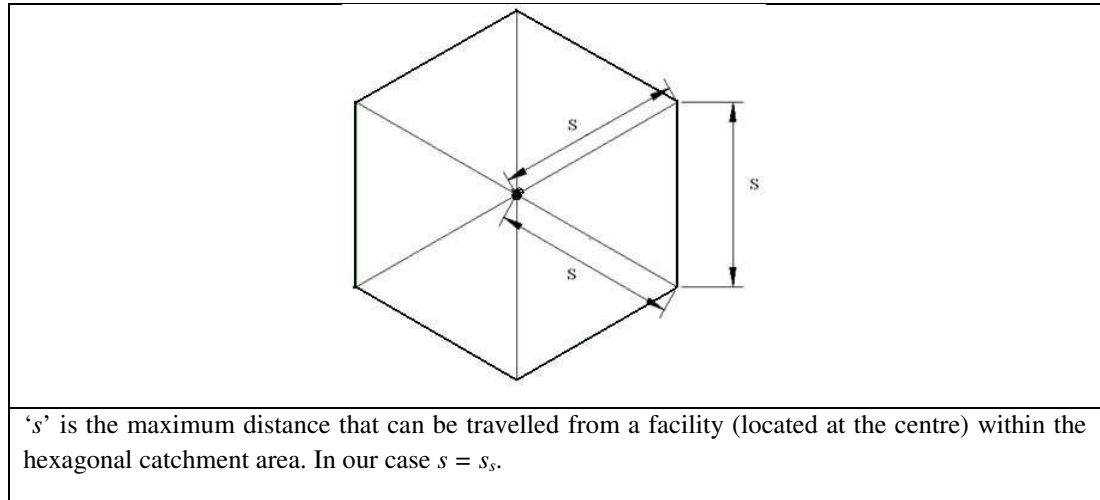
Figure 5.1: Facility catchment area

The setup consists of only one service facility type. Assuming that the system offers two service time options, the 4 hours service being the relaxed service and the 2 hours service being the strict service, the catchment areas of all service facilities are within the range of 2 hours from the facility points. A service facility provides both service types to the customers within its catchment area. Customer *C1* can get both 4 hours and 2 hours services from facility *F1*, and customer *C2* can get both 4 hours and 2 hours services from facility F2.

Figure 5.2: Non-hierarchical setup of service facilities

Let *n* be the number of service facilities. Then,

$n = \dfrac{A}{2.5981(s_s{}^2)}$, where '$2.5981(s_s{}^2)$' is the hexagonal catchment area of a service facility with an edge length of s_s.

As per the non-hierarchical setup (Figure 5.2), the number of service facilities providing the relaxed service and the number of service facilities providing the strict service are both equal to n.

$$Total\ demand\ (relaxed\ and\ strict\ service\ calls)\ served\ by\ one$$

$$service\ facility = \frac{(1-f)\lambda + f\lambda}{n} = \frac{\lambda}{n} \tag{5.5}$$

Note that in the above equation (5.5), the total demand served by one service facility represents the demand served by a facility with the full hexagonal service area (with an edge length of s_s). That is, '*one*' facility here and in the following formulations means a facility serving in the maximum service area under a distance constraint.

The distribution network setup cost (per unit time) is taken as the product of the number of facilities (n) and the fixed facility cost per unit time (F).

$$Distribution\ network\ setup\ cost = F \times n \tag{5.6}$$

By incorporating the number of service facilities (n) and the demand served by one facility (5.5) in the functions for total cyclic inventory cost (5.2) and total safety stock holding cost (5.3) in a decentralized system, we obtain the following cost functions for the non-hierarchical setup.

$$Total\ cyclic\ inventory\ cost = n\sqrt{2C_h C_o \frac{\lambda}{n}} = \sqrt{2C_h C_o \lambda n} \tag{5.7}$$

$$Total\ safety\ stock\ holding\ cost = nC_h\left(z\sqrt{L\frac{\lambda}{n}}\right) = C_h z\sqrt{L\lambda n} \qquad (5.8)$$

Based on (S-1, S) policy,

$$Total\ inventory\ (base\ stock)\ in\ the\ system = n \times S \qquad (5.9)$$

Considering the average distance to reach a customer within hexagonal catchment areas (5.4) and the maximum distance that can be covered from a service facility to provide the strict service (s_s),

$$Total\ expected\ transportation\ cost = C_t 0.60799\lambda s_s \qquad (5.10)$$

Under the non-hierarchical setup, considering the distribution network setup cost (5.6) cyclic inventory cost (5.7), safety stock holding cost (5.8), and transportation cost (5.10) per unit time, with the (R, Q) inventory policy,

$$Total\ cost\ per\ unit\ time = nF + \sqrt{2C_h C_o \lambda n} + C_h z\sqrt{L\lambda n} + C_t 0.60799\lambda s_s$$

5.3.1.1. Provision of more than two time-based service types under the non-hierarchical organization of service facilities

The functions formulated above, (5.6), (5.7), (5.8), (5.9) and (5.10), can also be considered for the cost analysis of a non-hierarchical system considering more than two time-based service types. This is because a service request is met by the closest facility, ignoring the requested service time window. Demand at facilities under the non-hierarchical setup remains the same regardless of the fractions of demand for different service types.

Let *Type 1, Type 2, Type 3* and *Type 4* services be four time-based service types in the order of their strictness, such that *Type 1* service is the most relaxed service and *Type*

4 service are the strictest. Let f_1, f_2, f_3 and f_4 be the fractions of demand for *Type 1, Type 2, Type 3* and *Type 4 services* respectively, such that $f_1+f_2+f_3+f_4= 1$. Then, assuming a uniform geographical distribution of customers in the entire area,

$$Total\ demand\ served\ by\ one\ service\ facility = \frac{f_1\lambda}{n} + \frac{f_2\lambda}{n} + \frac{f_3\lambda}{n} + \frac{f_4\lambda}{n} = \frac{\lambda}{n}$$

as $f_1+f_2+f_3+f_4= 1$.

Facility catchment areas, and hence the total number of facilities (*n*), will be based on the maximum distance that can be covered from a service facility to provide the *Type 4* service. It can be noticed in this case that the demand fractions for different service types in the total demand have no effect on the costs.

5.3.2. Time-differentiated distribution under hierarchical organization of service facilities

A nested hierarchical hexagonal pattern can be generated by locating the centres (facility points) of lower level hexagons 1) at the middle of the edges of the higher level hexagons (Figure 5.3), or 2) at the corner points of the higher level hexagons (Figure 5.4). These two approaches result in different ratios between the maximum distances (or service time constraints) within the higher and lower level hexagons.
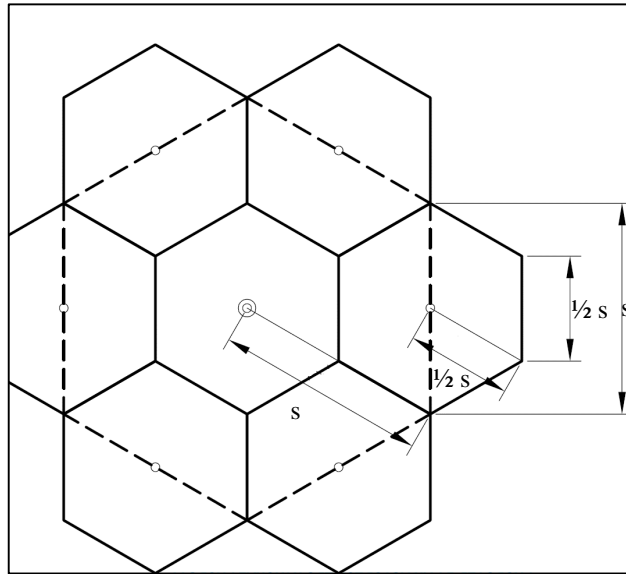
Figure 5.3: Hierarchical hexagonal pattern with the centres of the lower level hexagons at the middle of the edges of the higher level hexagons
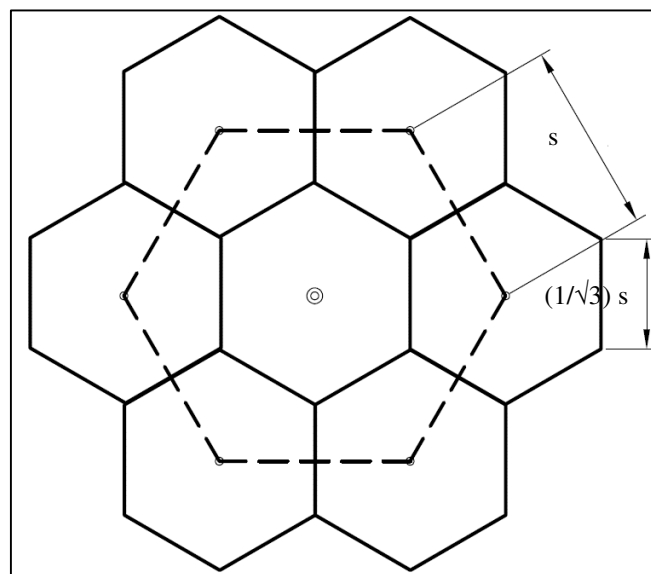
Figure 5.4: Hierarchical hexagonal pattern with the centres of the lower level hexagons at the corners of the higher level hexagons

Figure 5.5 presents a combination of the two approaches of locating lower level hexagons with respect to higher level hexagons.
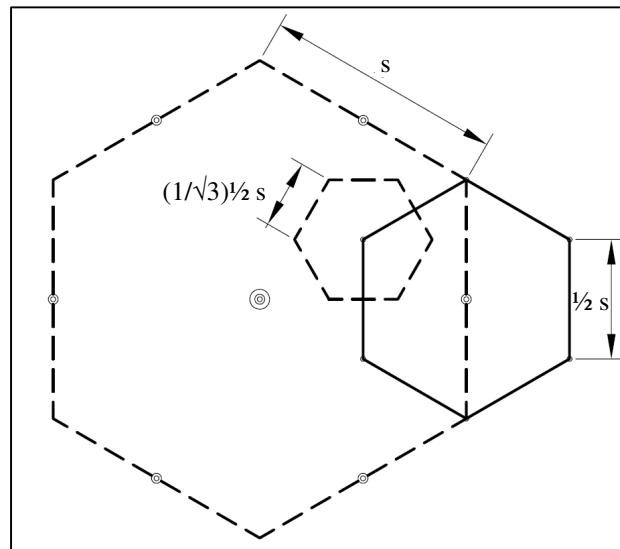
Figure 5.5: Hierarchical hexagonal pattern with the centres of the lower level hexagons at the corners and the edge midpoints of the higher level hexagons

The modelled system is described in Figures 5.6 and 5.7. Though these figures depict a hierarchical hexagonal pattern in which the centres of lower level hexagons are located at the middle of the edges of higher level hexagons, the formulation can also be used to study the service time ratios resulting from locating the centres of lower level hexagons at the corner points of higher level hexagons (Figure 5.4). However, the ratio between the time constraints for the relaxed and strict services that the pattern in Figures 5.6 and 5.7 allows is in line with the ones offered in the real world IT SPL systems that we have studied and those that are referred in the literature.

Assuming that a system offers two service time options, the 4 hours service being the relaxed services and the 2 hours service being the strict service, the setup consists of two types of service facilities, namely higher level service facilities and lower level service facilities. Higher level service facilities (e.g. *F1*) provide both the 4 hours and the 2 hours services. Lower level service facilities (e.g. *F2*) only provide the 2 hours services. Customer *C1* can get both the 4 hours and 2 hours services from *F1*, while customer *C2* can get the 4 hours service from *F1* and the 2 hours service from *F2* (as a part from *F1* can reach *C2* within 4 hours).

F1 has two service catchment areas, one within the range of 2 hours (marked by the continuous boundary) and the other within the range of 4 hours (marked by the dashed boundary). All customers within the first catchment area of *F1* (within the continuous boundary lines) can get both types of service from *F1*, whereas, the customers beyond the first catchment area and within the second catchment area (within the dashed boundary lines) can only get the 4 hours service from *F1*.

Unlike *F1, F2* has only one catchment area (marked with the continuous boundary) which is within the 2 hours range. *F2* can only provide the 2 hours service to the customers within its catchment area. For the 4 hours service, the customers within the catchment area of *F2* are served by *F1* or another higher level service facility.

Figure 5.6: Hierarchical organization of service facilities

Consider that a '*higher level*' service facility provides both the relaxed and the strict services, and that a '*lower level*' service facility only provides the strict service (Figure 5.6). We know s_r is the maximum distance that can be covered from a service facility to provide the relaxed service, and s_s is the maximum distance that can be covered from a service facility to provide the strict service. Consequently, s_r is equal to an edge length

of a hexagonal catchment area for providing the relaxed service, and s_s is equal to an

edge length of a hexagonal catchment area for providing the strict services (Figure 5.7).

In this setup, the maximum distance that can be travelled in a straight line from a lower level facility within its catchment area is half of the maximum distance that can be travelled from a higher level facility within its catchment area for the relaxed service. That is, assuming Euclidean distances, the next possible time constraint for the strict service that can be offered is half of the time constraint for the relaxed service. For instance, if the time constraint for the relaxed service is 8 hours, the next possible time constraint for the strict service would be 4 hours. If a service time option that is stricter than 4 hours has to be provided, then it can be 2 hours ($1/2 \times 4$ hours). This is achieved by locating lower level service facilities at the middle of continuous edges in this figure instead of the dashed edges.

Figure 5.7: A two level hierarchical organization of service facilities with the lower level service facilities located at the middle of the edges of the hexagonal catchment areas for the relaxed service

Let n_r be the number of service facilities providing the relaxed service and n_s be the

number of service facilities providing the strict service. Then,

$n_r = \frac{A}{2.5981 s_r^2}$, where $2.5981(s_r^2)$ is the hexagonal catchment area of a service facility

for providing the relaxed service, and

$n_s = \frac{A}{2.5981 s_s^2}$, where $2.5981(s_s^2)$ is the hexagonal catchment area of a service facility

for providing the strict service.

This gives the ratio: $\dfrac{n_r}{n_s} = \dfrac{\frac{A}{2.5981(s_r^2)}}{\frac{A}{2.5981(s_s^2)}} = \dfrac{s_s^2}{s_r^2}$.

In a successively inclusive hierarchy, in which a higher level service facility provides

both the relaxed and the strict services while a lower level service facility only provides

the strict service, service facilities can be categorized according to Table 5.1.

Table 5.1: Classification of service facilities

The types of service facilities	Service Types	
	Relaxed service	*Strict service*
Higher Level service facilities provide →	✓	✓
Lower Level service facilities provide →		✓

From the classification in Table 5.1,

Number of higher level service facilities $= n_r$

Number of lower level service facilities $= n_s - n_r$

Note that n_s is the total number of service facilities as all service facilities provide the

strict service. Considering the above classification of service facilities:

Total demand served by higher level facilities $= (1-f)\lambda + \frac{n_r}{n_s} f\lambda$

$$\text{Demand served by one higher level facility} = \left(\frac{(1-f)}{n_r} + \frac{f}{n_s}\right)\lambda \qquad (5.11)$$

$$Total\ demand\ served\ by\ lower\ level\ facilities = \left(\frac{n_s - n_r}{n_s}\right) f \lambda$$

$$Demand\ served\ by\ one\ lower\ level\ facility = \frac{f}{n_s} \lambda \qquad (5.12)$$

Incorporating the number of higher level service facilities (n_r), demand served by one higher level service facility (5.11), the number of lower level service facilities (n_s-n_r), and demand served by one lower level service facility (5.12) in the functions for total cyclic inventory cost (5.2) and total safety stock holding cost (5.3) in a decentralized system, the following cost (per unit time) functions and multiplication factors for the hierarchical setup are obtained.

$$Total\ cyclic\ inventory\ cost = n_r \sqrt{2 C_h C_o \left(\frac{(1-f)}{n_r} + \frac{f}{n_s}\right) \lambda} + (n_s - n_r) \sqrt{2 C_h C_o \frac{f}{n_s} \lambda}$$

$$= \sqrt{2 C_h C_o \lambda n_r} \left(\sqrt{(1-f) + \frac{n_r}{n_s} f} + \left(\frac{n_s}{n_r} - 1\right) \sqrt{\frac{n_r}{n_s} f} \right) \qquad (5.13)$$

$$Multiplication\ factor\ (cyclic\ inventory) = \left(\sqrt{(1-f) + \frac{n_r}{n_s} f} + \right.$$

$$\left. \left(\frac{n_s}{n_r} - 1\right) \sqrt{\frac{n_r}{n_s} f} \right) \qquad (5.14)$$

Here three extreme cases, or benchmarks, of the cyclic inventory cost can be identified. 1) When there is only one facility providing the service (i.e. completely centralized system), the cyclic inventory cost, as mentioned earlier, equals to $\sqrt{2 C_h C_o \lambda}$. 2) When there is no demand for the strict service ($f = 0$), the multiplication factor (5.14) equals to 1, reducing the total cyclic inventory cost function (5.13) to $\sqrt{2 C_h C_o \lambda n_r}$. That is, the cost increases by the factor of $\sqrt{n_r}$ compared to the completely centralized case. 3) When there is no demand for the relaxed service ($f = 1$), and hence the system becomes completely decentralized, the multiplication factor

(5.14) becomes $\sqrt{\frac{n_s}{n_r}}$, reducing the total cyclic inventory cost function (5.13) to

$\sqrt{2C_h C_o \lambda n_s}$. Hence the cost increases by the factor of $\sqrt{n_s}$ compared to the completely

centralized case.

$Total\ safety\ stock\ holding\ cost =$

$$C_h \sqrt{L\lambda} \left(n_r z_h \sqrt{\frac{(1-f)}{n_r} + \frac{f}{n_s}} + \left((n_s - n_r) z_l \sqrt{\frac{f}{n_s}} \right) \right)$$

(5.15)

where z_h and z_l are the safety factors from the Poisson distribution. z_h is based on

demand at a higher level service facility (5.11) and z_l is based on demand at a lower

level service facility (5.12).

Let S_h and S_l be the base stock levels determined according to the (S-1, S) policy for a

higher level and a lower level facility respectively. Given the minimum fill-rate level,

S_h is computed considering the demand at a higher level facility (5.11), and S_l is

computed considering the demand at a lower level facility (5.12). Then, under the (S-

1, S) inventory policy,

$Total\ inventory\ (base\ stock)\ in\ the\ system = n_r S_h + (n_s - n_r) S_l$ (5.16)

Finally, considering the average distance to reach a customer within hexagonal

catchment areas (5.4),

$Total\ average\ travelling = 0.60799(1 - f)\lambda s_r + 0.60799 f \lambda s_s$

$= 0.60799 \lambda s_r \left((1 - f) + \frac{s_s}{s_r} f \right) = 0.60799 \lambda s_r \left((1 - f) + \sqrt{\frac{n_r}{n_s}} f \right)$

$$Total\ average\ transportation\ cost = 0.60799\lambda s_r \left((1-f) + \right.$$

$$\left. \sqrt{\frac{n_r}{n_s}} f \right) C_t \tag{5.17}$$

$$Multiplication\ factor\ for\ transportation\ cost = \left((1-f) + \right.$$

$$\left. \sqrt{\frac{n_r}{n_s}} f \right) \tag{5.18}$$

When there is no demand for the strict service ($f = 0$) the multiplication factor for the transportation cost (5.18) equals to one and the transportation cost (5.17) becomes independent of the ratio between the strict and the relaxed service times. With the presence of demand for the strict service ($f > 0$), the higher the number of the facilities providing the strict service is (due to the smaller distance constraint for the strict service), the lower shall be the multiplication factor.

Considering the distribution network setup cost ($n_s{\times}F$) and the cyclic inventory cost (5.13), safety stock holding cost (5.15), and transportation cost (5.17) under the hierarchical setup, with the (R, Q) inventory policy,

$$Total\ cost\ per\ unit\ time = n_s F + \sqrt{2C_h C_o \lambda n_r} \left(\sqrt{(1-f) + \frac{n_r}{n_s} f} + \left(\frac{n_s}{n_r} - \right. \right.$$

$$\left. \left. 1 \right) \sqrt{\frac{n_r}{n_s}} f \right) + C_h \sqrt{L\lambda} \left(n_r z_h \sqrt{\frac{(1-f)}{n_r} + \frac{f}{n_s}} + \left((n_s - n_r) z_l \sqrt{\frac{f}{n_s}} \right) \right) +$$

$$0.60799\lambda s_r \left((1-f) + \sqrt{\frac{n_r}{n_s}} f \right) C_t$$

The extended formulation for cyclic inventory and transportation costs considering more than two service time options, i.e. considering more than two levels of the hierarchical system, is presented in Appendix 3. However, it does not give any further

analytical insights compared to the system with two time-based service types (i.e. two hierarchical levels). For this reason and for the tractability of the investigation, the analysis in Section 5.4 based on the system modelled for two service types.

5.3.3. Boundary effect

As stated earlier, a limitation in the above formulation is the ignorance of the boundary effect. The number of facilities is determined by dividing the total area by a full hexagonal catchment area determined according to the maximum distance that can be travelled within the committed service time. Packing an area in a plane with full identical hexagons may not provide a complete coverage, and typically, areas on the boundaries of the region may need to be covered by partial hexagons and therefore need additional facilities. Dividing the entire service area with a facility catchment area can underestimate the number of facilities required for the coverage. Bollobás and Stern (1972) and Morgan and Bolton (2002), without giving a proof, state that the boundary effect becomes less significant in case of a large overall area covered by a high number of regular hexagons. Below we prove that the boundary effect diminishes when the total area becomes large, keeping the size of packed hexagons constant. This is done by showing that the area covered by boundary hexagons divided by the total area tends to zero for a large area and by showing that the number of facilities on boundaries divided by the total number of facilities tends to zero for a large area.

Figure 5.8 shows the layers of hexagons representing the catchment areas of higher level facilities (providing relaxed service in larger service areas). The circular points on the boundaries of these hexagons (on edge midpoints) represent lower level facilities providing the stricter service (in a smaller service areas).
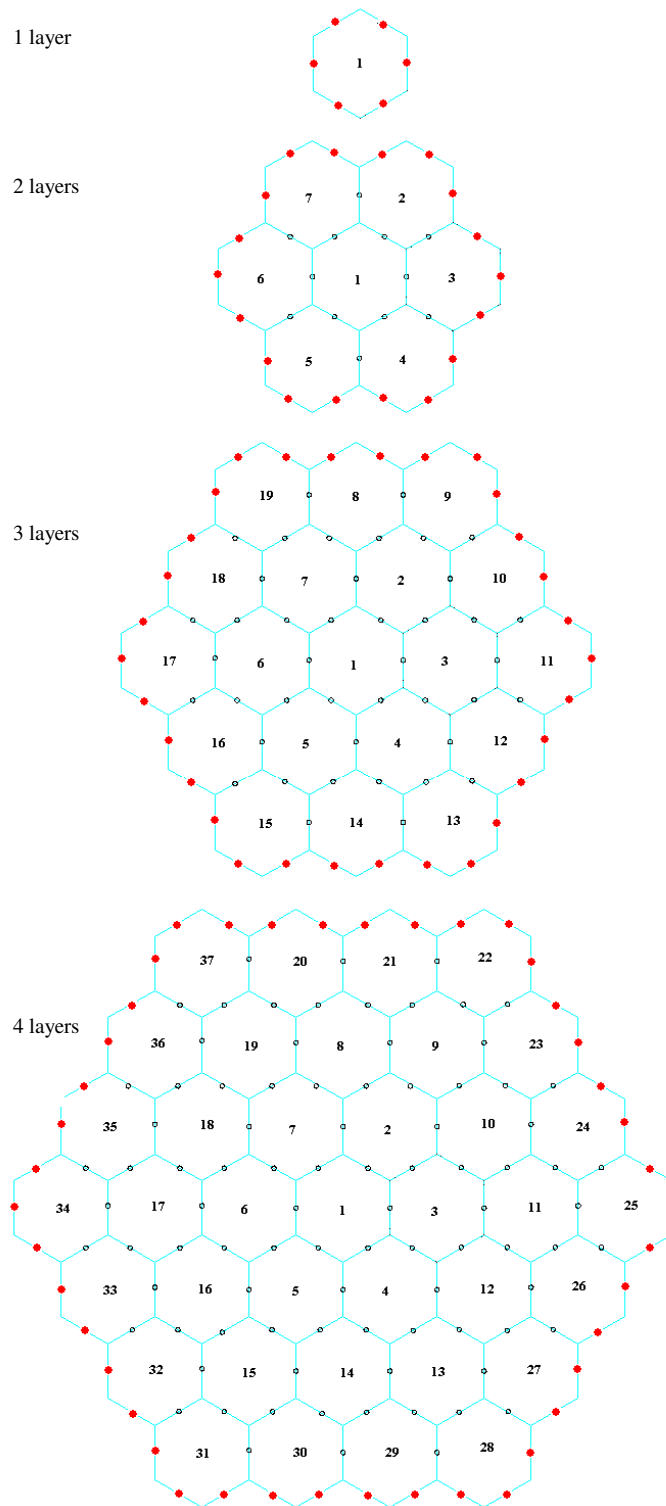
Placement of lower level facilities on edge midpoints of hexagonal catchment areas for the relaxed service

Figures 5.8: Hexagonal layers

Hexagons (higher level) on boundaries: We start with a hexagon and stepwise build

layers as shown in Figure 5.8 (from *1* to layer '*i*') to prove that,

$$\frac{Area\ covered\ by\ boundary\ hexagons}{Total\ area} \rightarrow 0, \ when\ i \uparrow$$

Table 5.2: Total number of hexagons in different number of hexagonal layers

Number of layers	Number of hexagons
1	1
2	1+**6** = 7
i 3	7+**12** = 19
4	19+**18** = 37
5	37+**24** = 61

Let, n_i be the number of hexagons in *i* layers, Δn_i be the difference between the number

of hexagons in *i* layers and *i-1* layers, and n_{bi} be the number of hexagons on the

boundaries in *i* layers. Then,

$$n_i = n_{i-1} + (i-1) \times 6 \qquad\qquad for\ i > 1, \qquad n_1 = 1$$

$$\Delta n_i = n_i - n_{i-1} = (i-1) \times 6$$

$$n_i = \left[\sum_{k=1}^{i} 6(i-1)\right] + 1$$

$$n_i = \frac{6(i-1)i}{2} + 1 \qquad\qquad , Quadratic\ in\ 'i'$$

$$n_{bi} = (i-1)6 \qquad\qquad , Linear\ in\ 'i'$$

Ratio: $\lim\limits_{i\to\infty} \dfrac{n_{bi}}{n_i} = \dfrac{6(i-1)}{\frac{6(i-1)i}{2}+1} = 0$

Hence, $\dfrac{area\ covered\ by\ boundary\ hexagons}{total\ area} \rightarrow 0, \ when\ i \uparrow.$

Lower level facilities on boundaries: Now we prove that,

$$\frac{Number\ of\ lower\ level\ facilities\ on\ boundary}{Total\ number\ of\ lower\ level\ facilities} \rightarrow 0, when,$$

$Total\ number\ of\ lower\ level\ facilities\ \uparrow$

Table 5.3: Number of lower level boundary and non-boundary facilities (see Figure 5.8)

Layers	Locations on edge midpoints on the boundaries	Non-boundary facilities
1 layer	6	0
2 layers	18	12
3 layers	30	42
4 layers	42	90
5 layers	54	156

Let, f_i be the number of facilities in i layers (where $f_0 = 0$), f_{bi} be the number of facilities on the boundaries in i layers, and f_{nbi} be the number of non-boundary facilities in i layers.

$$f_{nbi} = f_{i-1} + 6(i-1)$$

$$f_{bi} = 12i - 6 \qquad\qquad , Linear\ in\ 'i'$$

$$f_i = f_{bi} + f_{nbi}$$

$$f_i = f_{i-1} + 18i - 12$$

$$f_i = \sum_{k=1}^{i}(18i - 12)$$

$$f_i = \frac{18(i+1)i}{2} - 12i = 9i^2 + 9i - 12i$$

$$f_i = 3i(3i - 1) \qquad\qquad , Quadratic\ in\ 'i'$$

Ratio: $\lim_{i \to \infty} \frac{f_{bi}}{f_i} = \frac{12i-6}{3i(3i-1)} = 0$

5.4. Numerical analysis

The analysis is divided into two subsections. Section 5.4.1 analyses the costs considering the (R, Q) inventory policy and Section 5.4.2 analyses the costs under the (S-1, S) inventory policy. Before these subsections, we look at the effect of the service time constraint on the distribution network setup cost (5.6). As the total number of

facilities is same for the hierarchical and the non-hierarchical organizations of facilities in a network, the distribution network setup cost is considered to be the same. Figure 5.9 shows the setup cost for a system against different values of s_s considering the overall area size to be 200,000 length units^2 (the area of the mainland UK being 229,543 km^2) and a hypothetical fixed facility cost per unit time as £10,000. The reduction in the service time constraints when s_s is small sharply increases the number of required facilities and hence the setup cost. That is, the incremental marginal cost increases per unit decrease in the service time or distance constraint.
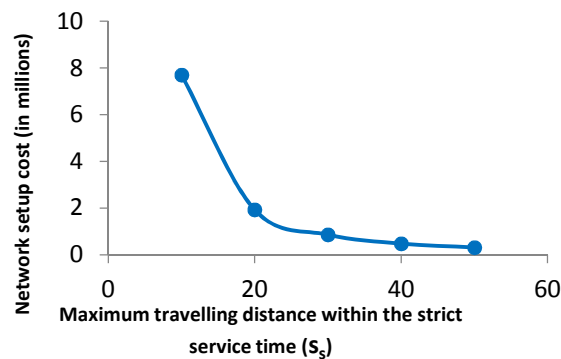
Figure 5.9: Impact of service distance constraint on system setup cost

5.4.1. Analysis considering the (R, Q) inventory policy

This section starts with an analysis of the multiplication factors for the cyclic inventory cost (5.14) and the transportation cost (5.18) under the hierarchical setup. Examples are then presented to investigate and compare the non-hierarchical and the hierarchical setups over different service distance constraints and demand fractions for the relaxed and the strict services.

Figure 5.10, based on the multiplication factor for the cyclic inventory cost (5.14), and Figure 5.11, based on the multiplication factor for the transportation cost (5.18), illustrate how cyclic inventory cost and transportation costs react to the changes in s_s/s_r

ratio (0.5, 0.25, … 0.002) and the demand fractions for the relaxed and strict services. Note that smaller s_s/s_r represents a greater time difference between the relaxed and the strict services and that there are comparatively more lower level service facilities, resulting in a higher level of decentralization. Likewise, a higher value of f (the fraction of demand for the strict service) means that more demand has to be fulfilled from lower level service facilities, which are more decentralized. It is important to mention that the non-hierarchical setup does not provide an opportunity to perform this type of analysis as the entire demand, whether for the strict service or the relaxed service, is met in a similar way (from the same facilities). This is discussed further in the examples.

Figure 5.10 shows that the inventory in a system increases when s_s/s_r decreases. An increase in f also results in a higher inventory level. The maximum inventory is maintained when s_s/s_r is at its minimum and f is maximum (i.e. $f = 1$). With s_r constant, smaller s_s/s_r means that more lower level service facilities have to be set-up to satisfy the service distance constraint for the strict service. With maximum f ($f = 1$) there is no demand for the relaxed service. This requires stocks to be maintained with the maximum decentralization as there is no allowance to meet demand from longer distances. In this scenario, a system under the hierarchical setup operates similar to when it is under the non-hierarchical setup. When f is at its minimum ($f = 0$), inventory levels are constant and minimum. This is the result of the absence of demand for the strict service, hence, the entire demand is met within larger service areas of more centralized higher level service facilities.
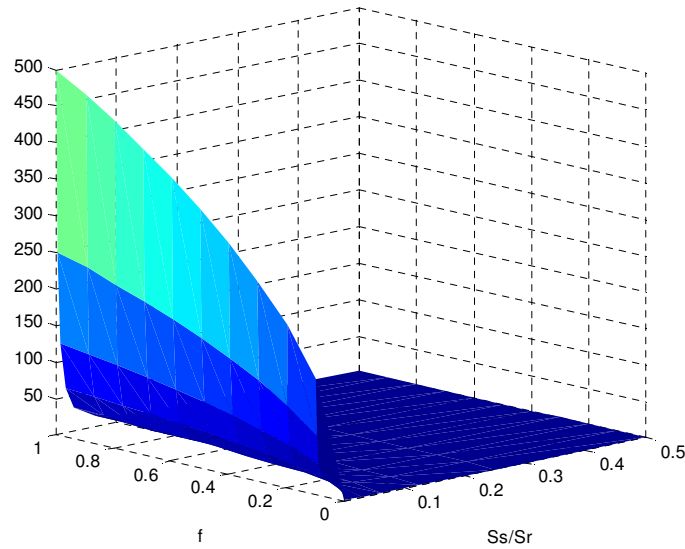
Figure 5.10: Multiplication factor for cyclic inventory

In contrast to inventory levels, Figure 5.11 shows that travelling reduces as s_s/s_r decreases. The minimum travelling results when s_s/s_r is at its minimum and f is at its maximum (i.e. $f = 1$). As stated earlier, smaller s_s/s_r results in more lower level facilities and maximum f results in entire demand being fulfilled by facilities in smaller catchment areas, which in turn results in lesser average travelling to serve customers. Travelling is at the maximum level when $f = 0$, irrespective of the s_s/s_r value, because, regardless of the number of lower level facilities, the entire demand is served by higher level service facilities within larger service areas. However, it should be noted that although a higher number of facilities reduces the average distance to reach customers, the average distance travelled for replenishments increases.

126
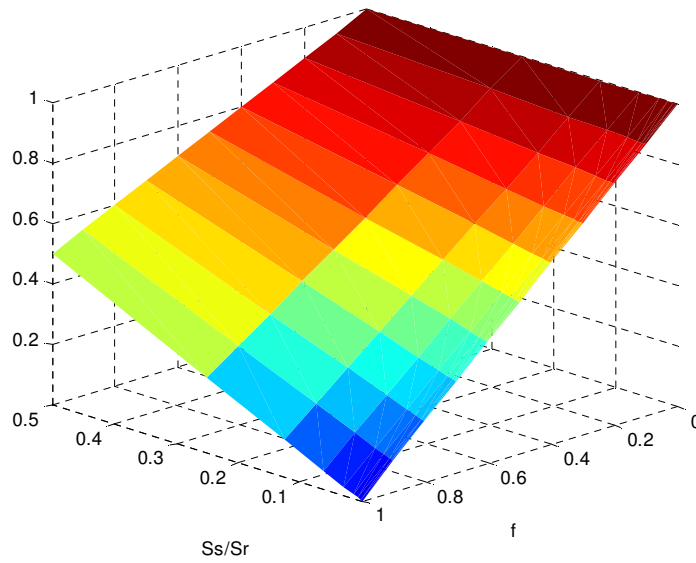
Figure 5.11: Multiplication factor for transportation

We now present the results from the computations based on the following hypothetical values to further illustrate and compare the hierarchical and non-hierarchical setups. Let, demand (λ) = 5000 per year, area (A) = 200000 unit length^2, inventory holding cost (C_h) = £1 per unit per year, inventory order cost (C_o) = £100, transportation cost per unit length (C_t) = £1, lead time (L) = 14 days, and the minimum probability of not stocking out = 0.98. The values for service distance constraints (s_s and s_r) and the demand fractions for both service types (*f and (1-f)*) are altered for the analysis.

Figure 5.12 presents the computation results for the cyclic inventory cost with different values of *f*, while keeping s_r=100 and s_s=50 unit lengths. The results show that the total cyclic inventory cost under the non-hierarchical setup (5.7) is unaffected by the changes in the fractions of demand for both service types. However, under the hierarchical setup, when the fraction of demand for the strict service decreases, the total cyclic inventory cost in the system (5.13) also decreases. The hierarchical system can exploit the opportunity to respond to relaxed service calls with a higher level of centralization as a longer distance can be travelled for a delivery. Under the non-hierarchical setup on

the other hand, the system cannot exploit this opportunity and supplies to customers with a constant level of decentralization, i.e. from the closest facility.
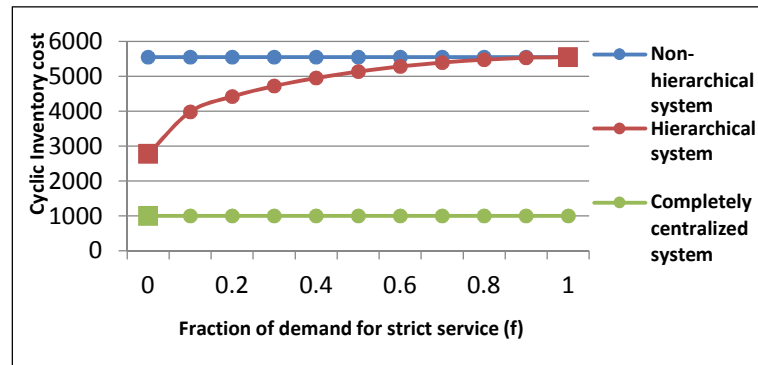
Figure 5.12: Impact of demand fractions for the strict and relaxed services on cyclic inventory cost ($s_r=100$, $s_s=50$)

Figure 5.12 shows two inventory cost benchmarks under the hierarchical setup (highlighted with square markers). The first relates to the inventory in the case of no demand for the strict service and hence the service is only provided from the higher level facilities. The second relate to the inventory in a completely decentralized operation, i.e. when there is no demand for the relaxed service and hence the system operates as a non-hierarchical system. The inventory costs under the hierarchical setup varies between these two levels depending on the value of *f*. Figure 5.12 also shows that the cyclic inventory cost is significantly lower if all service calls are satisfied in a completely centralized fashion (through only one facility) by ignoring the service distance constraints.

Figures 5.13 (a) and (b) present the results based on the inventory cost functions for the non-hierarchical setup (5.7) and the hierarchical setup (5.13) when the service distance constraints are altered while keeping the fractions of demand for both service types equal and constant ($f=0.5$). Figure 5.13 (a) presents the case where both s_r and s_s are changed, such that $s_r = 2s_s$, while Figure 5.13 (b) presents the case where s_r is kept constant and s_s is changed. When the distance constraint for the strict service decreases,

the difference between the cyclic inventory cost under the hierarchical and non-hierarchical setups increases (Figure 5.13 (b)). Under the non-hierarchical setup, a reduction in the time window for the strict service compels meeting both types of service demand with a higher level of decentralization. Under the hierarchical setup, only demand for the strict service, which is half of the total demand, is met with a higher level of decentralization.

(a) f =0.5; $s_r = 2s_s$.

(b) f =0.5, s_r=100.
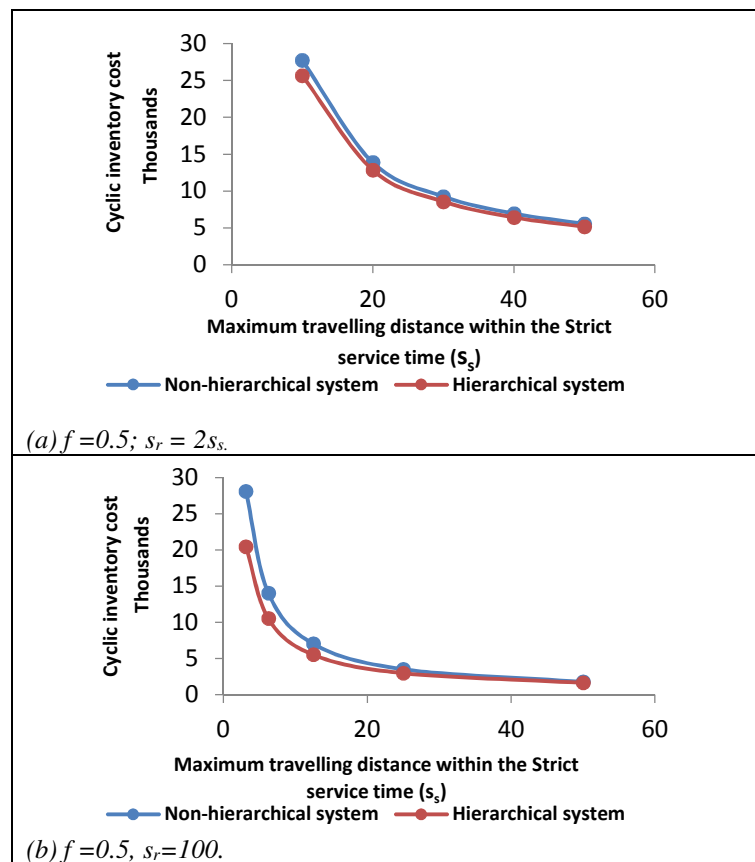
Figures 5.13 (a & b): Impact of service distance constraint on cyclic inventory cost

Figures 5.12 and 5.13 (a) and (b) show that the hierarchical setup typically results in a lower inventory cost compared to the non-hierarchical setup. The percentage difference between the inventory costs under the non-hierarchical and hierarchical setups is highest when the fraction of demand for the strict service *f* and s_s/s_r ratio are at their minimum (Figure 5.14).
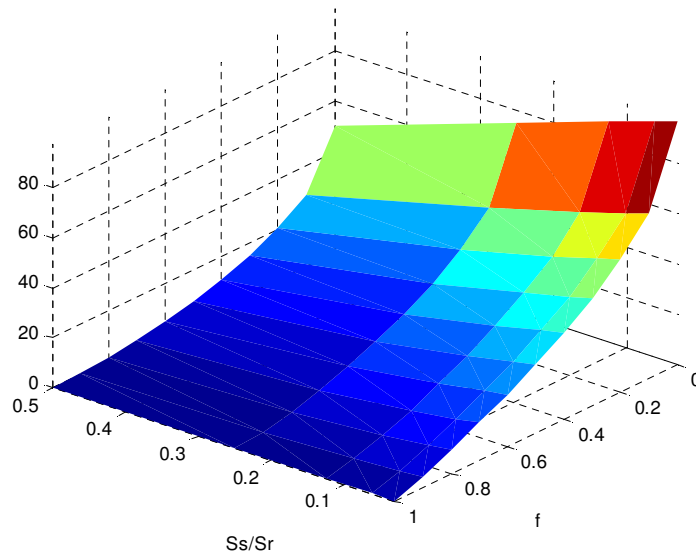
Figure 5.14: Percentage difference between inventory costs under the non-hierarchical and hierarchical setups

The results from the computations for the safety stock cost (Figures 5.15 and 5.16 (a) and (b)) generate similar insights to those generated by the computations for the cyclic inventory cost. However, the safety stock cost under the hierarchical system (5.15) increases stepwise when f increases. This is because of the discrete nature of the Poisson distribution used to determine safety stock levels. A certain number of items maintained as a safety stock can meet the minimum service level for a certain range of demand and decentralization. Changes in the value of f change the demand at both types of service facilities under the hierarchical setup, which can require discrete increments or decrements in the number items in stock. This results in fluctuations of the safety stock cost observed in Figure 5.15. On the other hand, under the non-hierarchical setup, changes in the value of f do not affect the demand at service facilities, hence the safety stock cost (5.8) and the availability level remains constant for different values of f (Figure 5.15). Besides, it can be observed in Figure 5.16 that in some cases, the difference in safety stock costs under the hierarchical and non-hierarchical setups is very low. This is also due to the discrete nature of the Poisson distribution. This

phenomenon is discussed in detail in the next section where the analysis is based on the

(S-1, S) inventory policy.
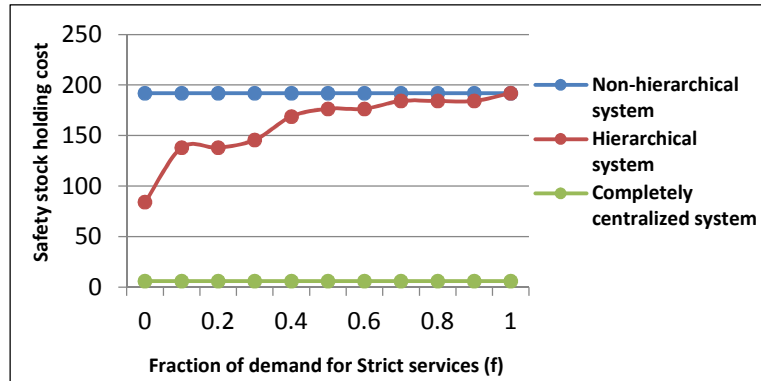
Figure 5.15: Impact of demand fractions for the strict and relaxed services on safety stocks (sr=100, ss=50, min. prob. of no-stockout = 0.98)
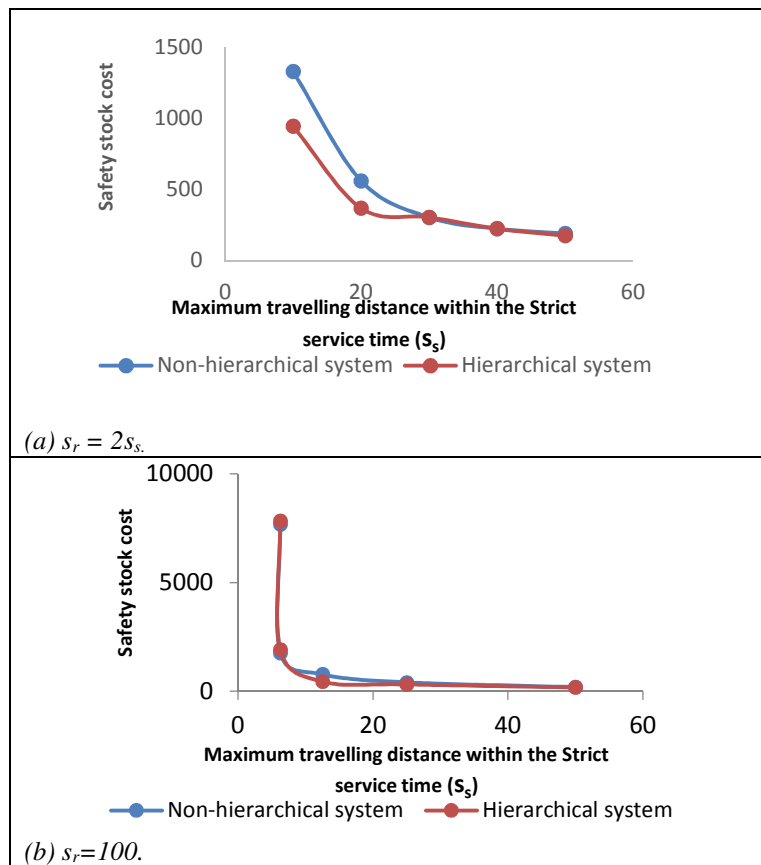
Figures 5.16 (a & b): Impact of service distance constraint on safety stocks (f = 0.5, min. prob. of no-stockout = 0.98)

The following results show that transportation is costlier under the hierarchical setup

compared to the non-hierarchical setup in presence of demand for the relaxed service.

With the decrease in *f*, and hence increase in the percentage of demand for the relaxed service, a system under the hierarchical setup provides services with a higher level of centralization, which increases the travelling (Figure 5.17). With fixed service distance constraints, the difference between the transportation cost under the non-hierarchical setup (5.10) and the transportation cost under the hierarchical setup (5.17) is largest when *f* is minimum, i.e. the fraction of demand for the relaxed service is maximum. With the minimum *f*, a system under the hierarchical setup meets the entire demand only through higher level service facilities (in larger catchment areas). Under the non-hierarchical setup, regardless of the time allowance, a system meets the entire demand with a higher level of decentralization, i.e. demand is always met from the closest facility.
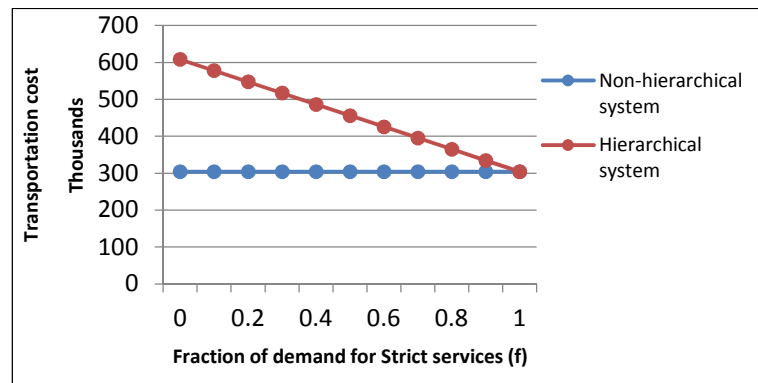
Figure 5.17: Impact of demand fractions for the strict and relaxed services on transportation cost ($s_r=100$, $s_s=50$)

Transportation costs under the hierarchical and non-hierarchical setups linearly decrease as the service distance constraints become stricter. Figure 5.18 (a) presents the resulting transportation costs when the service distance constraints for both the relaxed and the strict services decrease, such that the time window for the relaxed service is twice the time window for the strict service, i.e. $s_r=2s_s$. With equal demand fractions for both service types and a constant maximum service distance for the relaxed service, Figure 5.18 (b) shows that compared to the hierarchical setup, the transportation cost

under the non-hierarchical setup is more sensitive to the strict service distance constraint. As all service calls are met from the nearest facility under the non-hierarchical setup, an increase in the number of facilities reduces the average travelling distance to reach customers for both the relaxed and the strict services. Under the hierarchical setup, though the number of facilities increases with a reduction in the strict service distance constraint, the centralization level for providing the relaxed service does not change.

(a) f =0.5; s_r = 2s_s.

(b) f =0.5, s_r=100.
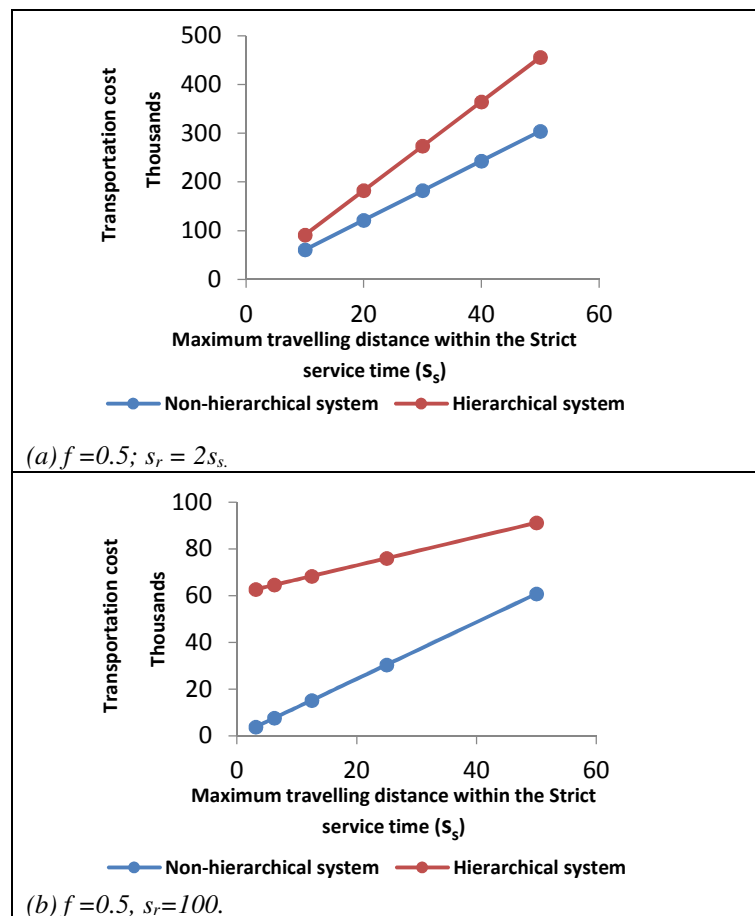
Figures 5.18 (a & b): Impact of service distance constraint on transportation cost

Figure 5.19 shows that the percentage difference between the transportation costs under the hierarchical and the non-hierarchical setups is highest when the fraction of demand for the strict service f and s_s/s_r ratio are at their minimum.
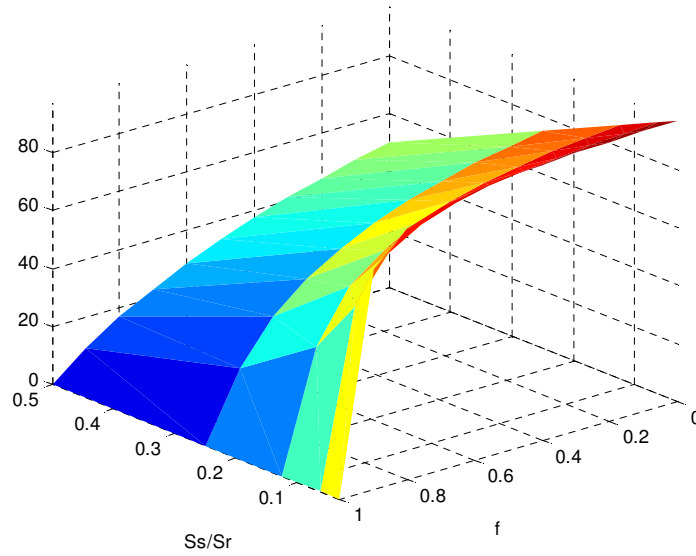
Figure 5.19: Percentage difference between transportation costs under the hierarchical and non-hierarchical setups

5.4.2. Analysis considering (S-1, S) inventory policy

The following analysis considers the (S-1, S) inventory policy to determine overall inventory levels (total base stock) under the modelled non-hierarchical setup (5.9) and the hierarchical setup (5.16). Figures 5.20 (a), (c) and (e) show the effect of varying demand fractions for the relaxed and strict services on inventory levels with three different demand rates. Figures 5.20 (b), (d) and (f) present the corresponding transportation costs under the non-hierarchical and hierarchical setups. The results in these figures are based on the demand rates of 1000, 2000 and 4000 parts per year, and considering the overall area as 200,000 length units^2, the distance constraints for the relaxed and strict services as 100 and 50 length units respectively, the transportation cost per unit length (C_t) as £1, the minimum fill-rate level (proportion of demand met from stock on hand) as 0.98, and the replenishment lead time as 2 days.
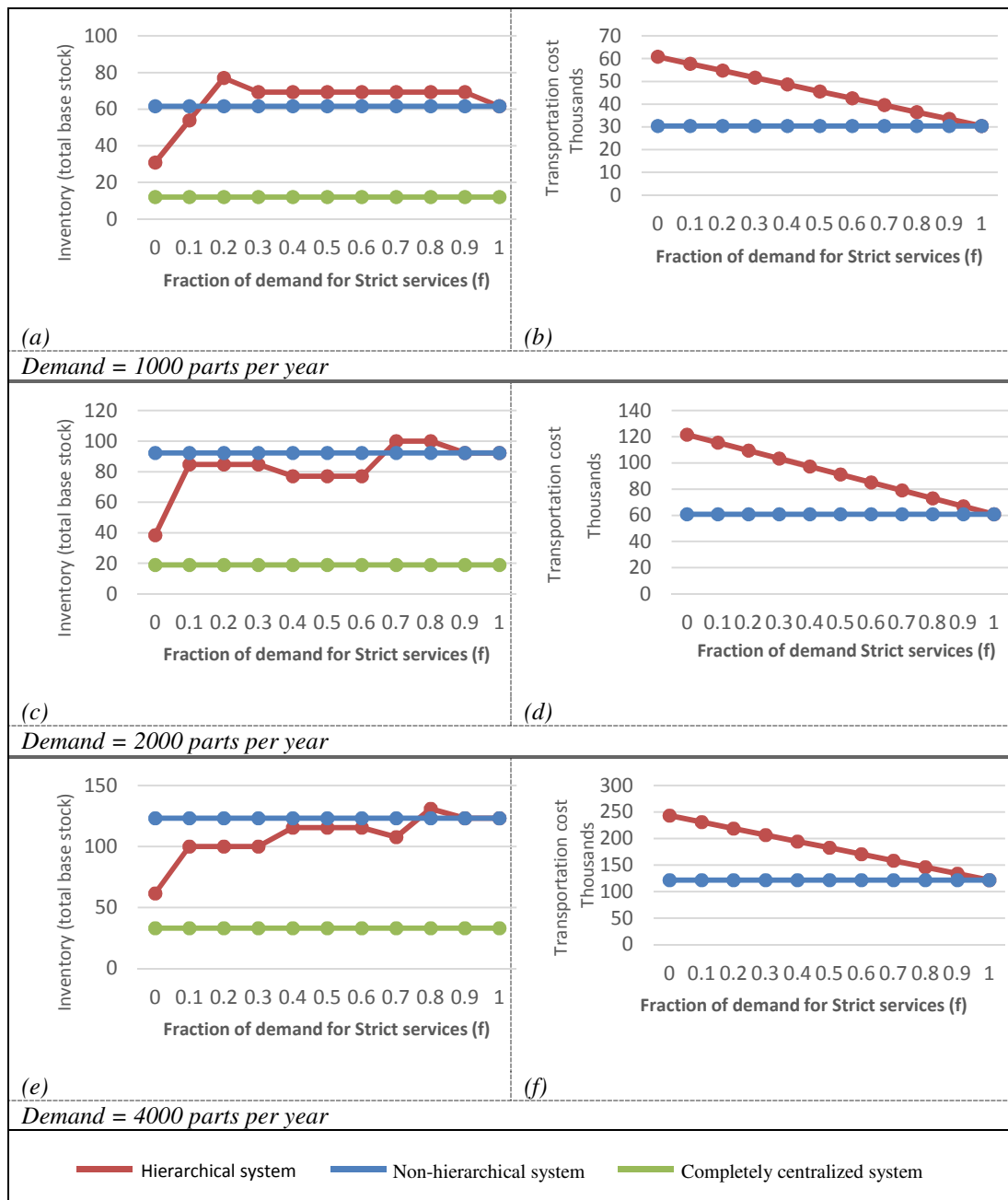
Figures 5.20 (a – f): Impact of demand fraction for the strict and relaxed services on inventory level and transportation cost

Figures 5.20 (a), (c) and (e) give no clear indication on whether a system under the hierarchical setup performs better in terms of inventory levels compared to under the non-hierarchical setup. The results show that the inventory levels under the hierarchical setup are not always lower than the inventory levels under the non-hierarchical setup, whereas, in the presence of demand for the relaxed service, the hierarchical setup always results in a higher transportation cost than the non-

hierarchical setup. As seen in the previous section as well, the improvement in the transportation cost by employing the non-hierarchical setup increases as the fraction of demand for the relaxed service increases (Figures 5.20 (b), (d) and (f)). It seems that unless the fraction of demand for the strict service is rather small, there might be no benefit from deploying a hierarchical setup. Note that the SPL cases do report low demand for the strictest service, which indicates that the hierarchical setup can be beneficial in these real life systems.

The observation with regards to inventory levels is interesting. The results suggest that even though the level of centralization is higher under the hierarchical setup, in many cases, the inventory levels under the completely decentralized non-hierarchical setup are lower. The reason for this behaviour is that a slight reduction in demand at a facility does not always allow to reduce the facility's base stock level.

Although transforming a system from the non-hierarchical setup to the hierarchical setup reduces demand at the majority of the facilities (i.e. at lower level facilities), the required base stock levels at these facilities cannot necessarily be reduced while maintaining the minimum fill-rate level. On the other hand, the transformation increases demand at higher level facilities and can potentially increase the required base stock levels at these facilities to maintain the minimum fill-rate level. Hence, on the whole, this can increase the stocks in the system. This explains why in several cases the inventory under the hierarchical setup, which is more centralized, is higher than the inventory under the more decentralized non-hierarchical setup. The following example can help to better comprehend this behaviour.

Let,

A = overall area = 200,000 Length unit^2

s_s = service distance constraint for strict service = 50 length units

s_r = service distance constraint for relaxed services = 100 length units

λ = demand rate per year = 4,000 units

L = replenishment lead time = 2 days = 0.005479 years

Demand over lead time in the system = 21.92 units

Minimum service availability level (fill-rate) = 0.98

f = fraction of demand for strict service = 0.8

$(1-f)$ = fraction of demand for relaxed services = 0.2

Table 5.4 shows the outcome of inventory functions for the non-hierarchical setup (5.9) and the hierarchical setup (5.16) considering the above values for the variables.

Table 5.4: (S-1, S) inventory policy example under the hierarchical and non-hierarchical setups

Non-hierarchical setup	Hierarchical setup
Number of facilities = 30.8 Demand over lead time at one facility = 0.71 units Required stock level at one facility = 4 (providing 99.4% fill-rate) Total stock in the system = 4*30.8 ≈ **123 units**	Number of Higher Level facilities = 7.7 Demand over lead time at one Higher Level facility = 1.14 units Required stock level at one Higher Level facility = 5 (providing 99.37% fill-rate) Total stock in Higher Level facilities = 7.7*5 = **38.49** Number of Lower Level facilities = 23.1 Demand over lead time at one Lower Level facility = 0.57 units Required stock level at one Lower Level facility = 4 (providing 99.7% fill-rate) Total stock in Lower Level facilities = 23.1*4 ≈ **92.4** Total stock in the system ≈ **131 units**

It can be seen in Table 5.4 that, under the non-hierarchical setup, demand at one facility is 0.71 units over the replenishment time, which requires 4 units to be maintained at a facility to meet the 98% minimum fill-rate. Under the hierarchical setup (where lower level facilities are in majority), the demand at a lower level facility reduces to 0.57 units over the replenishment time from 0.71 units, but we still need to

maintain 4 units in a facility to meet minimum 98% fill-rate requirement. If we reduce the base stock level at a facility from 4 to 3, the fill-rate becomes 97.974%. At the same time, demand at a higher level facility increases from 0.71 units over the replenishment time to 1.14 units. This increase also requires to increase the base stock level at a higher level facility from 4 to 5 to achieve the minimum 98% fill-rate level. Maintaining the base stock level of 4 at a higher level facility will provide a 97.1% fill-rate. Hence, overall the stock level in the system increases under the hierarchical setup compared to under the non-hierarchical setup.

For further investigation, Figure 5.22 explores the spike in the inventory level under the hierarchical setup in Figure 5.20 (e) when the fraction of demand for the strict service is 0.8 (highlighted in Figure 5.21). By dividing the demand fraction scale into smaller units, Figure 5.22 and Table 5.5 show that the increase in the inventory under the hierarchical setup is stepwise. As mentioned above, under the non-hierarchical setup, demand at one facility is 0.71 (independent of the fractions of demand for both service types), which requires 4 units in the inventory at a facility to meet the minimum fill-rate level of 98%. Total inventory under the non-hierarchical setup is equal to $4 \times 30.8 \approx 123$ units (where 30.8 is the total number of facilities). When f is equal to 0.80 the demand at a higher level and a lower level facility is equal to 1.14 and 0.57 units over the replenishment lead time respectively requiring 5 and 4 units in inventory respectively to meet the minimum fill-rate requirement. Demand at higher level facilities decreases as f increases, but the required stock level to meet the minimum fill-rate level remains the same at 5 till f equals to 0.85. Only when f increases to 0.86 the required stock level at a higher level facility can be reduced to 4 (with demand of 1.01 units over the replenishment lead time), reducing the overall inventory level in the system (Figure 5.22 and Table 5.5).
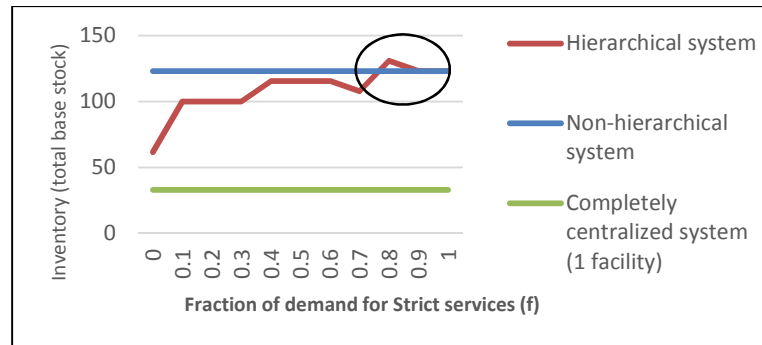
Figure 5.21: Impact of demand fraction for Strict and Relaxed services on inventory level (with demand of 4000 parts per year) – copy of Figure 5.20 (e)
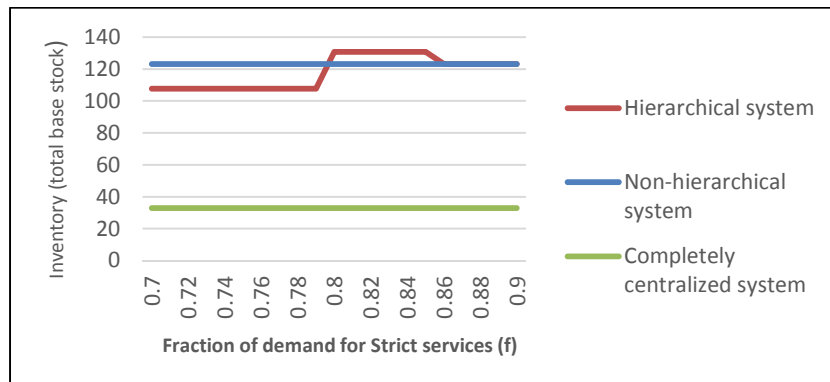
Figure 5.22: Impact of demand fraction for strict and relaxed services on inventory level (with demand of 4000 parts per year)

Table 5.5: Inventory under the hierarchical setup (corresponding to Figure 5.22)

(1-f)	f	Demand over lead time at a higher level facility	Demand over lead time at a lower level facility	Stock level at a higher level facility (S_h)	Stock level at a lower level facility (S_l)	Total inventory in the system = $n_r S_h + (n_s - n_r)S_l$. *
0.3	0.7	1.35	0.50	5	3	108
0.29	0.71	1.33	0.51	5	3	108
0.28	0.72	1.31	0.51	5	3	108
0.27	0.73	1.29	0.52	5	3	108
0.26	0.74	1.27	0.53	5	3	108
0.25	0.75	1.25	0.53	5	3	108
0.24	0.76	1.22	0.54	5	3	108
0.23	0.77	1.20	0.55	5	3	108
0.22	0.78	1.18	0.56	5	3	108
0.21	0.79	1.16	0.56	5	3	108
0.2	0.8	1.14	0.57	5	4	131
0.19	0.81	1.12	0.58	5	4	131
0.18	0.82	1.10	0.58	5	4	131
0.17	0.83	1.07	0.59	5	4	131
0.16	0.84	1.05	0.60	5	4	131
0.15	0.85	1.03	0.61	5	4	131
0.14	0.86	1.01	0.61	4	4	123
0.13	0.87	0.99	0.62	4	4	123
0.12	0.88	0.97	0.63	4	4	123
0.11	0.89	0.95	0.63	4	4	123
0.1	0.9	0.93	0.64	4	4	123

* $n_r = 7.7$ is the number of higher level facilities, and $n_s-n_r = 23.1$ is the number of lower level facilities

With equal fractions of demand for both service types (*f =0.5*), Figures 5.23 (a), (c) and (e) show the impact of reducing service distance constraints (such that $s_r = 2s_s$) on inventory levels. Inventory levels increase as the service distance constraints reduce, with the non-hierarchical setup performing better than the hierarchical setup in some of the cases. On the other hand, the transportation cost (Figures 5.23 (b), (d), and (f)) increases linearly when the service distance constraints become more relaxed with the non-hierarchical setup performing better than the hierarchical system in all cases. The gap in the transportation cost under the hierarchical and non-hierarchical setups increases with the relaxation in service distance constraints.
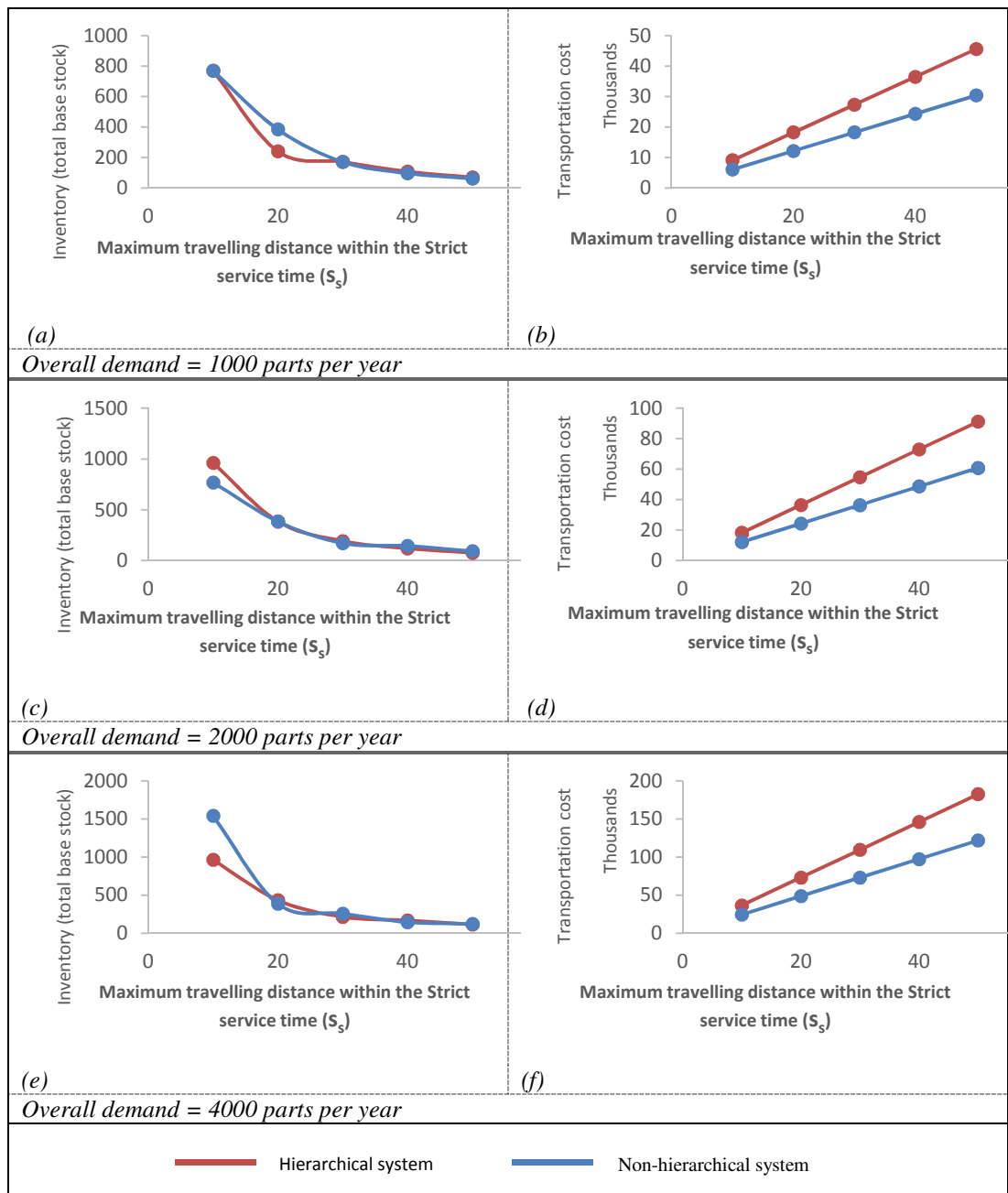
Figures 5.23 (a – f): Impact of service time constraint on inventory level and transportation cost

Figures 5.24 (a – c) show the impact of different fill-rate constraints on inventory levels when $s_s = 50$, $s_r = 100$ and $f = 0.5$. The inventory levels under the hierarchical and non-hierarchical setups increase stepwise when the minimum required fill-rate increases, with non-hierarchical setup performing better in several cases. The figures also show that the inventory level in a completely centralized system, where there is only one service facility, has very low sensitivity to the fill-rate constraint.

(a) Demand = 1000 parts per year

(b) Demand = 2000 parts per year

(c) Demand = 4000 parts per year
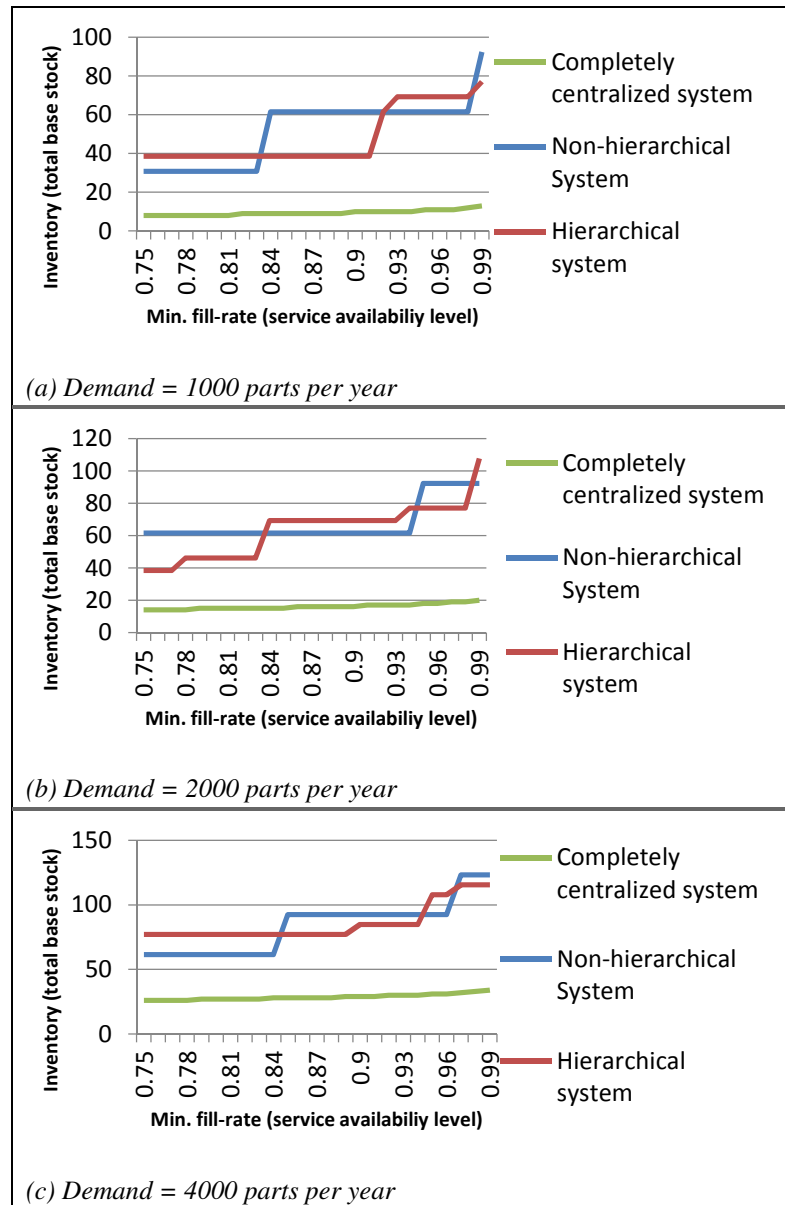
Figures 5.24 (a – c): Impact of required minimum fill-rate level on inventory

5.4.2.1.   Analysis based on a past demand of a part at Company B

This section presents an analysis based on the demand information of a Company B's average part and the company's warehouse locations in the mainland UK. The inventory and transportation levels are analysed with respect to the fractions of demand associated with two time-based service types considering a hierarchical and a non-hierarchical setup. The purpose is to realize how some of the research variables

can interact in a real-world setting. The analysis is based on the following assumptions and simplifications due to the limited information:

- The forward replenishment role of the central warehouse (at Milton Keynes) is ignored. That is, the system is considered to be a single echelon system.
- In the non-hierarchical setup, all of the local demand is fulfilled by the local warehouse, regardless of the requested service time.
- Under the hierarchical setup, one of the warehouses is considered as a higher level warehouse while the others are considered as lower level warehouses. All warehouses (both higher and lower level) fulfil the local requests for a part supply in the short service time(s) (strict services). A request for a part supply within a longer service time (relaxed service), e.g. next day service, is only fulfilled by the higher level warehouse to avail the opportunity to meet the relaxed service demand centrally (Table 5.6).

Table 5.6: Facility and service types

Facility type	Strict service(s)	Relaxed service
Higher level warehouse	✓	✓
Lower level warehouses	✓	

- The ratio of demand for the strict and relaxed services is same at all warehouses.
- All warehouses operate under the one-for-one (S-1, S) inventory policy assuming the Poisson demand process.
- Warehouse to warehouse distance (determined through Google maps) is considered if a local request of a warehouse is fulfilled by the higher level warehouse. Also, the transportation distance is neglected for a local service.

Table 5.7 presents the actual demand for a service part in the locality of each of Company B's mainland UK warehouses over approximately six months (195 days) and Figure 5.25 (a) shows the location of company's warehouses in the mainland UK.

Table 5.7: Mainland demand

Warehouse	Demand	% of Demand
Edinburgh	52	7.58
Manchester	131	19.10
Barming	105	15.31
Milton Keynes	152	22.16
West London	6	0.87
Aberdeen	9	1.31
Birmingham	112	16.33
Newcastle	32	4.66
Bristol	35	5.10
Southampton	4	0.58
Leeds	19	2.77
Cambridge	2	0.29
Glasgow	21	3.06
Nottingham	2	0.29
Plymouth	4	0.58
Total demand	686	

Figure 5.25 (b) presents the hierarchical setup which is considered for the investigation, with Birmingham as the higher level warehouse providing strict services in its locality and providing the relaxed service throughout the system. All other warehouses are considered as lower level warehouses only fulfilling the local requests for parts in short service times. Compared to other warehouses, selecting Birmingham as the higher level warehouse results in the lowest travelling level. That is, selecting Birmingham as the higher level warehouse minimizes $\sum_{i \in I} \left(X_i \sum_{j \in J_i} \lambda_j T_{ij} \right)$ subject to $\sum_{i \in I} X_i = 1$ and $X_i \in \{0,1\}$, where *I* is the set of all warehouses, X_i is one if warehouse *i* is the higher level warehouse (else zero), J_i is the set of warehouses other than warehouse *i*, λ_j is the demand at warehouse *j*, and T_{ij} is the distance between the warehouses *i* and *j*. See Appendix 4 for the distances between the warehouses.

a) Warehouse locations (non-hierarchical)	*b) Hierarchical setup*
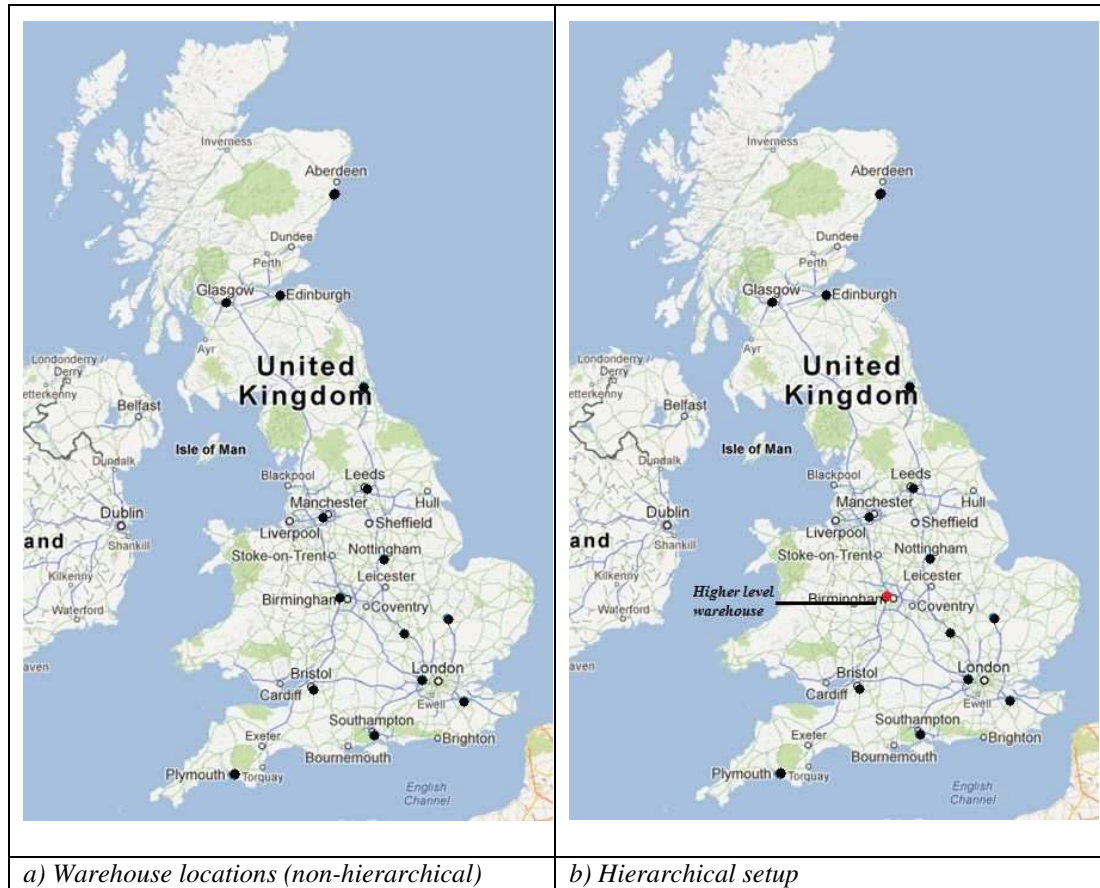
Figures 5.25 (a & b): Warehouse locations

The inventory levels in the non-hierarchical setup are determined under the (S-1, S) policy considering that each facility satisfies all the demand in its locality. Under the hierarchical setup, considering *f* to be the fraction of total demand for the strict services and *1-f* be the fraction of total demand for the relaxed service, the inventory levels are determined considering,

$Demand\ served\ by\ a\ local\ warehouse\ j = f \times \lambda_j$,

where λ_j is the total demand in the locality of the lower level warehouse *j*.

$Demand\ served\ by\ the\ higher\ level\ warehouse = (1 - f) \sum_{j \in J} (\lambda_j) + \lambda_{HW}$,

where *J* is the set of lower level warehouses, λ_j is demand in the locality of the lower level warehouse *j*, and λ_{HW} is demand in the locality of the higher level warehouse.

145

Neglecting the travelling distance to supply a part by a warehouse in its own locality,

$$Total\ distance\ travelled\ to\ reach\ demand\ locations = (1-f)\sum_{j\in J}(\lambda_j T_j),$$

where λ_j is demand in the locality of the lower level warehouse j and T_j is the distance between the higher level warehouse and the lower level warehouse j.

Considering the demand in the system (Table 5. 7), the minimum fill-rate level as 98%, and 1 day replenishment time, Figures 5.26 and 5.27 present the total inventory level and average per day travelling respectively against varying demand fractions for the relaxed and strict services.
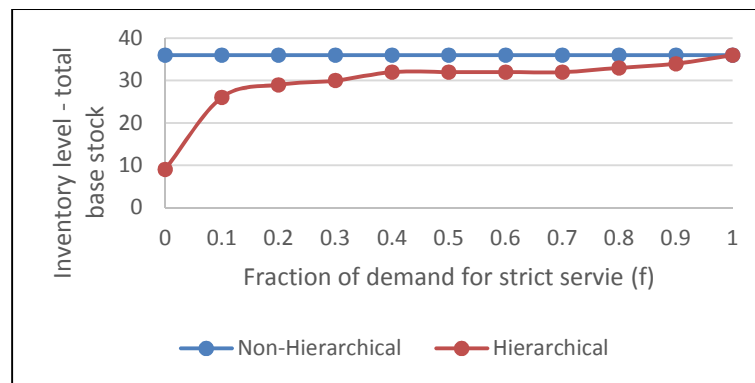
Figure 5.26: Inventory level under different demand fractions for strict and relaxed services
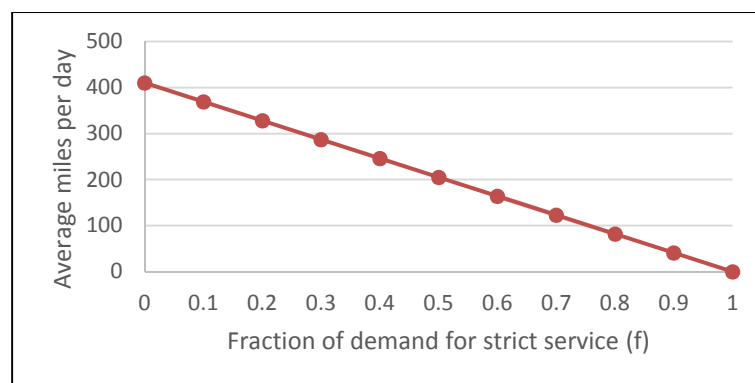
Figure 5.27: Average travelling per day to supply parts under the hierarchical setup

The results are similar to those in the previous section. The higher fractions of demand for the relaxed service result in higher travelling for the distribution. The inventory

levels on the other hand are reduced under the hierarchical setup and a higher demand fraction for the relaxed service. However, for a significant range of the demand fractions, the inventory level remains constant under the hierarchical setup. When the fraction of demand for the strict services *f* drops from 0.7 to 0.4, the inventory level under the hierarchical setup does not reduce (Figure 5.26), while the transportation level increases linearly throughout this change in *f* and almost doubles from 123 miles to 246.2 miles (Figure 5.27). However, in the low range of the demand fraction for the strict services (*f < 0.4*) the benefit on inventory cost from having the hierarchical setup can be significant. For example, when *f = 0.1*, there is a difference of 10 units between the inventory levels under the hierarchical and non-hierarchical setups. Considering the average cost of a part as £1,000 and 17,000 different parts in inventory at the company, the potential benefit of operating under the hierarchical setup can be substantial. Note that at Company B approximately 60% of demand is for the next day and eight hour service options and only around 10% of the demand is for the two hours (the strictest) service time option. When the demand for the strict services completely fades, the inventory under the hierarchical setup sharply drops, as the setup can supply parts in a completely centralized fashion from the higher level warehouse. When the fraction of demand for the strict services *f* reduces from 0.1 to 0, the inventory level under the hierarchical setup drops from 26 units to 9 units, a 65.4% decrease. The system under the non-hierarchical setup holds four time more stocks than under the hierarchical setup when *f = 0*.

5.5. Conclusion

The presented model and the analysis explore effects of different service time (or distance) constraints on some of the important components of distribution cost. The

results presented in the preceding sections confirm that there can be a significant impact of service time constraints on inventory and transportation costs. The proportions of the strict and relaxed service requests in overall demand also impact on the inventory and transportation costs if facilities in a distribution system have a hierarchical setup.

Under the (R, Q) inventory policy, the results show that the distribution through a hierarchical organization of facilities, where customers are not necessarily served from the nearest stocking point (in order to allow a higher level of centralization while adhering to service time commitments), can lower the overall inventory levels. However, this is at the expense of increased transportation costs.

A non-hierarchical organization of service facilities, in which all customers are served from the nearest stocking point, can result in a lower average distance to reach customers. However, when considering the (R, Q) inventory policy, resources associated with inventory management can be wasted under a non-hierarchical setup as the stocks are deployed in a more decentralized fashion even though a more centralized deployment can meet the service time constraints. As a non-hierarchical system treats all service calls in a similar fashion, overall similar inventory and transportation costs incur in serving customers with different service time requirements. The inventory and the transportation costs in a hierarchical system on the other hand can react to the changes in the demand fractions for different time-based service types. It can be beneficial to deploy stocks with high inventory related costs in a hierarchical fashion while deploying stocks with low inventory related costs in a non-hierarchical fashion.

The investigation based on the (S-1, S) inventory policy gives some counterintuitive outcomes. Besides the transportation cost being higher in a system under a hierarchical

setup as it serves in a more centralized fashion, a hierarchical setup does not necessarily result in a lower overall inventory level compared to that under a non-hierarchical setup. Because demand at facilities under a non-hierarchical setup is not affected when the fractions of demand for the relaxed and strict services change, the inventory level in the system also does not change. Under the hierarchical setup, due to the discrete nature of the (S-1, S) inventory policy, the inventory levels change stepwise when the demand fractions of the service types change; in several cases exceeding the required inventory level if the system was being operated as non-hierarchical. Results suggests that, apart from the cases where the fraction of demand for the relaxed service is very high, it can be financially better to distribute under a more decentralized (and hence more responsive) system as both inventory and transportation levels in the system can be lower. Hence, the presence of some demand for the service in the longer time window might not always result in lower the operational cost. However, demand for the service in the longer time window can increase the opportunity for inventory sharing between facilities. This is investigated in the next chapter, which also verifies the insights generated by the model presented in this chapter.

Chapter 6: Impact of inventory sharing on service availability and transportation levels

6.1.Introduction

The model presented in the previous chapter gives interesting insights but one limitation of the work is the boundary effect error. The model provides an analytical treatment of the problem and a flexibility to investigate several combinations of parameter values, but at the cost of some accuracy. Building on the study in the previous chapter, this chapter presents the analysis based on specific cases of facility point locations and their regions in a bounded square area. The following work verifies the insights generated so far. However, with the focus on items with low demand, the main extension is the simulation study which investigates the impact of different inventory sharing configurations considering the (S-1, S) inventory policy on the service availability level and transportation cost. When considering inventory sharing between facilities, estimating fill-rates becomes challenging even with strict assumptions on maximum stock levels, e.g. $S \leq 1$, and on the number of facilities in an inventory sharing pool, e.g. only two (Iyoob and Kutanoglu, 2013). In this study, the inventory sharing/transshipment is only considered if the demand point is within the coverage distance from the sharing facility. There can be several overlapping areas of different sizes and forms covered by multiple facilities having multiple covering ranges. This also makes it complex to estimate the average service distance.

The chapter is organized as follows. Section 6.2 states the description of the problem addressed in this chapter. In Section 6.3, issues in packing a rectangle with regular

hexagons are highlighted and the square partitioning is discussed as the alternative. Section 6.4 presents the cases for the analysis. Section 6.5 presents inventory and transportation level computations considering the cases in Section 6.4 with varying fractions of demand for two time-based service types. The computation results are not discussed in detail as they are similar to those generated in the previous chapter. The simulation study is presented in Section 6.6 to investigate the impact of different inventory sharing scenarios on the service availability and transportation levels against varying demand fractions for the time-based service types. Finally, the main insights are summarised in the conclusion in Section 6.7.

6.2. Problem description

Consider a square area in a Cartesian system. Customers with identical demand are uniformly distributed inside the area and the customers are supplied with parts within two different service time commitments. The service that ensures supply within the short time window is referred as the '*strict*' service, and the service that ensures supply in the longer time window is referred as the '*relaxed*' service. Travelling distances inside the area are Euclidean and are proportional to the time.

There exists a distribution system in the area comprising of facilities that, considering the distance (or time) constraint for the strict service, cover the entire area efficiently in terms of transportation by being located in the middle of their service regions. A service region comprises all the points that are closer to the respective facility than to any other facility (i.e. Voronoi cell). Under the non-hierarchical setup, all facilities provide both the strict and the relaxed services in these regions. To set-up a hierarchical system, a subset of facilities is designated as higher level facilities, which

are just sufficient to cover the entire area considering the relaxed service distance constraint. Under the hierarchical setup, all facilities provide the strict services, but, only higher level facilities provide the relaxed service for centralization. Facilities other than higher level facilities are lower level facilities (only providing the strict service).

The objectives of the study are to,

1) determine the effect of decentralization (the number of facilities) and the fractions of demand for the strict and relaxed services on the inventory level and the average distance to reach a demand point under the hierarchical and non-hierarchical setups, and

2) determine the effect of inventory sharing with different combinations of demand fractions for the strict and relaxed services on the service availability level (fill-rate) and the average distance to reach a demand point.

The terms '*availability level*' and '*fill-rate*' are used interchangeably and refer to the percentage of demand met from stocks on hand. The term '*average service distance*' is used to refer to the average distance to reach a demand point. The terms '*service distance constraint*' or just '*distance constraint*' is used to refer to the maximum distance that can be travelled for a time-based service type.

6.3. Packing a rectangle with regular polygon patterns

When serving customers with a uniform geographical distribution in a region, in terms of the average service distance, the optimum location for the facility point is the centroid of the region (Suzuki and Okabe, 1995) or the centre of the region in case the region is a regular polygon. The average distance from the origin is minimized

uniquely by the regular polygon centred at the origin (Morgan and Bolton, 2002). This cannot be achieved when regular hexagonal regions are to be bounded in a rectangle. Within a regular hexagonal pattern (Figure 6.1), the distance between two adjacent centre points in one layer is $\sqrt{3}s$, where s is the side (edge) length of hexagons and the maximum distance that can be travelled in a hexagon from its centre point. The vertical distance between two adjacent layers of centre points is *1.5s*. And, points on two adjacent layers are not aligned; there is a horizontal displacement of $\frac{\sqrt{3}}{2}s$.
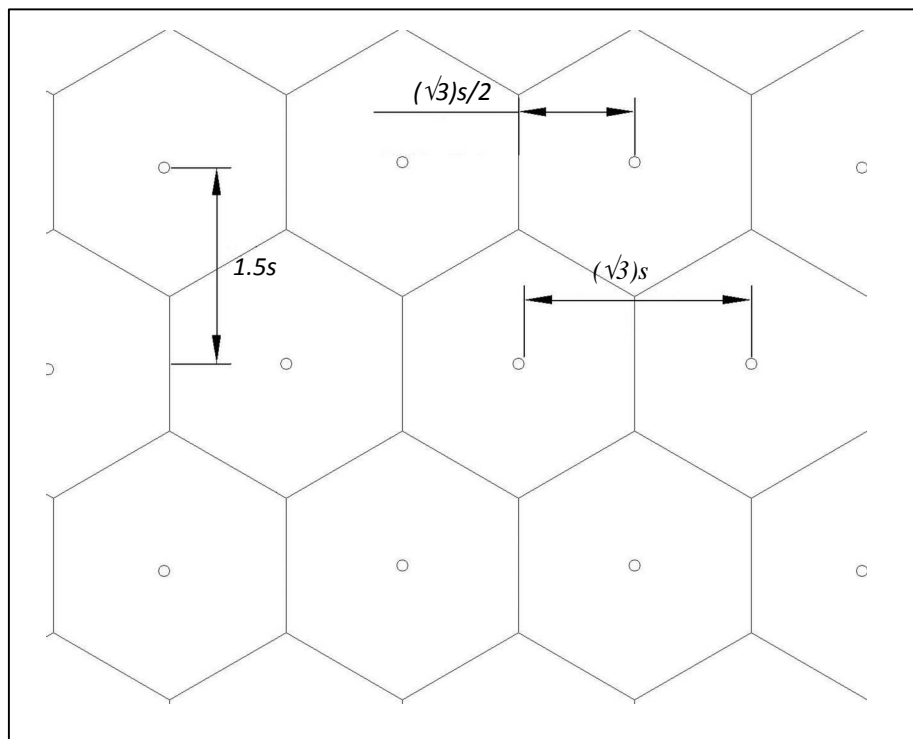
Figure 6.1: Regular hexagonal pattern

Let centre points of the hexagons in a regular hexagonal pattern (Figure 6.1) represent facility points. To bind a regular hexagonal pattern in a square/rectangle, starting placement from any corner point, the first layer of facilities can be started $\frac{\sqrt{3}}{2}s$ away from the adjacent vertical boundary and $\frac{1}{2}s$ away from the adjacent horizontal boundary (Figure 6.2 (a)). But, to retain the regular hexagonal pattern, the next layer

of facilities, *1.5s* away from the first row, has to start with a facility on the vertical boundary. As a result of this placement, all facilities adjacent to the horizontal boundary, and every second facility adjacent to the vertical boundary serve in partial hexagonal areas (Figure 6.2 (a)). Another approach to the placement is presented in Figure 6.2 (b), which is equivalent to rotating the pattern in Figure 6.2 (a) 90º to the right. In this placement, all facilities adjacent to the vertical boundary, and every second facility adjacent to the horizontal boundary serves a partial hexagonal area. These are clearly not good solutions with respect to the average travel distance and maximizing the coverage as some facilities are located at the regional boundaries.
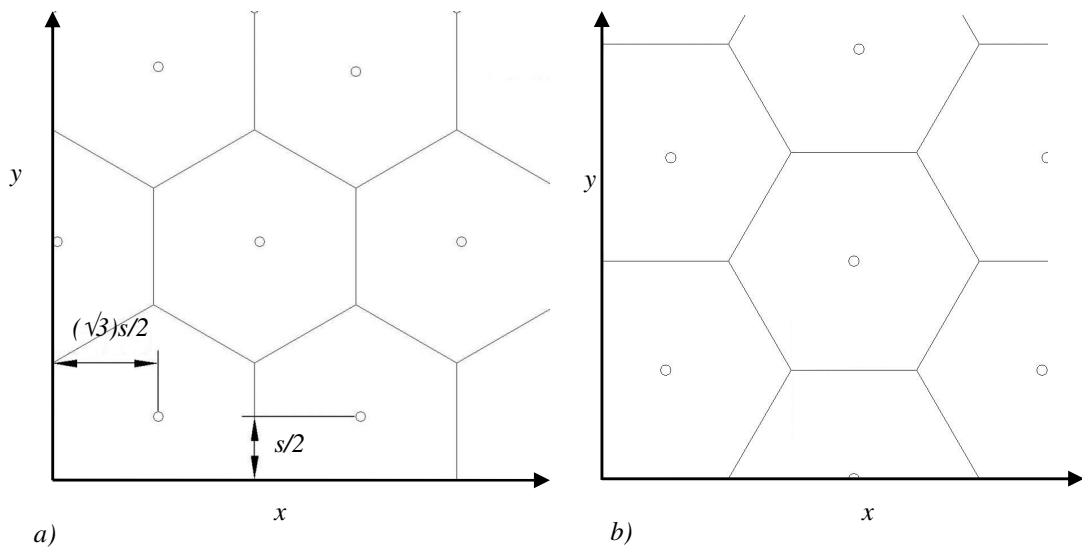
Figure 6.2 (a & b): Binding hexagonal pattern in a rectangle

There are additional issues in binding a hierarchical hexagonal pattern (Figure 6.3). Recalling from the previous chapter, in a nested hierarchical hexagonal pattern allowing the relaxed and strict service time constraints ratio of 1:0.5, lower level facilities are located on the middle of the edges of higher level facilities' hexagonal service areas for the relaxed service (Figure 5.7 Chapter 5). Such hierarchical hexagonal patterns comprise of two types of facility point layers; 1) higher level

layers, having alternating higher level and lower level facilities, where each higher level facility has both a smaller and a larger service area, and 2) lower level layers, having only lower level facilities with smaller service areas (Figure 6.3). Two adjacent higher level facility points on a higher level layer are $2 \times \sqrt{3}s$ away from each other, where s is the side length of smaller hexagons (representing service areas for the strict service). The vertical distance between two adjacent higher level layers is $3s$, and the horizontal displacement of between two adjacent higher level layers is equal to $\sqrt{3}s$.
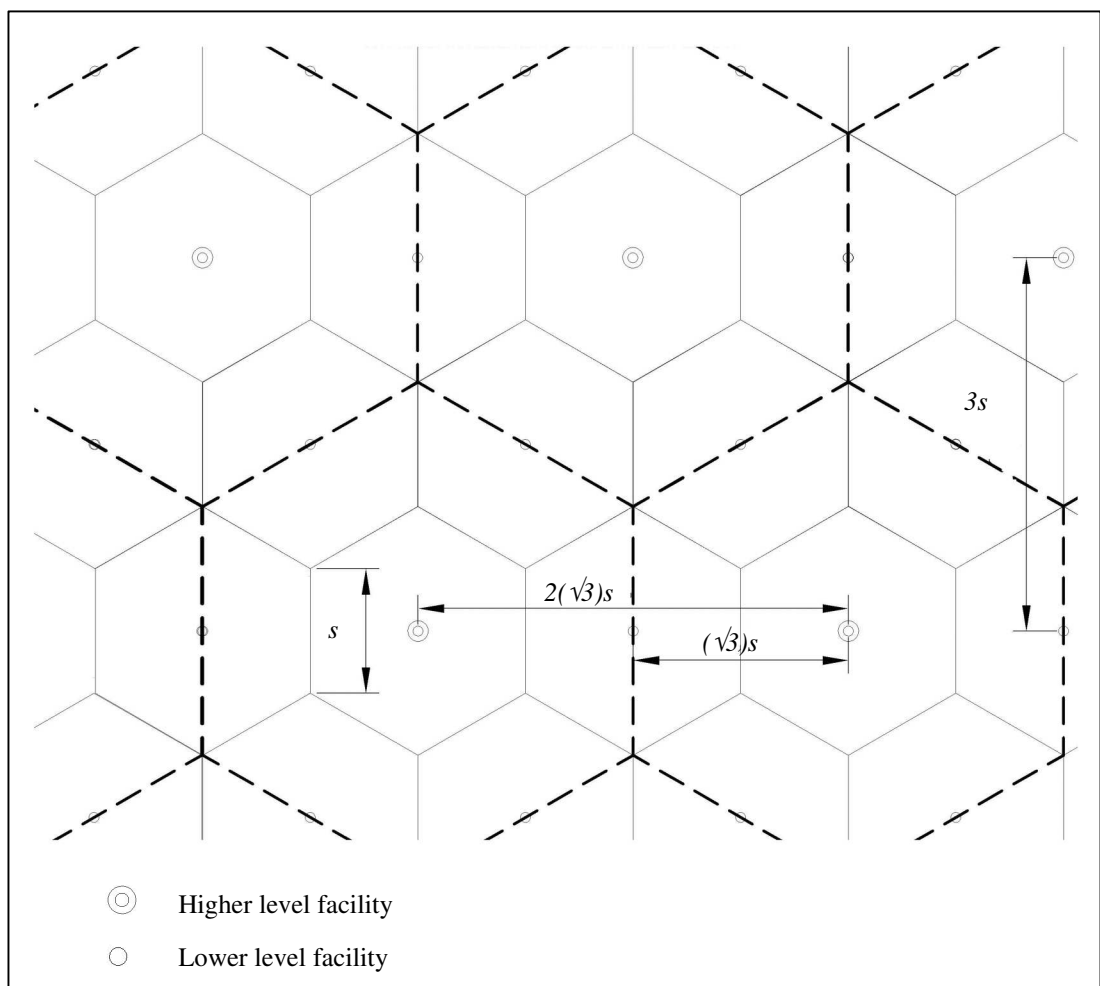
Figure 6.3: Hierarchical hexagonal pattern

Converting the single level hexagonal patterns in Figures 6.2 (a) and (b) into nested hierarchies results in some areas (on boundaries) that are not covered for the relaxed service by higher level facilities (Figures 6.4 (a) and (b)).
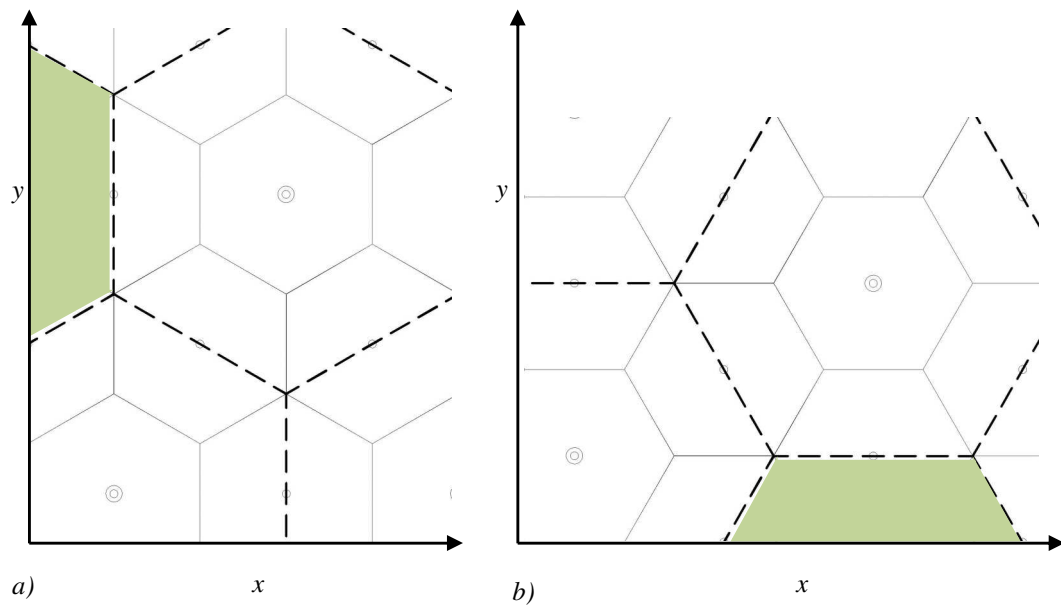
Figures 6.4 (a & b): Uncovered areas for relaxed service

Two approaches of fitting a hexagonal hierarchy that can allow covering for both service types are presented in Figures 6.5 (a) and (b). The patterns require the first facility point on the first layer (adjacent to the lower boundary) to be placed at the left vertical boundary. In these patterns, some higher level and lower level facility points are required to be located on boundaries. The patterns give a whole range of different service areas.
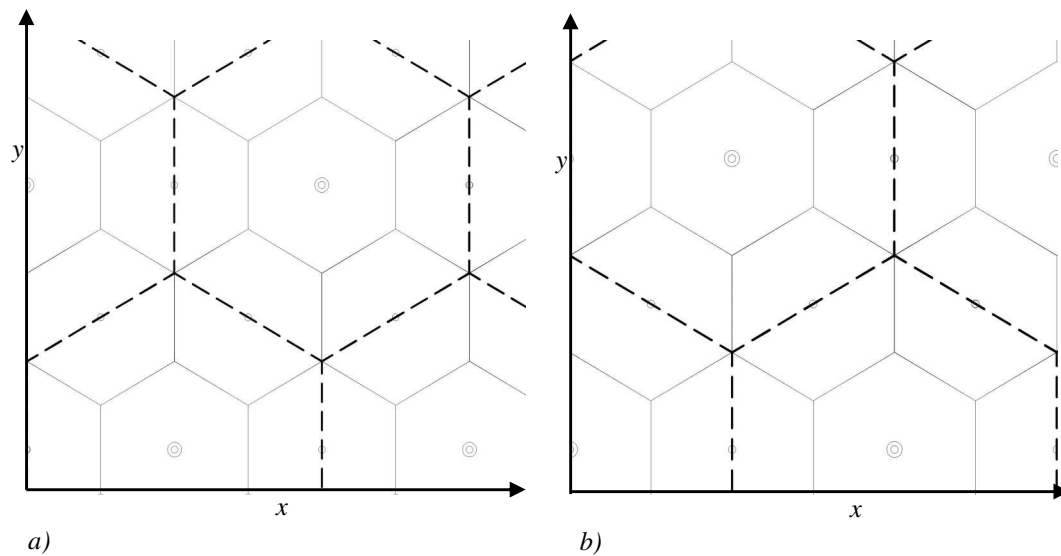
Figure 6.5 (a & b): Hierarchical hexagonal placement for full coverage

In view of the above mentioned issues with packing a regular hexagonal pattern in a rectangular area, we base our analysis on the square packing (Figures 6.6 and 6.7). Squares are one of the three regular polygons (others being hexagons and triangles) which can tile a plane to provide complete coverage. When tiling a rectangle with a hexagonal pattern of service areas, most facilities adjacent to the boundaries serve in partial hexagonal areas and are located away from the service area centres. Regular square tilling, on the single (strict) level under certain conditions, can represent a system where all facility points serve within full service areas from the centres. Also, it has been noted by Lösch, a seminal contributor to regional science and urban economics, that the superiority of a hexagon over a square region is small and of no practical importance in many instances. Square "is not much inferior to the hexagon" and is the second best region with an added advantage of simply drawn boundaries (Lösch, 1954).
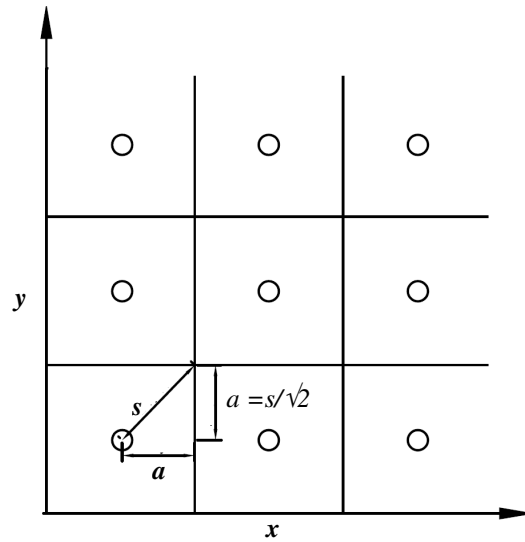
Figure 6.6: Square packing

As in a hexagonal hierarchical pattern (Figure 6.3), a hierarchical square pattern can be set-up by alternatively placing layers of alternating higher level and lower level facility points and layers of all lower level facilities (Figure 6.7). A hierarchical square packing can also result in the relaxed and strict service time constraints ratio of 1:0.5. That is, the maximum travelling distance in the catchment areas for the relaxed service is twice that in the catchment areas for the strict service. All rows of facility points are aligned and there are no facility points located on the boundaries.
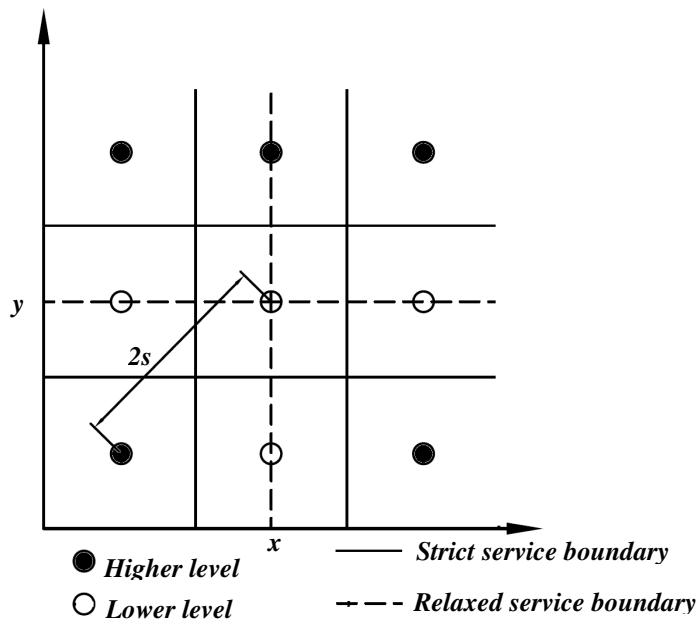
Figure 6.7: Hierarchical location in square pattern

A symmetrical nested hierarchical pattern with square packing is only possible when the number of layers and the number of facilities in a layer are odd. Also, as a limitation, although the strict service is provided from the centres of the service areas, the same is not true for the relaxed service provision in the areas adjacent to the boundaries.

6.4. Cases

We analyse four cases in a unit square. The distance constraints for the strict service are selected with the consideration that the unit square is perfectly tiled, i.e. without partial strict service regions (squares). Setting a hierarchy of these strict service regions however do result in relaxed service regions adjacent to boundaries being partial squares. The cases along with the maximum distances from the facilities in their service areas (the service distance constraints) are presented in Figures 6.8 – 6.11. The dashed lines --- in these figures represent the service area boundaries for relaxed service.

| Distance constraint for strict service = 0.2357 | Distance constraint for relaxed service = 0.4714 |
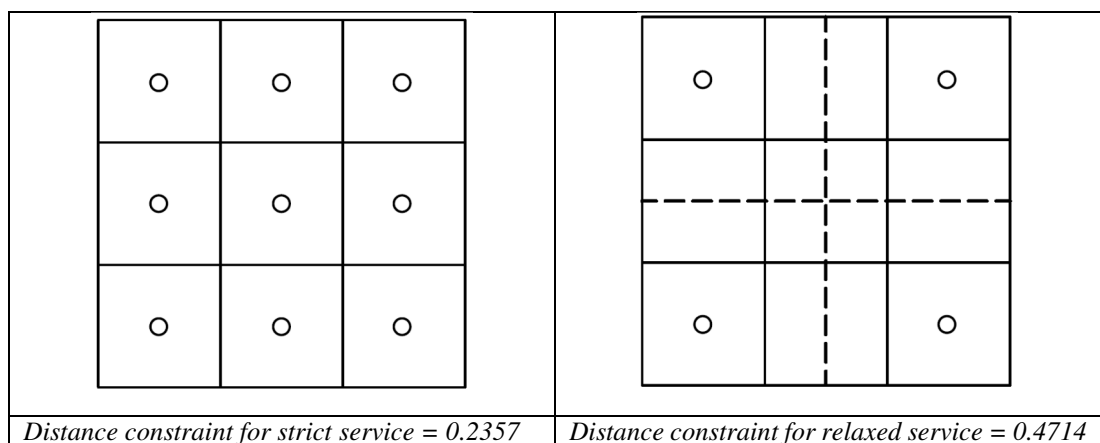
Figure 6.8: Case 1 – 9 facilities in a unit square

| Distance constraint for strict service = 0.1414 | Distance constraint for relaxed service = 0.2828 |
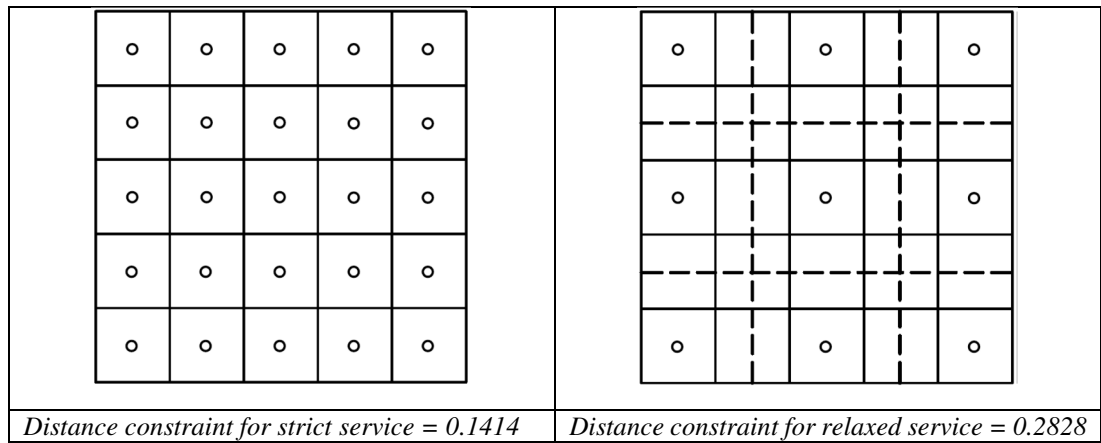
Figure 6.9: Case 2 – 25 facilities in a unit square

| Distance constraint for strict service = 0.1010 | Distance constraint for relaxed service = 0.202 |
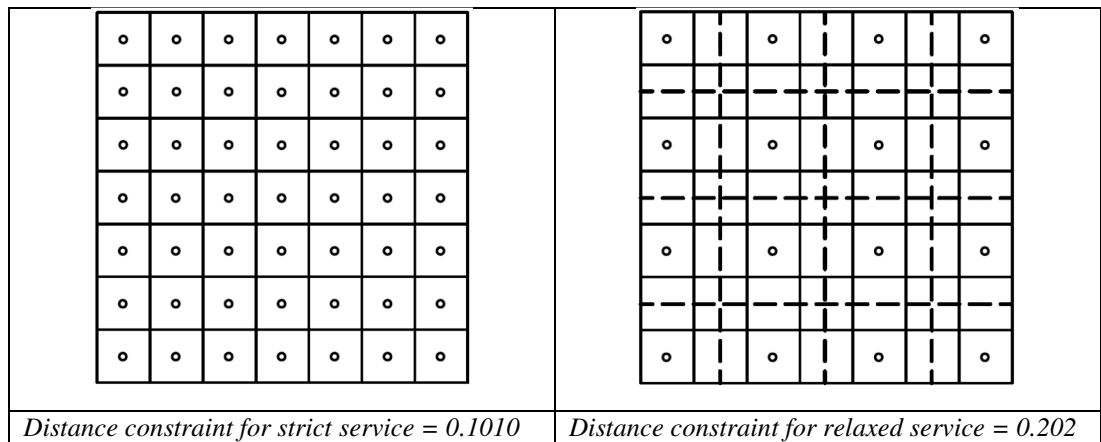
Figure 6.10: Case 3 – 49 facilities in a unit square

| Distance constraint for strict service = 0.0786 | Distance constraint for relaxed service = 0.1572 |
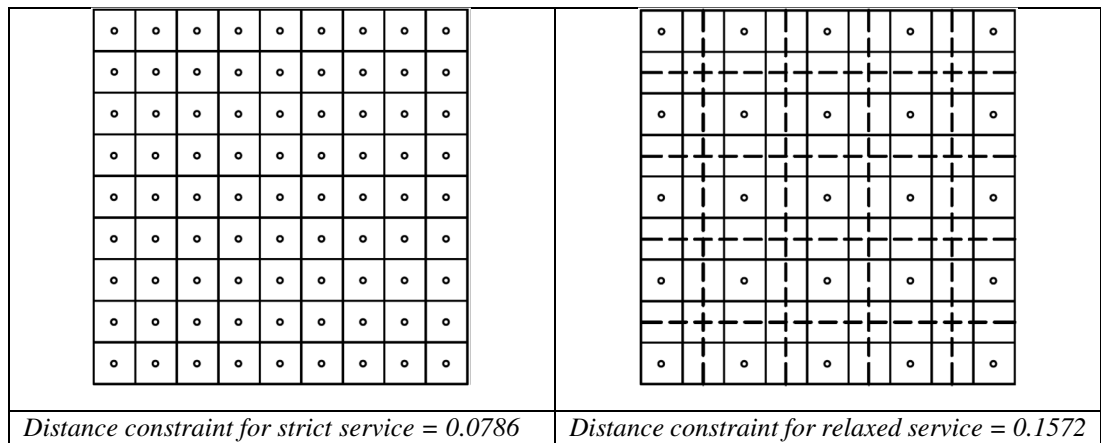
Figure 6.11: Case 4 – 81 facilities in a unit square

6.5. Computational analysis

This section presents the inventory level and average service distance computations based on the above cases and considering varying fractions of demand for the relaxed and strict services. Let,

$A =$ area of the rectangle (square) with uniform geographical distribution of customers

$\lambda =$ overall demand in the area over unit time

$L =$ replenishment lead time

$f =$ fraction of demand for the strict service

$1\text{-}f =$ fraction of demand for the relaxed service

$F_r =$ minimum service availability (fill-rate) level

$n =$ total number of facilities

$\lambda_i =$ demand served by facility i over unit time

$S_i =$ base stock level at facility i

$A_{ri} =$ size of service area covered by facility i for the relaxed service

$A_{si} =$ size of service area covered by facility i for the strict service

$T_{ri} =$ average distance from facility i to reach a demand point for the relaxed service in A_{ri}

$T_{si} =$ average distance from facility i to reach a demand point for the strict service in A_{si}

Note that $T_{ri} = T_{si}$ and $A_{ri} = A_{si}$ under the non-hierarchical setup.

$T_r =$ average service distance for the relaxed service

$T_s =$ average service distance for the strict service

The base stock level at facility i (S_i) under the (S-1, S) policy is determined according to the procedure in Section 5.3 (Chapter 5) considering the demand received by facility i over the replenishment lead time ($L\lambda_i$) and the minimum service availability level (F_r), where $\lambda_i = (1 - f)\lambda\left(\frac{A_{ri}}{A}\right) + f\lambda\left(\frac{A_{si}}{A}\right)$.

$$Total\ inventory\ (base\ stock)\ in\ system = \sum_{i=1}^{n} S_i \qquad (6.1)$$

The average service distances for the relaxed service (T_r) and for the strict service (T_s) are determined through the procedure in Appendix 5. The average service distance for the relaxed service (T_r) is computed as $\sum_{i=1}^{n} T_{ri} \left(\frac{A_{ri}}{A} \right)$, i.e. the weighted average of the average service distances in the service areas for the relaxed service. Similarly, the average service distance for the strict service (T_s) is computed as $\sum_{i=1}^{n} T_{si} \left(\frac{A_{si}}{A} \right)$. The overall average distance to reach a demand point is then determined considering the fractions of demand for both service types.

$$Average\ service\ distance = (1 - f)(T_r) + f(T_s) \tag{6.2}$$

The computational results are comparable to the results in the previous chapter. The inventory level computations under the (S-1, S) inventory policy are based on two different demand rates (1 unit and 4 units over the replenishment lead time) and considering the minimum service availability level of 98%. Figures 6.12 (a) and (b) present the inventory levels considering the (S-1, S) policy under the hierarchical and non-hierarchical setups in the four test cases considering equal demand fractions for the relaxed and strict services ($f=0.5$). It is confirmed that the hierarchical setup, offering a higher centralization level, does not necessarily result in a lower inventory level compared to the non-hierarchical setup under the same conditions. Another interesting observation is that, under the non-hierarchical setup with the demand rate of 1 unit over the replenishment lead time, the inventory level in case 4 is less than the inventory level in case 3 which is more centralized having fewer facilities (Figure 6.12 (a)). These behaviours are due to the nature of service/base stock level trade-offs in the (S-1, S) model for items with very low demand in which the inventory levels suddenly leap or drop with changes in demand.

(a) Demand = 1 unit over replenishment lead time

(b) Demand = 4 unit over replenishment lead time
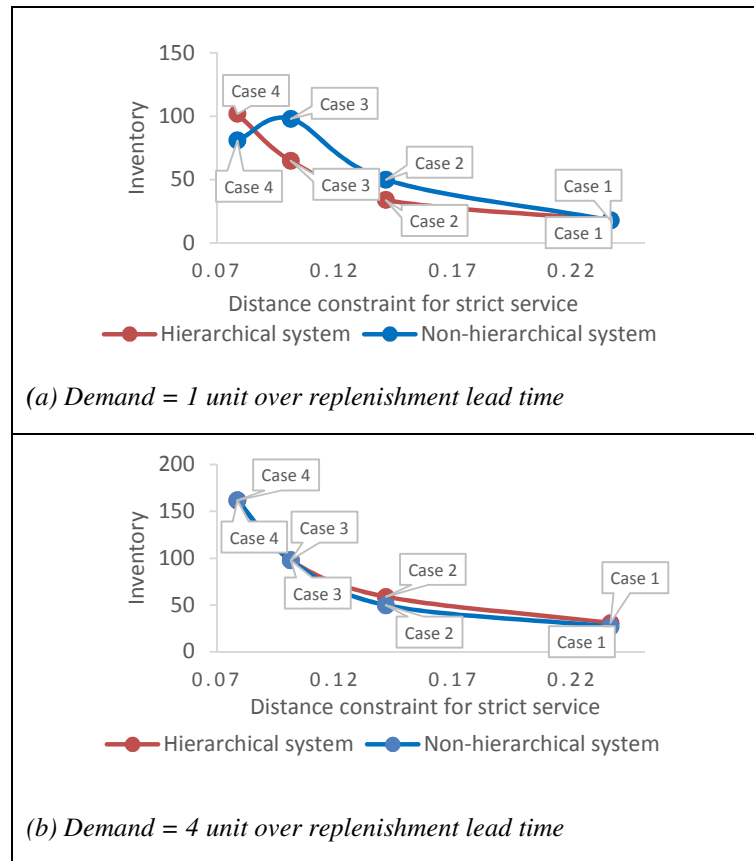
Figures 6.12 (a & b): Inventory (total base stock) under (S-1, S) policy - (f = 0.5)

Under the non-hierarchical setup, the demand rate at each facility equals to the total demand rate divided by the number of facilities. In case 3 there are 49 facilities, each receiving approximately 0.02041 requests over the replenishment lead time. Considering 98% minimum service availability level, each facility in case 3 requires the base stock level of 2 units, which makes the total inventory level in the system equal to 98 (=2×49). In case 4, there are 81 facilities, and hence each facility receives approximately 0.01235 requests over the replenishment lead time, i.e. 1/81. To satisfy the minimum service availability level of 98% each facility has to maintain the base stock level of 1, which makes the total inventory level in the system equal to 81 (less than that in case 3).

The average service distance under the hierarchical and non-hierarchical setups for all four cases is presented in Figure 6.13. The hierarchical setup results in a higher average service distance in all cases. The average service distance increases approximately linearly as the system becomes more centralized (from case 1 through to case 4), with the rate of change higher under the hierarchical setup than under the non-hierarchical setup.
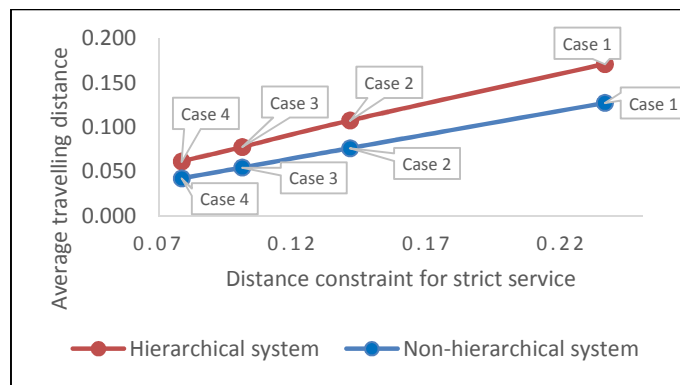
Figure 6.13: Average service distance - (f = 0.5)

Figures 6.14 and 6.15 show the impact of the demand fractions for the relaxed and strict services on the inventory (total base stock) level (considering the (S-1, S) inventory policy) and average service distance in case 1, while Figures 6.16 and 6.17 show the same for case 2. The finding confirm that an increase in the fraction of demand for the relaxed service does not necessarily decrease the inventory under the hierarchical setup. Also, operating in the non-hierarchical fashion, resulting in maximum decentralization, does not result in a higher inventory level in most instances compared to when the system operates under the hierarchical setup. While, with an increase in the demand fraction for the relaxed service, the average service distance increases under the hierarchical setup and remains lower and constant under the non-hierarchical setup.

(a) Demand = 1 unit over replenishment lead time

(b) Demand = 4 units over replenishment lead time
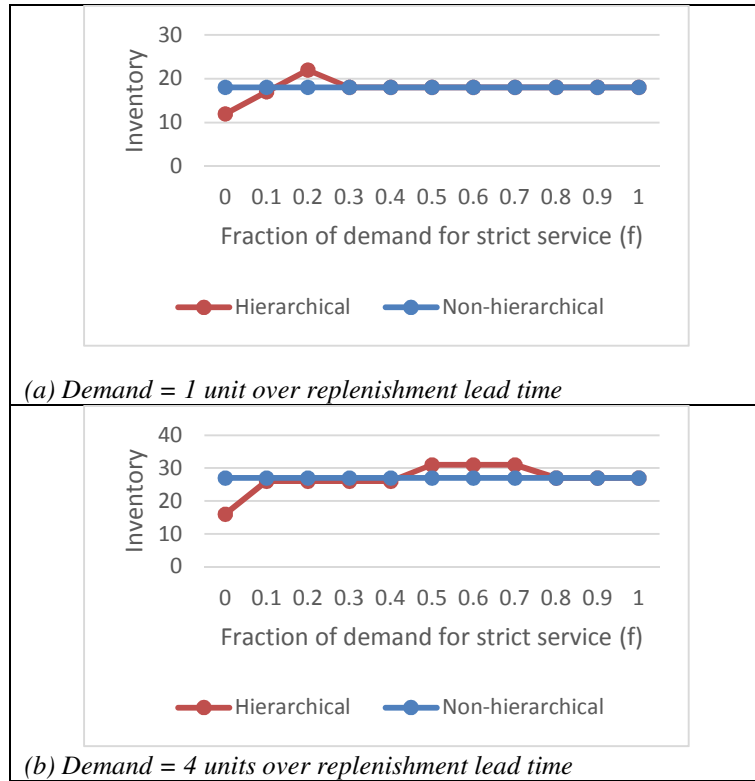
Figures 6.14 (a & b): Impact of demand fractions for strict and relaxed services on inventory (total base stock) under (S-1, S) policy – Case 1
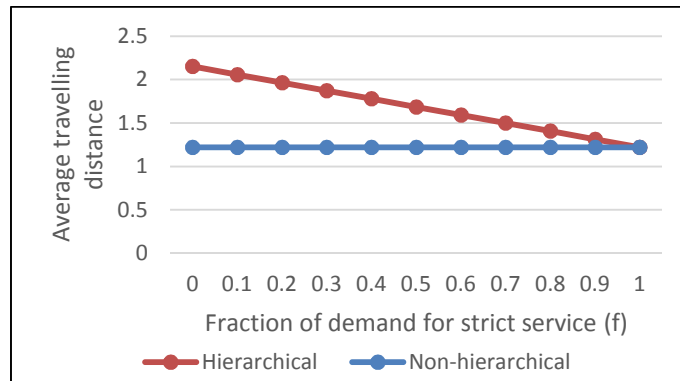
Figure 6.15: Impact of demand fractions for strict and relaxed services on average service distance – Case 1

165

(a) Demand = 1 unit over replenishment lead time

(b) Demand = 4 units over replenishment lead time
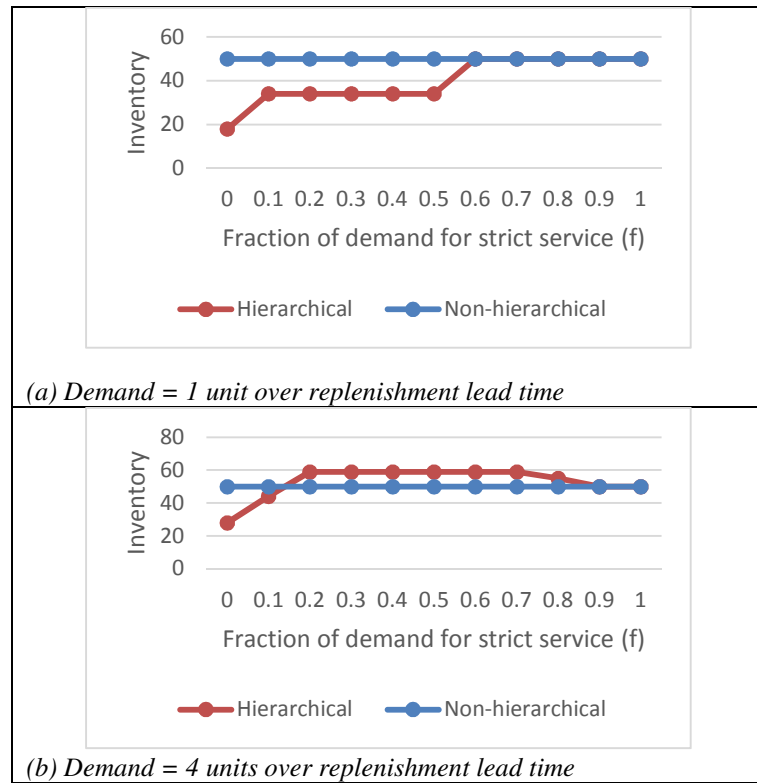
Figures 6.16 (a & b): Impact of demand fractions for strict and relaxed services on inventory (total base stock) under (S-1, S) policy – Case 2
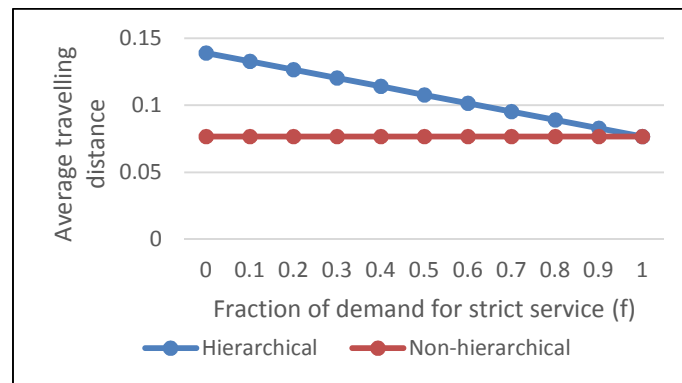
Figure 6.17: Impact of demand fractions for strict and relaxed services on average service distance – Case 2

Lastly, Figures 6.18 and 6.19 present the inventory level computations considering the EOQ inventory model with following hypothetical settings: demand (λ) = 1500 units per unit time, holding cost (C_h) = £10 per unit per unit time, ordering cost (C_o) = £100 per order. The average inventory levels are computed as $\sum_{i=1}^{n} \frac{1}{2} \sqrt{\frac{2 C_o \lambda_i}{C_h}}$ where n is the total number of facilities and λ_i is the demand at facility i. Figure 6.18 presents inventory levels with different fractions of demand for the relaxed and strict services

in case 1, and Figure 6.19 presents the inventory levels for the four test cases with the equal fractions of demand for the relaxed and strict services (*f=0.5*). The results confirm that the more centralized the system is, the lower is the inventory level under the EOQ inventory policy, i.e. hierarchical system performs better than the non-hierarchical system.
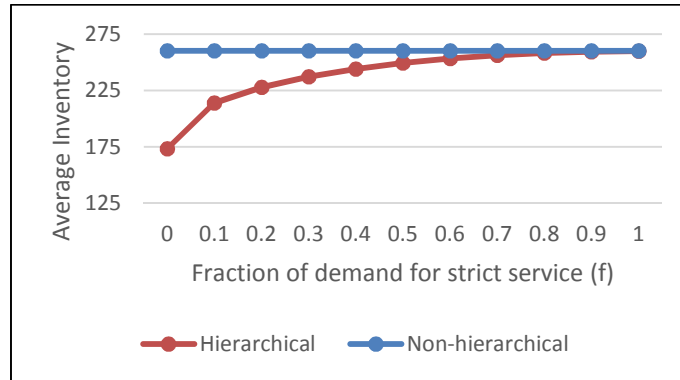
Figure 6.18: Impact of demand fractions for strict and relaxed services on inventory level under EOQ policy – Case 1
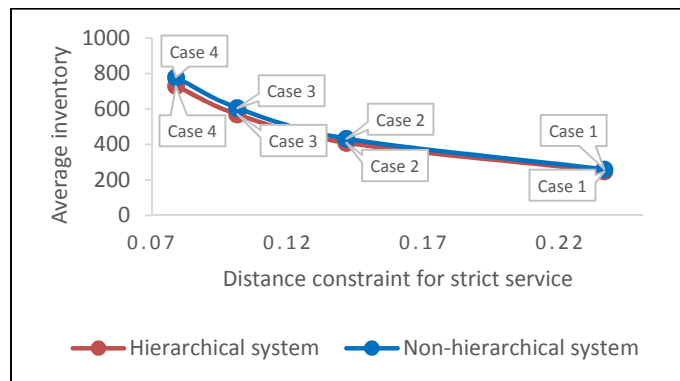
Figure 6.19: Average inventory level under EOQ policy – (f = 0.5)

Note that for the average inventory levels under the EOQ policy, the following insights persist and that the overall trends in the plots remain similar with the changes in demand λ, order cost C_o, and holding cost C_h:

- With the increase in the fraction of demand for the strict service, the inventory in the hierarchical setup increases, yet the marginal increase diminished. The inventory level under the non-hierarchical setup remains higher than the level

under the hierarchical setup except for the case where there is no demand for the relaxed service, in which case the inventory levels are the same under both setups.

- When the distance constraints reduce, the inventory levels increase and at a higher rate when the limiting distance is low. The inventory level and the rate of inventory increase under the non-hierarchical setup remain higher than the level and rate under the hierarchical setup.

As stated above, we compute the average inventory level as $\sum_{i=1}^{n} \frac{1}{2} \sqrt{\frac{2C_o \lambda_i}{C_h}}$, which is

equal to $\frac{1}{2} \sqrt{\frac{2C_o \lambda}{C_h A}} \sum_{i=1}^{n} \sqrt{(1-f)A_{ri} + fA_{si}}$, where $\frac{1}{2} \sqrt{\frac{2C_o \lambda}{C_h A}}$ is constant. The graph of

$\frac{1}{2} \sqrt{\frac{2C_o \lambda}{C_h A}} \sum_{i=1}^{n} \sqrt{(1-f)A_{ri} + fA_{si}}$ will retain the form of the graph of

$\sum_{i=1}^{n} \sqrt{(1-f)A_{ri} + fA_{si}}$ with respect to f or the service area sizes. The increase

(decrease) in the value of $\frac{1}{2} \sqrt{\frac{2C_o \lambda}{C_h A}}$, depending on the changes in inventory order and

holding costs and demand, raises (lowers) and stretches (compresses) the graph line along the vertical axis while retaining the characteristics stated above.

6.6. Simulation study

The study investigates the impact of inventory sharing on time-differentiated distribution under the hierarchical and non-hierarchical setups. Pooling the stocks reduces the risk from the uncertain demands, which results in lower costs and higher profits (Cherikh, 2000). However, sharing may also result in higher transportation costs as transshipments are made from nonlocal facilities. Hence the travelling factor should not be ignored in the evaluation of a sharing strategy. As mentioned in the

literature review, the studies related to inventory sharing normally consider a fixed transshipment cost, independent of the distance travelled. Also, when the services are time constrained, an obvious factor affecting the level of sharing is the range or distance that can be travelled in the constrained time window. Yang et al. (2013) note that though the transshipment time is not negligible, this aspect (the spatial consideration) is hardly considered in the existing service logistics literature. In this study, we only consider inventory sharing if an alternate supply can be made from within the service distance constraint, i.e. only in the areas where the service range of facilities overlap (Figure 6.20). This study provides some novel insights by considering two time (distance) constrained services, varying demand fractions of both service types, and the hierarchical and non-hierarchical setups.
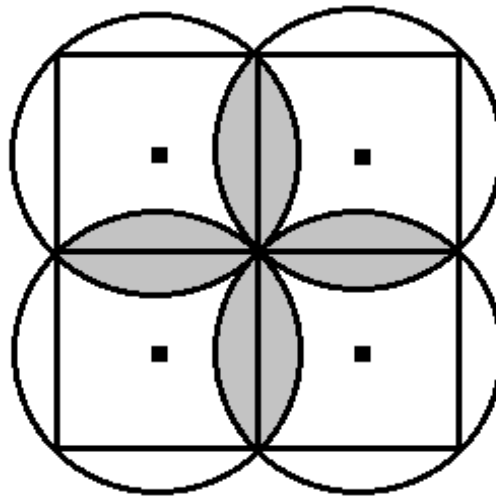
Figure 6.20: Overlapping ranges of facilities (sharing area on boundaries)

The simulation model is programmed to investigate a non-sharing and two sharing configurations described in the following section.

6.6.1. Configurations

Consider that the local facility for a request is the nearest facility offering the requested service type. That is, under the hierarchical setup, in case of a relaxed service request, the local facility is the nearest higher level facility, and, in the case of a strict service request, the local facility is the nearest facility, be it a higher level or a lower level facility. While, under the non-hierarchical setup, for either a relaxed or a strict service request, the local facility is the nearest facility (as all facilities provide the full range of service options). With these consideration we simulate the hierarchical and non-hierarchical setups under the following three configurations:

1) Non-sharing (under the hierarchical and the non-hierarchical setup)

A service request is met only by the local facility providing the required service type. In case of stock-out at the local facility at the time of request, the service request is back ordered.

2) Sharing with hierarchical restriction (under the hierarchical setup)

When a service request is registered, stocks are checked at the local facility providing the required service type. In case of stock-out at the local facility, service can be met from the closest facility that 1) **provides the required service type**, 2) has a positive stock level, and 3) is in the range of the demand location. That is, a relaxed service request can only be met by another higher level facility in the range if the local higher level facility has no stocks. If none of the facilities in the range has stocks at the time of a request, backorder is recorded at the local facility.

3) Full sharing (under the hierarchical and the non-hierarchical setups)

When a service request is registered, stocks are checked at the local facility providing the required service type. In case of stock-out at the local facility, service can be met from the closest facility of *any type* (higher level or lower level) in the distance range with stocks. If none of facilities in the distance range has stocks at the time of request, a backorder is recorded at the local facility.
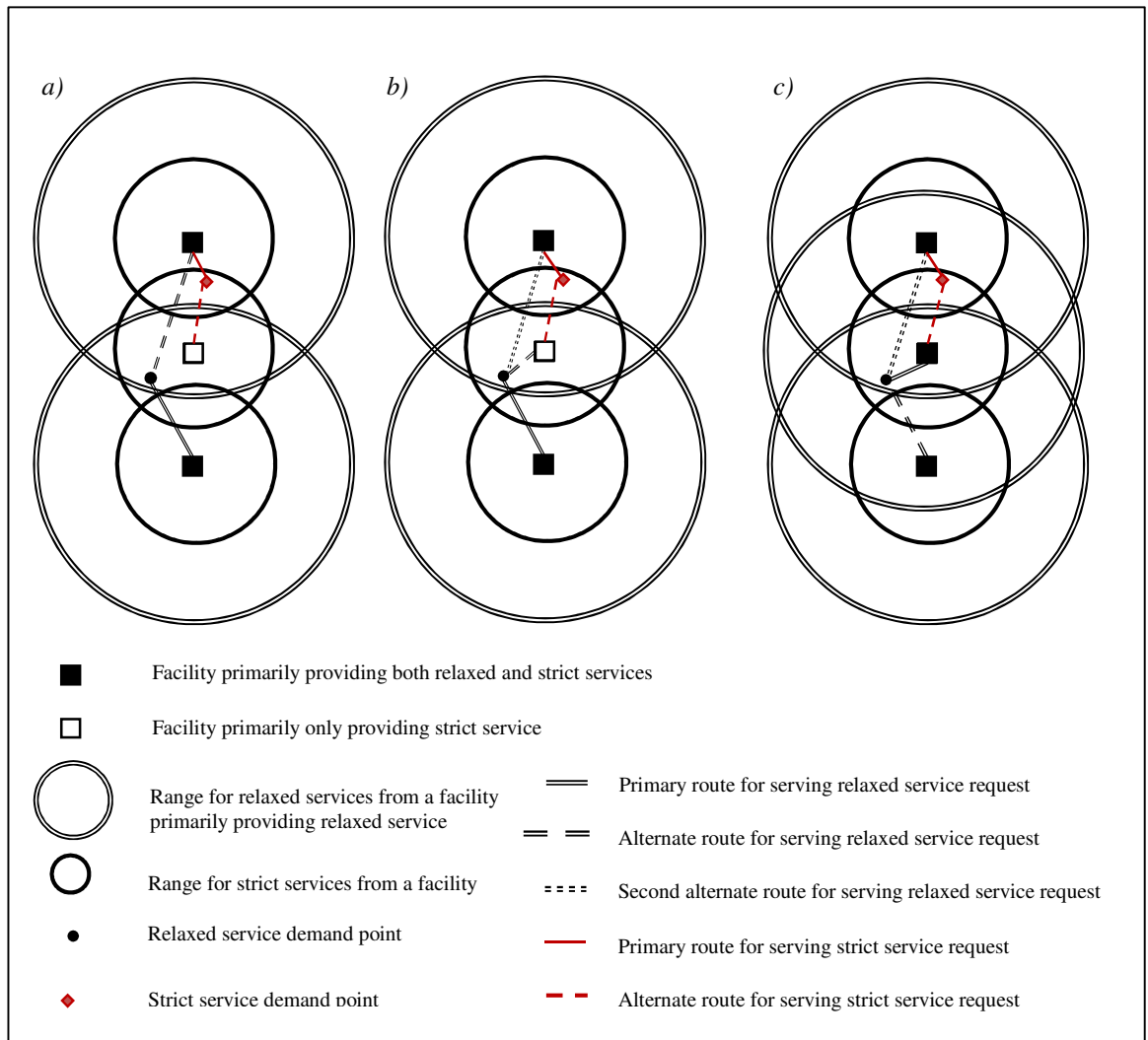
Figure 6.21 (a – c): Inventory sharing example. a) Sharing with hierarchical restriction (under the hierarchical setup), b) full-sharing under the hierarchical setup, c) full-sharing under the non-hierarchical setup

6.6.2. The model

The simulation model is based on the same general assumptions as made in the previous sections. We assume that services have to be provided to a uniformly distributed customer base in a bounded plane (unit square) under two service time commitments. The one-for-one (S-1, S) inventory policy with back orders and Euclidean distances are considered for meeting service demand. The following inputs are provided for the simulation:

1) Coordinates of the overall area in the Cartesian system: Unit square having corner point (0, 0), (0, 1), (1, 1) and (1, 0)
2) Facilities' coordinates and their types (higher level or lower level): from the cases in Section 6.4
3) Distance constraints for the relaxed and strict services: from the cases in Section 6.4
4) Lead time, demand rate and demand fractions for the relaxed and strict services
5) Starting inventory position: Base stock levels at the facilities (as computed in Section 6.5)

The randomness in the system is introduced by randomly generating inter-arrival times between two requests (according to an exponential distribution), demand type, i.e. relaxed or strict (using uniform distribution), and demand location (x and y coordinates in the overall area considering a uniform distribution). Service availability levels (fill-rates) at each facility and the overall average service distance are considered as the model's performance measures (or simulation output).

This simulation can be classified as a *termination simulation*. The end of a product's life cycle is the natural event that can specify the length of a period (simulation run) for which the service parts for the product are kept in inventory. After this point in time the system is *cleared out*.

172

6.6.2.1.   Simulation flow charts
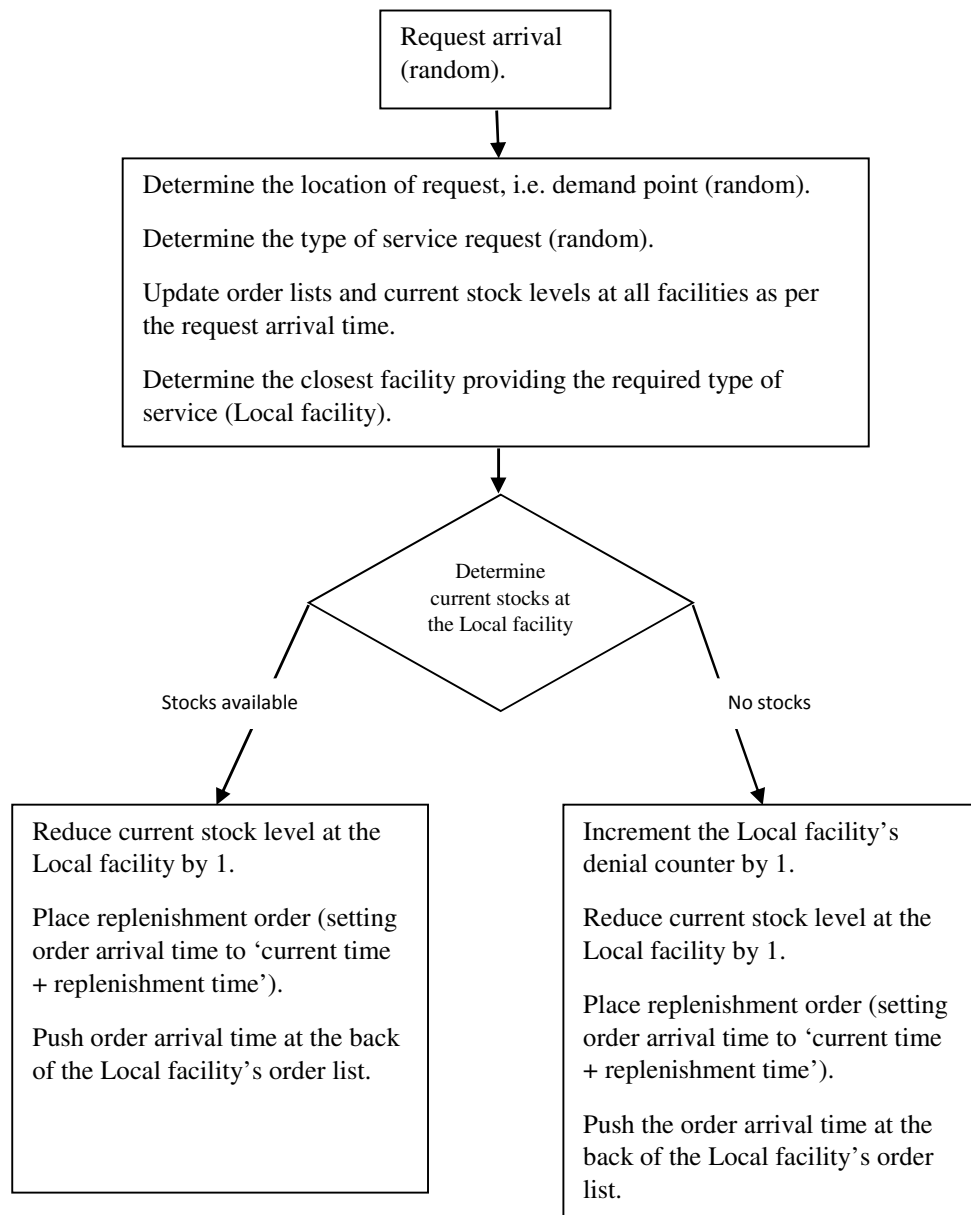
Figure 6.22: Flow chart of simulation without considering an inventory sharing mechanism (Configuration 1).

Request arrival (random).

Determine the location of request, i.e. demand point (random).

Determine the type of service request (random).

Update order lists and current stock levels at all facilities as per the request arrival time.

Determine candidate facilities - all the facilities, providing *the required service type,* in the range.

Determine the closest facility (Local facility) from the candidate facilities.

Determine current stocks at the Local facility

No stocks

Stocks available

Determine current stocks at other candidate facilities

Stocks available

No stocks

Reduce current stock at the Local facility by 1.

Place replenishment order.

Set order arrival time to 'current time + replenishment time'.

Push order arrival time at the back of Local facility's order list.

Reduce current stock level at the closest candidate facility with stocks by 1.

Place replenishment order (setting order arrival time to 'current time + replenishment time').

Push order arrival time at the back of the order list of the closest candidate facility with stocks.

Consider service distance as the distance between the closest candidate facility with stock and the demand point.

Increment the Local facility's denial counter by 1.

Reduce the current stock level at the Local facility by 1.

Place replenishment order (setting order arrival time to 'current time + replenishment time').

Push order arrival time at the back of Local facility's order list.

Consider service distance as the distance between the Local facility and the demand point.

Figure 6.23: Flow chart of simulation considering inventory sharing mechanism with hierarchical restriction (Configuration 2).

174
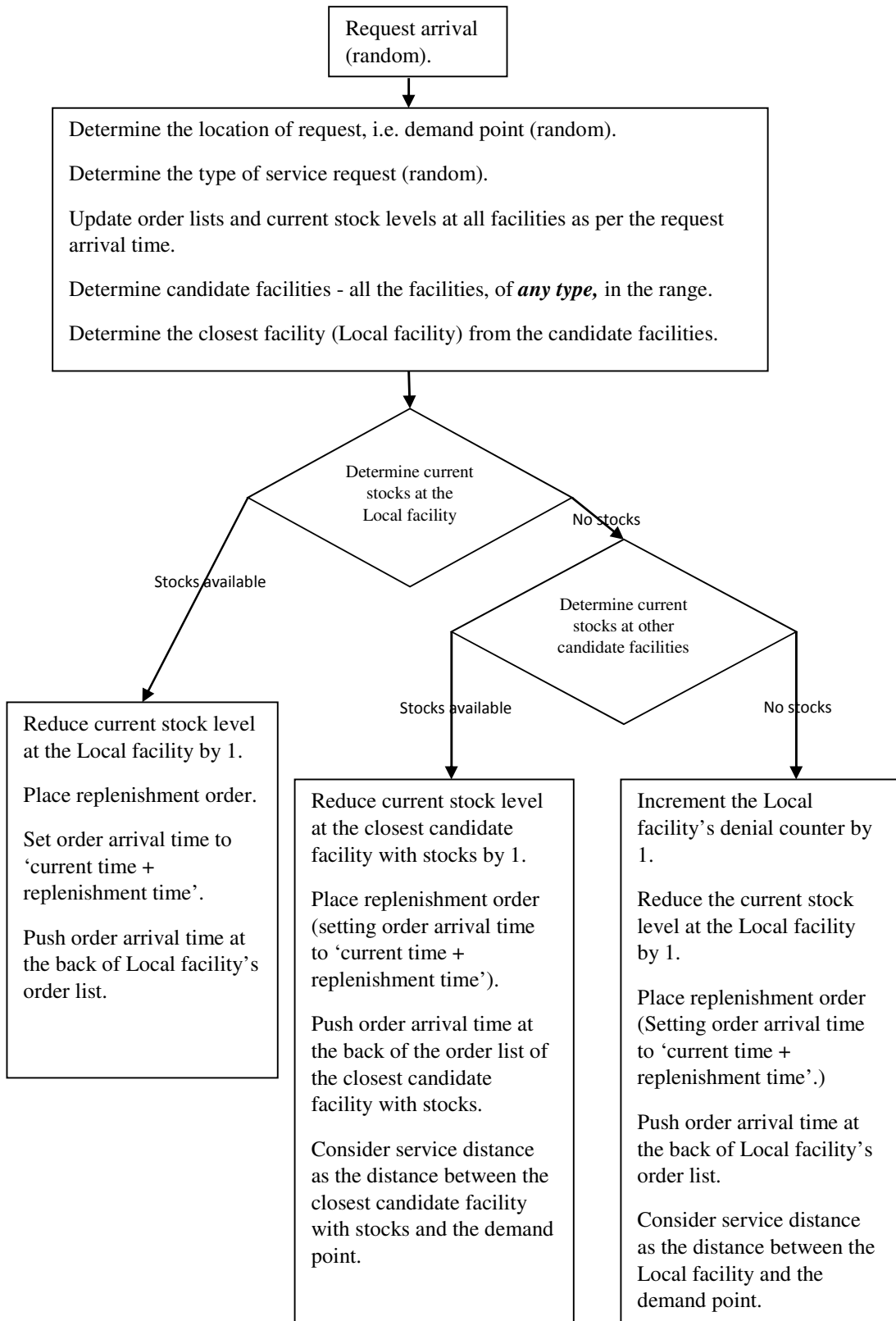
Figure 6.24: Flow chart of simulation considering full inventory sharing mechanism (Configuration 3).

6.6.3. Model programming and program verification

Implementing a simulation model in a general purpose programming language (GPL), special purpose simulation language, or simulation-software, each has its own advantages and disadvantages (Law, 2007). Advantages of using a GPL include greater program control and flexibility as there is no rigid modelling framework. However, using a GPL can require more programming time compared to a commercial simulation product which attempts to de-emphasize the programming aspect for a quicker model implementation (Watkins, 1993). As our model is rather compact (Figures 6.22-6.24), it can be implemented in a GPL within reasonable coding time.

The simulation program is coded in C++ and developed in Microsoft Visual C++ Express (2010) integrated development environment. Each facility is represented as an object, storing information about its x and y coordinates on the plane, base stock level, current stock level, number of requests received at the facility, number of requests not fulfilled by the facility from stock on hand, and orders in the pipeline and their arrival times (order list). This information is used to evaluate facility service availability performance measures. For the average service distance measure, the total distance covered to serve all service requests is divided by the total number of service requests.

To verify that the program components perform as anticipated, different aspects in the program, such as (S-1, S) inventory policy (pseudocode in Appendix 6), percentage of service requests for both types (relaxed and strict), and the percentage of demand received at each facility in a simulation run, are tested separately. Besides, information stored in the variables and the data members of interest are observed for a process walkthrough.

Common Random Numbers (CRN) are used as a Variance Reduction Technique (VRT) to compare the alternative system configurations. Despite a simple concept, CRN is a very useful and popular VRT (Law, 2007). The basic idea of the CRN technique is to compare the alternative configurations 'under similar experimental conditions' so that we can be more confident that any observed differences in a performance are due to the differences in the system configuration rather than due to the fluctuations of the 'experimental conditions'. To implement CRN properly, the random numbers across the different system configurations are synchronized or matched up. In simulation packages, the random numbers are not necessarily properly synchronized, which is critical for the success of CRN (Law, 2007). Using C++ for the model implementation gives us more control to ensure that the random numbers are synchronized for each configuration.

Global random number streams are generated for request arrival time, requested service type, and demand location. The configurations (non-sharing, sharing with hierarchical restriction, and full-sharing) are simulated in a sequence. Before starting the simulation runs for each configuration, the random number streams are reinitialized so that same random numbers are generated for each configuration. As a result, under each configuration, the sequence of service request arrival times and the associated service types and demand locations are same. Of course the difference is how the system under different configurations deals with the events.

6.6.4. Analysis

The following settings are used for the experiments

Overall area: Unit square (1 x 1)

Minimum service availability (fill-rate):   95%

Demand in the overall area:	4 units per day (close to actual demand of a part at Company B)
Replenishment time:	1 day
Run length:	6 years (average life cycle of products reported by the case companies)
Number of runs (n):	100 (providing stable performance measure outputs). Increasing the number of runs does not change the outputs significantly. For example, in the following experiments, increasing the runs from 100 to 150 provides same overall fill-rate at three decimal places and same average distance at four decimal places.

The experiments are based on cases 1 and 2 in Section 6.4 (also presented in Figures 6.25 (a) and (b) and Figures 6.31 (a) and (b) respectively). The output is presented in Appendix 7, which also shows that the simulation results with the non-sharing configuration (entries under 'Fr_{N-S}' and 'AD_{N-S}') closely match the computational results (entries under 'Fr' and 'AD') and hence verify the simulation output.

Case 1:

(a)

(b)
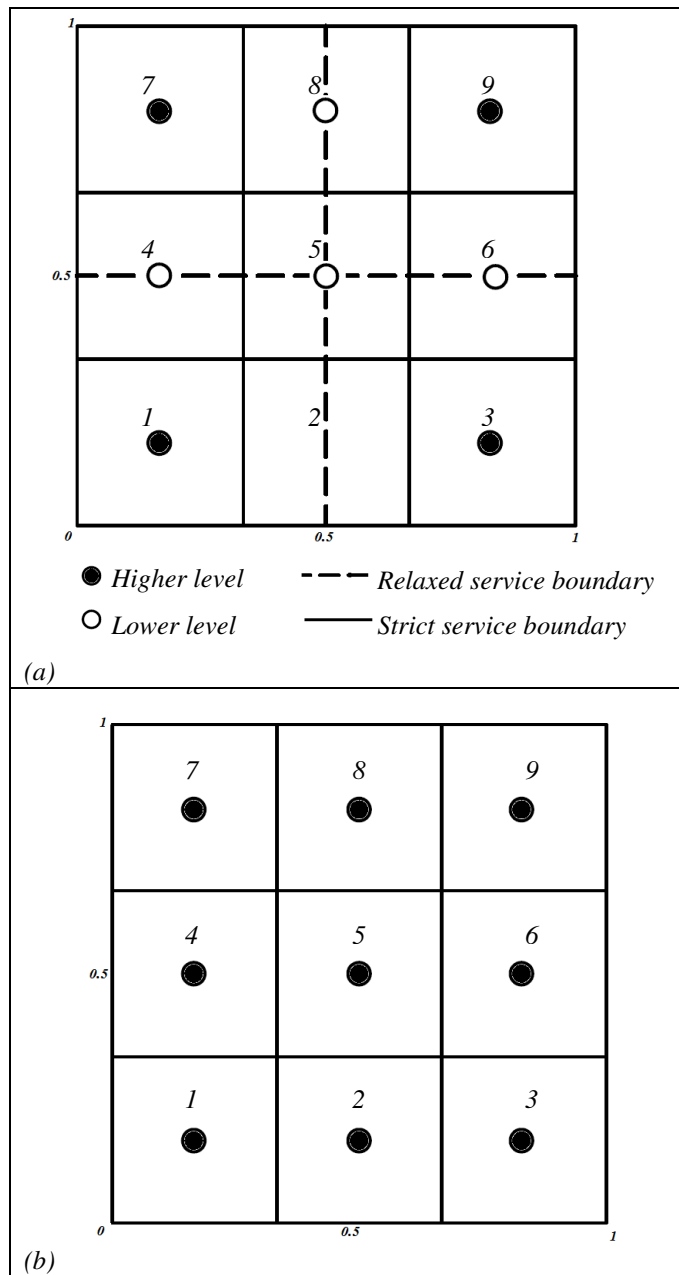
Figures 6.25 (a & b): Case 1 - a) Hierarchical setup. b) Non-hierarchical setup

Distance constraint for relaxed service =   0.472 unit length

Distance constraint for strict service =   0.236 unit length

Hierarchical setup:

Figures 6.26 (a – h) present the service availability levels and average service distances under the hierarchical setup with all three system configurations (Section 6.6.1). These figures are based on the data in Tables A7-1 to A7-5 in Appendix 7.

Demand fraction for relaxed services (1-f)	Service availability level (fill-rates)	Average service distance
0.2	 *(a)*	 *(b)*
0.4	 *(c)*	 *(d)*
0.6	 *(e)*	 *(f)*
0.8	 *(g)*	 *(h)*

■ Non-sharing　　■ Hierarchical-sharing　　■ Full-sharing
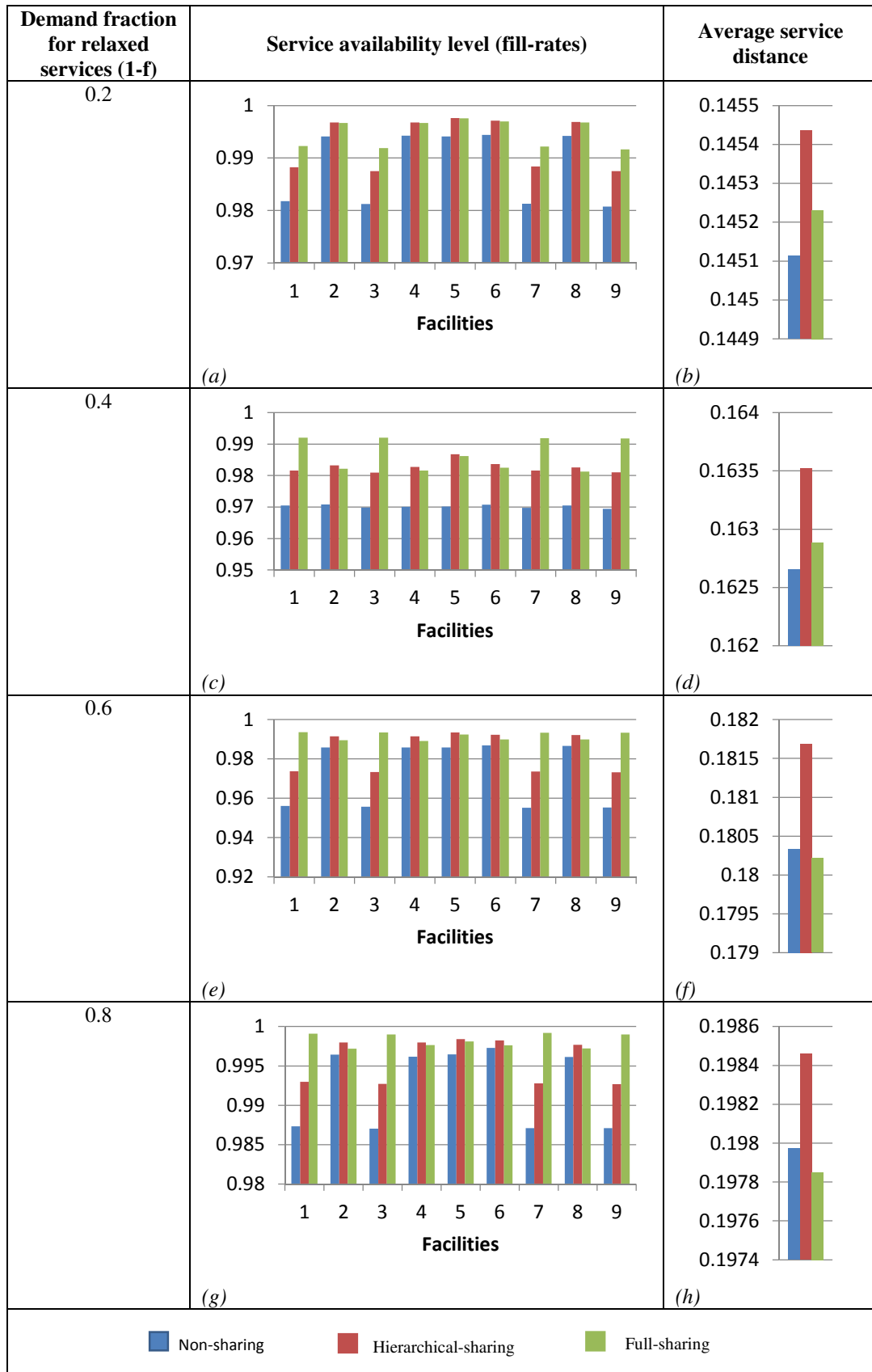
Figures 6.26 (a – h): Fill-rates and average service distance under the Hierarchical setup – Case 1

The outcome shows that both sharing configurations result in higher service availability levels compared to the non-sharing configuration. For the higher level facilities (facilities 1, 3, 7 and 9 (Figure 6.25)), the full-sharing scenario performs better in terms of service availability than the inventory sharing with the hierarchical restriction. While at the lower level facilities (2, 4, 5, 6 and 8), sharing under the hierarchical restriction performs slightly better than the full-sharing in terms of service availability. With the sharing mechanism under the hierarchical restriction, lower level facilities do not share their inventory with higher level facilities for the relaxed service requests. Hence the stock depletion at lower level facilities can be comparatively less, while alternate service opportunities for demand at lower level facilities are the same as under the full-sharing configuration.

With the full-sharing configuration under the hierarchical setup, in case of a stock-out at a local higher level facility when a relaxed service request is received, the demand can not only be met by another higher level facility in the range, but also by a lower level facility in the range. This means that the number of candidate facilities to meet a relaxed service request potentially increases. Therefore, it can be seen in Figures 6.26 (a), (c), (e) and (g) that when the fraction of demand for the relaxed service increases, the service availability levels at the higher level facilities increase under the full-sharing configuration with a higher rate compared to the sharing with the hierarchical restriction. Note that the service availability levels at facilities 1, 2, 3 and 4 are similar to the service availability levels at facilities 9, 8, 7 and 6 respectively as these facilities are equivalent in terms of service areas and inventory sharing opportunities.

In terms of the average service distance performance, sharing with the hierarchical restriction results in the highest average travelling distance to serve a demand in all

cases. Note that with sharing under the hierarchical restriction, in case of a stock-out, the transshipment to meet the demand is always from a facility that is at a longer distance (Figure 6.21 (a)). The results under the hierarchical setup also show that the full-sharing configuration not only performs overall better in terms of service availability, it can even results in a lower average service distance compared to under the non-sharing configuration (Figures 6.26 (f) and (h)).

Note that with the full-sharing configuration under the hierarchical setup (Figure 6.27), in case of a stock-out when a strict service request is received, compared to the local facility, an alternate facility to meet the demand is at a longer distance. However, in case of a stock-out when a relaxed service request is received, the first alternative to meet the demand is a lower level facility, which is likely to be at a shorter distance from the demand point than the higher level facility originally assigned to meet the request. With a higher demand fraction for the relaxed service, the shorter distances to meet the relaxed service requests from alternate facilities can offset the longer distances for alternative service for the strict service requests. Hence with a higher fraction of demand for the relaxed service under the full-sharing configuration, the alternate services for the relaxed service requests can not only positively impact the service availability levels, but also reduce the average service distance.

b)

■	Facility primarily providing both relaxed and strict services
□	Facility primarily only providing strict service
◎	Range for relaxed services from a facility primarily providing relaxed service
○	Range for strict services from a facility
●	Relaxed service demand
◆	Strict service demand point
═══	Primary route for serving relaxed service request
═ ═	Alternate route for serving relaxed service request
═══	Second alternate route for serving relaxed service request
——	Primary route for serving strict service request
▬ ▪	Alternate route for serving strict service request

Figure 6.27: Full-sharing in Hierarchical setup

Non-hierarchical setup:

Figures 6.28 (a – h) (based on the data in Tables A7-6 to A7-9 and A7-14 in Appendix 7) show the service availability levels and the average service distances with the non-sharing and the full-sharing configurations under the non-hierarchical setup.

Demand fraction for relaxed services (1-f)	Service availability level (fill-rates)	Average service distance
0.2	 *(a)*	 *(b)*
0.4	 *(c)*	 *(d)*
0.6	 *(e)*	 *(f)*
0.8	 *(g)*	 *(h)*

Non-sharing   Sharing
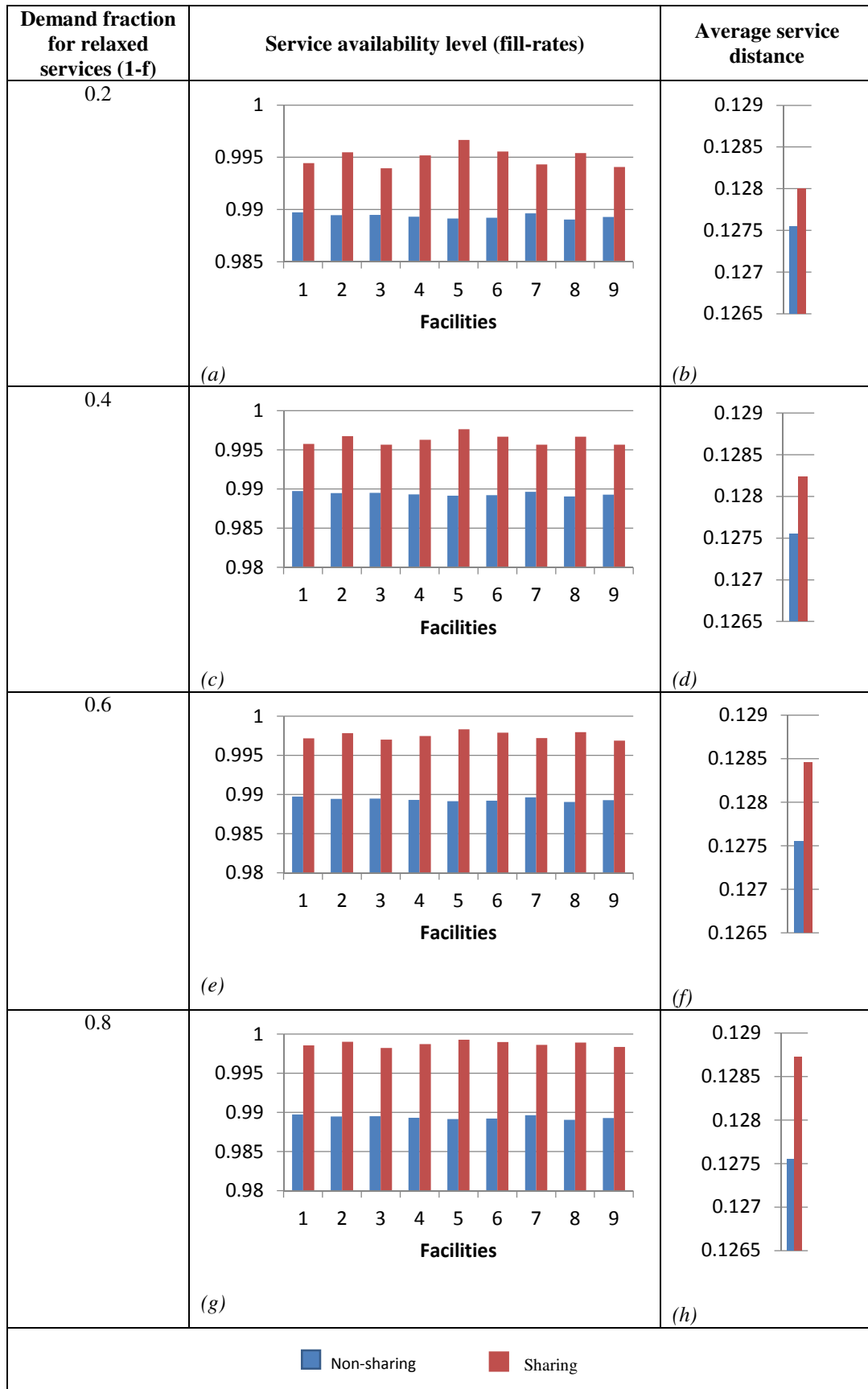
Figures 6.28 (a – h): Fill-rates and average service distance under the non-hierarchical setup – Case 1

Under the non-hierarchical setup with the sharing configuration, an increase in the demand fraction for the relaxed service increases both the service availability levels and the average service distance (Figures 6.28, 6.29 and 6.30). Inventory sharing in this case can allow to reduce the total stock in the system from 27 (3 units at each facility (Tables A7-6 to A7-9 in Appendix 7) to 18 (2 units at each facility) while still meeting the minimum availability level of 95%; a 33% reduction in the total inventory.

When the service distance constraint is slightly relaxed (from 0.472 to 0.5 and from 0.236 to 0.25 for the relaxed and strict services respectively), the service availability levels and the average service distance both increase under the sharing configuration. This improvement in the service availability levels and the increase in the average service distance fades as the fraction of demand for the relaxed services increases (Figures 6.29 and 6.30). It can also be observed that, the highest increase in the service availability level due to inventory sharing is at facility 5, which has the highest number of neighbouring facilities (Figures 6.28 and 6.29).
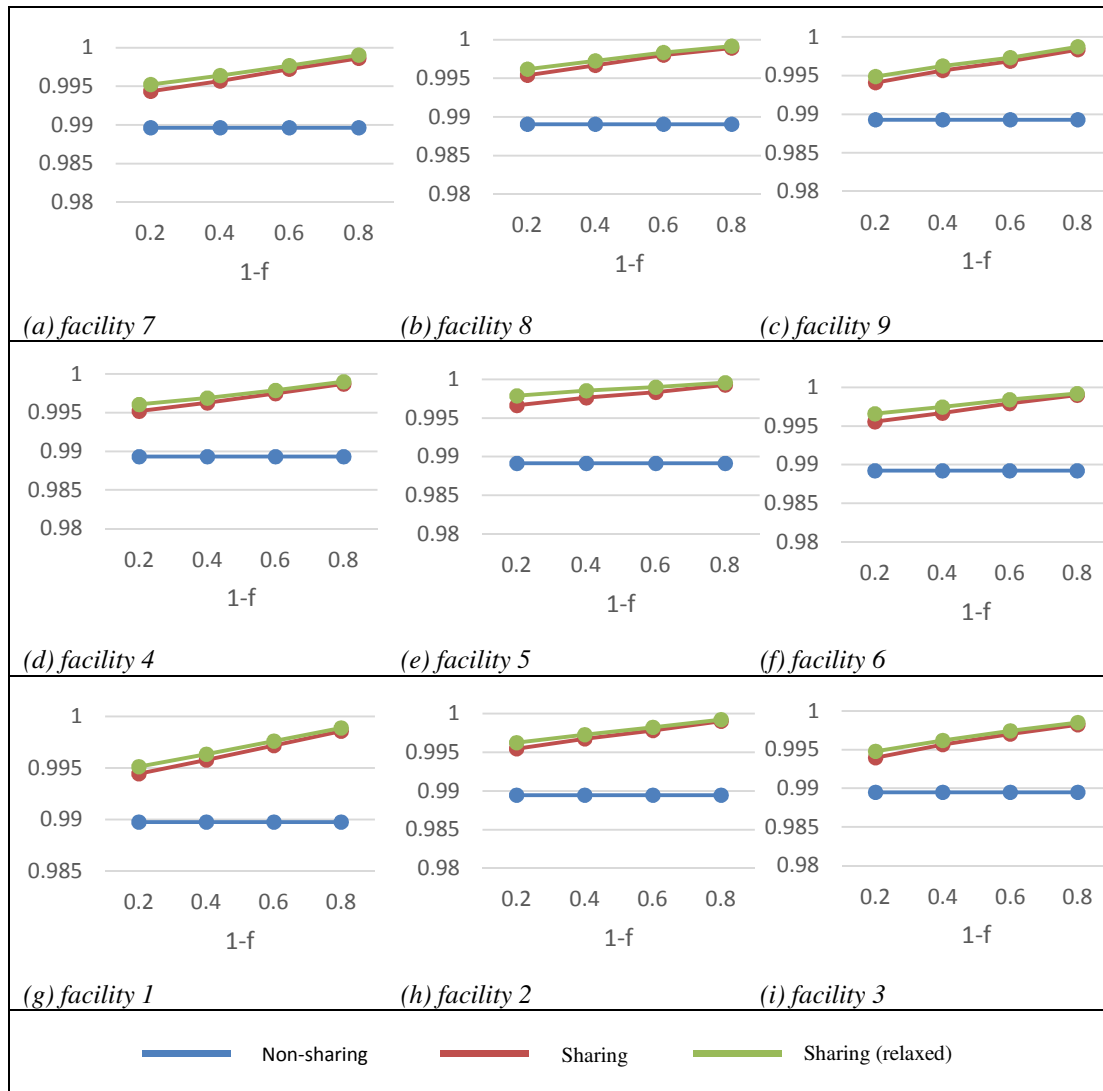
Figures 6.29 (a – i): Fill-rates under the non-hierarchical setup – Case 1
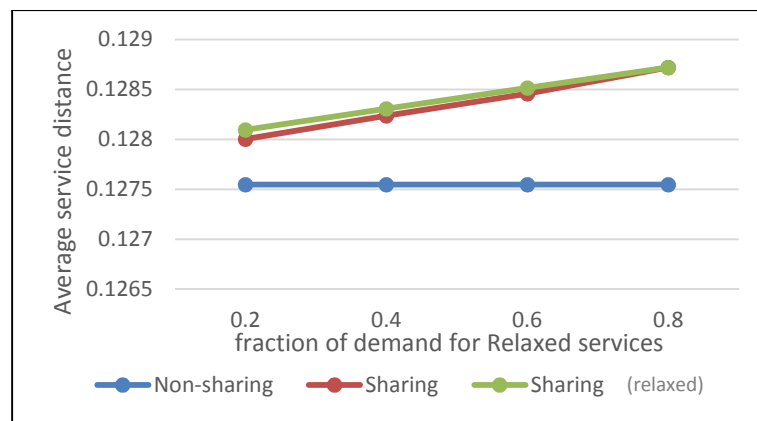
Figure 6.30: Average service distance under the non-hierarchical setup – Case 1

Contrary to the non-sharing case, a system under the non-hierarchical setup with an inventory sharing mechanism reacts to the changes in the fractions of demand for different time-based services. Inventory sharing under the non-hierarchical setup increases the service availability levels when the fraction of demand for the relaxed service increases. This may allow stock reduction and hence the changes in the demand fractions for different time-based services can result in different service costs.

Case 2:

(a)

(b)
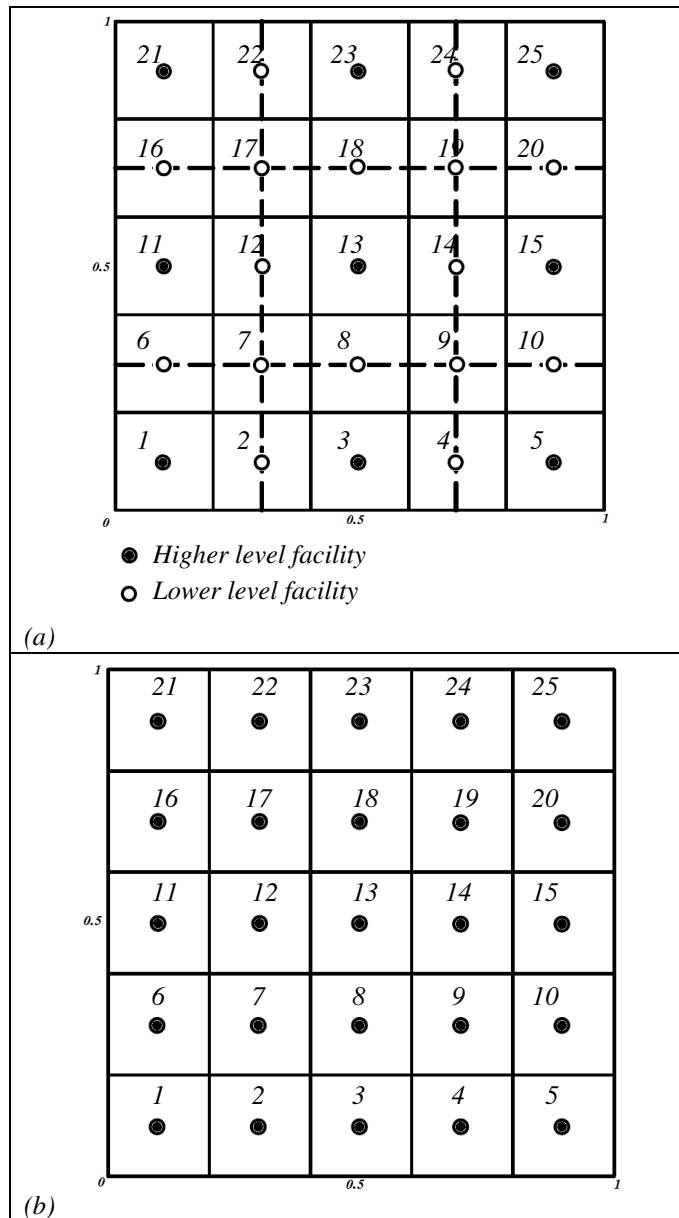
Figures 6.31 (a & b): Case 2 - a) Hierarchical setup. b) Non-hierarchical setup

Distance constraint for relaxed service =   0.283 unit length

Distance constraint for strict service =   0.142 unit length

Figures 6.32 and Figure 6.33 present the system wide service availability level and average service distance measures respectively for sharing and non-sharing configurations under the hierarchical setup, while Figures 6.34 and Figure 6.35 present the measures under the non-hierarchical setup. These figures correspond to the data in Tables A7-16 to A7-25 in Appendix 7.

Hierarchical setup:

As per the results under the hierarchical setup, the full-sharing configuration performs the best in terms of service availability level while the non-sharing configuration performs the worst (Figures 6.32). When the fraction of demand for the relaxed service increases from 0.2 to 0.4 and 0.6, more demand is concentrated at higher level facilities, however the stock levels at the higher level facilities remain same (Tables A7-16 – A7-18 in Appendix 7) as the minimum service availability level is still achievable with same stocks. This drops the overall service availability levels under the non-sharing and the hierarchical sharing configurations. The rate at which the system wide service availability level drops under the non-sharing configuration is higher than that under the hierarchical-sharing which has a limited opportunity for alternative services in stock-out cases. Under the full-sharing configuration however, the service availability slightly increases in this range of the demand fraction for the relaxed service (0.2 to 0.6). An increase in the demand fraction for the relaxed service provides more opportunity for alternate services under the full-sharing configuration. The sudden change in the service availability (Figure 6.32) when the demand fraction for the relaxed service increases from 0.6 to 0.8 is due to the change in the required

levels of stocks at the facilities to meet the minimum service availability level (Table A7-19 in Appendix 7). The required stock level when the demand fraction for the relaxed service (*1-f*) is equal to 0.2, 0.4 and 0.6 are approximately same (Tables A7-16 to A7-18 in Appendix 7)). When (*1-f*) is 0.8, the base stock levels increase at the higher level facilities (Table A7-19), which also considerably increases the service availability levels at these facilities without a sharing mechanism.
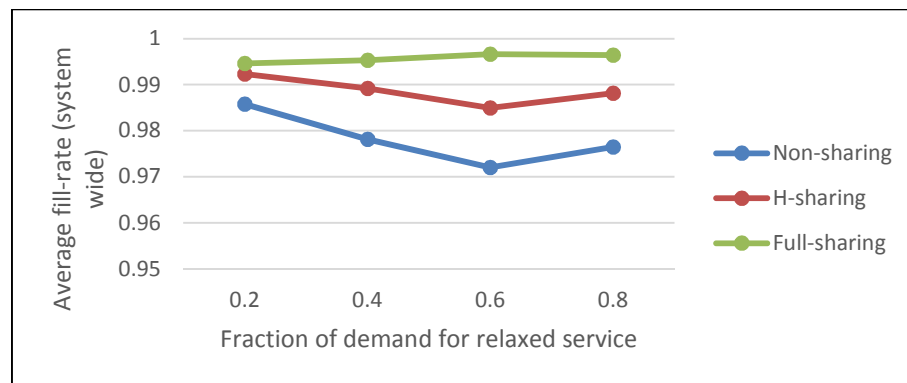
Figures 6.32: System wide fill-rate under hierarchical setup with sharing and non-sharing configurations – Case 2

With regards to the average service distance performance under the hierarchical setup, again the full-sharing configuration performs better than the other two configurations in most of the simulated instances (Figure 6.33). The non-sharing configuration results in slightly lower average service distance than the full-sharing configuration only when the demand fraction for the relaxed service is equal to *0.2*. While, the hierarchical-sharing configuration results in the highest average service distance under all tested instances of the demand fractions. However the difference between the average service distances under sharing and non-sharing configurations is small.
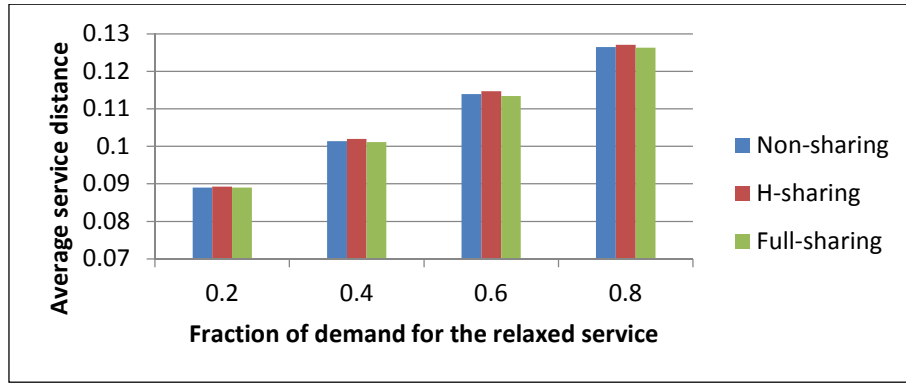
Figure 6.33: Average service distance under hierarchical setup with sharing and non-sharing configurations – Case 2

Non-hierarchical setup:

As the base stock levels remain the same under the non-hierarchical setup, it is easier to notice the impact of inventory sharing with different demand fractions for the relaxed and strict services. As in case 1, the results show that both the service availability and the average service distance increase with an increase in the demand fraction for the relaxed service under the non-hierarchical setup with the inventory sharing configuration (Figures 6.34 and 6.35). While without inventory sharing under the non-hierarchical setup, the service availability and average service distance remain constant over the changes in the demand fractions for both service types.
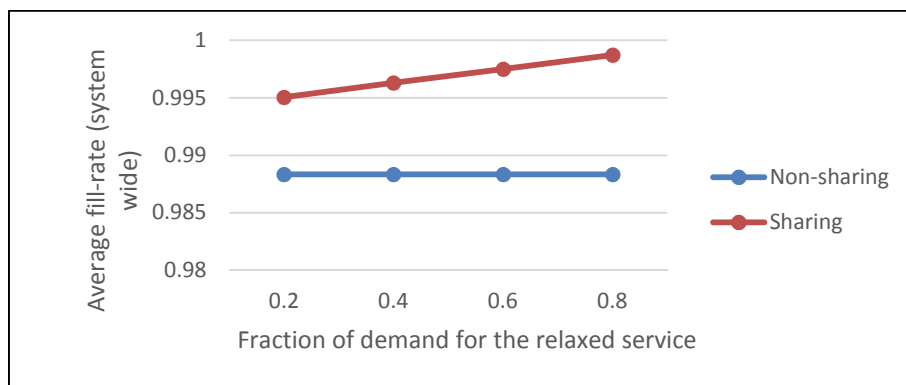
Figures 6.34: System wide fill-rate under non-hierarchical setup with sharing and non-sharing configurations – Case 2
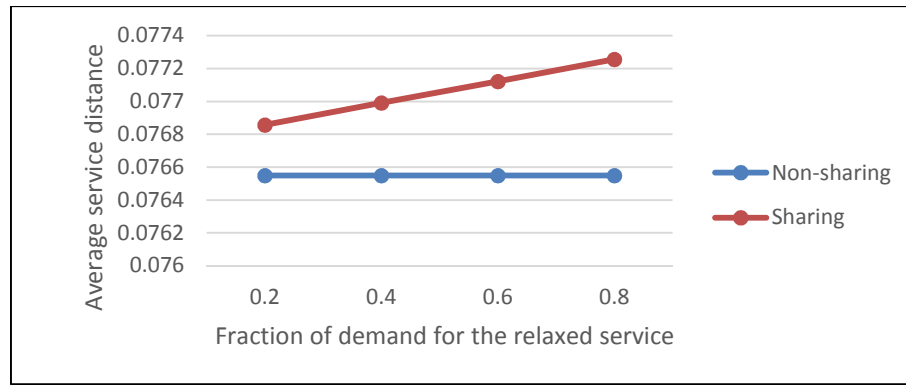
Figure 6.35: Average service distance under non-hierarchical setup with sharing and non-sharing configurations – Case 2

6.7. Conclusion

The computational results presented in this chapter verify the insights generated by the modelling study in Chapter 5. The simulation study gives further insights by investigating the impact of the fractions of demand for different time-based service types on the service availability and average service distance performance in presence of inventory sharing mechanisms under service distance constraints. Three different configurations were simulated under the hierarchical setup. These are non-sharing (with no alternate service in stock-out cases), hierarchical sharing (with sharing in a stock-out case, but only from a facility offering the required service type in the range), and full-sharing (with sharing from any facility in the range in a stock-out case). Under the non-hierarchical setup, non-sharing and full-sharing configurations were simulated.

Characteristically, inventory sharing among facilities should positively impact on the availability level while increasing the transportation cost as sharing results in shipments from nonlocal facilities. This is confirmed by the simulation of the sharing and non-sharing configurations under the non-hierarchical setup, where a higher fraction of demand for the service in the longer time window increases the inventory

sharing opportunity. This, in turn, increases the service availability levels (which may allow stocks reductions) and the average distance to serve a demand. However, the case of full-sharing under the hierarchical setup is interesting as it as such does not reveal a trade-off between service availability and average service distance performance. Under the hierarchical setup, sharing without hierarchical restriction can perform better in terms of both the service availability and the average service distance performance measures, especially when the demand fraction for services in the longer time window is high. That is, the sharing can also reduce the average service distance besides increasing the service availability levels. Compared to the full-sharing configuration under the hierarchical setup, sharing with a hierarchical restriction, where requests for the service in the longer time window are only met by higher level facilities in ranges, does not exhibit any special benefit in terms of both performance measures.

Though the simulation study provides interesting insights into the impact of inventory sharing on the service level and average service distance under various settings, it does not as such focus on the adjustment of base stock levels when the service level increases. Extending the study to consider the base stock level adjustment can be more insightful with regards to the impact of inventory sharing on service costs.

Chapter 7: Impact of inventory and transportation costs on optimum zoning for time-differentiated distribution: A unidimensional analysis

7.1.Introduction

So far we have investigated the impact of service time constraints and demand fractions for different time-based service types on inventory and transportation costs considering that facilities are located efficiently, in terms of average distance, to cover demand. In this chapter we take a slightly different approach and investigate the impact of the demand fractions for different time-based service types and inventory and transportation costs on the optimum facility locations and their service areas to cover demand. Note that inventory costs and the proportions of demand for different time-based service types are not considered by the case companies for locating facilities.

Locating facilities to reach customer sites within certain time limits has traditionally been tackled as a covering problem where the objective can be to locate the minimum number of facilities to cover the demand, or to cover the maximum demand with a given number of facilities. The recent reviews on covering problems by Li et al. (2011) and Farahani et al. (2012) both highlight the need for more work on the continuous customer location domain and on the covering scenarios where there are more than one service time. The existing models for covering problems have also generally overlooked the inventory management factors. There is a lack of focus in the literature on providing a general understanding of the impacts of different service costs and

demand profiles on the optimum location setups under time/distance constraints. As discussed in the Literature review (Chapter 2), setting the level of (de)centralization in a distribution network is a well-argued strategic decision and has various trade-offs. One of these trade-offs is between inventory and transportation costs. The more centralized the system is and the bigger the service areas are, the higher will be the transportation costs and the lower the inventory cost. The work in this chapter essentially investigates this trade-off in determining facility locations and service zones in the presence of multiple time-constrained services as in SPL and emergency services systems.

Through a Mixed Integer Nonlinear Programming (MINLP) model, this chapter looks into the problem of setting up an optimum hierarchical system to meet time-differentiated services with the minimum number of facilities. The model considers transportation cost, inventory cost (considering the EOQ model), and the fraction of demand for two time-based service types. We are not aware of any study in the literature that has studied these factors in conjunction. The model assumes continuous customer and possible facility locations, and hence can be classified as a continuous location model (Plastria, 2002). However, facility location problems considering continuous customer location are complex, even when considering single clear objective. Due to the problem being multifaceted, considering several factors simultaneously, the problem is explored on a line segment. Considering a two dimensional plane with the multiple and conflicting objectives being considered makes the problem mathematically very complex. We could not find any study in the huge wealth of location studies that looks into both minimizing the number of the required facilities to cover a continuous location and also minimizing the average distance to reach customers. To add to this, the investigation seeks to optimize

locations considering not only the transportation cost, but also the inventory cost, and set up a hierarchical system. On a practical note, the problem can be related to providing time-constrained services along one road with the demand spread uniformly.

The rest of the chapter is structured as follows. The problem description and assumptions are stated in Section 7.2. The cost formulations, analysis of the behaviour of transportation and inventory costs with respect to a service area size, and the optimum location patterns considering a single distance constraint are presented in Section 7.3. Sections 7.4 and 7.5 respectively present the MINLP model and the analysis based on numerical examples considering two service distance constraints. Finally the conclusion and a summary of the findings is provided in Section 7.6.

7.2. Problem description and assumptions

Consider a service system in which the service calls have to be responded within certain time limit(s) by reaching requesting locations with the required part(s) or material. In order to meet a service time commitment for the entire customer base, every client location should be within the maximum distance, which can be covered within the committed service time, from at least one service facility. We refer to the service area of a facility, i.e. the area in which the facility is responsible to serve exclusively, as its '*service zone*' or just '*zone*'. The terms '*setup*' and '*system*' are used interchangeably to refer to the organization of service facilities and their service zones. The following assumptions and simplifications are made to study the problem:

1) The system comprises the minimum number of facilities that can provide full demand coverage considering a service time constraint. The optimum setup is then

determined considering the minimization of the sum of inventory and transportation costs.

2) The inventory cost is determined based on the EOQ model.

3) It is assumed that customers are continuously distributed over a line segment and that there is an equal probability of demand being generated from any location on the line segment. The travelling distances are assumed to be Euclidean.

4) A travelling distance is considered to be proportional to the travel time. Hence, a service time constraint can be translated into a service distance constraint.

5) The transportation cost is determined considering the average distance between a facility point and a customer location.

7.3. Cost formulation and analysis of the problem on single level

In this section, considering one service type, we formulate the inventory, transportation, and total cost functions. The multiplication factors of these cost functions are then analysed to establish the behaviour of the service costs in relation to the size of service zones. A non-linear programming (NP) model is also presented with the goal to understand the optimum facility location patterns with a service distance constraint. The following notations are defined for the modelling:

$L =$ length of a line segment on which the locations of facility points and their service zones have to be decided considering a uniform spread of demand. For normalization, we consider $L = 1$ in the analysis. By the definition, as the line segment $\overline{01}$ is the set of every point between 0 and 1, there are unlimited possibilities for locating of a facility point.

$d_{max} =$ service distance constraint.

$n =$ number of facilities to be located.

$Z_i =$ length of facility i's service zone on the line segment, such that $i = 1 \ldots n$.

$\lambda =$ demand for service parts/material per unit time.

$C_t =$ transportation cost per unit length.

$C_h =$ inventory holding cost per unit per unit time.

$C_o =$ setup cost per inventory replenishment order.

$Q =$ inventory order quantity.

The total operating cost is considered as the sum of inventory and transportation costs, where the total inventory cost includes inventory holding and order costs.

Inventory cost: With the EOQ inventory policy, we know that considering both inventory holding and ordering costs,

$Total\ inventory\ cost\ in\ a\ centralized\ system\ per\ unit\ time =$

$$\sqrt{2C_h C_o \lambda} \qquad\qquad\qquad (7.1)$$

In context of our problem, the fraction of demand served by a facility i on a line segment $\overline{0L}$ is equal Z_i/L, where Z_i is the length of the zone of facility i. Considering the fraction of demand served by a facility i as Z_i/L and the total number of facilities in the system (n).

$Total\ inventory\ cost\ in\ system\ per\ unit\ time = \sum_{i=1}^{n}\sqrt{2C_h C_o \frac{Z_i}{L}\lambda} =$

$$\sqrt{2C_h C_o \lambda} \times \sum_{i=1}^{n}\sqrt{\frac{Z_i}{L}} \qquad\qquad\qquad (7.2)$$

Where $\sqrt{2C_h C_o \lambda}$ is the constant and $\sum_{i=1}^{n}\sqrt{\frac{Z_i}{L}}$ is the inventory cost increase factor (multiplication factor) and the ratio between the inventory costs in a centralized system

(7.1) and in a decentralized system (7.2). The behaviour of the inventory cost increase factor with respect to the service zone size of a facility can be seen by plotting $\sqrt{\frac{Z_1}{L}}$, where Z_1 is the length of a facility's service zone in the line segment (Figure 7.1).
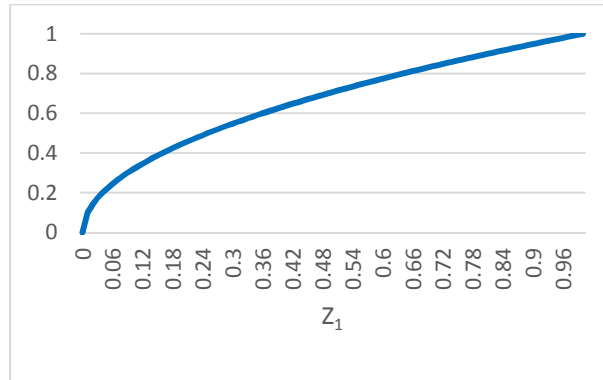
Figure 7.1: Inventory cost in Z_1 with $L = 1$

Due to the square root factor, there exists an economy of scale in relation to the size of a zone (or service area). With the increase in the size of a zone within the line segment, the variation in the unit cost decreases, i.e. the marginal cost decreases per unit length. Hence, when the inventory cost is dominant, it should be better to have bigger service areas.

Transportation cost: We use average distance to reach a point on the line segment to determine the transportation cost as the customers are assumed to be uniformly distributed. Note that locating a facility point at the middle of its service zone minimizes the travelling in the zone without affecting the inventory cost. Also, in an optimal solution, each service zone should be contiguous, i.e. not split into smaller segments. Considering that a facility is located in the middle of its service zone Z, the length of zone on each side of the facility is equal to $Z/2$. Formally, the average distance to reach a demand point within Z from the facility point can be defined as the average value of a continuous function $f(x) = x$ on the closed interval $[0, Z/2]$, i.e.

$\frac{1}{\frac{Z}{2}-0}\int_0^{\frac{Z}{2}} x\,dx$, which equals to $\frac{Z}{4}$. With n facilities located on the line segment and in

middle of their respective zones,

$$Average\ distance\ to\ serve\ a\ demand\ on\ line\ segment = \sum_{i=1}^n \frac{Z_i}{L} \times \frac{Z_i}{4} \qquad (7.3)$$

where $\frac{Z_i}{4}$ is the average distance to reach a demand point in facility i's zone and $\frac{Z_i}{L}$ is

the probability that a demand originates from facility i's zone within the line segment.

Considering the total demand λ, cost of transportation per unit length C_t, and the

average distance to serve a demand (7.3),

$$Total\ transportation\ cost = C_t\lambda \times \sum_{i=1}^n \frac{Z_i^2}{4L} = \frac{C_t\lambda}{4} \times \sum_{i=1}^n \frac{Z_i^2}{L} \qquad (7.4)$$

where $\frac{C_t\lambda}{4}$ is the constant and $\sum_{i=1}^n \frac{Z_i^2}{L}$ can be considered as the multiplication factor

for the transportation cost. Plotting $\frac{Z_1^2}{L}$ (or Z_1^2 as $L=1$), such that Z_1 belongs to Z_i,

illustrates the manner in which the transportation cost at a facility, with respect to its

service zone size, contributes to the total transportation cost in the system (Figure 7.2).
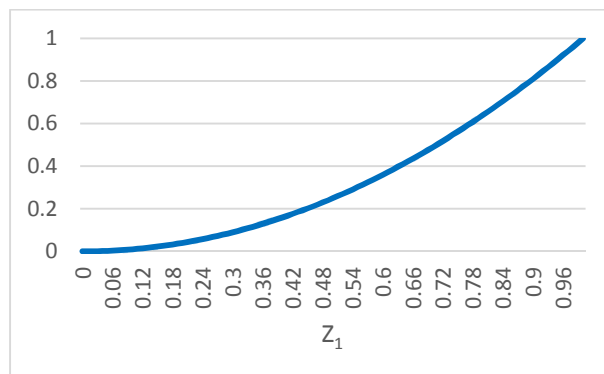
Figure 7.2: Transportation cost in Z_1 with $L = 1$

The transportation cost has a diseconomy of scale behaviour in relation to the zone

sizes. As the zone size within the line segment increases, the variation in the unit cost

increases, asserting that a decentralized system with smaller zones is better when it comes to the transportation cost.

Total Cost: Separately, the inventory and the transportation cost functions demonstrate concave and convex properties respectively with respect to the service zone sizes. Adding total inventory cost (7.2) and total transportation cost (7.4) per unit time, we define

$$Total\ service\ cost = \left(\sqrt{2C_hC_o\lambda} \times \sum_{i=1}^{n}\sqrt{\frac{Z_i}{L}}\right) + \left(\frac{C_t\lambda}{4} \times \sum_{i=1}^{n}\frac{Z_i^2}{L}\right) \qquad (7.5)$$

To focus on the impact of the size of zones on the total cost, we simplify the above cost function by transforming it and incorporating a constant $\boldsymbol{\alpha}$ to contain all the cost constants and the demand. Dividing the total cost (7.5) by $\frac{C_t\lambda}{4}$,

$$Transformed\ cost\ function = \frac{1}{\frac{C_t\lambda}{4}}\left(\sqrt{2C_hC_o\lambda} \times \sum_{i=1}^{n}\sqrt{\frac{Z_i}{L}}\right) + \frac{1}{\frac{C_t\lambda}{4}}\left(\frac{C_t\lambda}{4} \times\right.$$

$$\left.\sum_{i=1}^{n}\frac{Z_i^2}{L}\right)$$

$$= \frac{1}{C_t}\sqrt{\frac{32C_hC_o}{\lambda}}\left(\sum_{i=1}^{n}\sqrt{\frac{Z_i}{L}}\right) + \sum_{i=1}^{n}\frac{Z_i^2}{L}$$

Substituting $\frac{1}{C_t}\sqrt{\frac{32C_hC_o}{\lambda}}$ with α,

$$Transformed\ cost\ function = \alpha\left(\sum_{i=1}^{n}\sqrt{\frac{Z_i}{L}}\right) + \sum_{i=1}^{n}\frac{Z_i^2}{L} \qquad (7.6)$$

To examine the impact of a facility's service zone size on the total cost, we reduce the transformed cost function (7.6) to $F(Z_1) = \alpha\sqrt{\frac{Z_1}{L}} + \frac{Z_1^2}{L}$, where Z_1 is a service zone

length in the line segment $\overline{0L}$. The plot of $F(Z_1)$ (Figure 7.3) displays the behaviour of total cost in a zone with respect to different lengths of the zone within the line segment considering α and L equal to 1. For certain value of Z_1 there is a point (the point of inflection) where $F(Z_1)$ changes from concave to convex.
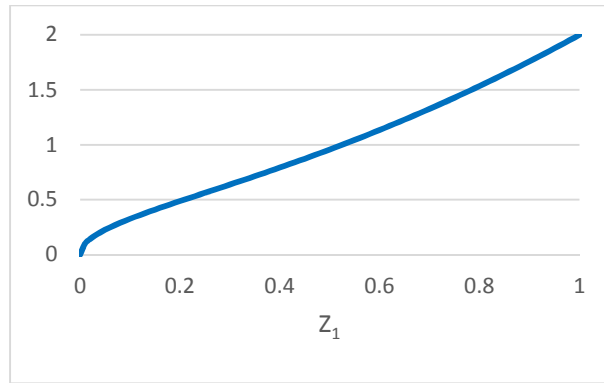
Figure 7.3: Total cost in Z_1, $F(Z_1)$, with $\alpha = L = 1$

Let $(Z_{inf}, F(Z_{inf}))$ be the inflection point on the graph of $F(Z_1)$. The second derivative of $F(Z_1)$ can be used to find Z_{inf}.

$$F'(Z_1) = \frac{1}{2}\frac{\alpha}{\sqrt{Z_1 L}} + \frac{2Z_1}{L} \geq 0, F(Z_1) \text{ is strictly increasing}$$

$$F''(Z_1) = -\frac{1}{4}\alpha\frac{Z_1^{-3/2}}{\sqrt{L}} + \frac{2}{L}$$

$$-\frac{1}{4}\alpha\frac{Z_1^{-3/2}}{\sqrt{L}} + \frac{2}{L} = 0$$

$$Z_{inf} = \left(\frac{\alpha\sqrt{L}}{8}\right)^{\frac{2}{3}}$$

With the increase in the value of α, the inflection point on the graph of $F(Z_1)$ moves towards the right (larger Z_{inf}), increasing the concave part of the function and making the inventory cost more relevant in the total cost (Figure 7.4). Note that α, which equals to $\frac{1}{C_t}\sqrt{\frac{32C_hC_o}{\lambda}}$, is higher when the inventory order cost C_o and the inventory holding

cost per unit per unit time C_h are high compared to the transportation cost per unit length C_t.
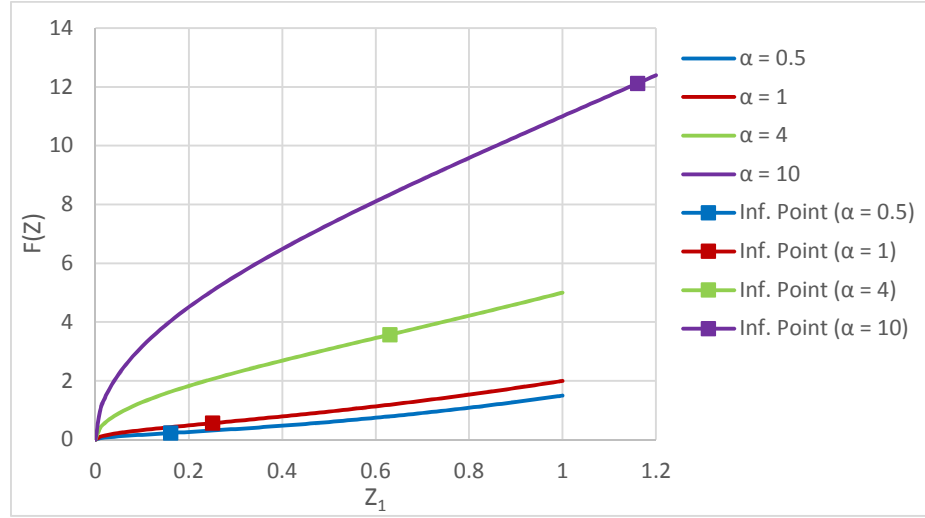
Figure 7.4: $F(Z_1)$ for different values of α. ($L = 1$)

Considering that n facilities are located at the middle of their service zones on the line segment $\overline{0L}$, the following Nonlinear Programming (NP) model determines the optimal lengths of service zones Z_i, where $i = 1 \ldots n$, which minimize the total cost (7.5) while covering the line segment considering the maximum service distance d_{max}. The number of facilities, n, is minimum when it is equal to $\left[\frac{L}{2d_{max}}\right]$.

Objective: *Minimize* $\alpha \left(\sum_{i=1}^{n} \sqrt{\frac{Z_i}{L}} \right) + \sum_{i=1}^{n} \frac{Z_i^2}{L}$

Subject to,

$$\sum_{i=1}^{n} Z_i = L \tag{7.7}$$

$$Z_i \leq 2 \times d_{max} \tag{7.8}$$

$$Z_i \geq 0 \tag{7.9}$$

The objective is to minimize the transformed cost function (7.6). Note that the values of Z_i (the decision variables) that minimize the transformed cost function (7.6) also minimize the total service cost function (7.5). The first constraint (7.7) ensures that the sum of the lengths of service zones is equal to the length of the line segment so that the entire line segment is covered. The second constraint (7.8) restricts the service zone lengths to a maximum $2d_{max}$ so that all the points in a service zone are within the maximum allowable distance (d_{max}) from the facility point (Figure 7.5). Finally, the length of a zone cannot be negative (7.9).
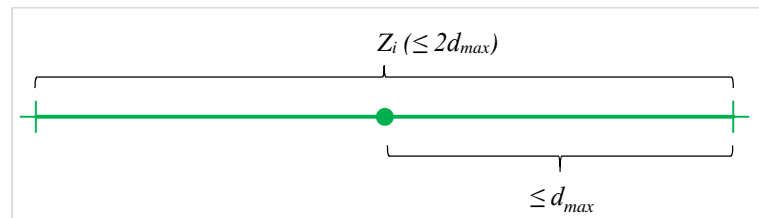
Figure 7.5: Placement of a facility on a line segment

Two distinct location patterns can be observed by running the optimization model for different values of α and d_{max}. These two patterns, *Pattern 1* and *Pattern 2*, favouring diseconomy and economy of scale respectively, are defined below and are illustrated in Figures 7.6 and 7.7 for $n = 3$.

Pattern-1 (occurs when the transportation cost is dominant): All zones are of equal length, i.e. for *n* facilities covering a line segment, $Z_i = 1/n$, where Z_i is the zone length of facility *i*, for $i = 1... n$. This brings the centralization level to the minimum and hence minimizes the transportation cost.
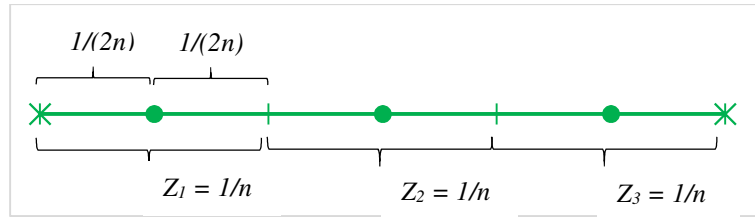
Figure 7.6: Location Pattern-1 for three facilities

Pattern-2 (occurs when the inventory cost is dominant): All zones except one are of the maximum length ($2d_{max}$). One of the facilities has a smaller zone of length equalling to $L-2d_{max}(n-1)$. Assuming the last facility on the line segment has the smaller zone, for $i = 1...(n-1)$, $Z_i = 2d_{max}$, while for $i = n$, $Z_i = L - 2d_{max}(n-1)$, where Z_i is the zone length of facility i. Maximizing the centralization level under the distance constraint d_{max}, this pattern minimizes the inventory cost.
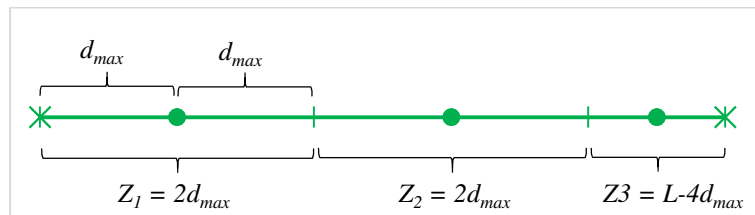
Figure 7.7: Location Pattern-2 for three facility points

For a particular distance constraint, d_{max}, there is a threshold values of α *(and Z_{inf})* below which *Pattern 1* turns out to be optimum and above which the optimum pattern appears to be *Pattern 2*. As shown previously, the increase in the value of α increases Z_{inf} and makes the concave part of the cost curve (and the inventory cost) more significant (illustrated in Figure 7.8). Figure 7.9 shows the cost difference between *Pattern 1* and *Pattern 2* over different values of α when $L = 1$, $d_{max} = 0.24$ and hence $n = 3$.
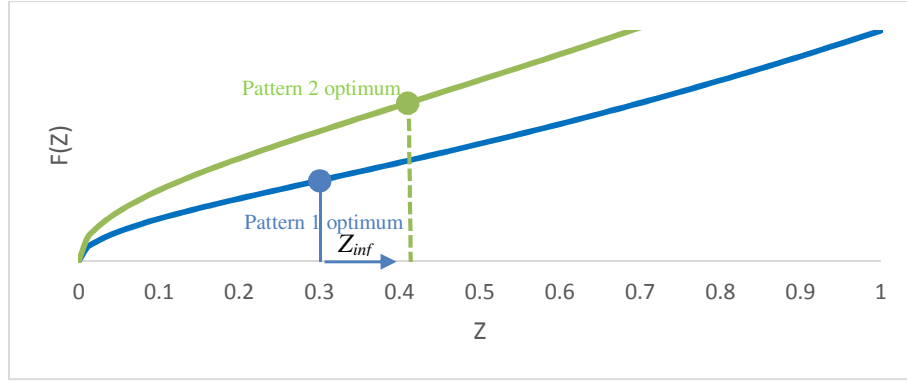
Figure 7.8: Optimality of Pattern 1 and Pattern 2 in relation to Z_{inf}
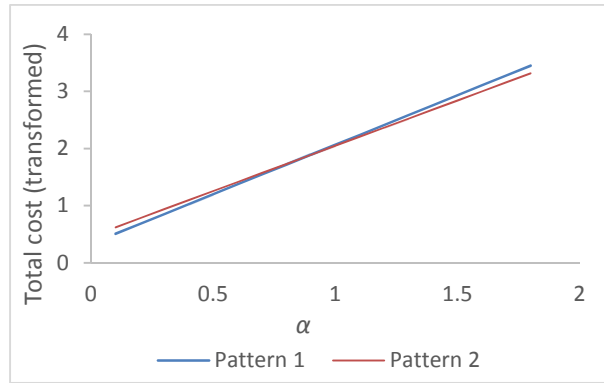
Figure 7.9: Cost difference between Pattern 1 and Pattern 2 with L=1, d_{max}=0.24 and n=3

Let α_t be the value of α at which the cost in *Pattern 1* is equal to the cost in *Pattern 2*. For α below α_t, *Pattern 1* results in lower costs than in *Pattern 2*, in contrast to when α is greater than α_t. As $z_i = \frac{L}{n}$ for *i = 1 … n* in *Pattern 1*, and $z_i = 2d_{max}$, for *i = 1 … (n-1)*, and $z_n = L - (n-1)2d_{max}$ in *Pattern 2*, considering the transformed cost function (7.6),

$Transformed\ cost\ function\ (Pattern\ 1) = \alpha\sqrt{n} + \frac{L}{n}$, and

$Transformed\ cost\ function\ (Pattern\ 2) = \alpha \sum_{i=1}^{n-1} \sqrt{\frac{2d_{max}}{L}} + \alpha \sqrt{\frac{L-(n-1)2d_{max}}{L}} +$

$\sum_{i=1}^{n-1} \frac{(2d_{max})^2}{L} + \frac{(L-(n-1)2d_{max})^2}{L}$

205

Equating the transformed cost functions for *Pattern 1* and *Pattern 2*,

$$\alpha\sqrt{n} + \frac{L}{n} = \alpha \sum_{i=1}^{n-1}\sqrt{\frac{2d_{max}}{L}} + \alpha\sqrt{\frac{L-(n-1)2d_{max}}{L}} + \sum_{i=1}^{n-1}\frac{(2d_{max})^2}{L} + \frac{(L-(n-1)2d_{max})^2}{L}$$

$$\alpha_t = \frac{\left[\frac{(n-1)(2d_{max})^2}{L} + \frac{(L-(n-1)2d_{max})^2}{L} - \frac{L}{n}\right]}{\left[\sqrt{n}-(n-1)\sqrt{\frac{2d_{max}}{L}} - \sqrt{\frac{L-(n-1)2d_{max}}{L}}\right]}$$

The value of α_t increases with an increase in service distance constraint d_{max} (Figure 7.10). Note that with a given number of facilities, zone sizes in *Pattern 2* change with the change in the service distance constraint d_{max}. With a higher d_{max}, the system becomes more centralized, making it more favourable in terms of the inventory cost. On the other hand, with a given number of facilities, any increase in the service distance constraint does not affect *Pattern 1* as all zones have same length, hence the costs remain the same.
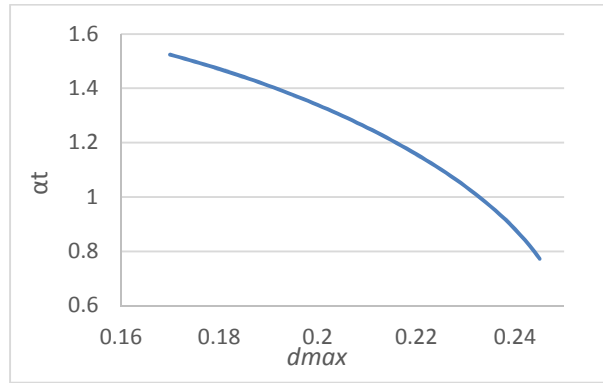
Figure 7.10: α_t (with L=1 and n=3)

7.4. A MINLP model for hierarchical optimization

In this section a MINLP model is presented to study the problem of setting up an optimum hierarchical system on a line segment considering two service distance constraints. As in previous chapters, we refer to the service that ensures the supply within the short maximum time window as the '*strict*' service, and the service that

ensures the supply within the longer maximum time window as the '*relaxed*' service. The nested hierarchical system consists of '*higher level*' facilities, which provide both the relaxed and strict services, and '*lower level*' facilities, which only provide the strict service. In order to be covered for both service types, every demand point has to be within the maximum service distance for the strict service from a higher or a lower level facility, and within the maximum service distance for the relaxed service from a higher level facility. The objective is to determine the hierarchical setup, based on the minimum number of service facilities of each type to cover the demand, which minimizes the sum of transportation and inventory costs.

In view of the above description and the assumptions in section 7.2, we introduce the following notations in addition to those defined in the previous section:

$f =$ fraction of demand for the strict service

$1-f =$ fraction of demand for the relaxed service

$d_{maxS} =$ service distance constraint for the strict service

$d_{maxR} =$ service distance constraint for the relaxed service

$n_S =$ total number of facilities in the system determined considering d_{maxS}. n_S is also the total number of facilities providing the strict service. To be minimum, n_S is taken as $\left\lceil \dfrac{L}{2d_{maxS}} \right\rceil$ where $2d_{maxS}$ is the maximum length a facility can cover to provide the strict service.

$n_R =$ minimum number of facilities out of n_S to cover the demand considering d_{maxR}. It is the number of facilities providing both the strict and relaxed services and, therefore, it is the number of higher level facilities in the system.

$n_S - n_R =$ number of lower level facilities = number of facilities only providing the strict service.

The above definitions of the facility numbers ensure the lowest total number of facilities and the lowest number of higher level facilities in the system. The categorization of the service facilities is summarized in Table 7.1.

Table 7.1: Classification of service facilities

The types of service facilities	Service Types	
	Relaxed service	*Strict service*
Higher level service facilities provide →	✓	✓
Lower level service facilities provide →		✓

The model considers separate variables to represent the length of the zone to the left and to the right of a facility point for each time based service type. The following is the set of decision variables for each facility i, with an associated location coordinate x_i over the line segment $\overline{0L}$, where $i = 1…n_S$,:

$Z_{Rli} =$ length of relaxed service zone to the left

$Z_{Rri} =$ length of relaxed service zone to the right

$Z_{Sli} =$ length of strict service zone to the left

$Z_{Sri} =$ length of strict service zone to the right

$X_i =$ 1 if facility i is a higher level facility, else 0

Z_{Rli} and Z_{Rri} are greater than zero only if facility i is a higher level facility, i.e. if $X_i = 1$, otherwise, Z_{Rli}, Z_{Rri} and X_i equal to zero. Variables/parameters n_R and n_S, Z_{Rli}, Z_{Rri}, Z_{Sli} and Z_{Sri}, and d_{maxR} and d_{maxS} respectively supersede n, Z, and d_{max} defined formerly.

If a facility i, for $i = 1…n_S$, is a higher level facility, then '$\frac{(1-f)(Z_{Rli}+Z_{Rri})}{L} + \frac{f(Z_{Sli}+Z_{Sri})}{L}$', is the fraction of demand it serves under the hierarchical setup. Considering this fraction of demand in the EOQ model (Section 7.3),

$$Total\ inventory\ cost\ at\ Higher\ Level\ facilities =$$

$$\sqrt{2C_h C_o \lambda} \sum_{i=1}^{n_S} \sqrt{\frac{(1-f)(Z_{Rli}+Z_{Rri})}{L} + \frac{f(X_i(Z_{Sli}+Z_{Sri}))}{L}} \qquad (7.10)$$

$\sqrt{\frac{(1-f)(Z_{Rli}+Z_{Rri})}{L} + \frac{f(X_i(Z_{Sli}+Z_{Sri}))}{L}}$ drops out in the above function (7.10) if facility i is

not a higher level facility, because $X_i = Z_{Rli} = Z_{Rri} = 0$.

Similarly, '$\frac{f(Z_{Sli}+Z_{Sri})}{L}$' is the fraction of demand served by facility i if it is a lower

level facility. This gives,

$$Total\ inventory\ cost\ at\ lower\ level\ facilities =$$

$$\sqrt{2C_h C_o \lambda} \sum_{i=1}^{n_S} \sqrt{\frac{f((1-X_i)(Z_{Sli}+Z_{Sri}))}{L}} \qquad (7.11)$$

When facility i is a higher level facility, i.e. $X_i = 1$, '$\sqrt{\frac{f((1-X_i)(Z_{Sli}+Z_{Sri}))}{L}}$' in the above

function (7.11) drops out. Adding total inventory costs at higher level facilities (7.10)

and at lower level facilities (7.11), we get,

$$Total\ inventory\ cost =$$

$$\sqrt{2C_h C_o \lambda} \left(\sum_{i=1}^{n_S} \sqrt{\frac{(1-f)(Z_{Rli}+Z_{Rri})}{L} + \frac{f(X_i(Z_{Sli}+Z_{Sri}))}{L}} + \right.$$

$$\left. \sum_{i=1}^{n_S} \sqrt{\frac{f((1-X_i)(Z_{Sli}+Z_{Sri}))}{L}} \right) \qquad (7.12)$$

The transportation cost for a decentralized service system with n_S facilities over a line

segment $\overline{0L}$ can be defined as '$\lambda C_t \sum_{i=1}^{n_S} \frac{Z_{li}}{2}\left(\frac{Z_{li}}{L}\right) + \frac{Z_{ri}}{2}\left(\frac{Z_{ri}}{L}\right)$', where $\frac{Z_{li}}{2}$ and $\frac{Z_{ri}}{2}$ are the

average distances in facility $i's$ service zone on the left and right respectively, and $\frac{Z_{li}}{L}$

and $\frac{Z_{ri}}{L}$ are the proportions of the line segment $\overline{0L}$ covered by facility $i's$ zone on the

left and right respectively. Considering this,

$$Total\ tranportation\ cost = \frac{\lambda C_t}{2}\left(\sum_{i=1}^{n_S}\frac{(1-f)(Z_{Rli}{}^2+Z_{Rri}{}^2)}{L}+\right.$$

$$\left.\sum_{i=1}^{n_S}\frac{f(Z_{Sli}{}^2+Z_{Sri}{}^2)}{L}\right) \tag{7.13}$$

Adding total inventory cost (7.12) and total transportation cost (7.13),

$$Total\ service\ cost =$$

$$\sqrt{2C_hC_o\lambda}\left(\sum_{i=1}^{n_S}\sqrt{\frac{(1-f)(Z_{Rli}+Z_{Rri})}{L}+\frac{f(X_i(Z_{Sli}+Z_{Sri}))}{L}}+\right.$$

$$\left.\sum_{i=1}^{n_S}\sqrt{\frac{f((1-X_i)(Z_{Sli}+Z_{Sri}))}{L}}\right)+\frac{\lambda C_t}{2}\left(\sum_{i=1}^{n_S}\frac{(1-f)(Z_{Rli}{}^2+Z_{Rri}{}^2)}{L}+\right.$$

$$\left.\sum_{i=1}^{n_S}\frac{f(Z_{Sli}{}^2+Z_{Sri}{}^2)}{L}\right) \tag{7.14}$$

To simplify the formulation, we transform the total cost function (7.14) by dividing it

with $\frac{\lambda C_t}{2}$.

$$=\frac{\sqrt{2C_hC_o\lambda}}{\frac{\lambda C_t}{2}}\left(\sum_{i=1}^{n_S}\sqrt{\frac{(1-f)(Z_{Rli}+Z_{Rri})}{L}+\frac{f(X_i(Z_{Sli}+Z_{Sri}))}{L}}+\sum_{i=1}^{n_S}\sqrt{\frac{f((1-X_i)(Z_{Sli}+Z_{Sri}))}{L}}\right)+$$

$$\frac{\frac{\lambda C_t}{2}}{\frac{\lambda C_t}{2}}\left(\sum_{i=1}^{n_S}\frac{(1-f)(Z_{Rli}{}^2+Z_{Rri}{}^2)}{L}+\sum_{i=1}^{n_S}\frac{f(Z_{Sli}{}^2+Z_{Sri}{}^2)}{L}\right)$$

Replacing '$\frac{\sqrt{2C_hC_o\lambda}}{\frac{\lambda C_t}{2}}$' in the above by a constant $\boldsymbol{\alpha}$, we formulate the *Objective* as,

Minimize:

$$\alpha \left(\sum_{i=1}^{n_S} \sqrt{\frac{(1-f)(Z_{Rli}+Z_{Rri})}{L} + \frac{f(X_i(Z_{Sli}+Z_{Sri}))}{L}} + \sum_{i=1}^{n_S} \sqrt{\frac{f((1-X_i)(Z_{Sli}+Z_{Sri}))}{L}} \right) +$$

$$\left(\sum_{i=1}^{n_S} \frac{(1-f)(Z_{Rli}{}^2+Z_{Rri}{}^2)}{L} + \sum_{i=1}^{n_S} \frac{f(Z_{Sli}{}^2+Z_{Sri}{}^2)}{L} \right) \tag{7.15}$$

(Note that the values of the decision variables that minimize the objective function (7.15) also minimize the total service cost (7.14).)

Subject to constraints,

$$\sum_{i=1}^{n_S} X_i \le n_R \tag{7.16}$$

$$Z_{Rli} \le d_{maxR} \times X_i \tag{7.17}$$

$$Z_{Rri} \le d_{maxR} \times X_i \tag{7.18}$$

$$\sum_{i=1}^{n_S} Z_{Rli} + Z_{Rri} = L \tag{7.19}$$

$$\sum_{i=1}^{n_S} Z_{Sli} + Z_{Sri} = L \tag{7.20}$$

$$Z_{Sli} \le d_{maxS} \tag{7.21}$$

$$Z_{Sri} \le d_{maxS} \tag{7.22}$$

$$x_1 = Z_{Sl1} \tag{7.23}$$

$$x_i = x_{i-1} + Z_{Sri-1} + Z_{Sli} \text{ , } for \ i > 1 \tag{7.24}$$

$$x_{n_S} + Z_{Srn_S} = L \tag{7.25}$$

$$x_1 \geq Z_{Rl1} \tag{7.26}$$

$$x_1 - Z_{Rl1} \leq L(1 - X_1) \tag{7.27}$$

$$x_i \geq \left(\sum_{k<i} Z_{Rlk} + Z_{Rrk} \right) + Z_{Rli} - (L(1 - X_i)), for\ i > 1 \tag{7.28}$$

$$x_i - \left(\sum_{k<i} Z_{Rlk} + Z_{Rrk} \right) - Z_{Rli} \leq (L(1 - X_i)), for\ i > 1 \tag{7.29}$$

$$X_i = [0, 1]. (Binary) \tag{7.30}$$

$$Z_{Rli}, Z_{Rri}, Z_{Sli}, Z_{Sri} \geq 0 \tag{7.31}$$

A facility can either be a higher level facility or a lower level facility, which is indicated by the value of X_i as a binary variable (constraint (7.31)); *1* if facility *i* is a higher level, else *0* if facility *i* is a lower level facility. The number of higher level facilities is restricted to n_R, which is the minimum number of facility points out of n_S required to cover the line segment for the relaxed service considering d_{maxR} (constraint (7.16)). A facility is allowed zones for providing the relaxed service only if the facility is a higher level facility (constraints (7.17) and (7.18)). The sum of service zone lengths for the relaxed service and the sum service zone lengths for the strict service are equal to the line segment's length L (constraints (7.19) and (7.20)), where the zones for the relaxed service and the strict services on either side of a facility point cannot be longer than d_{maxR} and d_{maxS} respectively (constraints (7.17), (7.18), (7.21) and (7.22)).

Constraints (7.23) to (7.29) check the consistency of the facility location coordinates and the length of service zones. Constraint (7.23) ensures that the coordinate of the

first facility point is equal to the length of its strict services zone on its left. Constraint (7.24) ensures that the coordinate of a facility, other than the first facility, is equal to the sum of the coordinate of the adjacent facility on the left, the length of the adjacent facility's strict services zone on right, and the facility's own strict services zone on its left. Constraint (7.25) ensures that the sum of the coordinate of the last facility and the last facility's strict services zone on its right is equal to the length of the line segment (7.25). Constraints (7.26) to (7.29) define the relationships between the location of facility points and the relaxed service zones. If the first facility is a higher level facility, its coordinate on the line segment is equal to the length of its relaxed service zone on the left, else, if it is a lower level facility, the length of its relaxed service zone is equal to zero. Similarly, if the first facility is a higher level facility, the difference between its coordinate and its relaxed service zone length on the left is equal to zero, otherwise the difference equals to its coordinate on the line segment. Constraint (7.27) reduces to '$Z_{Rl1} - x_1 \leq 0$' if '$X_1 = 1$'. Constraints (7.26) and (7.27) together ensure that '$x_1 = Z_{Rl1}$', if '$X_1 = 1$'. Constraints (7.28) and (7.29) perform a similar function as the constraints (7.26) and (7.27), but for the facilities other than the first. If '$X_i = 1$', the last component of the constraints (7.28) and (7.29), $(L \times (1-X_i))$, drops out, ensuring that x_i equals to $(\sum_{k<i} Z_{Rlk} + Z_{Rrk}) + Z_{Rli}$.

7.5. Numerical examples

This section reports the optimum setup solutions for scenarios based on different combinations of the demand fractions for the strict and relaxed services (*f* and *(1-f)*), and the cost constant α for two sets of service distance constraints over a line segment of unit length ($\overline{01}$, i.e. *L = 1*). The model allows rather small problems, in terms of facility number, to be solved to optimality for low α and high *f* values (discussed later

in this section). The set of constraints and the minimum facility numbers are stated in Table 7.2.

Table 7.2: Experimental settings

	Setting 1	Setting 2
d_{maxR}	0.48	0.24
d_{maxS}	0.24	0.12
$n_R = \lceil L/2d_{maxR} \rceil$	2	3
$n_S = \lceil L/2d_{maxS} \rceil$	3	5
$L = 1$		

For *Settings 1* and *2* in Table 7.2, the number of higher level facilities determined as $\lceil L/2d_{maxR} \rceil$, i.e. $n_R = 2$ and $n_R = 3$ respectively, gives a feasible hierarchical solution. However, note that the minimum number of higher level facilities (n_R) determined as $\lceil L/2d_{maxR} \rceil$, might not always be feasible. For example, consider $L = 1$, $d_{maxR} = 0.5$ and $d_{maxS} = 0.25$. The minimum total number of facilities (ns) equals to $\lceil L/2d_{maxS} \rceil = 2$. For providing full coverage for the strict service on a $\overline{01}$ line segment with the distance constraint $d_{maxS} = 0.25$, a facility point has to be exactly located at the points *0.25* and *0.75* each. While the minimum number of higher level facilities (n_R), determined as $\lceil L/2d_{maxR} \rceil$, is 1. However, to cover the line segment for the relaxed service through just one higher level facility, the facility has to be located exactly at the middle of the line segment, i.e. at point *0.5*, as $d_{maxR} = 0.5$. Clearly this cannot be achieved as the only location options are points 0.25 and 0.75. In such cases, to determine the minimum number of higher level facilities, we can find the initial value of n_R as equal to $\lceil L/2d_{maxR} \rceil$, and then increment n_R by one until the model gives a feasible solution.

The chosen d_{maxR} and d_{maxS} values (Table 7.2) provide a level of flexibility in locating the minimum number of facility points. For example, the maximum distance constraint as *0.167* and *0.24* both require minimum three facility points to cover the line segment of a unit length. However, *0.167* as the distance constraint is very restrictive in terms

of where the facilities can be located, while considering *0.24* as the distance constraint results in a greater flexibility to move the facility points while providing the full coverage on the line. This makes the effect of different settings on the location pattern more noticeable.

The value of α is considered as *0.05* and *8* under *Setting 1* and *0.02* and *2* under *Setting 2*. The values of α as *0.05* and *0.02* in *Setting 1* and *Setting 2* respectively represent the scenarios where the transportation cost dominates. Whereas, the values of α as *8* and *2* in *Setting 1* and *Setting 2* respectively represent the scenarios where the inventory cost dominates. For each value of α, the model is run with f = *0.05, 0.2, 0.4, 0.6, 0.8* and *0.95*. Note that optimization for $f = 0$ and $f = 1$ reduces the problem to locating facility points considering one service level (Section 7.3). The model is programmed in LINGO and solved using LINGO's '*Global Solver*'. The solver combines a series of range bounding (e.g. interval analysis and convex analysis) and range reduction techniques (e.g. linear programming and constraint propagation) within a branch-and-bound framework to find proven global solutions to nonconvex nonlinear programs. Gau and Schrage (2004) describe the global solver, discussing its fundamental algorithm, techniques and performance in detail. The solver converts the original nonlinear/nonconvex problem into several linear/convex sub-problems, uses a Convex, Interval, and Algebraic (CIA) analysis, and applies a branch-and-bound technique to exhaustively search over the sub-problems for the global solution.

The solver struggles to find (or confirm) the global optimum solution when the value of α is low and the fraction of demand for the strict service f is high. With Setting 2 (Table 7.2) the solver could not confirm the global optimum for $f \geq 0.8$ when $\alpha = 0.02$ within considerable runtime. Appendix 8 provides details on the computational effort

for Settings 1 and 2 along with a case with higher number of facilities considering the maximum runtime of 30 minutes. With $\alpha = 2$ and $f = 0.6$, the problem was solved for 9 facilities; while decreasing α to 0.02 or increasing f to 0.8 did not allow the problem with 9 facilities to be solved.

Optimum solutions with low α values (dominant travelling cost): Below a certain value of α, the hierarchical pattern demonstrates characteristics that are comparable to the Pattern 1 in Section 7.3. The zones for a service type tend to be equal in length and the facility points tend to be in the middle of the service zones. With the increase in the fraction of demand for the relaxed service (*1-f*), the higher level facility points in the optimum solution appear closer to the middle of their relaxed service zones. This reduces the difference between the left and right relaxed service zones of higher level facilities. As a result, the sum of squared lengths of the facilities' left and right relaxed service zones reduces, while the same increases for the strict service zones. That is, with the change in the value of *f*, the zones for one service type can become more symmetrical at the cost of a higher asymmetry in the zones of the other service type. Another observation is that the higher level and lower level facilities are placed alternatively and the first and the last facilities are setup as higher level facilities. These behaviours can be seen in Tables 7.3 and 7.4 which present cases with *Setting 1* and *Setting 2* respectively with low α values. The smaller *f* is, the more equally spread the facility locations providing the relaxed service are. The larger *f* is, the more equally spread the facility locations providing the strict service are.

Table 7.3: Test cases for running the optimization model with Setting 1 ($d_{maxR} = 0.48$, $d_{maxS} = 0.24$), and $\alpha = 0.05$

Case	f	Location
1	0	
2	0.05	**Relaxed service zone** **Strict service zone**
3	0.2	**Relaxed service zones** **Strict service zones**
4	0.4	**Relaxed service zones** **Strict service zones**
5	0.6	**Relaxed service zones** **Strict service zones**
6	0.8	**Relaxed service zones** **Strict service zones**

| 7 | 0.95 | *Relaxed service zone*
 0.172495 ... 0.827505
 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

 Strict service zone
 0.172495 ... 0.5 ... 0.827505
 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 |
| 8 | 1 | 0.167 ... 0.5 ... 0.833
 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 |

Table 7.4: Test cases for running the optimization model with Setting 2
($d_{maxR} = 0.24$, $d_{maxS} = 0.12$), and $\alpha = 0.02$

Case	f	Location
9	0	
10	0.05	**Relaxed service zone** **Strict service zone**
11	0.2	**Relaxed service zones** **Strict service zones**
12	0.4	**Relaxed service zones** **Strict service zones**
13	0.6	**Relaxed service zones** **Strict service zones**
14*	0.8	**Relaxed service zones** **Strict service zones**

15*	0.95	**Relaxed service zone** 0.104, 0.500, 0.896 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 **Strict service zone** 0.104, 0.302, 0.500, 0.698, 0.896 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
16	1	0.100, 0.3, 0.5, 0.7, 0.900 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

*** Local optimum**
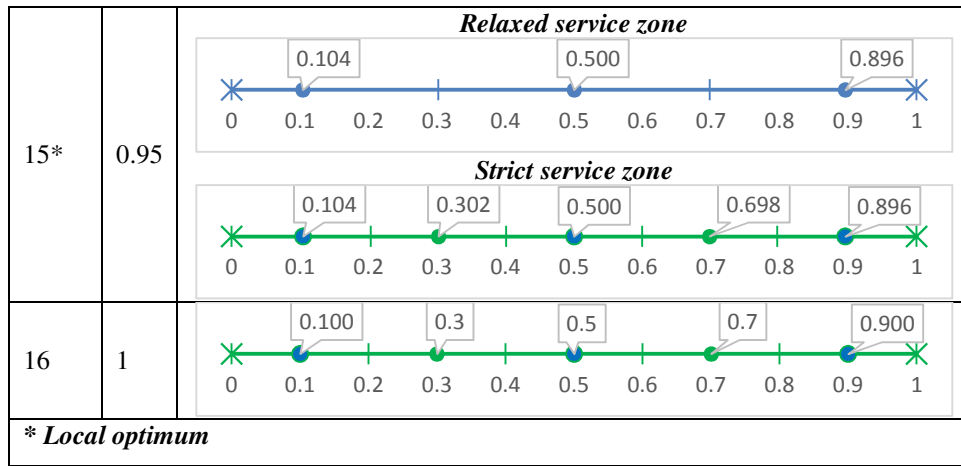
Optimum solution with high α values (dominant inventory cost): Beyond a certain value of α, the optimum hierarchical solutions show characteristics of Pattern 2 (Section 7.3), and the alternating higher level and lower level facility pattern tends to disappear (Tables 7.5 and 7.6). With a higher fraction of demand for the relaxed service (_1-f_), the sum of the square roots of the lengths of the relaxed service zones reduces, which in turn reduces the inventory for relaxed service provision.

Unlike several of the cases where the value of α is low, i.e. where travelling cost dominates (Tables 7.3 and 7.4), in all the cases in Tables 7.5 and 7.6, there are facilities with zones that comprise demand points that are closer to an adjacent facility. This violates the key assumption of Voronoi diagrams. It can also be observed in several of the reported cases (Cases 18, 19 and 20 in Table 7.5, and 26, 27, 28 and 29 in Table 7.6) that one of the higher level facilities provides the relaxed service from (very close to) a boundary of its zone. This is an interesting insight because the inventory cost is only affected by the size of a service zone and not where the facility is placed inside the zone, whereas, the transportation cost is affected by the position of a facility point in its zone. It is not favourable in terms of the average service distance if the facility serves from an extreme point of its zone. Looking on one level, the facility point should always be located at the centre of its zone as it reduces the transportation cost

without compromising the inventory cost. On two levels however, moving a facility point on one level to improve the transportation cost may mean compromising the transportation cost on the other level. For example, in Cases 18 and 19 (Table 7.5), assuming that the service zones are fixed, moving the second higher level facility located at point *0.96* to the point *0.98* will bring the facility to the centre of its relaxed services zone and would reduce the transportation cost for providing the relaxed service without increasing the inventory cost. This will however move the same facility point further away from its strict services zone's centre and hence increase the costs for the strict service provision. On the other hand, assuming that the facility positions are fixed in Cases 18, 19 and 20 (Table 7.5), having no relaxed service zone on the second higher level facility's left side allows more relaxed service demand to be consolidated at the first higher level facility, which increases the centralization level and reduces the inventory. Hence, though a solution where a facility serves from a boundary of its service area may look unreasonable, when inventory cost is dominant and the fraction of demand for the relaxed service is high, it can in fact be optimum to have a service facility providing the relaxed service from a boundary of its service area. This illustrates that determining the optimal service areas considering two distance constrained services is highly complex even for small problems on a line segment.

Table 7.5: Test cases for running the optimization model
with Setting 1 ($d_{maxR} = 0.48$, $d_{maxS} = 0.24$), and $α = 8$
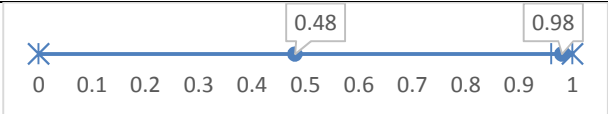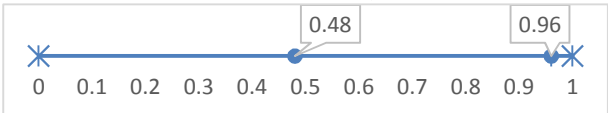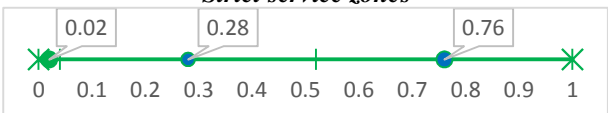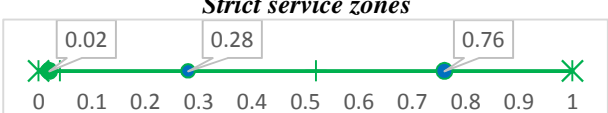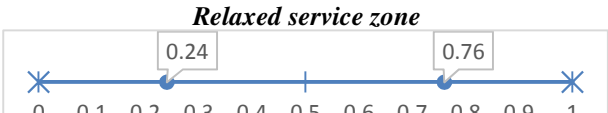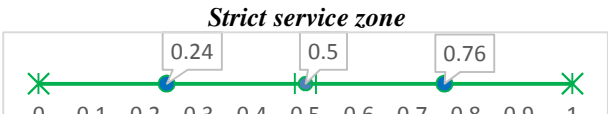
Case	f	Location
17	0	
18	0.05	**Relaxed service zones**
19	0.2	**Strict service zones**
20	0.4	**Relaxed service zones** **Strict service zones**
21	0.6	**Relaxed service zones**
22	0.8	**Strict service zones**
23	0.95	**Relaxed service zone** **Strict service zone**
24	1	

Table 7.6: Test cases for running the optimization model with Setting 2
(d_{maxR} = 0.24, d_{maxS} = 0.12), and α = 2

Case	f	Location
25	0	
26	0.05	*Relaxed service zone* *Strict service zone*
27	0.2	*Relaxed service zones* *Strict service zones*
28	0.4	*Relaxed service zones*
29	0.6	*Strict service zones*
30	0.8	*Relaxed service zones* *Strict service zones*
31	0.95	*Relaxed service zone* *Strict service zone*
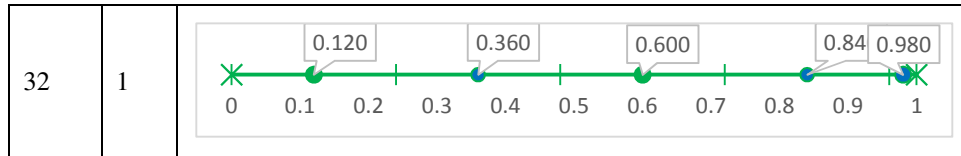
Relaxing the hierarchy restriction: Not considering the constraint for the maximum number of higher level facilities (7.16) in effect allows the system to be non-hierarchical since all facilities can potentially provide both types of service. Analysing the problem without this restriction confirms that it might be better to set-up a non-hierarchical system when the transportation cost dominates, while, a hierarchical system, providing a higher centralization level, can be beneficial when the inventory cost dominates. When the transportation cost is dominant, i.e. when α is small, it is observed that without the restriction on maximum number of higher level facilities (7.16) the optimum setup turns out to be of a non-hierarchical nature, where all facility points provide both types of service. Figure 7.11 presents the optimum solution for Case 4 (Table 7.3) when the constraint for the maximum number of higher level facilities (7.16) is removed. All facilities provide both service types in equal service zones and from the centre of the zones. This increases the level of decentralization in the system and reduces the transportation cost.
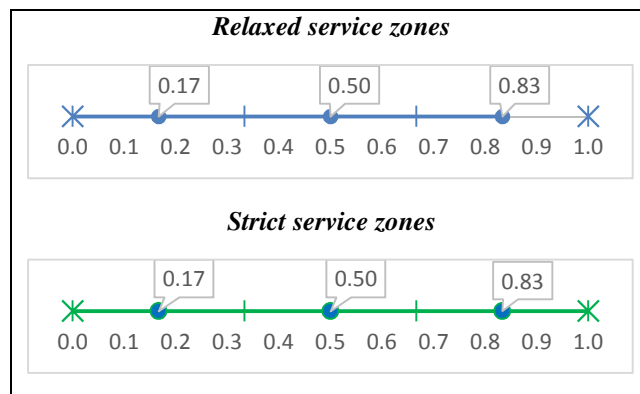
Figure 7.11: Case 4 without constraint on the number of higher level facilities (α = 0.05, f = 0.4)

For the reported cases, when the inventory cost is dominant, the optimum setup without the constraint on the number of higher level facilities (7.16) remains

hierarchical. Figure 7.12 presents the optimum setup for Case 20 (Table 7.5) when the number of higher level facilities is not constrained. The solution remains similar with exactly the same objective function value as when the maximum number of higher level facilities is constrained.
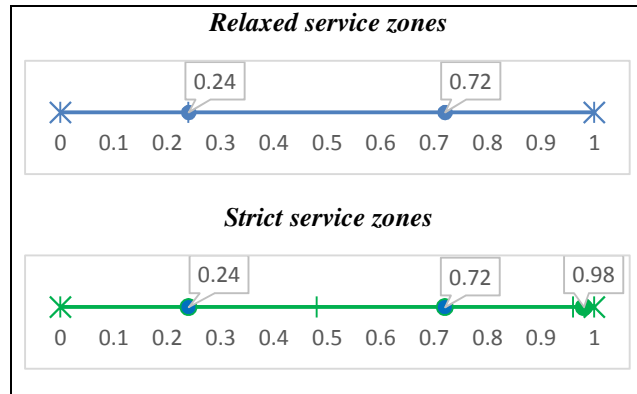
Figure 7.12: Case 20 without constraint on the number of higher level facilities (α = 8, f = 0.4)

The following analysis compares the performance of the optimum hierarchical setups determined earlier with the performance of the non-hierarchical setup providing the maximum decentralization. Under the non-hierarchical setup, all facility points have the service zones of the same size (*1/n*) and each facility point provides both the relaxed and strict services in its service zone (Figures 7.13 (a) and (b)). This results in the maximum level of decentralization and all demand points being served from the nearest facility point regardless of the required service type. As the non-hierarchical setup is not sensitive to the changes in the demand fractions for both service types, the inventory and transportation costs remain constant in the system over the changes in *f*.
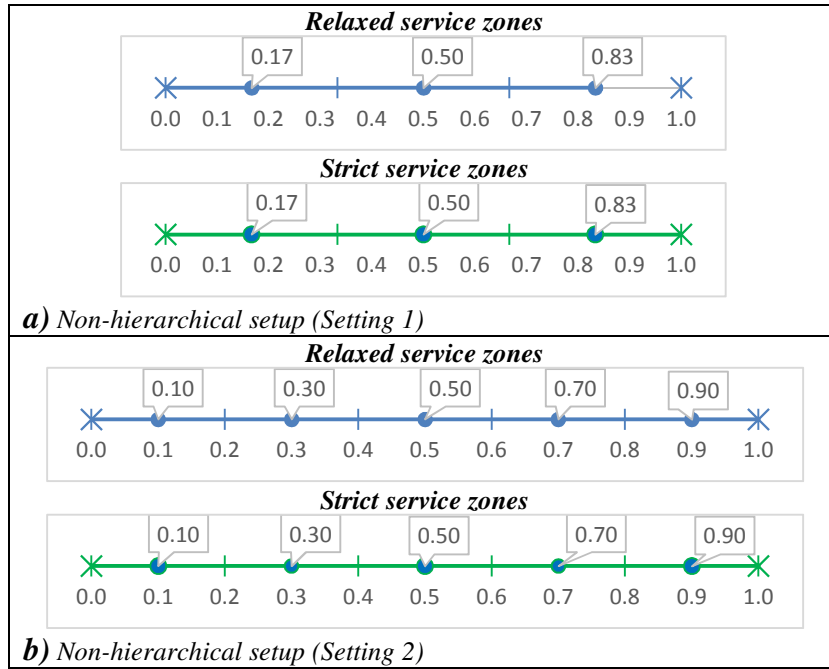
Figures 7.13 (a & b): Non-hierarchical (maximum decentralized) setup

Figure 7.14 (a) and (b) present the objective function (transformed cost function (7.15)) values based on the optimum hierarchical setups for Cases 3, 4, 5, 6, 11, 12, 13 and 14 (cases with dominant transportation cost), and the objective function value when the setup is non-hierarchical (Figures 7.13 (a) and (b)).

a): Setting 1 (d_{maxR} = 0.48, d_{maxS} = 0.24), and α = 0.05. f=0.2 (Case 3), f=0.4 (Case 4), f=0.6 (Case 5), f=0.8 (Case 6)

b): Setting 2 (d_{maxR} = 0.24, d_{maxS} = 0.12), and α = 0.02. f=0.2 (Case 11), f=0.4 (Case 12), f=0.6 (Case 13), f=0.8 (Case 14)
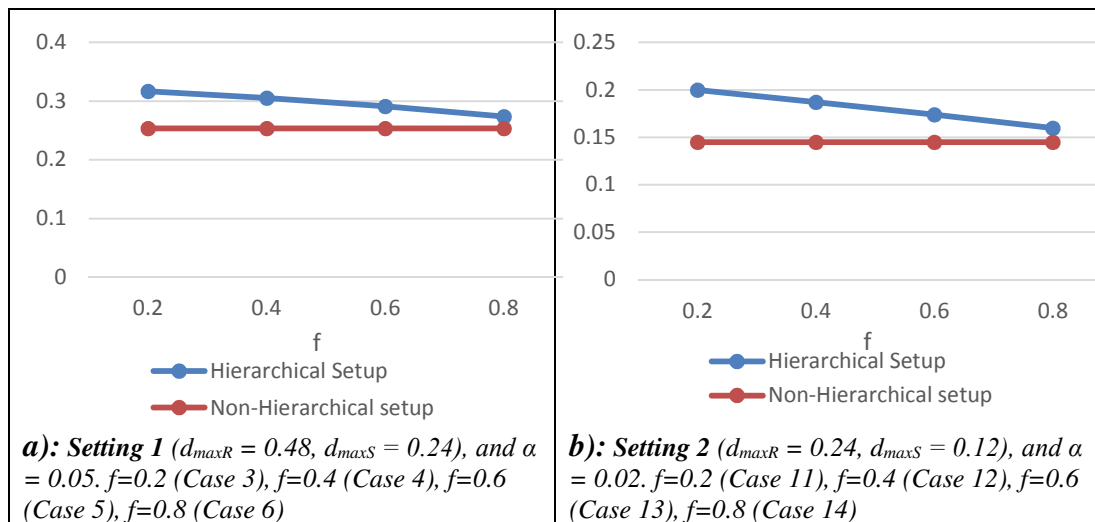
Figures 7.14 (a & b): Objective function values based on the optimum hierarchical setups for the cases with dominant transportation cost against the objective function value under the non-hierarchical setup.

Figures 7.15 (a) – (d) compare the optimum hierarchical setups for Cases 3, 6, 11 and 14, in which the transportation cost dominates, with the non-hierarchical setup

226

(Figures 7.13 (a) and (b)) over varying fractions of demand for both service types in each case. The results show that the non-hierarchical setup performs better than the optimum hierarchical setups for the cases with dominant transportation costs and the saving by setting up the non-hierarchical setup increases as the fraction of demand for the relaxed service *(1-f)* increases (Figures 7.14 and 7.15). With the increase in the fraction of demand for the relaxed service, more demand is met from more centralized higher level facilities under the hierarchical setup. This results in higher transportation cost, which dominates in these cases. Under the non-hierarchical setup on the other hand, demand is met with the maximum decentralization from the nearest facility, which results in a lower transportation cost.

a) Optimum hierarchical setup for Case 3 (setting 1, f=0.2, α=0.05) versus non-hierarchical setup

b) Optimum hierarchical setup for Case 11 (setting 2, f=0.2, α=0.02) versus non-hierarchical setup

c) Optimum hierarchical setup for Case 6 (setting 1, f=0.8, α=0.05) versus non-hierarchical setup

d) Optimum hierarchical setup for Case 14 (setting 2, f=0.8, α=0.02) versus non-hierarchical setup
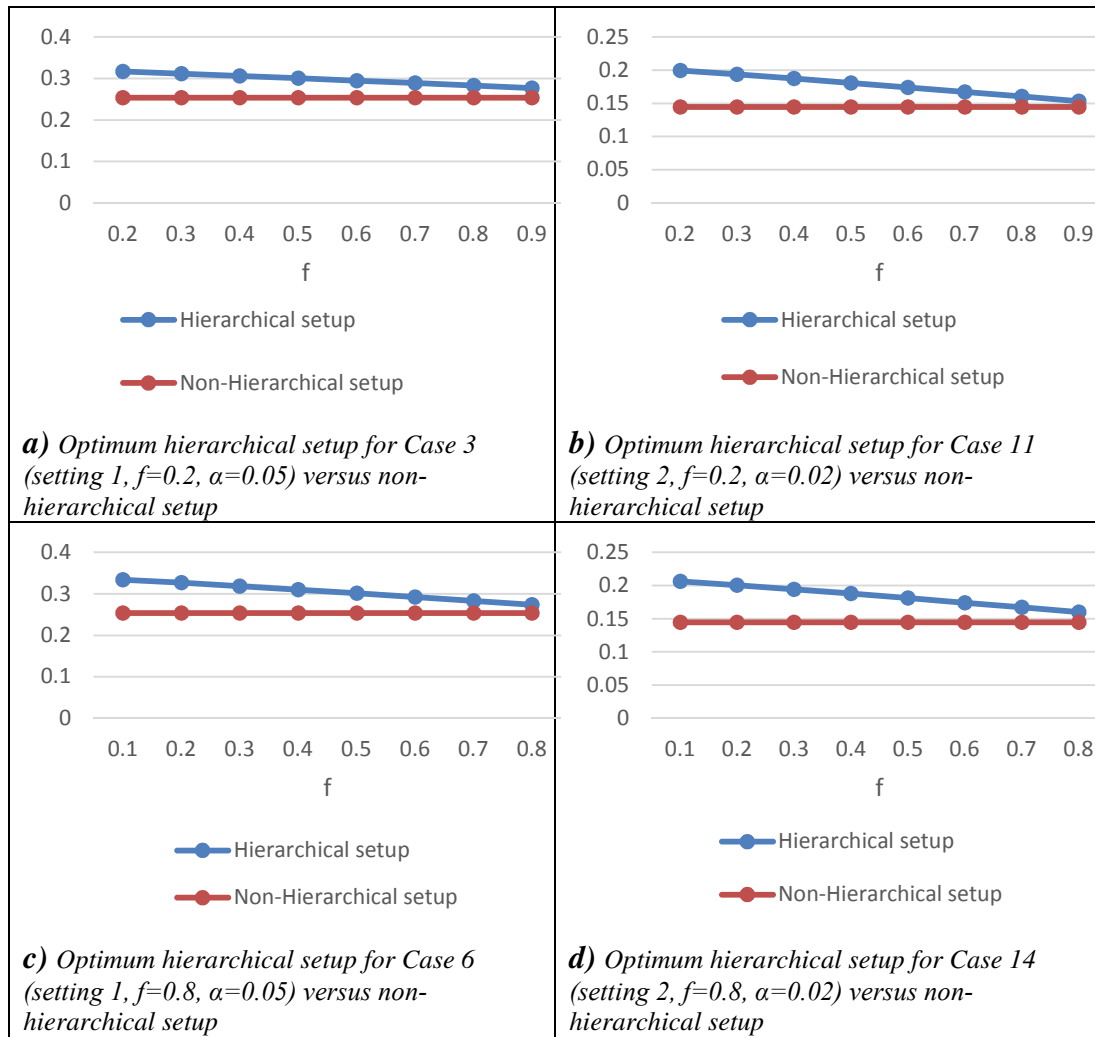
Figures 7.15 (a – d): Objective function values for varying f under the optimum hierarchical setups and the non-hierarchical setups – Cases with dominant transportation cost

As opposed to the cases where the transportation cost dominates, for the cases where the inventory cost dominates, the hierarchical setups performs better than the non-hierarchical setup (Figures 7.16 (a) and (b) and 7.17 (a) – (d)). The higher the fraction of demand for the relaxed service (*1-f*) gets, the lower the service costs becomes under the hierarchical setups. Under the hierarchical setups, with the increase in the fraction of demand for the relaxed service, the supply gets more centralized through the higher level facilities, and in turn, the inventory cost, which is dominant in these cases, reduces. On the other hand, the non-hierarchical setup, meeting both types of demand

from the nearest facility, does not avail the opportunity to increase the level of centralization when the fraction of demand for the relaxed service increases.

a): Setting 1 (*d$_{maxR}$ = 0.48, d$_{maxS}$ = 0.24*) with α = 8. f=0.2 (Case 19), f=0.4 (Case 20), f=0.6 (Case 21), f=0.8 (Case 22)

b): Setting 2 (*d$_{maxR}$ = 0.24, d$_{maxS}$ = 0.12*) with α = 2. f=0.2 (Case 27), f=0.4 (Case 28), f=0.6 (Case 29), f=0.8 (Case 30)
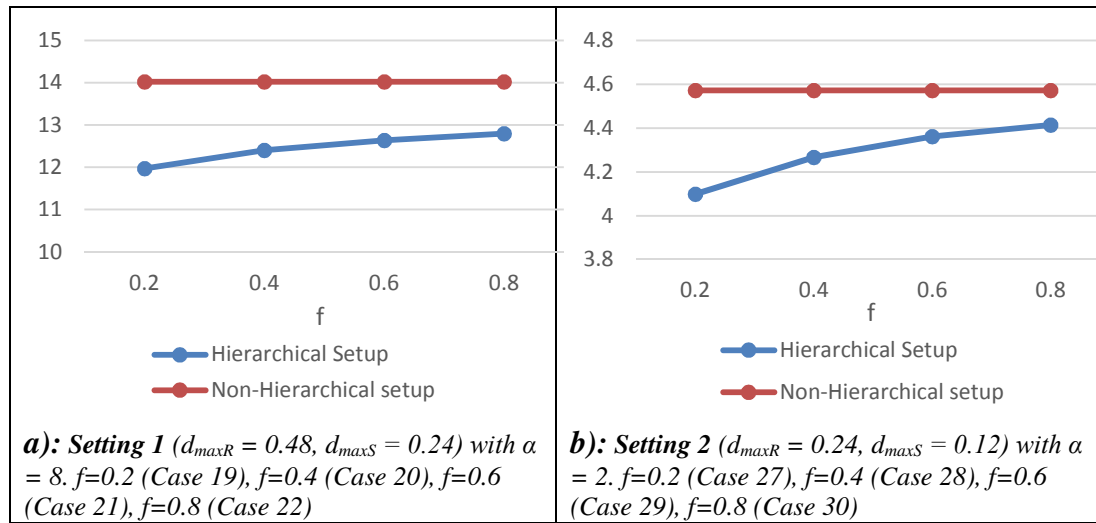
Figures 7.16 (a & b): Objective function values based on the optimum hierarchical setups for the cases with dominant inventory cost versus the objective function values for the non-hierarchical setup

a) *Optimum hierarchical setup for Case 19 (setting 1, f=0.2, α=8) versus non-hierarchical setup*

b) *Optimum hierarchical setup for Case 27 (setting 2, f=0.2, α=2) versus non-hierarchical setup*

c) *Optimum hierarchical setup for Case 22 (setting 1, f=0.8, α=8) versus non-hierarchical setup*

d) *Optimum hierarchical setup for Case 30 (setting 2, f=0.8, α=2) versus non-hierarchical setup*
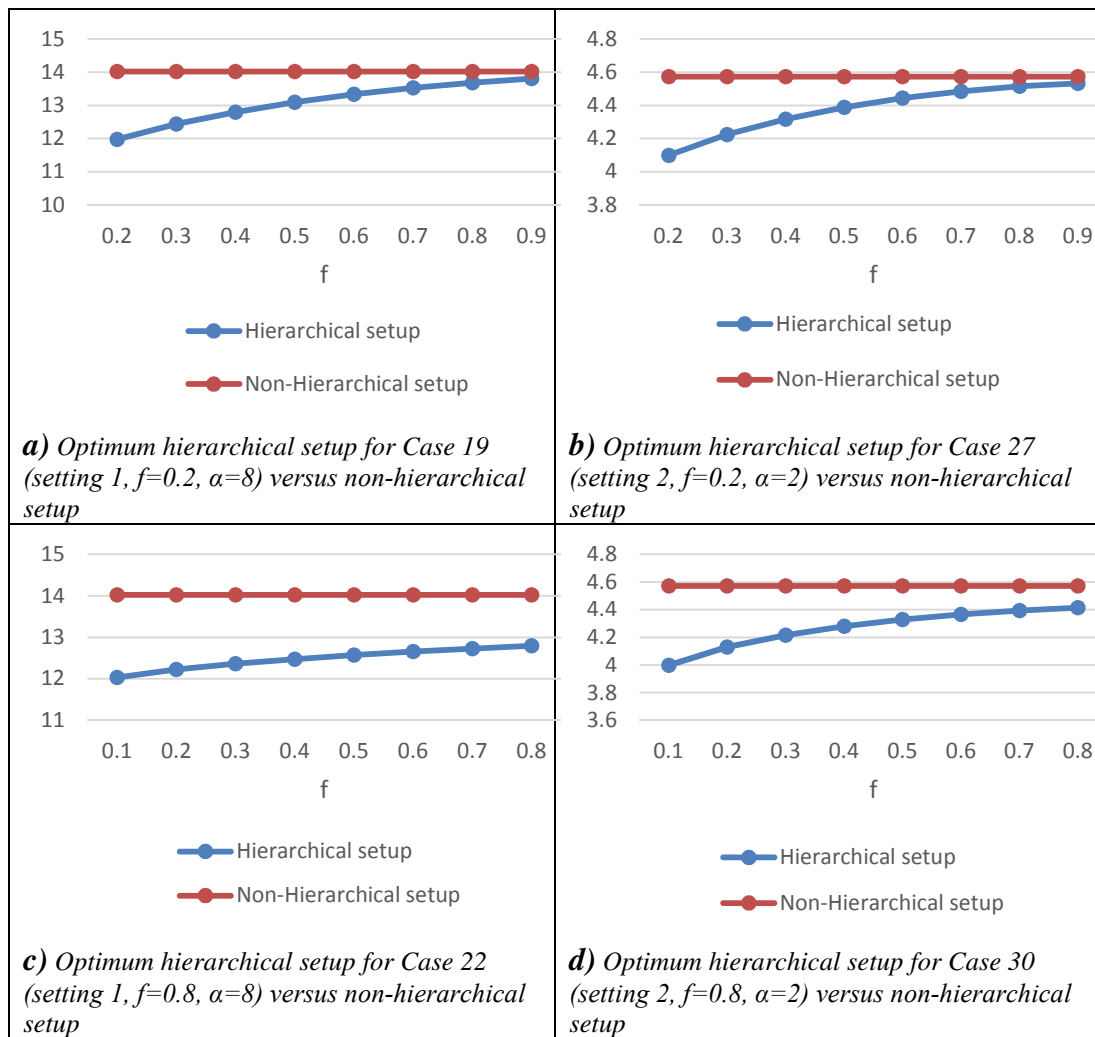
Figures 7.17 (a – d): Objective function values for varying f under the optimum hierarchical setups and the non-hierarchical setups – Cases with dominant inventory cost

7.6. Summary and Conclusion

The chapter looks into the impact of the inventory and transportation costs and the fractions of demand for two time-based service types, one having a shorter service time window, on the optimum hierarchical location setup with minimum number of facilities. The investigation starts with the analysis of the behaviour of service cost as a combination of the inventory cost, possessing economy of scale characteristics, and the transportation cost, possessing diseconomy of scale characteristics, in relation to a service area (zone) size. Two distinct location patterns considering a distance constrained service are identified. One of the patterns, which favours transportation

230

cost reduction, comprises of service zones of equal length, while in the other pattern, favouring inventory cost reduction, all zones but one are of maximum allowable length whereas the remaining zone is of a smaller length. The optimum solution switches between these two patterns depending on the comparative values of inventory and transportation cost constants.

A MINLP model is then presented that determines the optimum hierarchical location setup given the distance constraints for two time-based service types, the fractions of total demand for the service types, and the minimum number of facilities of each type. The solutions generated by the MINLP model can be loosely linked to the optimum patterns for a system with single time based service. When the transportation cost is dominant, the service zones tend to be equal and facility points tend to be in the middle of their zones. In the cases where the increase in the demand fraction for the relaxed service changes the optimum solution, facility points tend to move away from the centres of the strict service zones to let the higher level facilities move towards the centre of the relaxed service zones. On the contrary, when the inventory cost is dominant, resulting in a solution where one facility has a comparatively smaller service zone, if the optimum solution changes with the higher fraction of demand for the relaxed service (strict service), the smaller zone for the relaxed service (strict service) becomes even smaller. Also, with dominant inventory cost, in most cases, not all facilities cover the demand that is closer to them rather than to an adjacent facility, i.e. the Voronoi assumption does not fit well in determining the service areas. There are even cases where in the optimum hierarchical setup a facility serves only on one side, i.e. from a boundary of its zone.

In the cases tested for dominant transportation cost, when there is no restriction on the total number of facilities that can provide the relaxed service, the optimum setup becomes non-hierarchical, in which each facility serves both the relaxed and the strict services in the proximities. However, in cases where the inventory cost is dominant, relaxing the hierarchy restriction still gives hierarchical setup solutions, which are more centralized compared to the non-hierarchical system and hence better for inventory reduction. Note that if the strict service time commitments are for a specific time of day, as in the case of Highways Agency, the hierarchical setup can have an added benefit. If the system is setup as hierarchical, the lower level facilities (providing only the strict service) can be closed during the times when the strict service is not required.

The work in this chapter indicates that setting up an optimum location pattern for time-differentiated distribution involves delicate balances with respect to different costs and demand profiles. A small change in the costs or the fractions of demand for different service types can radically change the optimum facility locations and service area configuration. Counterintuitive placements, such as serving the demand from an extreme location point of the service area, allocating demand to a distant facility point rather than to a closer facility providing the required service type, and setting up a system with uniform capability to meet different types of demand can be better options in certain cases. This is a complex area that needs further exploration.

Chapter 8: Conclusions and further research directions

8.1. Conclusions

In this research, we address some of the distribution cost issues in context of time-based service differentiation. The analysis considers a system that offers supply of a required item at demand locations within two different service time windows. The factors explored in the work include service time/distance constraints, fractions of demand for different time-based service types, inventory and transportation costs, inventory sharing, service availability levels, and organizations of service facilities and their service zones. These factors are explored under hierarchical and non-hierarchical setups. Under the hierarchical setup, selected facilities (i.e. higher level facilities) provide service within the longer time window so that the longer service time (and distance) can be exploited to consolidate demand at higher level facilities and meet demand in a more centralized fashion. While under the non-hierarchical setup, service is provided in a uniform fashion such that all facilities cater for services within the shorter as well as the longer time windows in their vicinities, hence making the system completely decentralized.

This section provides the research conclusions in view of some empirical findings and the following questions that define the main objectives of this research:

What is the impact of the service time window lengths associated with different time-based service types on inventory and transportation costs?

What is the impact of the demand fractions associated with different time-based service types on inventory and transportation costs?

What is the impact of inventory sharing with varying fractions of demand for different time-based service types on transportation and service availability levels?

What is the impact of transportation and inventory costs and the demand fractions for different time-based service types on the optimum facility locations and service zones?

Impact of the service time window lengths and the demand fractions associated with different time-based service types on inventory and transportation costs*:* The analysis in Chapter 5 and 6 suggests that, though improving the service time capability can have significant impact on the setup cost, once a system has been designed, an increase in the percentage of demand for the service within the shorter time window does not necessarily mean that the cost of service provision becomes higher. The setup cost as well might not be high for service providers due to the outsourcing of storage. Besides, though normally inventory increases as the number of facilities in a system increases for covering demand in a shorter time, the average travelling to reach demand locations reduces.

The analysis, considering the EOQ based (R, Q) and the (S-1, S) inventory policies, shows how, unlike under the non-hierarchical setup, a system under the hierarchical setup reacts to the changes in the proportions of demand for different time-based service types. Considering the EOQ model, when demand for the service within the longer time window gets comparatively higher, the hierarchical setup can result in lower inventory levels. However the higher transportation cost can offset this

reduction and make the non-hierarchical setup more feasible, especially when transportation cost are significant as reported by the case companies.

Under the (S-1, S) inventory policy (employed by the SPL case companies), there is an indication that, even though demand under the hierarchical setup is more centralized than under the non-hierarchical setup, the required inventory can be lower under the non-hierarchical setup in some cases. Hence, as the transportation costs is also lower under the non-hierarchical setup, in these cases it is more reasonable to serve to all customers under the non-hierarchical system, i.e. in a uniform fashion. Nevertheless, in several cases, when the fraction of demand for the service within the longer time window is significantly high, the hierarchical setup does lower the inventory levels. Demand for the service options offering longer service time windows is high at the SPL case companies, hence there can be an opportunity to increase the level of centralization and reduce the inventory level through a hierarchical setup. Where the hierarchical setup reduces inventory, it might be beneficial to deploy the stocks with high inventory related costs in a hierarchical fashion while deploying the stocks with low inventory related costs in a non-hierarchical fashion. Significant saving can be achieved when the entire demand for an item is consolidated at one facility. Both SPL case companies have indicated that exceptionally expensive parts are only maintained centrally. If a part is provided from all facilities in the system, even if most of the demand is consolidated at one facility and the remaining demand at other facilities is very low, the system will still be required to maintain some stocks at every facility. This can be costly if the part is expensive as the number of facilities is normally high in SPL systems.

Impact of inventory sharing with varying fractions of demand for different time-based service types on transportation and service availability levels: Chapter 6 presents a simulation study considering the (S-1, S) inventory policy to compare inventory sharing and non-sharing configurations under varying fractions of demand for the two time-based service types. The comparison is based on the fill-rate and average service distance performance measures under the hierarchical and non-hierarchical setups.

Under the non-hierarchical setup, which is favourable in terms of transportation cost reduction and is deployed by the case companies, inventory sharing results in a higher service availability level (which may allow stock reductions) and a higher average service distance compared to when inventory is not shared. Unlike in the absence of inventory sharing, a system under the non-hierarchical setup with sharing mechanism reacts to the changes in the demand fractions for different time-based service types. An increase in the fraction of demand for the service in the longer time window increases the inventory sharing opportunity. This in turn increases the service availability level along with the average distance to serve a demand. Hence a consideration should be given to the inventory and transportation cost trade-off in deciding whether the inventory sharing mechanism is beneficial or not.

On the other hand, under the hierarchical setup, which can favour inventory level reduction by providing more centralization in the system, sharing without hierarchical restriction as such does not reveal a trade-off between the service availability level and the average service distance performance. Under the hierarchical setup, transshipment from the nearest facility in the range with positive stocks in case of a stock-out can not only increase the service availability level, but can also reduce the average service

distance. The simulation study also investigates inventory sharing under a hierarchical restriction where for a request for the service within the longer time window a transshipment is only allowed from a higher level facility in the range. Inventory sharing with the hierarchical restriction does not exhibit any special benefit over inventory sharing without this restriction.

Impact of transportation and inventory costs and the demand fractions for different time-based service types on optimum service zones: Chapter 7 looks into the impact of inventory and transportation costs and the demand fractions for two different time-based service types on the optimum hierarchical location with minimum number of facilities. The investigation is done through optimization models considering that customers are uniformly spread along a line segment or one route.

Considering a time (distance) constraint, two distinct types of service zone patterns are observed depending on whether transportation or inventory cost dominates. These patterns suggest that the optimum facility locations and service area zoning can be significantly different for situations where transportation cost is dominant and where inventory cost is dominant. When transportation cost dominates, it can be more efficient to have all service zones with similar sizes. Whereas, when inventory cost is dominant, it can be more efficient to have all but one facilities with the maximum service areas in order to attain an economy of scale.

With multiple distance constraints, the optimum hierarchical facility locations and service zones can in addition depend on the demand fractions for different time-based service types. When the inventory cost is dominant, the better organization of facility locations and service zones can be the ones where economy of scale is achieved for

the service type which has the higher fraction in overall demand. The analysis confirms that the hierarchical setup might not be optimum when transportation cost dominates. In this situation, it can be better to setup a non-hierarchical setup providing the maximum decentralization.

The study shows that determining facility locations and service areas is a very complex problem in the context of time-differentiated distribution. There are several factors on which the optimum organization of facilities and service areas depend, and balancing these factors can be challenging. There can be situations in which it can be better to serve demand from an extreme location point of a service area (i.e. from a service area boundary), or allocate demand to a distant facility point rather than to a closer facility providing the required service type, or set up a system with uniform capability to meet different types of demand.

Limitations: It is important to emphasize the limitations of our research approach. Many of the assumptions in this work are not strictly appropriate in real life settings. For example, in many cases, customers do not have a uniform geographical distribution in an area and there are several factors affecting the possible facility locations. Besides, there can be many capacity constraints in service operations. Also, many real world distribution systems, especially SPL systems, are complex and include multiple echelons. We clearly cannot claim that our models, based on stylized systems, can be applied directly to a real world system. However, the analysis here is extensive and provides important generic insights which are unlikely to be generated through studying real life setups or instances. The insights can be useful in understanding likely impacts of different scenarios in time-differentiated distribution and can aid in decision making.

8.2. Directions for further research

Some of the related issues that deserve further investigation are summarised below:

- This research mainly focuses on inventory and transportation costs. Considering other costs and resources such as the number of service engineers and vehicles, can give further insights.

- The simulation study does not as such focus on the adjustment of base stock levels when the service level increases due to inventory sharing. Extending the study to consider the base stock level adjustment can be more insightful with regards to the impact of inventory sharing on service costs.

- Inventory sharing is a popular theme of research, however, sharing of engineers and other resources by different facilities/regions, which can be impacted by the allowed service time, has not received much attention.

- Serving multiple customers in one trip is a realistic consideration. Longer service time commitments can increase the opportunity for serving multiple customers in a trip and can possibly reduce the required resources for satisfying demand.

- We have only considered time-based service differentiation. Incorporating service availability level differentiation in the investigation can also be insightful.

- Very few location studies in the current literature address continuous customer location problems. Our multi-objective investigation considering unidimensional demand spread is of an exploratory nature and should motivate further work for tackling more than one objectives. The following facility

location problems can be explored considering continuous geographical distribution of customers in two dimensions:

- o Minimizing average distance subject to a distance constraint

- o Multi-level (hierarchical) covering where there are more than one distance constraints

- o Minimizing the sum of transportation and inventory costs subject to a distance constraint

Our optimization model can potentially be used as a part of a constructive heuristic for a related solution in a two dimensional area.

- It can also be interesting to study the design of service zones considering resource sharing, i.e. to find the optimal service areas when sharing is possible. A higher level of centralization, where service areas are large and facilities are far apart, can be beneficial for inventory costs, considering that inventory is not shared. When inventory is shared, it might be beneficial to locate facilities closer to each other and have more overlapping of facility service ranges.

Appendix 1 – Interview questions

ICT Cases

Range of services/service contracts

- What service time contracts for IT (hardware) support services are being offered to the customers?

- What is the scope of activities within the contracted service times?

- What is the approximate percentage of calls for each response time option?

- Can there be different service time commitments for different types of equipment with one customer?

Network structure and capabilities of supplying service parts

- How many service part stocking facilities are there in the UK?

- How many echelons are there in the service parts distribution network?

- How many stocking facilities are there in each echelon?

- Is the number of service part stocking facilities sufficient for meeting customers' requirements?

- Does the central warehouse also serve the customers directly?

- What are the considerations in deciding the location of service parts stocking facilities?

- What is the response time capability to reach UK customers?

- Do all of the warehouses provide full range of service response times?

Procedure for satisfying service requests

- How are the service calls received and logged?

- What is the criteria to select service part stocking facility to supply the part(s) in response to a service call?

- Do customers mostly perform diagnosis themselves to determine which part to order?

- Is the servicing at customer site being outsourced to third party service engineers who provide service on your behalf?

- Is the transportation function for supplying the parts outsourced?

Service parts characteristics

- What is the number of Stock Keeping Units (SKUs)?

- What is the range of lifecycle durations of products under service contracts?

- What is the proportion of repairable parts and consumable parts?

- Do customers always receive new service parts?

- How are the obsolete service parts in the inventory dealt with (e.g. they are scrapped, resold etc.)?

- What is the proportion of fast moving, normal moving and slow moving parts?

- Is it likely that other major companies in this sector hold similar service parts in their inventories?

Demand/customer characteristics

- Can the location of customers be classified as clustered (i.e. customers are more concentrated in fewer places)?

- Is the demand at certain stocking facilities significantly higher compared to others?

- On average how many service calls are received per month?

- What is the approximate percentage of calls for each response time option?

- Is supplying only one part per service request common?

- What is the installed base (number of units)?

- What is the total number of customers?

Inventory policies and stocking rules

- Do different parts have different availability levels (e.g. 95% availability level for some parts while 97% for others)?

- Are there different part availability levels set for different customers/equipment? If yes - what are the different part availability options being offered?

- What software package, if any, is being used to manage the inventory levels?

- What inventory control policies are being employed for inventory management/inventory replenishment?

- Is there central visibility of stock levels at all stocking points?

- Can you explain replenishment procedures at local warehouses (e.g. automatic replenishment requests are generated by local warehouses)?

- Approximately what is the proportion of stocks being maintained/deployed centrally?

- Are expensive parts only stored centrally?

Sourcing of service parts

- What is the average replenishment lead time from suppliers for most products (excluding products with exceptionally long lead times)?

- From where are service parts sourced?

Service cost characteristics/cost structure

- Can the service business be considered as a high revenue/profit generating function of the organization rather than just a support function which is necessary to be provided to the customers?

- What is the cost of delivering part(s) to the customers on same day?

- What are the percentages of different costs constituting overall service costs?

243

- What is the average cost of service parts, cost of most expensive service part, and cost of cheapest service part?

Issues in managing service parts logistics

- What are the system vulnerabilities and the key issues in managing Service Parts Logistics effectively?

Management trends

- What are the new trends in managing Service Parts Logistics?

- Where do you think there are opportunities for improvement in service operations?

The Highways Agency

- Can you describe scope and scale of Highways Agency's service delivery operations?

Range of services

- Are there different response times for different types of incidents?

- What are the response time windows?

- Are the response time windows just to reach at the site of an incident?

Network structure and capabilities

- How many stocking facilities are there in your region?

- Is there any central warehouse?

- What are the considerations in deciding the location of service/stocking facilities?

- Do all of the warehouses provide full range of service response times?

- Do all facilities store same kind of material?

Procedure for satisfying service requests

- How are the service calls received and logged?

- How are the contractors coordinated?

- What is the criteria to select service/stocking facility to respond to a service call?

- Is the servicing being outsourced to third party service providers?

Inventory and demand characteristics

- What is the number of different materials maintained in the inventory?

- Can you give examples of some materials used in the repair as part of incident response?

- Do the stocks become obsolete in any case and is it a major concern?

- How is the obsolete inventory dealt with?

- Can the demand locations be classified as clustered?

- Is the demand at certain facilities significantly higher compared to others?

- On average how many service calls are received per day?

- Is the demand predictable?

Inventory policies and sourcing of material

- Is there any target 'on-time service availability' level?

- What inventory control policies are being employed for inventory management/inventory replenishment?

- Do you know the stocking levels at all times?

- Can you explain replenishment procedures at the warehouses?

Issues in managing service parts logistics

- Can you highlight the system vulnerabilities and the issues?

Management trends

- Where do you think are the opportunities for improvement in the service operations?

Appendix 2 – Expected distance to serve uniformly distributed customers within a hexagonal catchment area

A hexagonal catchment area with an edge length of s can be divided into six equilateral triangles each having an edge length s (*Figure A2-1*). Hence the expected distance to serve a customer from the centre of the hexagonal catchment area, and the distance between a vertex of an equilateral triangle in the hexagon and a random point inside the triangle can be considered as same.
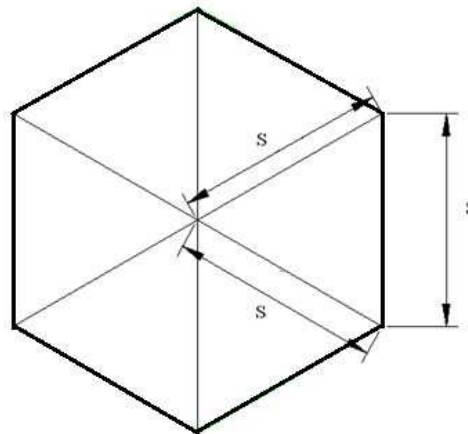
Figure A2-1: Facility catchment area (with 's' as the maximum distance that can be travelled from the facility (located at the centre))

To compute the expected distance to serve a customer:

Let O (0,0) be the vertex where the service facility is located in a Cartesian coordinates system. Then the distance between the service facility and another vertex is s (where s is equal to the length of an edge of the hexagon and the equilateral triangles forming the catchment area of the facility).

A equilateral triangle in the hexagon can be further divided into two right angle triangles of equal dimensions. The resulting right angle triangles are exhibited in *Figure A2-2*. Let *R* represent one of these triangles.
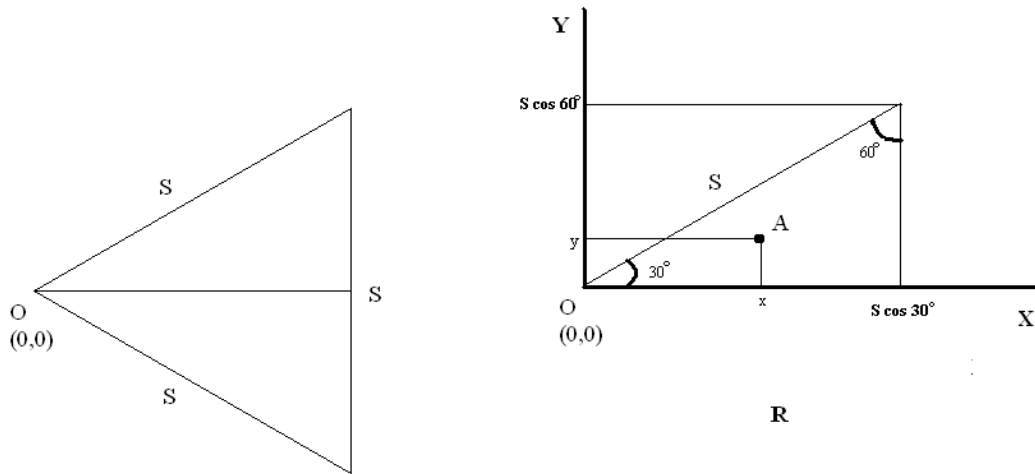
Figure A2-2: Average distance in an equilateral triangle and hexagon with edge length s

Let *A* (x, y) be a random point inside the right angle triangle *R*. Then, the vector *OA* represents the expected distance from the facility point (located at *O*) to a randomly selected point inside R, and the expected distance from the facility point to a point inside the equilateral triangle and the hexagon as well.

Considering the Pythagorean theorem, the distance between *A*, having coordinates *x* and *y*, and *O* is $\sqrt{(x-0)^2 + (y-0)^2}$.

The probability of the occurrence of point *A* is given by:

$$Prob.(x,y) = \frac{1}{Total\ Area} = \frac{1}{\frac{s\cos30^0 \times s\cos60^0}{2}} = \frac{2}{s^2\cos30^0\cos60^0}$$

As the coordinates (x, y) can take values between the coordinates of the vertices of R,

the following integral can be formulated to compute the expected distance:

$$Expected\ distance = \int_0^{s\cos30^0} \int_0^{x\tan30^0} Prob.(x,y)\sqrt{x^2+y^2}\ dx\ dy$$

$$Expected\ distance = \frac{2}{s^2\cos30^0\cos60^0} \int_0^{s\cos30^0} \int_0^{x\tan30^0} \sqrt{x^2+y^2}\ dx\ dy$$

$$as \int \sqrt{x^2+y^2}\ dy = \frac{1}{2}y\sqrt{x^2+y^2} + \frac{1}{2}x^2\ln\left(y+\sqrt{x^2+y^2}\right) + C$$

$$Expected\ distance = \frac{2}{s^2\cos30^0\cos60^0} \int_0^{s\cos30^0} \left[\frac{1}{2}y\sqrt{x^2+y^2} + \frac{1}{2}x^2\ln\left(y+\sqrt{x^2+y^2}\right)\right]_0^{xtan30^0} dx$$

Substituting $\tan 30^0$ by $\frac{1}{\sqrt{3}}$ and solving the integral further gives us:

$$Expected\ distance = \left(\frac{1}{3} + \frac{\ln(3)}{4}\right)s = \mathbf{0.60799\ s}$$

Appendix 3 – Model formulation considering more than two service time options under the hierarchical setup

The formulation presented in Section 5.3.2 can be extended to consider more than two levels under the hierarchical setup (providing more than two service time options). For a demonstration, cyclic inventory and transportation cost functions are formulated below for such system.

Let Type k service be a stricter service than all Type $k - i$ services, and let a Type k service facility be a lower level service facility than all Type $k - i$ service facilities (where $2 \leq k \leq m,\ 1 \leq i < k$ and m is the total number of service types and types of service facilities). In addition,

Let:

$A =$ total area to be served (a large geographical area)

$\lambda =$ total demand in the area A, i.e. the total number of service calls per unit time

$C_o =$ setup cost per inventory replenishment order

$C_h =$ inventory holding cost per unit per unit time

$C_t =$ travel cost per unit distance

$f_1 =$ fraction of the total demand for Type 1 (most relaxed) service calls

$f_k =$ fraction of the total demand for Type k service calls. Where $2 \leq k \leq m$. such that $\sum_{i=1}^{m} f_i = 1$

$s_1 =$ maximum distance that can be covered from a service facility to provide the Type *1* service = the edge length of a hexagonal catchment area for Type *1* service

$s_k =$ maximum distance that can be covered from the service facility to provide Type *k* service = the edge length of a hexagonal catchment area for Type *k* service

$n_1 =$ number of service facilities providing Type *1* service $= \dfrac{A}{2.5981 s_1{}^2}$, where *2.5981(s₁²)* $2.5981(s_1{}^2)$ is the hexagonal catchment area (with an edge length of *s₁*) of a service facility for Type *1* service provision

$n_k =$ number of service facilities providing Type *k* service $= \dfrac{A}{2.5981(s_k{}^2)}$, where $2 \leq k \leq m$ and $2.5981(s_k{}^2)$ is the hexagonal catchment area (with an edge length of *s_k*) of a service facility for Type *k* service provision.

The ratio between n_1 and n_k can be stated as:

$$\frac{n_1}{n_k} = \frac{\frac{A}{2.5981 s_1{}^2}}{\frac{A}{2.5981 s_k{}^2}} = \frac{s_k{}^2}{s_1{}^2}$$

Considering that a Type *1* service facility provides all types of service while a Type *k* service facility provides Type *k* service and all the other service types that are stricter than Type *k* service, the number of Type *1* and Type *k* service centres are determined as,

Number of Type *1* service facilities = n_1.

Number of Type *k* service facilities = $n_k - n_{k-1}$.

Where 'n_{k-1}' is the number of service facilities providing Type $k-1$ service.

For instance, *Table A3-1* and *A3-2* show the classification of service facilities assuming a successively inclusive hierarchical system providing four service types through four types of service facilities. Here, Type *1* service is the most relaxed service, Type 2 service is stricter than Type *1* service, Type *3* service is stricter than Type *2* service, and Type *4* service is the strictest service. In terms of facility levels, Type *1* facilities are the highest level facilities, Type *2* facilities are lower level facilities than Type *1* facilities, Type *3* facilities are lower level facilities than Type *2* facilities, and Type *4* facilities are the lowest level facilities.

Table A3-1: Classification of service facilities in four-level nested-hierarchical system

	Type 1 service	Type 2 service	Type 3 service	Type 4 service
Type 1 service facilities provide: →	✓	✓	✓	✓
Type 2 service facilities provide: →		✓	✓	✓
Type 3 service facilities provide: →			✓	✓
Type 4 service facilities provide: →				✓

Table A3-2: Number of service facilities for each service type in four-level nested-hierarchical system

	Type 1 service facilities	Type 2 service facilities	Type 3 service facilities	Type 4 service facilities
Number of service facilities providing Type 1 service, i.e. n_1 include: →	✓			
Number of service facilities providing Type 2 service, i.e. n_2 include: →	✓	✓		
Number of service facilities providing Type 3 service, i.e. n_3 include: →	✓	✓	✓	
Number of service facilities providing Type 4 service, i.e. n_4 include: →	✓	✓	✓	✓

In order to determine the number of Type *4* service facilities (i.e. $k = 4$), the number of service facilities providing Type *3* service is subtracted from the number of service facilities providing Type *4* service:

Number of Type *4* facilities $= n_4 - n_3$.

Where, n_3 is the sum of the numbers of Type *1*, Type *2* and Type *3* service facilities.

Reverting back to the example of the distribution network providing four types of services in a successively inclusive hierarchical system, the total demand can be considered as $\lambda = f_1(\lambda) + f_2(\lambda) + f_3(\lambda) + f_4(\lambda)$. Where $f_1(\lambda), f_2(\lambda), f_3(\lambda)$ and $f_4(\lambda)$ represent demand for Type *1*, Type *2*, Type *3* and Type *4* services respectively, such that $f_1+f_2+f_3+f_4=1$. Considering the above classification of service facilities, *Table A3-3* shows the expressions for the demand served by the different types of service facilities.

Table A3-3: Demand served by different types of service facilities in the four-level nested-hierarchical system

	Type 1 service demand		Type 2 service demand		Type 3 service demand		Type 4 service demand
Total demand served by Type 1 service facilities =	$f_1\lambda$	$+$	$\frac{n_1}{n_2}f_2\lambda$	$+$	$\frac{n_1}{n_3}f_3\lambda$	$+$	$\frac{n_1}{n_4}f_4\lambda$
Total demand served by Type 2 service facilities =			$\frac{n_2-n_1}{n_2}f_2\lambda$	$+$	$\frac{n_2-n_1}{n_3}f_3\lambda$ $+$		$\frac{n_2-n_1}{n_4}f_4\lambda$
Total demand served by Type 3 service facilities =					$\frac{n_3-n_2}{n_3}f_3\lambda$	$+$	$\frac{n_3-n_2}{n_4}f_4\lambda$
Total demand served by Type 4 service facilities =							$\frac{n_4-n_3}{n_4}f_4\lambda$
Total demand =	$f_1.\,\lambda$	$+$	$f_2.\,\lambda$	$+$	$f_3.\,\lambda$	$+$	$f_4.\,\lambda$

The demand served by Type *1* service facilities is equal to the sum of entire demand for Type *1* service and the part of demand for the services stricter than the Type *1* service.

$$Total\ demand\ served\ by\ Type\ 1\ service\ facilities = \sum_{i=1}^{m} \frac{n_1}{n_i} f_i \lambda \qquad (A3.1)$$

Where $i = 1, 2 \dots m$ and m is the total number of service and service facility types.

$$Total\ demand\ served\ by\ one\ Type\ 1\ service\ facility = \frac{1}{n_1} \sum_{i=1}^{m} \frac{n_1}{n_i} f_i \lambda$$

$$Total\ demand\ served\ by\ one\ Type\ 1\ service\ facility = \sum_{i=1}^{m} \frac{f_i}{n_i} \lambda \qquad (A3.2)$$

Similarly, the total demand that is served by Type k service facilities is equal to the sum of the fractional demand for Type k and the stricter services that are facilitated by Type k service facilities.

$$Total\ demand\ served\ by\ Type\ k\ service\ facilities =$$

$$\sum_{i=k}^{m} \frac{n_k - n_{k-1}}{n_i} f_i \lambda \qquad (A3.3)$$

Where $i = 1, 2 \dots m$ and m is the total number of service and service facility types.

$$Total\ demand\ served\ by\ one\ Type\ k\ service\ facility =$$

$$\frac{1}{(n_k - n_{k-1})} \sum_{i=k}^{m} \frac{n_k - n_{k-1}}{n_i} f_i \lambda$$

$$Total\ demand\ served\ by\ one\ Type\ k\ service\ facility = \sum_{i=k}^{m} \frac{f_i}{n_i} \lambda \qquad (A3.4)$$

Cyclic inventory cost in multilevel (>2) hierarchical organization

We know that the total cyclic inventory cost in a decentralized system assuming the EOQ model equals to $\sum_{i=1}^{n} \sqrt{2 C_h C_o \lambda_i}$, where λ_i is the demand served by stocking location i ($i=1, 2 \dots n$) and n is the total number of locations. Considering the total demand served by one Type *1* facility (A3.2), the total demand served by one Type *k*

facility (A3.4), the number of Type 1 facilities n_1, and the number of Type k facilities

$n_k - n_{k-1}$, we obtain

$$Total\ cyclic\ inventory\ cost = \left(n_1 \sqrt{2 C_h C_o \sum_{i=1}^{m} \frac{f_i}{n_i} \lambda} \right) + \sum_{k=2}^{m} \left((n_k - \right.$$

$$\left. n_{k-1}) \sqrt{2 C_h C_o \sum_{i=k}^{m} \frac{f_i}{n_i} \lambda} \right) = \sqrt{2 C_h C_o \lambda n_1} \left(\sqrt{\sum_{i=1}^{m} \frac{n_1 \cdot f_i}{n_i}} + \sum_{k=2}^{m} \frac{(n_k - n_{k-1})}{\sqrt{n_1}} \sqrt{\sum_{i=k}^{m} \frac{f_i}{n_i}} \right)$$

$$Multiplication\ factor\ (cyclic\ inventoy\ cost) = \left(\sqrt{\sum_{i=1}^{m} \frac{n_1 \cdot f_i}{n_i}} + \right.$$

$$\left. \sum_{k=2}^{m} \frac{(n_k - n_{k-1})}{\sqrt{n_1}} \sqrt{\sum_{i=k}^{m} \frac{f_i}{n_i}} \right)$$

Transportation cost in multilevel (>2) hierarchical organization

Considering the average distance inside a hexagonal area as *0.60799(s)* from its centre

point, where '*s*' is the length of edges of the hexagon,

$$Total\ transportation\ cost = C_t(0.60799) \sum_{i=1}^{m} f_i \lambda s_i = C_t \lambda (0.60799) s_i \left(f_1 + \right.$$

$$\left. \sum_{i=2}^{m} f_i \frac{s_i}{s_1} \right)$$

$$Multiplication\ factor\ (transportation\ cost) = \left(f_1 + \sum_{i=2}^{m} f_i \sqrt{\frac{n_1}{n_i}} \right)$$

Appendix 4 – Distances between warehouses (Company B)

Table A4-1: From to chart - Distance between warehouses in miles

	Ed	Man	Bar	MK	WL	Abr	Bir	Nc	Bri	St	Leds	Cam	Gla	Not	Ply
Ed		217	450	356	407	123	289	115	375	432	211	333	34	270	487
Man	217		246	152	203	346	85	147	171	228	50	176	205	94	283
Bar	450	246		96	58	588	174	305	166	108	226	87	447	171	279
MK	356	152	96		48.3	485	75.2	228	153	114	144	42.4	344	77.4	266
WL	407	203	58	48.3		536	122	274	120	76.5	190	69.2	395	123	233
Abr	123	346	588	485	536		417	237	503	560	334	455	148	393	615
Bir	289	85	174	75.2	122	417		209	89	148	124	99.3	276	58.1	201
Nc	115	147	305	228	274	237	209		291	318	97.5	219	145	157	404
Bri	375	171	166	153	120	503	89	291		111	206	178	362	140	118
St	432	228	108	114	76.5	560	148	318	111		234	135	419	168	149
Leds	211	50	226	144	190	334	124	97.5	206	234		141	215	72.5	319
Cam	333	176	87	42.4	69.2	455	99.3	219	178	135	141		336	86.5	289
Gla	34	205	447	344	395	148	276	145	362	419	215	336		274	474
Not	270	94	171	77.4	123	393	58.1	157	140	168	72.5	86.5	274		253
Ply	487	283	279	266	233	615	201	404	118	149	319	289	474	253	

Key: Edinburgh (Ed), Manchester (Man), Barming (Bar), Milton Keynes (MK), West London (WL), Aberdeen (Abr), Birmingham (Bir), Newcastle (Nc), Bristol (Bri), Southampton (St), Leeds (leds), Cambridge (Cam), Glasgow (Gla), Nottingham (Not), Plymouth (Ply)

Appendix 5 – Computation of the average distance in a Voronoi diagram

Consider a right angled triangle laid out in a Cartesian system with one of its vertex at the origin (0, 0). The adjacent side to the vertex on origin overlaps with the X axis and has a length 'α', whereas the opposite side is parallel to Y axis and has a length 'β' (Figure A5-1). The average distance in this right angled triangle from its vertex at the origin can be determined through the following formula (Stone, 1991):

$$\frac{\alpha^2}{3\beta}\left(\left(\frac{\beta}{\alpha}\right)\sqrt{1 + \left(\frac{\beta}{\alpha}\right)^2} + \sinh^{-1}\left(\frac{\beta}{\alpha}\right)\right) \qquad \text{(A5-1)}$$
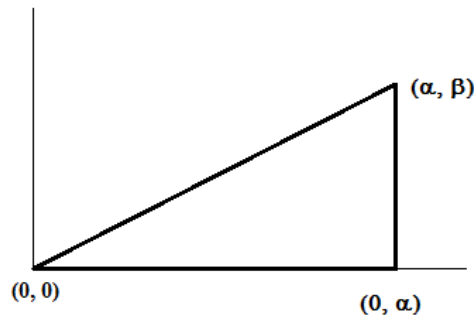
Figure A5-1

The average distance in a Voronoi diagram representing the service areas is computed based on the above formula (A5-1). For the computation, each Voronoi cell is taken in turn. Each cell is broken into triangles (Figure A5-2) and, depending on the form of the triangle, each triangle is further divided or extended into right angled triangles to compute the average distance (Figures A5-3 and A5-4). Weighted average distances of triangles (according to the percentage of triangle's size in the cell) are then summed up to determine the overall average distance in the cell. Finally the weighted average

distances of cells (according to the percentage of cell's size in the overall area) are summed to determine the overall average distance in the Voronoi diagram (average distance to reach a point in the overall area from the nearest facility point).

Adapting geometrical data structures and a divide-and-conquer algorithm to determine Voronoi cells from Laszlo (1996), the following procedure is programmed in C++ to compute the average distance in a Voronoi diagram:

Decompose each *Cell* in the *Rectangle* (overall area) into triangles as shown in Figure A5-2. The number of triangles is equal to the number of vertexes of the *Cell*.

Let *i* be the number of Voronoi *Cells* in the *Rectangle* and *j* be the number of vertexes (corner points) of a *Cell*. Let *A* be the facility point, *B* be the current vertex of the current Cell and *C* be the clockwise neighbour vertex of *B*.

Pseudocode:

```
RArea = Total area
For i
{
        Cell = current Voronoi cell
        CArea = area of Cell
        For j
        {
                Triangle T = Δ ABC
                TArea = area of T
                AvgDinT = average distance in T

                If (∠ B and ∠ C = 0)        (i.e. facility on the boundary of overall area, and the
                edge defined by current vertex and is clockwise neighbour overlaps the boundary)
                {
                        Average distance in T = 0
                }
                Else if (∠ ABC or ∠ ACB = 90º)
                {
                        AvgDinT = average distance in right angled triangle Δ ABC from A.
                }
                Else if (∠ ABC or ∠ ACB < 90º)
                {
                        Break Δ ABC into two right angled triangles Δ ADB and Δ ADC (as in
                Figure A5-3).
                        AvgDinT = weighted average of average distance in right angled triangles
                ΔADB and ΔADC.
                }
                Else if (∠ ABC or ∠ ACB > 90º)
```

{
    Extend ΔABC into two right angled triangles ΔADB and ΔADC (as in Figure A5-4).
    AvgDinT = weighted average distance in right angled ΔADB – weighted average
     distance in right angled ΔADC
}
Sum of weighted average distance in Triangles + = TArea/CArea × (AvgDinT)

   Advance current vertex of Cell clockwise
 }
 Sum of weighted average distance in Cells + = CArea/BArea × (Sum of weighted average distance in
  Triangles)

 Advance to next Cell

}
Average distance in Voronoi diagram Box = Sum of weighted average distance in Cells
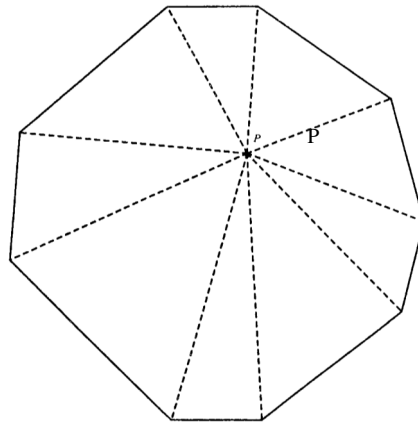
Figure A5-2: Decomposition of Voronoi cell into triangles for average distance calculation; where P is the facility point (from Stone, 1991).
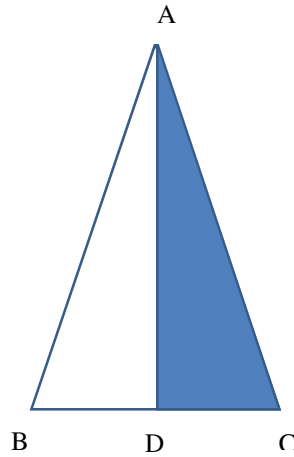
Figure A5-3: Decomposition of triangle (ABC) for average distance calculation when both base angles (at B and C) are less than 90º; where A the is facility point.
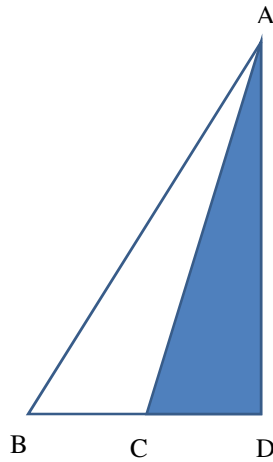
Figure A5-4: Extension of triangle (ABC) for average distance calculation when one of the base angles (at B or C) is more than 90º; where A is the facility point.

Appendix 6 – Simulation procedure for one-for-one (S-1, S) inventory policy

Below is a pseudocode for (S-1, S) inventory policy simulation, considering a Poisson demand process, to determine the fill-rate at a facility, given the demand rate, replenishment lead-time, and base stock level.

Table A6-1: (S-1, S) inventory policy simulation variables

Label	Description
MeanReqIntrArvlTm	Mean 'part request' inter arrival time – Stochastic element according to the exponential distribution
ReplenishmentTm	Replenishment time – mean time required for part to arrive after placing order
S	Base stock level in the inventory policy
i	Request number ($i = 1, 2, 3 \ldots$)
ReqArvlTm	Discrete event in the simulation
CS	Current stock level
PartArvlTm	Part arrival time –based on ReplenishmentTm
ServDeny	Counter for backorders
OrderLst	List of arrival times for parts in the pipeline for replenishments
SimTmLength	Simulation Time length
PWarmUpPrd	Warm-up period in simulation – certain percentage of SimTmLength
WarmUpCntr	Counter for requests within warm-up period
FillRt	Fill-rate: Percentage of demand fulfilled from stock on hand (for one simulation run)
Reps	Number of simulation runs
AvgFillRt	Average of fill-rates in all simulation runs

```
For number of Reps
{
        i = 1
        WarmUpCntr = 0
        CS = S
        While ReqArvlTm < SimTmLength
        {
                If (i = 1)
                {
                        ReqArvlTm = Random MeanReqIntrArvlTm
                        If (ReqArvlTm <= (PWarmUpPrd × SimTmLength)
                                { WarmUpCntr = 1 }
                        CS – = 1
                        PartArvlTm = ReqArvlTm + ReplenishmentTm
                        Initialize OrderLst with PartArvlTm
                }
                Else
                {
                        ReqArvlTm += Random MeanReqIntrArvlTm
                        If (ReqArvlTm <= (PWarmUpPrd × SimTmLength)
```

```
                            { WarmUpCntr += 1 }
                Check OrderLst from the start. Delete every node with PartArvlTm <
                ReqArvlTm and add 1 to CS for every node deleted

                If (CS < 1, and, ReqArvlTm > (PWarmUpPrd × SimTmLength))
                {
                        ServDeny += 1
                }
                CS – = 1
                PartArvlTm = ReqArvlTm + ReplenishmentTm
                Add PartArvlTm at end of OrderLst
        }
        FillRt = 1 – ((ServDeny) / ( i – WarmUpCntr))
        i += 1


    } // end of While
    Empty Orderlst
    AvgFillRt = (AvgFillRt + FillRt) / Reps
} // End of For
```

Appendix 7 – Simulation output

Notations

F#:	Facility number.
λ:	Local demand over lead time at a facility.
1-f:	fraction of demand for relaxed service
S:	Base stock level (computed through the procedure in Section 5.3).
Fr:	Fill-rate (availability level) (computed through the formula in Section 5.3) considering no inventory sharing
Fr_{N-S}:	Fill-rate determined through simulation considering non-sharing mechanism (Configuration 1).
Fr_{H-S}:	Fill-rate determined through simulation considering sharing with hierarchical restriction (Configuration 2).
Fr_{F-S}:	Fill-rate determined through simulation considering full sharing (Configuration 3).
AD:	Average distance to reach a demand location for service (computed numerically through the procedure explained in Section 6.5) considering no inventory sharing.
AD_{N-S}:	Average distance to reach a demand location for service determined through simulation without considering sharing mechanism (Configuration 1).
AD_{H-S}:	Average distance to reach a demand location for service determined through simulation considering sharing with hierarchical restriction (Configuration 2).
AD_{F-S}:	Average distance to reach a demand location for service determined through simulation considering full sharing (Configuration 3).

A7-1 - Case 1: Hierarchical

Fill-rates

Table A7-1: Fill-rates with configurations 1, 2 and 3 under hierarchical setup with (1-f) = 0.2 – Case 1

F#	λ	S	Fr	Fr$_{N-S}$	±*	Fr$_{H-S}$	±*	Fr$_{F-S}$	±*
1	0.56	3	0.98105	0.98180	0.00081	0.988244	0.000664	0.992277	0.00049
2	0.36	3	0.99425	0.99410	0.00099	0.996787	0.000781	0.996697	0.00064
3	0.56	3	0.98105	0.98122	0.00128	0.987508	0.000973	0.991902	0.00079
4	0.36	3	0.99425	0.99427	0.00141	0.99678	0.001053	0.996664	0.00089
5	0.36	3	0.99425	0.99412	0.00151	0.997632	0.001104	0.997567	0.00095
6	0.36	3	0.99425	0.99444	0.00158	0.997133	0.001147	0.996977	0.00101
7	0.56	3	0.98105	0.98129	0.00175	0.988416	0.001269	0.992193	0.0011
8	0.36	3	0.99425	0.99420	0.00183	0.996871	0.001313	0.99678	0.00116
9	0.56	3	0.98105	0.98073	0.00199	0.987507	0.001415	0.991634	0.00126

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-2: Fill-rates with configurations 1, 2 and 3 under hierarchical setup with (1-f) = 0.4 – Case 1

F#	λ	S	Fr	Fr$_{N-S}$	±*	Fr$_{H-S}$	±*	Fr$_{F-S}$	±*
1	0.67	3	0.96979	0.97051	0.00990	0.981559	0.00069	0.99203	0.00041
2	0.27	2	0.97018	0.97083	0.00164	0.983235	0.001192	0.982189	0.00106
3	0.67	3	0.96979	0.96983	0.00192	0.980959	0.001373	0.992045	0.00114
4	0.27	2	0.97018	0.97007	0.00239	0.98277	0.001783	0.981606	0.00166
5	0.27	2	0.97018	0.97015	0.00272	0.986766	0.001942	0.98618	0.00183
6	0.27	2	0.97018	0.97074	0.00300	0.983664	0.00223	0.982486	0.00214
7	0.67	3	0.96979	0.96971	0.00310	0.981564	0.002311	0.991845	0.00218
8	0.27	2	0.97018	0.97045	0.00340	0.982559	0.002511	0.98129	0.0024
9	0.67	3	0.96979	0.96938	0.00353	0.980996	0.002592	0.991767	0.00243

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-3: Fill-rates with configurations 1, 2 and 3 under hierarchical setup with (1-f) = 0.6 – Case 1

F#	λ	S	Fr	Fr$_{N-S}$	±*	Fr$_{H-S}$	±*	Fr$_{F-S}$	±*
1	0.78	3	0.95572	0.95616	0.00113	0.973747	0.0008	0.99361	0.00033
2	0.18	2	0.98595	0.98577	0.00166	0.991528	0.001165	0.989527	0.00094
3	0.78	3	0.95572	0.95573	0.00197	0.973289	0.0014	0.99344	0.001
4	0.18	2	0.98595	0.98581	0.00226	0.991465	0.001646	0.989113	0.0014
5	0.18	2	0.98595	0.98586	0.00248	0.993437	0.001772	0.992435	0.00157
6	0.18	2	0.98595	0.98682	0.00269	0.992249	0.001989	0.989841	0.00185
7	0.78	3	0.95572	0.95517	0.00289	0.973634	0.002103	0.993383	0.00189
8	0.18	2	0.98595	0.98666	0.00308	0.992135	0.002268	0.989922	0.00209
9	0.78	3	0.95572	0.95525	0.00324	0.973257	0.002391	0.993316	0.00212

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-4: Fill-rates with configurations 1, 2 and 3 under hierarchical setup with (1-f) = 0.8 – Case 1

F#	λ	S	Fr	Fr_N-S	±*	Fr_H-S	±*	Fr_F-S	±*
1	0.89	4	0.98708	0.98733	0.00063	0.992966	0.000416	0.9991	0.00011
2	0.09	2	0.99628	0.99643	0.00100	0.997974	0.000695	0.997187	0.00065
3	0.89	4	0.98708	0.98704	0.00110	0.992715	0.000785	0.99901	0.00066
4	0.09	2	0.99628	0.99616	0.00134	0.997986	0.000945	0.99764	0.00088
5	0.09	2	0.99628	0.99645	0.00151	0.9984	0.001063	0.998104	0.00103
6	0.09	2	0.99628	0.99728	0.00164	0.998237	0.001168	0.99761	0.00118
7	0.89	4	0.98708	0.98709	0.00173	0.992773	0.001219	0.999187	0.00118
8	0.09	2	0.99628	0.99614	0.00191	0.997676	0.001376	0.997218	0.00137
9	0.89	4	0.98708	0.98709	0.00200	0.992687	0.001442	0.999005	0.00137

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Average distances

Table A7-5: Average distance to reach a demand location for service with configurations 1, 2 and 3 under the hierarchical setup – Case 1

(1-f)	AD	AD_N-S	±*	AD_H-S	±*	AD_F-S	±*
0.2	0.145110	0.145113	0.000125	0.145436	0.000124902	0.145231	0.000125487
0.4	0.162687	0.162655	0.000145	0.163526	0.000146058	0.162886	0.000148333
0.6	0.180265	0.180338	0.000167	0.181691	0.00017009	0.180218	0.000169252
0.8	0.197842	0.197973	0.000187	0.19846	0.000191143	0.19785	0.000180155

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

A7-2 – Case 1: Non-Hierarchical

Fill-rates

Table A7-6: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.2 – Case 1

F#	λ	S	Fr	Fr_N-S	±*	Fr_F-S	±*
1	0.44	3	0.98948	0.98974	0.00068	0.994432	0.00048
2	0.44	3	0.98948	0.98946	0.00098	0.995478	0.00064
3	0.44	3	0.98948	0.98950	0.00120	0.993955	0.00078
4	0.44	3	0.98948	0.98930	0.00140	0.995191	0.0009
5	0.44	3	0.98948	0.98915	0.00152	0.996654	0.00097
6	0.44	3	0.98948	0.98921	0.00163	0.995542	0.00103
7	0.44	3	0.98948	0.98964	0.00175	0.994332	0.00112
8	0.44	3	0.98948	0.98905	0.00184	0.995398	0.00117
9	0.44	3	0.98948	0.98929	0.00196	0.994081	0.00127

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-7: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.4 – Case 1

F#	λ	S	Fr	Fr$_{N-S}$	±*	Fr$_{F-S}$	±*
1	0.44	3	0.98948	0.98974	0.00068	0.995764	0.00038
2	0.44	3	0.98948	0.98946	0.00098	0.99673	0.00052
3	0.44	3	0.98948	0.98950	0.00120	0.995653	0.00063
4	0.44	3	0.98948	0.98930	0.00140	0.996277	0.00072
5	0.44	3	0.98948	0.98915	0.00150	0.997621	0.00078
6	0.44	3	0.98948	0.98921	0.00163	0.99667	0.00031
7	0.44	3	0.98948	0.98964	0.00180	0.995668	0.00092
8	0.44	3	0.98948	0.98905	0.00180	0.996666	0.00097
9	0.44	3	0.98948	0.98929	0.00196	0.99565	0.00106

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-8: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.6 – Case 1

F#	λ	S	Fr	Fr$_{N-S}$	±*	Fr$_{F-S}$	±*
1	0.44	3	0.98948	0.98974	0.00068	0.997179	0.00032
2	0.44	3	0.98948	0.98946	0.00098	0.997831	0.00042
3	0.44	3	0.98948	0.98950	0.00120	0.997021	0.00052
4	0.44	3	0.98948	0.98930	0.00140	0.99747	0.00061
5	0.44	3	0.98948	0.98915	0.00150	0.998321	0.00066
6	0.44	3	0.98948	0.98921	0.00163	0.997909	0.0007
7	0.44	3	0.98948	0.98964	0.00180	0.997204	0.00076
8	0.44	3	0.98948	0.98905	0.00180	0.997973	0.0008
9	0.44	3	0.98948	0.98929	0.00196	0.996877	0.00087

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table 87-9: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.8 – Case 1

F#	λ	S	Fr	Fr$_{N-S}$	±*	Fr$_{F-S}$	±*
1	0.44	3	0.98948	0.98974	0.00068	0.998566	0.00023
2	0.44	3	0.98948	0.98946	0.00098	0.999027	0.00031
3	0.44	3	0.98948	0.98950	0.00120	0.998216	0.00039
4	0.44	3	0.98948	0.98930	0.00140	0.998726	0.00044
5	0.44	3	0.98948	0.98915	0.00150	0.999277	0.00047
6	0.44	3	0.98948	0.98921	0.00163	0.998994	0.0005
7	0.44	3	0.98948	0.98964	0.00180	0.998624	0.00054
8	0.44	3	0.98948	0.98905	0.00180	0.998913	0.00057
9	0.44	3	0.98948	0.98929	0.00196	0.998339	0.00062

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-10: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.2 - Case 1, (0.25 and 0.5 as distance constraints for Strict and Relaxed services respectively)

F#	λ	S	Fr	Fr_{N-S}	±*	Fr_{F-S}	±*
1	0.44	3	0.98948	0.98974	0.00068	0.99512	0.00043
2	0.44	3	0.98948	0.98946	0.00098	0.99626	0.00057
3	0.44	3	0.98948	0.98950	0.00120	0.99479	0.00071
4	0.44	3	0.98948	0.98930	0.00140	0.99609	0.00081
5	0.44	3	0.98948	0.98915	0.00152	0.99788	0.00084
6	0.44	3	0.98948	0.98921	0.00163	0.99662	0.00090
7	0.44	3	0.98948	0.98964	0.00175	0.99521	0.00099
8	0.44	3	0.98948	0.98905	0.00184	0.99619	0.00100
9	0.44	3	0.98948	0.98929	0.00196	0.99486	0.00110

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-11: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.4 - Case 1, (0.25 and 0.5 as distance constraints for Strict and Relaxed services respectively)

F#	λ	S	Fr	Fr_{N-S}	±*	Fr_{F-S}	±*
1	0.44	3	0.98948	0.98974	0.00068	0.99633	0.00036
2	0.44	3	0.98948	0.98946	0.00098	0.99730	0.00047
3	0.44	3	0.98948	0.98950	0.00120	0.99619	0.00058
4	0.44	3	0.98948	0.98930	0.00140	0.99691	0.00066
5	0.44	3	0.98948	0.98915	0.00150	0.99853	0.00070
6	0.44	3	0.98948	0.98921	0.00163	0.99747	0.00074
7	0.44	3	0.98948	0.98964	0.00180	0.99637	0.00083
8	0.44	3	0.98948	0.98905	0.00180	0.99723	0.00088
9	0.44	3	0.98948	0.98929	0.00196	0.99623	0.00095

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-12: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.6 - Case 1, (0.25 and 0.5 as distance constraints for Strict and Relaxed services respectively)

F#	λ	S	Fr	Fr_{N-S}	±*	Fr_{F-S}	±*
1	0.44	3	0.98948	0.98974	0.00068	0.99760	0.00028
2	0.44	3	0.98948	0.98946	0.00098	0.99824	0.00037
3	0.44	3	0.98948	0.98950	0.00120	0.99745	0.00048
4	0.44	3	0.98948	0.98930	0.00140	0.99791	0.00054
5	0.44	3	0.98948	0.98915	0.00150	0.99898	0.00057
6	0.44	3	0.98948	0.98921	0.00163	0.99843	0.00060
7	0.44	3	0.98948	0.98964	0.00180	0.99768	0.00066
8	0.44	3	0.98948	0.98905	0.00180	0.99834	0.00070
9	0.44	3	0.98948	0.98929	0.00196	0.99732	0.00076

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-13: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.8 - Case 1, (0.25 and 0.5 as distance constraints for Strict and Relaxed services respectively)

F#	λ	S	Fr	Fr_{N-S}	±*	Fr_{F-S}	±*
1	0.44	3	0.98948	0.98974	0.00068	0.99887	0.00021
2	0.44	3	0.98948	0.98946	0.00098	0.99922	0.00028
3	0.44	3	0.98948	0.98950	0.00120	0.99851	0.00035
4	0.44	3	0.98948	0.98930	0.00140	0.99900	0.00040
5	0.44	3	0.98948	0.98915	0.00150	0.99959	0.00041
6	0.44	3	0.98948	0.98921	0.00163	0.99921	0.00043
7	0.44	3	0.98948	0.98964	0.00180	0.99901	0.00047
8	0.44	3	0.98948	0.98905	0.00180	0.99914	0.00048
9	0.44	3	0.98948	0.98929	0.00196	0.99872	0.00053

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Average distances

Table A7-14: Average distance to reach a demand location for service with configurations 1 and 3 under the non-hierarchical setup – Case 1

(1-f)	AD	AD_{N-S}	±*	AD_{F-S}	±*
0.2	0.127533	0.127548	0.000087	0.128003	0.0000877739
0.4	0.127533	0.127548	0.000087	0.128236	0.0000881775
0.6	0.127533	0.127548	0.000087	0.128459	0.0000898959
0.8	0.127533	0.127548	0.000087	0.128722	0.0000913686

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-15: Average distance to reach a demand location for service with configurations 1 and 3 under the non-hierarchical setup - Case 1, (0.25 and 0.5 as distance constraints for Strict and Relaxed services respectively)

(1-f)	AD	AD_{N-S}	±*	AD_{F-S}	±*
0.2	0.127533	0.127548	0.000087	0.128097	0.000089
0.4	0.127533	0.127548	0.000087	0.128307	0.000089
0.6	0.127533	0.127548	0.000087	0.128515	0.000090
0.8	0.127533	0.127548	0.000087	0.128722	0.000091

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

A7-3 - Case 2: Hierarchical

Fill-rates

Table A7-16: Fill-rates with configurations 1, 2 and 3 under hierarchical setup with (1-f) = 0.2 – Case 2

F#	λ	S	Fr	Fr_{N-S}	±*	Fr_{H-S}	±*	Fr_{F-S}	±*
1	0.20	2	0.98248	0.98222	0.0012	0.987912	0.0007	0.992214	0.0006
2	0.13	2	0.99247	0.99157	0.0014	0.995707	0.0009	0.995597	0.0008
3	0.22	2	0.97836	0.97860	0.0018	0.988433	0.0012	0.993018	0.0010
4	0.13	2	0.99247	0.99303	0.0020	0.995975	0.0012	0.995975	0.0011
5	0.20	2	0.98248	0.98398	0.0023	0.989313	0.0016	0.992965	0.0014
6	0.13	2	0.99247	0.99165	0.0025	0.995546	0.0017	0.995337	0.0015
7	0.13	2	0.99247	0.99219	0.0027	0.996788	0.0016	0.996715	0.0015
8	0.13	2	0.99247	0.99173	0.0029	0.99639	0.0017	0.996287	0.0016
9	0.13	2	0.99247	0.99165	0.0028	0.996419	0.0018	0.996419	0.0017
10	0.13	2	0.99247	0.99207	0.0029	0.995479	0.0019	0.995403	0.0018
11	0.22	2	0.97836	0.97832	0.0028	0.988442	0.0021	0.99308	0.0017
12	0.13	2	0.99247	0.99215	0.0027	0.996806	0.0017	0.996806	0.0020
13	0.26	2	0.97232	0.97159	0.0028	0.98842	0.0019	0.994412	0.0021
14	0.13	2	0.99247	0.99173	0.0029	0.996561	0.0018	0.996426	0.0015
15	0.22	2	0.97836	0.97655	0.0036	0.987631	0.0017	0.993215	0.0018
16	0.13	2	0.99247	0.99242	0.0039	0.995329	0.0016	0.995293	0.0019
17	0.13	2	0.99247	0.99197	0.0037	0.99619	0.0021	0.996154	0.0023
18	0.13	2	0.99247	0.99366	0.0037	0.997268	0.0019	0.997231	0.0020
19	0.13	2	0.99247	0.99231	0.0038	0.997005	0.0023	0.997005	0.0020
20	0.13	2	0.99247	0.99203	0.0034	0.995108	0.0015	0.995073	0.0020
21	0.20	2	0.98248	0.98206	0.0037	0.988072	0.0022	0.991982	0.0016
22	0.13	2	0.99247	0.99249	0.0038	0.995644	0.0018	0.995682	0.0019
23	0.22	2	0.97836	0.97749	0.0040	0.987951	0.0026	0.993264	0.0024
24	0.13	2	0.99247	0.99265	0.0046	0.995894	0.0025	0.995823	0.0022
25	0.20	2	0.98248	0.98343	0.0046	0.98904	0.0031	0.9931	0.0028

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-17: Fill-rates with configurations 1, 2 and 3 under hierarchical setup with (1-f) = 0.4–Case 2

F#	λ	S	Fr	Fr$_{N-S}$	±*	Fr$_{H-S}$	±*	Fr$_{F-S}$	±*
1	0.24	2	0.97542	0.97567	0.0012	0.983807	0.0009	0.993027	0.0005
2	0.10	2	0.99568	0.99604	0.0015	0.997901	0.0011	0.997675	0.0007
3	0.29	2	0.96569	0.96625	0.0019	0.982678	0.0012	0.993938	0.0008
4	0.10	2	0.99568	0.99599	0.0020	0.997702	0.0013	0.997656	0.0008
5	0.24	2	0.97542	0.97603	0.0025	0.985016	0.0022	0.993832	0.0009
6	0.10	2	0.99568	0.99512	0.0026	0.997319	0.0018	0.996998	0.0007
7	0.10	2	0.99568	0.99550	0.0027	0.998248	0.0019	0.998205	0.0009
8	0.10	2	0.99568	0.99530	0.0028	0.997968	0.0019	0.997925	0.0015
9	0.10	2	0.99568	0.99598	0.0029	0.998238	0.0017	0.99814	0.0015
10	0.10	2	0.99568	0.99518	0.0028	0.99683	0.0016	0.996645	0.0007
11	0.29	2	0.96569	0.96585	0.0021	0.982368	0.0019	0.993928	0.0007
12	0.10	2	0.99568	0.99485	0.0024	0.997688	0.0018	0.99746	0.0011
13	0.35	2	0.95083	0.95044	0.0023	0.979203	0.0018	0.994814	0.0012
14	0.10	2	0.99568	0.99547	0.0022	0.997976	0.0017	0.997848	0.0014
15	0.29	2	0.96569	0.96423	0.0037	0.981222	0.0015	0.994361	0.0013
16	0.10	2	0.99568	0.99630	0.0031	0.997674	0.0017	0.997629	0.0019
17	0.10	2	0.99568	0.99539	0.0032	0.997705	0.0014	0.997655	0.0017
18	0.10	2	0.99568	0.99638	0.0031	0.998365	0.0018	0.998213	0.0019
19	0.10	2	0.99568	0.99576	0.0034	0.998503	0.0028	0.998503	0.0021
20	0.10	2	0.99568	0.99489	0.0032	0.997084	0.0028	0.996889	0.0019
21	0.24	2	0.97542	0.97526	0.0033	0.998414	0.0027	0.993131	0.0020
22	0.10	2	0.99568	0.99553	0.0033	0.99734	0.0026	0.997151	0.0021
23	0.29	2	0.96569	0.96491	0.0034	0.981176	0.0030	0.993683	0.0022
24	0.10	2	0.99568	0.99588	0.0031	0.997568	0.0031	0.997427	0.0019
25	0.24	2	0.97542	0.97619	0.0041	0.984851	0.0029	0.993674	0.0021

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-18: Fill-rates with configurations 1, 2 and 3 under hierarchical setup with (1-f) = 0.6–Case 2

F#	λ	S	Fr	Fr$_{N-S}$	±*	Fr$_{H-S}$	±*	Fr$_{F-S}$	±*
1	0.28	2	0.96740	0.96783	0.0013	0.979425	0.0009	0.994864	0.0003
2	0.06	2	0.99804	0.99846	0.0014	0.999432	0.0010	0.998949	0.0005
3	0.35	2	0.95083	0.95146	0.0020	0.975793	0.0013	0.995342	0.0006
4	0.06	2	0.99804	0.99816	0.0021	0.999131	0.0013	0.998985	0.0007
5	0.28	2	0.96740	0.96836	0.0026	0.980519	0.0017	0.995068	0.0008
6	0.06	2	0.99804	0.99810	0.0027	0.999075	0.0017	0.998661	0.0011
7	0.06	2	0.99804	0.99776	0.0027	0.99906	0.0018	0.99906	0.0009
8	0.06	2	0.99804	0.99769	0.0028	0.998755	0.0018	0.99869	0.0011
9	0.06	2	0.99804	0.99850	0.0029	0.999354	0.0011	0.999137	0.0011
10	0.06	2	0.99804	0.99747	0.0030	0.998301	0.0019	0.997758	0.0012
11	0.35	2	0.95083	0.95100	0.0033	0.976152	0.0018	0.995372	0.0011
12	0.06	2	0.99804	0.99751	0.0034	0.998793	0.0021	0.99864	0.0008
13	0.45	3	0.98925	0.98887	0.0035	0.995248	0.0023	0.999475	0.0012
14	0.06	2	0.99804	0.99848	0.0035	0.99904	0.0024	0.998786	0.0015
15	0.35	2	0.95083	0.94989	0.0038	0.97493	0.0019	0.995753	0.0013
16	0.06	2	0.99804	0.99803	0.0038	0.998722	0.0019	0.99838	0.0014
17	0.06	2	0.99804	0.99820	0.0039	0.999162	0.0019	0.999162	0.0012
18	0.06	2	0.99804	0.99885	0.0039	0.999297	0.0019	0.999223	0.0016
19	0.06	2	0.99804	0.99776	0.0039	0.999451	0.0025	0.999382	0.0010
20	0.06	2	0.99804	0.99767	0.0038	0.99898	0.0021	0.998692	0.0014
21	0.28	2	0.96740	0.96724	0.0043	0.979551	0.0022	0.994871	0.0019
22	0.06	2	0.99804	0.99832	0.0043	0.998	0.0021	0.998498	0.0012
23	0.35	2	0.95083	0.95130	0.0045	0.975481	0.0019	0.995514	0.0017
24	0.06	2	0.99804	0.99863	0.0046	0.999109	0.0024	0.99898	0.0019
25	0.28	2	0.96740	0.96812	0.0048	0.979415	0.0026	0.995011	0.0018

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-19: Fill-rates with configurations 1, 2 and 3 under hierarchical setup with (1-f) = 0.8–Case 2

F#	λ	S	Fr	Fr_{N-S}	±*	Fr_{H-S}	±*	Fr_{F-S}	±*
1	0.32	2	0.95852	0.95899	0.0013	0.976336	0.0009	0.996604	0.0003
2	0.03	1	0.96851	0.96983	0.0039	0.982292	0.0027	0.979213	0.0027
3	0.42	3	0.99119	0.99122	0.0040	0.995414	0.0028	0.999587	0.0027
4	0.03	1	0.96851	0.97326	0.0052	0.985874	0.0033	0.98246	0.0036
5	0.32	2	0.95852	0.95891	0.0054	0.976953	0.0035	0.996718	0.0036
6	0.03	1	0.96851	0.96840	0.0064	0.981572	0.0042	0.97809	0.0044
7	0.03	1	0.96851	0.96830	0.0074	0.985167	0.0048	0.983893	0.0049
8	0.03	1	0.96851	0.96807	0.0080	0.986789	0.0050	0.985832	0.0052
9	0.03	1	0.96851	0.96710	0.0088	0.984292	0.0053	0.983018	0.0056
10	0.03	1	0.96851	0.96660	0.0090	0.979714	0.0060	0.975737	0.0058
11	0.42	3	0.99119	0.99135	0.0087	0.995404	0.0061	0.999518	0.0055
12	0.03	1	0.96851	0.96708	0.0091	0.988519	0.0063	0.986995	0.0058
13	0.54	3	0.98206	0.98119	0.0087	0.992138	0.0060	0.999584	0.0056
14	0.03	1	0.96851	0.96970	0.0099	0.989051	0.0060	0.988064	0.0059
15	0.42	3	0.99119	0.99097	0.0085	0.995602	0.0066	0.999641	0.0057
16	0.03	1	0.96851	0.97319	0.0101	0.984607	0.0070	0.981471	0.0064
17	0.03	1	0.96851	0.97034	0.0094	0.987292	0.0070	0.9865	0.0067
18	0.03	1	0.96851	0.96924	0.0119	0.986729	0.0070	0.985897	0.0062
19	0.03	1	0.96851	0.96673	0.0117	0.985168	0.0070	0.984082	0.0079
20	0.03	1	0.96851	0.96876	0.0121	0.981044	0.0080	0.9779	0.0077
21	0.32	2	0.95852	0.95781	0.0126	0.975262	0.0080	0.996828	0.0080
22	0.03	1	0.96851	0.96843	0.0129	0.980443	0.0080	0.976258	0.0087
23	0.42	3	0.99119	0.99084	0.0123	0.99549	0.0086	0.999607	0.0088
24	0.03	1	0.96851	0.96973	0.0128	0.982215	0.0089	0.977947	0.0090
25	0.32	2	0.95852	0.95842	0.0135	0.985296	0.009	0.996512	0.0086

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Average distances

Table A7-20: Average distance to reach a demand location for service with configurations 1, 2 and 3 under the hierarchical setup – Case 2

(1-f)	AD	AD_{N-S}	±*	AD_{H-S}	±*	AD_{F-S}	±*
0.2	0.089010	0.089032	0.000084	0.0893029	0.000086	0.0890355	0.000087
0.4	0.101501	0.101430	0.000088	0.101954	0.000090	0.101096	0.000086
0.6	0.113992	0.113939	0.000101	0.114678	0.000107	0.113392	0.000097
0.8	0.126482	0.126483	0.000097	0.127082	0.000098	0.12629	0.000099

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

A7-4 – Case 2: Non-Hierarchical

Fill-rates

Table A7-21: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.2 – Case 2

F#	λ	S	Fr	Fr_{N-S}	±*	Fr_{F-S}	±*
1	0.16	2	0.98849	0.98786	0.00101	0.993197	0.00067
2	0.16	2	0.98849	0.98749	0.00146	0.994702	0.00093
3	0.16	2	0.98849	0.98822	0.00175	0.994435	0.00011
4	0.16	2	0.98849	0.98896	0.00204	0.994817	0.00125
5	0.16	2	0.98849	0.98983	0.00224	0.994119	0.00144
6	0.16	2	0.98849	0.98799	0.00250	0.995149	0.00156
7	0.16	2	0.98849	0.98833	0.00268	0.996082	0.00163
8	0.16	2	0.98849	0.98863	0.00284	0.995848	0.00171
9	0.16	2	0.98849	0.98718	0.00299	0.995675	0.00178
10	0.16	2	0.98849	0.98867	0.00318	0.994802	0.00188
11	0.16	2	0.98849	0.98789	0.00336	0.994496	0.00197
12	0.16	2	0.98849	0.98805	0.00347	0.995864	0.00202
13	0.16	2	0.98849	0.98832	0.00362	0.995472	0.00210
14	0.16	2	0.98849	0.98700	0.00376	0.996008	0.00213
15	0.16	2	0.98849	0.98771	0.00392	0.994762	0.00223
16	0.16	2	0.98849	0.98886	0.00403	0.994777	0.00231
17	0.16	2	0.98849	0.98843	0.00415	0.995501	0.00237
18	0.16	2	0.98849	0.99055	0.00426	0.996969	0.00240
19	0.16	2	0.98849	0.98809	0.00439	0.996142	0.00245
20	0.16	2	0.98849	0.98833	0.00452	0.994465	0.00253
21	0.16	2	0.98849	0.98817	0.00464	0.993561	0.00263
22	0.16	2	0.98849	0.98867	0.00475	0.99502	0.00268
23	0.16	2	0.98849	0.98733	0.00486	0.994873	0.00274
24	0.16	2	0.98849	0.98897	0.00494	0.995086	0.00280
25	0.16	2	0.98849	0.98937	0.00503	0.994377	0.00286

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-22: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.4 – Case 2

F#	λ	S	Fr	Fr$_{N-S}$	±*	Fr$_{F-S}$	±*
1	0.16	2	0.98849	0.98786	0.00101	0.994547	0.00059
2	0.16	2	0.98849	0.98749	0.00146	0.995967	0.00079
3	0.16	2	0.98849	0.98822	0.00175	0.995977	0.00092
4	0.16	2	0.98849	0.98896	0.00204	0.996317	0.00107
5	0.16	2	0.98849	0.98983	0.00224	0.995945	0.00119
6	0.16	2	0.98849	0.98799	0.00250	0.996436	0.00130
7	0.16	2	0.98849	0.98833	0.00268	0.997214	0.00136
8	0.16	2	0.98849	0.98863	0.00284	0.997083	0.00141
9	0.16	2	0.98849	0.98718	0.00299	0.996773	0.00147
10	0.16	2	0.98849	0.98867	0.00318	0.996006	0.00155
11	0.16	2	0.98849	0.98789	0.00336	0.995931	0.00163
12	0.16	2	0.98849	0.98805	0.00347	0.996907	0.00167
13	0.16	2	0.98849	0.98832	0.00362	0.996465	0.00173
14	0.16	2	0.98849	0.98700	0.00376	0.996892	0.00177
15	0.16	2	0.98849	0.98771	0.00392	0.99609	0.00185
16	0.16	2	0.98849	0.98886	0.00403	0.996348	0.00192
17	0.16	2	0.98849	0.98843	0.00415	0.996694	0.00196
18	0.16	2	0.98849	0.99055	0.00426	0.997685	0.00199
19	0.16	2	0.98849	0.98809	0.00439	0.996903	0.00204
20	0.16	2	0.98849	0.98833	0.00452	0.995626	0.00211
21	0.16	2	0.98849	0.98817	0.00464	0.995116	0.00220
22	0.16	2	0.98849	0.98867	0.00475	0.996059	0.00226
23	0.16	2	0.98849	0.98733	0.00486	0.996366	0.00231
24	0.16	2	0.98849	0.98897	0.00494	0.99634	0.00236
25	0.16	2	0.98849	0.98937	0.00503	0.995569	0.00241

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-23: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.6 – Case 2

F#	λ	S	Fr	Fr$_{N-S}$	±*	Fr$_{F-S}$	±*
1	0.16	2	0.98849	0.98786	0.00101	0.996141	0.00048
2	0.16	2	0.98849	0.98749	0.00146	0.997292	0.00063
3	0.16	2	0.98849	0.98822	0.00175	0.997059	0.00077
4	0.16	2	0.98849	0.98896	0.00204	0.997718	0.00087
5	0.16	2	0.98849	0.98983	0.00224	0.997185	0.00097
6	0.16	2	0.98849	0.98799	0.00250	0.997593	0.00105
7	0.16	2	0.98849	0.98833	0.00268	0.998151	0.00109
8	0.16	2	0.98849	0.98863	0.00284	0.997742	0.00115
9	0.16	2	0.98849	0.98718	0.00299	0.997758	0.00120
10	0.16	2	0.98849	0.98867	0.00318	0.997461	0.00127
11	0.16	2	0.98849	0.98789	0.00336	0.997231	0.00135
12	0.16	2	0.98849	0.98805	0.00347	0.997696	0.00138
13	0.16	2	0.98849	0.98832	0.00362	0.997668	0.00143
14	0.16	2	0.98849	0.98700	0.00376	0.997581	0.00148
15	0.16	2	0.98849	0.98771	0.00392	0.997423	0.00154
16	0.16	2	0.98849	0.98886	0.00403	0.997514	0.00159
17	0.16	2	0.98849	0.98843	0.00415	0.997887	0.00162
18	0.16	2	0.98849	0.99055	0.00426	0.998625	0.00164
19	0.16	2	0.98849	0.98809	0.00439	0.997886	0.00168
20	0.16	2	0.98849	0.98833	0.00452	0.997141	0.00174
21	0.16	2	0.98849	0.98817	0.00464	0.996382	0.00183
22	0.16	2	0.98849	0.98867	0.00475	0.99769	0.00188
23	0.16	2	0.98849	0.98733	0.00486	0.997641	0.00192
24	0.16	2	0.98849	0.98897	0.00494	0.997816	0.00197
25	0.16	2	0.98849	0.98937	0.00503	0.996887	0.00202

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Table A7-24: Fill-rates with configurations 1 and 3 under non-hierarchical setup with (1-f) = 0.8 – Case 2

F#	λ	S	Fr	Fr_{N-S}	±*	Fr_{F-S}	±*
1	0.16	2	0.98849	0.98786	0.00101	0.997812	0.00038
2	0.16	2	0.98849	0.98749	0.00146	0.99871	0.00048
3	0.16	2	0.98849	0.98822	0.00175	0.998455	0.00056
4	0.16	2	0.98849	0.98896	0.00204	0.999041	0.00061
5	0.16	2	0.98849	0.98983	0.00224	0.998622	0.00068
6	0.16	2	0.98849	0.98799	0.00250	0.998731	0.00076
7	0.16	2	0.98849	0.98833	0.00268	0.999061	0.00079
8	0.16	2	0.98849	0.98863	0.00284	0.998984	0.00082
9	0.16	2	0.98849	0.98718	0.00299	0.998971	0.00085
10	0.16	2	0.98849	0.98867	0.00318	0.998616	0.00090
11	0.16	2	0.98849	0.98789	0.00336	0.998729	0.00096
12	0.16	2	0.98849	0.98805	0.00347	0.998775	0.00098
13	0.16	2	0.98849	0.98832	0.00362	0.998818	0.00102
14	0.16	2	0.98849	0.98700	0.00376	0.999037	0.00105
15	0.16	2	0.98849	0.98771	0.00392	0.998684	0.00109
16	0.16	2	0.98849	0.98886	0.00403	0.998731	0.00113
17	0.16	2	0.98849	0.98843	0.00415	0.998931	0.00116
18	0.16	2	0.98849	0.99055	0.00426	0.999372	0.00118
19	0.16	2	0.98849	0.98809	0.00439	0.998798	0.00121
20	0.16	2	0.98849	0.98833	0.00452	0.99843	0.00126
21	0.16	2	0.98849	0.98817	0.00464	0.998309	0.00131
22	0.16	2	0.98849	0.98867	0.00475	0.998641	0.00135
23	0.16	2	0.98849	0.98733	0.00486	0.998932	0.00138
24	0.16	2	0.98849	0.98897	0.00494	0.998895	0.00140
25	0.16	2	0.98849	0.98937	0.00503	0.998231	0.00144

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Average distances

Table A7-25: Average distance to reach a demand location for service with configurations 1 and 3 under the non-hierarchical setup – Case 2

(1-f)	AD	AD_{N-S}	±*	AD_{F-S}	±*
0.2	0.076520	0.076549	0.000054	0.0768567	0.000055
0.4	0.076520	0.076549	0.000054	0.0769916	0.000056
0.6	0.076520	0.076549	0.000054	0.0771209	0.000057
0.8	0.076520	0.076549	0.000054	0.0772555	0.000058

* Approximate 90 % confidence interval. ($t_{99,0.5}$)

Appendix 8 – MINLP computation times

Notations

α: cost constant (higher when the inventory cost is high compared to the transportation cost)

f: fraction of demand for the strict service

L: length of line segment

d_{maxS}:   service distance constraint for the strict service

d_{maxR}:   service distance constraint for the relaxed service

n_s: total number of facilities

n_r: number of higher level facilities

Table A8-1: Experimental settings

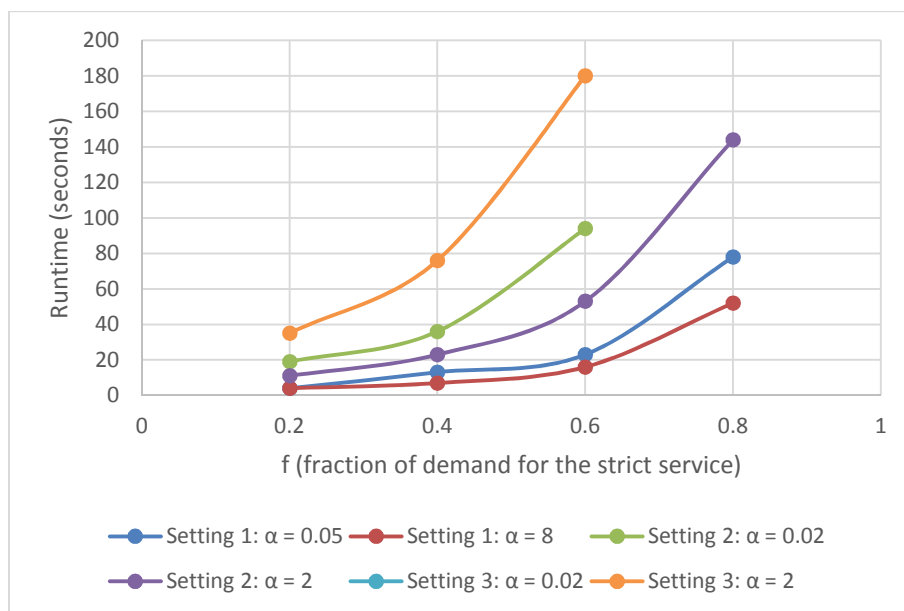|	Setting 1	Setting 2	Setting 3
d_{maxS}	0.24	0.12	0.06
d_{maxR}	0.48	0.24	0.12
n_s	3	5	9
n_r	2	3	5
$L = 1$			

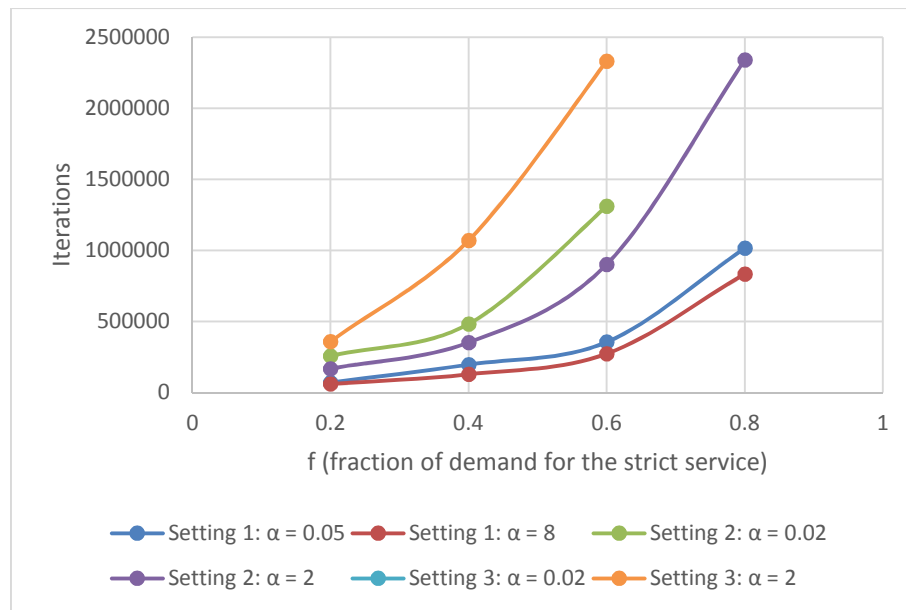*Figure A8-1: Solution runtime (values from Table A8-2)*

Figure A8-2: Solution iterations (values from Table A8-2)

Table A-8.2: Computation iterations (runtime)

	Setting 1		Setting 2		Setting 3	
	α		*α*		*α*	
	0.05	8	0.02	2	0.02	2
f						
0.2	68831 (4 sec.)	59008 (4 sec.)	257179 (19 sec.)	166649 (11 sec.)	*	358057 (35 sec.)
0.4	195517 (13 sec.)	127219 (7 sec.)	481385 (36 sec.)	351307 (23 sec.)	*	1069590 (1 min. 16 sec.)
0.6	354326 (23 sec)	271788 (16 sec.)	1310069 (1 min. 34 sec.)	901667 (53 sec.)	*	2332817 (3 mins.)
0.8	1015063 (1 min. 18 sec.)	833892 (52 sec.)	*	2340901 (2 min. 24 sec.)	*	*
* Global optimum not established within 30 minutes runtime (Processor: Intel Core i3 @ 2.4 GHz).						

References

Alvarez, E.M., van der Heijden, M.C. and Zijm, W.H.M. (2013), "The selective use of emergency shipments for service-contract differentiation", *International Journal of Production Economics*, Focusing on Inventories: Research and Applications, Vol. 143 No. 2, pp. 518–526.

Amini, M.M., Retzlaff-Roberts, D. and Bienstock, C.C. (2005), "Designing a reverse logistics operation for short cycle time repair services", *International Journal of Production Economics*, Vol. 96 No. 3, pp. 367–380.

Anton, H. (2013), *Calculus: early transcendentals*, Wiley, Singapore, 10th ed., International Student version.

Ashayeri, J., Heuts, R., Jansen, A. and Szczerba, B. (1996), "Inventory management of repairable service parts for personal computers: A case study", *International Journal of Operations & Production Management*, Vol. 16 No. 12, pp. 74–97.

Ash, P.F. and Bolker, E.D. (1985), "Recognizing Dirichlet tessellations", *Geometriae Dedicata*, Vol. 19 No. 2, pp. 175–206.

Aurenhammer, F. (1991), "Voronoi diagrams—a survey of a fundamental geometric data structure", *ACM Computing Surveys (CSUR)*, Vol. 23 No. 3, pp. 345–405.

Beckmann, M.J. (1968), *Location theory*, Random House, New York.

Beckmann, M.J. and Thisse, J.-F. (1987), "The location of production activities", *Handbook of regional and urban economics*, Vol. 1, pp. 21–95.

Benbasat, I., Goldstein, D.K. and Mead, M. (1987), "The case research strategy in studies of information systems", *MIS quarterly*, Vol. 11 No. 3, pp. 369–386.

Bendoly, E., Blocher, D., Bretthauer, K.M. and Venkataramanan, M.A. (2007), "Service and cost benefits through clicks-and-mortar integration: Implications for the centralization/decentralization debate", *European Journal of Operational Research*, Vol. 180 No. 1, pp. 426–442.

Benjaafar, S., ElHafsi, M., Chung-Yee Lee and Weihua Zhou. (2011), "Optimal control of an assembly system with multiple stages and multiple demand classes", *Operations Research*, Vol. 59 No. 2, pp. 522–529.

Berry, B.J.L. (1967), *Geography of market centers and retail distribution*, Prentice-Hall, Englewood Cliffs, NJ.

Bertrand, J.W.M. and Fransoo, J.C. (2002), "Operations management research methodologies using quantitative modeling", *International Journal of Operations & Production Management*, Vol. 22 No. 2, pp. 241–264.

Bertsimas, D.J. and Simchi-Levi, D. (1996), "A new generation of vehicle routing research: robust algorithms, addressing uncertainty", *Operations Research*, Vol. 44 No. 2, pp. 286–304.

Blais, M., Lapierre, S.D. and Laporte, G. (2003), "Solving a home-care districting problem in an urban setting", *Journal of the Operational Research Society*, Vol. 54 No. 11, pp. 1141–1147.

Bollobás, B. and Stern, N. (1972), "The optimal structure of market areas", *Journal of Economic Theory*, Vol. 4 No. 2, pp. 174–179.

Bolumole, Y.A. (2001), "The supply chain role of third-party logistics providers", *The International Journal of Logistics Management*, Vol. 12 No. 2, pp. 87–102.

Boyaci, T. and Ray, S. (2006), "The impact of capacity costs on product differentiation in delivery time, delivery reliability, and price", *Production and Operations Management*, Vol. 15 No. 2, pp. 179–197.

Bozkaya, B., Erkut, E. and Laporte, G. (2003), "A tabu search heuristic and adaptive memory procedure for political districting", *European Journal of Operational Research*, Vol. 144 No. 1, pp. 12–26.

Braglia, M., Grassi, A. and Montanari, R. (2004), "Multi-attribute classification method for spare parts inventory management", *Journal of Quality in Maintenance Engineering*, Vol. 10 No. 1, pp. 55–65.

Brotcorne, L., Laporte, G. and Semet, F. (2003), "Ambulance location and relocation models", *European Journal of Operational Research*, Vol. 147 No. 3, pp. 451–463.

Brown, R.G. (1967), *Decision rules for inventory management*, Holt, Rinehart & Winston, New York.

Candas, M.F. and Kutanoglu, E. (2007), "Benefits of considering inventory in service parts logistics network design problems with time-based service constraints", *IIE Transactions*, Vol. 39 No. 2, pp. 159–176.

Cattani, K.D. and Souza, G.C. (2002), "Inventory rationing and shipment flexibility alternatives for direct market firms", *Production & Operations Management*, Vol. 11 No. 4, pp. 441–457.

Chang, P.-L. and Lin, C.-T. (1991), "On the effect of centralization on expected costs in a multi-location newsboy problem", *Journal of the Operational Research Society*, Vol. 42 No. 11, pp. 1025–1030.

Chen, M.-S. and Lin, C.-T. (1989), "Effects of centralization on expected costs in a multi-location newsboy problem", *Journal of the Operational Research Society*, Vol. 40 No. 6, pp. 597–602.

Chen, M.-S. and Lin, C.-T. (1990), "An Example of Disbenefits of Centralized Stocking", *Journal of the Operational Research Society*, Vol. 41 No. 3, pp. 259–262.

Cherikh, M. (2000), "On the effect of centralisation on expected profits in a multi-location newsboy problem", *Journal of the Operational Research Society*, Vol. 51 No. 6, pp. 755–761.

Christaller, W. (1933), *Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*, University Microfilms.

Christaller, W. (1966), *Central places in Southern Germany*, Prentice-Hall, Englewood Cliffs, N.J.

Cobbaert, K. and Van Oudheusden, D. (1996), "Inventory models for fast moving spare parts subject to 'sudden death' obsolescence", *International Journal of Production Economics*, Vol. 44 No. 3, pp. 239–248.

Cohen, M. (2005), "Tapping the service supply chain", Line56, available at: http://www.line56.com/articles/default.asp?ArticleID=6615 (accessed 1 May 2010).

Cohen, M.A., Agrawal, N. and Agrawal, V. (2006), "Winning in the aftermarket", *Harvard Business Review*, Vol. 84 No. 5, pp. 129–138.

Cohen, M.A., Cull, C., Lee, H.L. and Willen, D. (2000), "Saturn's supply-chain innovation: High value in after-sales service", *Sloan Management Review*, Vol. 41 No. 4, pp. 93–101.

Cohen, M.A., Kleindorfer, P.R. and Lee, H.L. (1988), "Service constrained (s, S) inventory systems with priority demand classes and lost sales", *Management Science*, Vol. 34 No. 4, pp. 482–499.

Cohen, M.A., Kleindorfer, P.R. and Lee, H.L. (1989), "Near-optimal service constrained stocking policies for spare parts", *Operations Research*, Vol. 37 No. 1, pp. 104.

Cohen, M.A. and Lee, H.L. (1990), "Out of touch with customer needs? Spare parts and after sales service", *Sloan Management Review*, Vol. 31 No. 2, pp. 55.

Cohen, M.A., Yu-Sheng Zheng and Yunzeng Wang. (1999), "Identifying opportunities for improving Teradyne's service-parts logistics system", *Interfaces*, Vol. 29 No. 4, pp. 1–18.

Cohen, M.A., Zheng, Y.-S. and Agrawal, V. (1997), "Service parts logistics: a benchmark analysis", *IIE Transactions*, Vol. 29 No. 8, pp. 627–639.

Cohen, M., Kamesam, P.V., Kleindorfer, P., Hau Lee and Tekerian, A. (1990), "Optimizer: IBM's multi-echelon inventory for managing service logistics", *Interfaces*, Vol. 20 No. 1, pp. 65–82.

Cohn, A.M. and Barnhart, C. (2006), "Composite-variable modeling for service parts logistics", *Annals of Operations Research*, Vol. 144 No. 1, pp. 17–32.

Cooper, J.C. (1983), "The Use of straight line distances in solutions to the vehicle scheduling problem", *Journal of the Operational Research Society*, Vol. 34 No. 5, pp. 419–424.

Current, J.R. and Schilling, D.A. (1990), "Analysis of errors due to demand data aggregation in the set covering and maximal covering location problems", *Geographical Analysis*, Vol. 22 No. 2, pp. 116–126.

D'Amico, S.J., Wang, S.-J., Batta, R. and Rump, C.M. (2002), "A simulated annealing approach to police district design", *Computers & Operations Research*, Location Analysis, Vol. 29 No. 6, pp. 667–684.

Das, C. and Tyagi, R. (1997), "Role of inventory and transportation costs in determining the optimal degree of centralization", *Transportation Research: Part E*, Vol. 33 No. 3, pp. 171.

Daskin, M.S., Haghani, A.E., Khanal, M. and Malandraki, C. (1989), "Aggregation effects in maximum covering models", *Annals of Operations Research*, Vol. 18 No. 1, pp. 113–139.

Dekker, R., Hill, R.M., Kleijn, M.J. and Teunter, R. h. (2002), "On the (S − 1, S) lost sales inventory model with priority demand classes", *Naval Research Logistics (NRL)*, Vol. 49 No. 6, pp. 593–610.

Dekker, R., Kleijn, M. j. and de Rooij, P. j. (1998), "A spare parts stocking policy based on equipment criticality", *International Journal of Production Economics*, Vol. 56/57 No. 3, pp. 69–77.

Deloitte. (2006), "The service revolution in global manufacturing industries", Deloitte, available at: http://www.deloitte.com/view/en_TR/tr/industries/manufacturing/bcff5c968b0fb110VgnVCM100000ba42f00aRCRD.htm (accessed 7 November 2013).

Dennis, M.J. and Kambil, A. (2003), "Service management: Building profits after sale", *Supply Chain Management Review, (JAN./FEB. 2003), pp. 42-48,*

available at: http://trid.trb.org/view.aspx?id=623922 (accessed 7 November 2013).

Deshpande, V., Cohen, M.A. and Donohue, K. (2003a), "A threshold inventory rationing policy for service-differentiated demand classes", *Management Science*, Vol. 49 No. 6, pp. 683–703.

Deshpande, V., Cohen, M.A. and Donohue, K. (2003b), "An empirical study of service differentiation for weapon system service parts", *Operations Research*, Vol. 51 No. 4, pp. 518–530.

Dökmeci, V.F. (1973), "An optimization model for a hierarchical spatial system", *Journal of Regional Science*, Vol. 13 No. 3, pp. 439.

Drexl, A. and Haase, K. (1999), "Fast approximation methods for sales force deployment", *Management Science*, Vol. 45 No. 10, pp. 1307–1323.

Drèze, J., Le Breton, M., Savvateev, A. and Weber, S. (2008), "'Almost' subsidy-free spatial pricing in a multi-dimensional setting", *Journal of Economic Theory*, Vol. 143 No. 1, pp. 275–291.

Drezner, T. and Drezner, Z. (1997), "Replacing continuous demand with discrete demand in a competitive location model", *Naval Research Logistics (NRL)*, Vol. 44 No. 1, pp. 81–95.

Drezner, Z. and Hamacher, H. (2002), *Facility location: applications and theory*, Springer, Berlin.

Drezner, Z., Klamroth, K., Schöbel, A. and Wesolowsky, G. (2002), "The Weber problem", *Facility Location: Applications and Theory*, Springer, Berlin.

Ellram, L.M. (1996), "The use of the case study method in logistics research", *Journal of Business Logistics*, Vol. 17 No. 2, pp. 93–138.

Eppen, G.D. (1979), "Effects of centralization on expected costs in a multi-location newsboy problem", *Management Science*, Vol. 25 No. 5, pp. 498–501.

Evans, R.V. (1968), "Sales and restocking policies in a single item inventory system", *Management Science*, Vol. 14 No. 7, pp. 463–472.

Evers, P.T. (1995), "Expanding the square root law: An analysis of both safety and cycle stocks", *Logistics and Transportation Review*, Vol. 31 No. 1, pp. 1.

Evers, P.T. and Frederick J. Beier. (1993), "The portfolio effect and multiple consolidation points: A critical assessment of the square root law", *Journal of Business Logistics*, Vol. 14 No. 2, pp. 109–125.

Fadıloğlu, M.M. and Bulut, Ö. (2010), "A dynamic rationing policy for continuous-review inventory systems", *European Journal of Operational Research*, Vol. 202 No. 3, pp. 675–685.

Farahani, R.Z., Asgari, N., Heidari, N., Hosseininia, M. and Goh, M. (2012), "Covering problems in facility location: A review", *Computers & Industrial Engineering*, Vol. 62 No. 1, pp. 368–407.

Farahani, R.Z., Hekmatfar, M., Fahimnia, B. and Kazemzadeh, N. (2014), "Hierarchical facility location problem: Models, classifications, techniques, and applications", *Computers & Industrial Engineering*, Vol. 68, pp. 104–117.

Ferland, J.A. and Guénette, G. (1990), "Decision support system for the school districting problem", *Operations Research*, Vol. 38 No. 1, pp. 15–21.

Fleischmann, B. and Paraschis, J.N. (1988), "Solving a large scale districting problem: a case report", *Computers & Operations Research*, Vol. 15 No. 6, pp. 521–533.

Flynn, B.B., Sakakibara, S., Schroeder, R.G., Bates, K.A. and Flynn, E.J. (1990), "Empirical research methods in operations management", *Journal of Operations Management*, Vol. 9 No. 2, pp. 250–284.

Fortuin, L. (1980), "The all-time requirement of spare parts for service after sales— Theoretical analysis and practical results", *International Journal of Operations & Production Management*, Vol. 1 No. 1, pp. 59–70.

Fortuin, L. and Martin, H. (1999), "Control of service parts", *International Journal of Operations & Production Management*, Vol. 19 No. 9, pp. 950–971.

Fowler, R.J., Paterson, M.S. and Tanimoto, S.L. (1981), "Optimal packing and covering in the plane are NP-complete", *Information processing letters*, Vol. 12 No. 3, pp. 133–137.

Frankel, R., Naslund, D. and Bolumole, Y. (2005), "The 'white space' of logistics research: A look at the role of methods usage", *Journal of Business Logistics*, Vol. 26 No. 2, pp. 185–208.

Frank, K.C., Zhang, R.Q. and Duenyas, I. (2003), "Optimal policies for inventory systems with priority demand classes", *Operations Research*, Vol. 51 No. 6, pp. 993–1002.

Gau, C.-Y. and Schrage, L.E. (2004), "Implementation and testing of a branch-and-bound based method for deterministic global optimization: Operations research applications", *Frontiers in global optimization*, Springer, pp. 145–164.

Gebauer, H., Edvardsson, B., Gustafsson, A. and Witell, L. (2010), "Match or mismatch: Strategy-structure configurations in the service business of manufacturing companies", *Journal of Service Research*, Vol. 13 No. 2, pp. 198–215.

Granot, D. and Sosic, G. (2003), "A three-stage model for a decentralized distribution system of retailers", *Operations Research*, Vol. 51 No. 5, pp. 771–784.

Graves, S.C. (1985), "A multi-echelon inventory model for a repairable item with one-for-one replenishment", *Management Science*, Vol. 31 No. 10, pp. 1247–1256.

Guba, E.G. (1990), *The paradigm dialog*, SAGE Publications.

Gusein-Zade, S. m. (1992), "Comments on 'a note on the location of medical facilities' by Z. Drezner", *Journal of Regional Science*, Vol. 32 No. 2, p. 229.

Gzara, F., Nematollahi, E. and Dasci, A. (2014), "Linear location-inventory models for service parts logistics network design", *Computers & Industrial Engineering*, Vol. 69, pp. 53–63.

Haimovich, M. and Magnanti, T.L. (1988), "Extremum properties of hexagonal partitioning and the uniform distribution in Euclidean location", *SIAM Journal on Discrete Mathematics*, Vol. 1 No. 1, pp. 50–64.

Hanafi, S., Freville, A. and Vaca, P. (1999), "Municipal solid waste collection: An effective data structure for solving the sectorization problem with local search methods", *Information Systems & Operational Research*, Vol. 37 No. 3, pp. 236–254.

Hanson, S. (1997), *Ten geographic ideas that changed the world*, Rutgers University Press, New Brunswick, N.J.

Hartvigsen, D. (1992), "Recognizing Voronoi diagrams with linear programming", *ORSA Journal on Computing*, Vol. 4 No. 4, p. 369.

Heskett, J.L. (1973), *Business logistics; physical distribution and materials management*, Ronald Press Co., New York.

Hillier, F.S. (2005), *Introduction to operations research*, McGraw-Hill, New York ; London, 8th ed., International ed.

Hodgson, M.J. (1986), "A hierarchical location-allocation model with allocations based on facility size", *Annals of Operations Research*, Vol. 6 No. 8, pp. 273–289.

Huiskonen, J. (2001), "Maintenance spare parts logistics: Special characteristics and strategic choices", *International Journal of Production Economics*, Vol. 71 No. 1–3, pp. 125–133.

Iannoni, A.P., Morabito, R. and Saydam, C. (2009), "An optimization approach for ambulance location and the districting of the response segments on highways", *European Journal of Operational Research*, Vol. 195 No. 2, pp. 528–542.

Inderfurth, K. and Kleber, R. (2013), "An advanced heuristic for multiple-option spare parts procurement after end-of-production", *Production & Operations Management*, Vol. 22 No. 1, pp. 54–70.

Iri, M., Murota, K. and Ohya, T. (1984), "A fast Voronoi-diagram algorithm with applications to geographical optimization problems", *System Modelling and Optimization*, Springer, pp. 273–288.

Iyoob, I.M. and Kutanoglu, E. (2013), "Inventory sharing in integrated network design and inventory optimization with low-demand parts", *European Journal of Operational Research*, Vol. 224 No. 3, pp. 497–506.

Jalil, M.N., Zuidwijk, R.A., Fleischmann, M. and Van Nunen, J.A.E.E. (2011), "Spare parts logistics and installed base information", *The Journal of the Operational Research Society*, Vol. 62 No. 3, pp. 442–457.

Jat, M.N. and Muyldermans, L. (2013), "Time-differentiated service parts distribution: Costs under hierarchical and non-hierarchical setups", *2013 IEEE Workshop on Computational Intelligence In Production And Logistics Systems (CIPLS)*, Presented at the IEEE Symposium Series on Computational Intelligence 2013, Singapore, pp. 17–24.

Jayaraman, V. (1998), "Transportation, facility location and inventory issues in distribution network design An investigation", *International Journal of Operations & Production Management*, Vol. 18 No. 5, pp. 471–494.

Jayaraman, V., Gupta, R. and Pirkul, H. (2003), "Selecting hierarchical facilities in a service-operations environment", *European Journal of Operational Research*, Vol. 147 No. 3, pp. 613–628.

Jeet, V., Kutanoglu, E. and Partani, A. (2009), "Logistics network design with inventory stocking for low-demand parts: Modeling and optimization", *IIE Transactions*, Vol. 41 No. 5, pp. 389–407.

Jin, T. and Tian, Y. (2012), "Optimizing reliability and service parts logistics for a time-varying installed base", *European Journal of Operational Research*, Vol. 218 No. 1, pp. 152–162.

Van Kampen, T.J., Akkerman, R. and van Donk, D.P. (2012), "SKU classification: a literature review and conceptual framework", *International Journal of Operations & Production Management*, Vol. 32 No. 7, pp. 850–876.

Kaplan, A. (1969), "Stock rationing", *Management Science*, Vol. 15 No. 5, pp. 260–267.

Kennedy, W.J., Wayne Patterson, J. and Fredendall, L.D. (2002), "An overview of recent literature on spare parts inventories", *International Journal of Production Economics*, Vol. 76 No. 2, pp. 201–215.

Khawam, J., Hausman, W.H. and Cheng, D.W. (2007), "Warranty inventory optimization for Hitachi global storage technologies, Inc", *Interfaces*, Vol. 37 No. 5, pp. 455–471.

Klein, H.K. and Lyytinen, K. (1985), "The poverty of scientism in information systems", *Research methods in information systems*, pp. 131–161.

Klose, A. and Drexl, A. (2005), "Facility location models for distribution system design", *European Journal of Operational Research*, Logistics: From Theory to Application, Vol. 162 No. 1, pp. 4–29.

Kranenburg, A.A. and van Houtum, G.J. (2007a), "Cost optimization in the (S-1,S) lost sales inventory model with multiple demand classes", *Operations Research Letters*, Vol. 35 No. 4, pp. 493–502.

Kranenburg, A.A. and van Houtum, G.J. (2007b), "Effect of commonality on spare parts provisioning costs for capital goods", *International Journal of Production Economics*, Vol. 108 No. 1–2, pp. 221–227.

Kranenburg, A.A. and van Houtum, G.J. (2008), "Service differentiation in spare parts inventory management", *The Journal of the Operational Research Society*, Vol. 59 No. 7, pp. 946–955.

Kranenburg, A.A. and van Houtum, G.J. (2009), "A new partial pooling structure for spare parts networks", *European Journal of Operational Research*, Vol. 199 No. 3, pp. 908–921.

Kukreja, A., Schmidt, C.P. and Miller, D.M. (2001), "Stocking decisions for low-usage items in a multi-location inventory system", *Management Science*, Vol. 47 No. 10, pp. 1371–1383.

Kutanoglu, E. (2008), "Insights into inventory sharing in service parts logistics systems with time-based service levels", *Computers & Industrial Engineering*, Vol. 54 No. 3, pp. 341–358.

Kutanoglu, E. and Mahajan, M. (2009), "An inventory sharing and allocation method for a multi-location service parts logistics network with time-based service levels", *European Journal of Operational Research*, Vol. 194 No. 3, pp. 728–742.

Laszlo, M.J. (1996), *Computational geometry and computer graphics in C++*, Prentice Hall, Upper Saddle River, NJ.

Law, A.M. (2007), *Simulation modeling and analysis*, McGraw-Hill series in Industrial Engineering and Management Science, McGraw-Hill, Boston, 4th ed.

Lele, M.M. and Karmarkar, U.S. (1983), "Good product support is smart marketing", *Harvard Business Review*, Vol. 61 No. 6, pp. 124–132.

Levitt, T. (1983), "After the sale is over", *Harvard business review*, Vol. 61 No. 5, pp. 87–93.

Li, X., Zhao, Z., Zhu, X. and Wyatt, T. (2011), "Covering models and optimization techniques for emergency response facility location and planning: a review", *Mathematical Methods of Operations Research*, Vol. 74 No. 3, pp. 281–310.

Lösch, A. (1944), *Die räumliche ordnung der wirtschaft*, Verlag von Gustav Fischer.

Lösch, A. (1954), *The economics of location*, Yale University Press, New Haven ; London.

Lotan, T., Cattrysse, D. and Van Oudheusden, D. (1996), "Winter gritting in the province of Antwerp-a combined location and routing problem", *Belgian*

Journal of Operations Research, Statistics and Computer Science, Vol. 36 No. 3, pp. 141-157.

Maister, D.H. (1976), "Centralisation of inventories and the 'square root law'", *International Journal of Physical Distribution & Logistics Management*, Vol. 6 No. 3, pp. 124–134.

Male, J.W. and Liebman, J.C. (1978), "Districting and routing for solid waste collection", *Journal of the Environmental Engineering Division*, Vol. 104 No. 1, pp. 1–14.

Mangan, J., Lalwani, C. and Gardner, B. (2004), "Combining quantitative and qualitative methodologies in logistics research", *International Journal of Physical Distribution & Logistics Management*, Vol. 34 No. 7, pp. 565–578.

Marianov, V. and ReVelle, C. (1995), "Siting emergency services", *Facility Location: a survey of applications and methods*, pp. 199–223.

Marianov, V. and Serra, D. (2001), "Hierarchical location–allocation models for congested systems", *European Journal of Operational Research*, Vol. 135 No. 1, pp. 195–208.

Masuyama, S., Ibaraki, T. and Hasegawa, T. (1981), "The computational complexity of the m-center problems on the plane", *IEICE TRANSACTIONS (1976-1990)*, Vol. E64-E No. 2, pp. 57–64.

McCutcheon, D.M. and Meredith, J.R. (1993), "Conducting case study research in operations management", *Journal of Operations Management*, Vol. 11 No. 3, pp. 239–256.

McKay, K.N., Safayeni, F.R. and Buzacott, J.A. (1988), "Job-shop scheduling theory: What is relevant?", *Interfaces*, Vol. 18 No. 4, pp. 84–90.

Megiddo, N. and Supowit, K.J. (1984), "On the complexity of some common geometric location problems", *SIAM Journal on Computing*, Vol. 13 No. 1, pp. 182–196.

Mehrotra, A., Johnson, E.L. and Nemhauser, G.L. (1998), "An optimization based heuristic for political districting", *Management Science*, Vol. 44 No. 8, pp. 1100–1114.

Meller, R.D. (1995), "The impact of multiple stocking points on system profitability", *International Journal of Production Economics*, Vol. 38 No. 2–3, pp. 209–214.

Mello, J. and Flint, D.J. (2009), "A refined view of grounded theory and its application to logistics research", *Journal of Business Logistics*, Vol. 30 No. 1, pp. 107–125.

Mentzer, J.T. and Kahn, K.B. (1995), "A framework of logistics research", *Journal of Business Logistics*, Vol. 16 No. 1, pp. 231–250.

Meredith, J.R., Raturi, A., Amoako-Gyampah, K. and Kaplan, B. (1989), "Alternative research paradigms in operations", *Journal of Operations Management*, Vol. 8 No. 4, pp. 297–326.

Mingers, J. (2000), "The contribution of critical realism as an underpinning philosophy for OR/MS and systems", *The Journal of the Operational Research Society*, Vol. 51 No. 11, pp. 1256–1270.

Mirchandani, P.B. (1987), "Generalized hierarchical facility locations", *Transportation Science*, Vol. 21 No. 2, pp. 123–125.

Moinzadeh, K. and Schmidt, C.P. (1991), "An (S - 1, S) inventory system with emergency orders", *Operations Research*, Vol. 39 No. 2, pp. 308.

Moore, G.C. and ReVelle, C. (1982), "The hierarchical service location problem", *Management Science*, Vol. 28 No. 7, pp. 775–780.

Morgan, F. and Bolton, R. (2002), "Hexagonal economic regions solve the location problem", *The American mathematical monthly*, Vol. 109 No. 2, pp. 165–172.

Morris, M. (2013), "Forecasting challenges of the spare parts industry", *Journal of Business Forecasting*, Vol. 32 No. 3, pp. 22–27.

Muckstadt, J.A. (2005), *Analysis and algorithms for service parts supply chains*, Springer, New York, NY.

Muckstadt, J.A. (2010), *Principles of inventory management: when you are down to four, order more*, Springer series in operations research and financial engineering, Springer, New York.

Murray, A.T., O'Kelly, M.E. and Church, R.L. (2008), "Regional service coverage modeling", *Computers & Operations Research*, Vol. 35 No. 2, pp. 339–355.

Muyldermans, L., Cattrysse, D. and Oudheusden, D.V. (2003), "District design for arc-routing applications", *Journal of the Operational Research Society*, Vol. 54 No. 11, pp. 1209–1221.

Muyldermans, L., Cattrysse, D., Van Oudheusden, D. and Lotan, T. (2002), "Districting for salt spreading operations", *European Journal of Operational Research*, Vol. 139 No. 3, pp. 521–532.

Nahmias, S. and Demmy, W.S. (1981), "Operating characteristics of an inventory system with rationing", *Management Science*, Vol. 27 No. 11, pp. 1236–1245.

Narula, S.C. (1984), "Hierarchical location-allocation problems: A classification scheme", *European Journal of Operational Research*, Vol. 15 No. 1, pp. 93–99.

Näslund, D. (2002), "Logistics needs qualitative research–especially action research", *International Journal of Physical Distribution & Logistics Management*, Vol. 32 No. 5, pp. 321–338.

Nozick, L.K. and Turnquist, M.A. (2001), "Inventory, transportation, service quality and the location of distribution centers", *European Journal of Operational Research*, Vol. 129 No. 2, pp. 362–371.

Okabe, A., Boots, B., Sugihara, K. and Chiu, S.N. (2000), *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, Wiley-Blackwell, Chichester ; New York, 2nd Edition edition.

O'Kelly, M.E. and Storbeck, J.E. (1984), "Hierarchical location models with probabilistic allocation", *Regional Studies*, Vol. 18 No. 2, pp. 121–129.

Owen, S.H. and Daskin, M.S. (1998), "Strategic facility location: A review", *European Journal of Operational Research*, Vol. 111 No. 3, pp. 423–447.

Paterson, C., Kiesmüller, G., Teunter, R. and Glazebrook, K. (2011), "Inventory models with lateral transshipments: A review", *European Journal of Operational Research*, Vol. 210 No. 2, pp. 125–136.

Patton, J.D. (1986), *Logistics technology and management: the new approach : a comprehensive handbook for commerce, industry, government*, Solomon Press, New York.

Perrier, N., Langevin, A. and Campbell, J.F. (2006), "A survey of models and algorithms for winter road maintenance. Part I: system design for spreading

and plowing", *Computers & Operations Research*, Vol. 33 No. 1, pp. 209–238.

Perrier, N., Langevin, A. and Campbell, J.F. (2008), "The sector design and assignment problem for snow disposal operations", *European Journal of Operational Research*, Vol. 189 No. 2, pp. 508–525.

Pezzella, F., Bonanno, R. and Nicoletti, B. (1981), "A system approach to the optimal health-care districting", *European Journal of Operational Research*, Vol. 8 No. 2, pp. 139–146.

Plastria, F. (1995), "Continuous location problems", *Facility Location: a survey of applications and methods*, Springer, New York.

Plastria, F. (2002), "Continuous covering location problems", *Facility Location: Applications and Theory*, Springer, Berlin.

Poole, K. (2003), "Seizing the potential of the service supply chain", *Supply Chain Management Review*, Vol. 7 No. 4, pp. 54–61,3.

Pourakbar, M. and Dekker, R. (2012), "Customer differentiated end-of-life inventory problem", *European Journal of Operational Research*, Vol. 222 No. 1, pp. 44–53.

Pourakbar, M., Frenk, J.B.G. and Dekker, R. (2012), "End-of-life inventory decisions for consumer electronics service parts", *Production & Operations Management*, Vol. 21 No. 5, pp. 889–906.

Reese, J. (2006), "Solution methods for the p-median problem: An annotated bibliography", *Networks*, Vol. 48 No. 3, pp. 125–142.

ReVelle, C.S., Williams, J.C. and Boland, J.J. (2002), "Counterpart models in facility location science and reserve selection science", *Environmental Modeling & Assessment*, Vol. 7 No. 2, pp. 71–80.

Ronen, D. (1990), "Inventory centralization/decentralization--the 'square root law' revisited again", *Journal of Business Logistics*, Vol. 11 No. 2, pp. 129–138.

Rumyantsev, S. and Netessine, S. (2007), "What can be learned from classical inventory models? A cross-industry exploratory investigation", *Manufacturing & Service Operations Management*, Vol. 9 No. 4, pp. 409–429.

Rustenburg, W.D., van Houtum, G.J. and Zijm, W.H.M. (2001), "Spare parts management at complex technology-based organizations: An agenda for research", *International Journal of Production Economics*, Vol. 71 No. 1–3, pp. 177–193.

Şahin, G. and Süral, H. (2007), "A review of hierarchical facility location models", *Computers & Operations Research*, Vol. 34 No. 8, pp. 2310–2331.

Sang-Hyun Kim, Cohen, M.A. and Netessine, S. (2007), "Performance Contracting in After-Sales Service Supply Chains", *Management Science*, Vol. 53 No. 12, pp. 1843–1858.

Sargent, T.A. and Kay, M.G. (1995), "Implementation and utilization of a decentralized storage system: Costing model", *International Journal of Operations & Production Management*, Vol. 15 No. 9, pp. 210.

Satır, B., Savasaneril, S. and Serin, Y. (2012), "Pooling through lateral transshipments in service parts systems", *European Journal of Operational Research*, Vol. 220 No. 2, pp. 370–377.

Schilling, D.A., Jayaraman, V. and Barkhi, R. (1993), "A review of covering problems in facility location", *Location Science*, Vol. 1 No. 1, pp. 25–55.

Schrady, D.A. (1967), "A deterministic inventory model for reparable items", *Naval Research Logistics Quarterly*, Vol. 14 No. 3, pp. 391–398.

Schultz, C.R. (2004), "Spare parts inventory and cycle time reduction", *International Journal of Production Research*, Vol. 42 No. 4, pp. 759–776.

Silver, E.A., Pyke, D.F. and Peterson, R. (1998), *Inventory management and production planning and scheduling*, John Wiley & Sons, New York, 3rd ed.

Simchi-Levi, D. (1992), "Hierarchical planning for probabilistic distribution systems in Euclidean spaces", *Management Science*, Vol. 38 No. 2, pp. 198–211.

Skiera, B. and Albers, S. (1998), "COSTA: Contribution Optimizing Sales Territory Alignment", *Marketing Science*, Vol. 17 No. 3, pp. 196–213.

Smykay, E.W. and Bowersox, D.J.P. (1973), *Physical Distribution Management*, Macmillan, New York, 3rd ed.

De Souza, R., Tan, A.W.K., Othman, H. and Garg, M. (2011), "A proposed framework for managing service parts in automotive and aerospace industries", *Benchmarking*, Vol. 18 No. 6, pp. 769–782.

Starr, M.K. and Miller, D.W. (1962), *Inventory control: Theory and practice*, Prentice-Hall, New Jersey.

Stern, N. (1972), "The optimal size of market areas", *Journal of Economic Theory*, Vol. 4 No. 2, pp. 154–173.

Stone, R.E. (1991), "Technical notes. Some average distance results", *Transportation Science*, Vol. 25 No. 1, p. 83.

Stuart, I., McCutcheon, D., Handfield, R., McLachlin, R. and Samson, D. (2002), "Effective case research in operations management: a process perspective", *Journal of Operations Management*, Vol. 20 No. 5, pp. 419–433.

Suomala, P., Sievänen, M. and Paranko, J. (2002), "The effects of customization on spare part business: A case study in the metal industry", *International Journal of Production Economics*, Vol. 79 No. 1, pp. 57–66.

Suzuki, A. and Drezner, Z. (1996), "The p-center location problem in an area", *Location Science*, Vol. 4 No. 1–2, pp. 69–82.

Suzuki, A. and Okabe, A. (1995), "Using Voronoi diagrams", *Facility Location: A Survey of Applications and Methods*, Springer, New York.

Tallon, W.J. (1993), "The impact of inventory centralization on aggregate safety stock: The variable supply lead time case", *Journal of Business Logistics*, Vol. 14 No. 1, pp. 185–203.

Tempelmeier, H. (2006), "Supply chain inventory optimization with two customer classes in discrete time", *European Journal of Operational Research*, Vol. 174 No. 1, pp. 600–621.

Teunter, R.H. and Klein Haneveld, W.K. (2008), "Dynamic inventory rationing strategies for inventory systems with two demand classes, Poisson demand and backordering", *European Journal of Operational Research*, Vol. 190 No. 1, pp. 156–178.

Thompson, W.W. (1967), *Operations research techniques*, Merrill's mathematics and quantitative methods series, C.E. Merrill Books, Columbus, Ohio.

Topkis, D.M. (1968), "Optimal ordering and rationing policies in a nonstationary dynamic inventory model with n demand classes", *Management Science*, Vol. 15 No. 3, pp. 160–176.

Toregas, C., Swain, R., ReVelle, C. and Bergman, L. (1971), "The location of emergency service facilities", *Operations Research*, Vol. 19 No. 6, pp. 1363–1373.

Veinott Jr, A.F. (1965), "Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes", *Operations Research*, Vol. 13 No. 5, p. 761.

Vigoroso, M. and Gecker, R. (2005), "The service parts management solution selection report", Aberdeen Group, available at: http://v1.aberdeen.com/access/research_studies/Default.asp?tfm_order=ASC &tfm_orderby=AuthorLastName&selection=ch_svc&offset=105 (accessed 1 May 2010).

Wagner, S.M. and Lindemann, E. (2008), "A case study-based analysis of spare parts management in the engineering industry", *Production Planning and Control*, Vol. 19 No. 4, pp. 397–407.

Wang, Y., Cohen, M.A. and Zheng, Y.-S. (2002), "Differentiating customer service on the basis of delivery lead-times", *IIE Transactions*, Vol. 34 No. 11, pp. 979–989.

Watkins, K. (1993), *Discrete Event Simulation in C*, McGraw-Hill, Inc., New York, NY, USA.

Whitt, W. (1999), "Partitioning customers into service groups", *Management Science*, Vol. 45 No. 11, pp. 1579–1592.

Wong, K. and Beasley, J. (1984), "Vehicle routing using fixed delivery areas", *Omega*, Vol. 12 No. 6, pp. 591–600.

Yang, G., Dekker, R., Gabor, A.F. and Axsäter, S. (2013), "Service parts inventory control with lateral transshipment and pipeline stockflexibility", *International Journal of Production Economics*, Anticipation of risks impacts and industrial performance evaluation in distributed organizations life cycles, Vol. 142 No. 2, pp. 278–289.

Yin, R.K. (1994), *Case study research: design and methods*, Applied social research methods series, Sage, Thousand Oaks ; London, 2nd ed.

Zhen, L. (2012), "An analytical study on service-oriented manufacturing strategies", *International Journal of Production Economics*, Supply Chain Risk Management, Vol. 139 No. 1, pp. 220–228.

Zinn, W., Levy, M. and Bowersox, D.J. (1989), "Measuring the effect of inventory centralization/decentralization on aggregate safety stock: The 'square root law' revisited", *Journal of Business Logistics*, Vol. 10 No. 1, pp. 1–14.

Zipkin, P.H. (2000), *Foundations of Inventory Management*, McGraw-Hill Higher Education, Boston.

Zoller, K. (2005), "In-stock service differentiation using nested protection", *International Journal of Logistics: Research and Applications*, Vol. 8 No. 3, pp. 193–210.

Zoller, K. (2006), "Differentiating in-stock service by customer value", *International Journal of Logistics: research and applications*, Vol. 9 No. 2, pp. 95–110.