Capacity flexibility of a maintenance service provider in specialized and commoditized system environments

DOI:
10.6100/IR738099

Document status and date:
Published: 01/01/2012

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Capacity Flexibility of a Maintenance Service Provider in Specialized and Commoditized System Environments

This thesis is number D158 of the thesis series of the Beta Research School for Operations Management and Logistics. The Beta Research School is a joint effort of the departments of Industrial Engineering & Innovation Sciences, and Mathematics and Computer Science at Eindhoven University of Technology and the Center for Production, Logistics and Operations Management at the University of Twente.

A catalogue record is available from the Eindhoven University of Technology Library.

ISBN: 978-90-8891-479-9

Printed by Proefschriftmaken.nl

Cover Design by Niels Groenendijk

Capacity Flexibility of a Maintenance Service Provider in Specialized and Commoditized System Environments

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. C.J. van Duijn, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op woensdag 3 oktober 2012 om 16.00 uur

door

Nasuh Çagdas Büyükkaramikli

geboren te Ankara, Turkije

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr.ir. J.W.M. Bertrand
en
prof.dr. A.G. de Kok

Copromotor:
dr.ir. H.P.G. van Ooijen

Table of Contents

1 Introduction

This thesis studies the integrated down-time service and capacity management of a **m**aintenance **s**ervice **p**rovider (MSP), which is running a repair shop in an environment with numerous operating systems that are prone to failure. The MSP is responsible for keeping all systems in an environment up and working. We mainly focus on two types of environments: 1) Specialized System Environment 2) Commoditized System Environment.

The systems in the first, specialized system environment are highly customized. They are designed and built specifically, following the owners' precise requirements. Mostly, these specialized systems have a modular design and consist of several smaller subsystems. The same sub-system type can be a common part of several different specialized systems. Complex defense systems, specific lithography systems, mission aircrafts or other advanced/complex, engineer-to-order capital goods are examples of such specialized systems. Due to the diversity of owners' requirements, each system develops many unique characteristics, which make it hard, if not impossible, to find a substitute for the system, in the market as a whole.

In the second environment, the systems are more generic in terms of their functionality. Trucks, cranes, printers, copy machines, forklifts, computer systems, cooling towers, power systems are examples of such more commoditized systems. Due to the more generic features of the owners' requirements, it is easier to find a substitute for a system in the market, with more or less the same functionality.

Upon a system breakdown, the defective unit (system/subsystem) is sent to the repair shop, which is operated by the MSP. The MSP is not only responsible for the repair of the defective units, but also liable for the costs related to the down-time. In order to alleviate the down-time costs, there are chiefly two different down-time service strategies that MSP can follow, depending on the environment the repair shop is operating in.

First strategy is ideal for the systems in the specialized system environment. In this strategy, MSP holds a spare unit inventory for the critical subsystem that causes most of the failures. The down-time service related decision in such a case would be the inventory level of the critical spare subsystems.

On the other hand, in the commoditized system environment, rather than keeping a spare unit inventory, the MSP hires a substitute system from an agreed rental store/3rd party supplier. The down-time service related decision in this second strategy is the hiring duration.

Next to the down-time service decisions above (spare unit inventory level in the specialized system environment and the hiring duration in the commoditized system environment), the repair shop's capacity level is the other primary determinant of the systems' uptime/availability. The increasing role of the after-sales services and the pressure for profitability makes the efficient use of maintenance capacity more than an obligation, which inspires us to investigate the prospects of cost savings from capacity volume flexibility in repair shops.

In light of the discussions above, in this thesis, we focus on the integration of the repair shop capacity related decisions and the down-time service related decisions under both specialized system and commoditized system environments with different capacity volume flexibility alternatives. In the specialized system environment, the down-time service related decision is the spare unit inventory level, whereas in the commoditized system environment, the down-time service related decision is the hiring duration.

The remainder of this introductory chapter is organized as follows. In section 1.1, we introduce the main concepts that shaped our motivation, elaborate on them in more detail and provide the relevant literature therein. Afterwards, in section 1.2, we explain the characteristics of the system environment, problem and the main players in our study setting. Finally, in section 1.3, we discuss the research questions, used methodologies and the further outline of this thesis.

1.1 Problem Context and Key Concepts

1.1.1 After-Sales Services

There are many industrial research reports and academic studies in the literature that advice product manufacturers to add after-sales service to their product offerings. For instance, the Aberdeen Group, a research consultancy firm, reported that the spare parts sales accounted for 8% of the annual gross domestic product in United States and global spending on after-sales services added up to $1.5 trillion annually (Aberdeen Group 2003). Similarly, a study conducted by Deloitte Consulting reveals that the average growth of the service businesses of the companies is 10% higher than for the business units overall (Deloitte 2006).

Many production companies are following this trend and shift from pure manufacturing to an integrated approach that includes the servicing of their products. (Oliva & Kallenberg 2003) study and analyze 11 capital equipment manufacturers that realized this transition and developed service offerings for their products.

However, the transition of a manufacturing company from its core business to an integrated after-sales service business may not be a smooth process. In a recent study, the major challenges that firms may face in starting their aftermarket operations are listed (Cohen et al. 2006). In the center of these challenges lie the differences between a manufacturing supply chain and an after-sales service chain. Table 1-1 summarizes the main differences between these two chains.

Unlike physical products, businesses cannot produce services in advance of demand. The service is demanded only when an unpredictable event, such as a system failure, triggers a need. Furthermore, fulfilling demand in the after-sales services supply chain involves the customer for its realization. The key performance metrics are also different: for manufacturing supply chain, the product fill rate is of concern, whereas for the after-sales services supply chain, the system availability or in another form, the system uptime is the central focus.

Parameter	Manufacturing Supply Chain	After-Sales Services Supply Chain
Nature of Demand	Can be forecasted to some extent	More unpredictable, sporadic
Required Response	Standard, scheduled	ASAP (same day or next day)
Inventory Management Aim	Maximize the velocity of materials	Pre-position resources
Reverse Logistics	Doesn't handle	Handles return, repair and disposal of failed components
Performance Metric	Fill rate	Product-availability (uptime)
Inventory Turns	6-50 per year	1-4 per year

Table 1-1 Differences between manufacturing and after-sales services supply chains (Cohen et al. 2006)

These differences and other complications may squander the manufacturers' economic potential in after-sales service market. A survey study published in Mckinsey Quarterly also substantiated the underperformance of production firms transition to after-sales services in terms of revenues (Alexander et al. 2002). This under-utilized revenue potential impels either the emergence of new service providers in the aftermarket or the evolution of the after sales departments of manufacturers as semi-autonomous business units/organizations. There are many after-sales services that might be provided by these autonomous/semi-autonomous units such as system user support/technical education, in exchange for a service fee. However, in this thesis, we focus on the maintenance aspect of the after sales services, since the majority of the after-sales costs are due to the system down-time, and the down-time of the systems can be controlled primarily by the maintenance activities. Therefore, in this thesis, we analyze the operations of a maintenance service provider (MSP), which is responsible for the uptime of its customers' systems in exchange for a service fee.

1.1.2 Maintenance

Maintenance can be defined as the total of activities required to retain the systems in, or restore them to, the state necessary for the fulfillment of the production function (Gits 1992). In this definition, the activities to "retain in" are considered to be under the umbrella of "preventive maintenance", whereas the activities to "restore" are considered to be under the umbrella of "corrective maintenance". In contrast to preventive maintenance, corrective maintenance actions are taken after a failure/ breakdown, which make them difficult to plan in advance. In addition, previous studies report that the responsiveness of the corrective maintenance activities is decisive on the duration of down-time (Coetzee 2004). Motivated by the planning challenges that the uncertainty brings as well as the observations in the academia/industry that most of the maintenance actions that are performed are corrective maintenance (Vliegen 2009), we narrow down our focus on corrective maintenance activities in this thesis.

Generally speaking, maintenance activities necessitate three types of resources: capacity (manpower), tools and materials. It has been observed (e.g. (Keizers 2000) and (Schmenner 1995)) that tools generally do not appear to be bottleneck resources in maintenance activities that take place mostly in repair shops (i.e. when there is no/limited field repair activity). On the other hand, capacity management is quite critical, because maintenance itself is labor intensive and workforce capacity is needed during the entire processing time of a maintenance job. Similarly, materials management is the other critical pillar. Especially, after the industrial development of interchangeable parts, sound management of the availability of the parts can reduce the down-time of the systems drastically, due to the *repair by replacement* concept. Repair by replacement infers the following: if a critical part fails and leads to a system failure, the system is restored by replacing the defective part with a new, ready for use one. The decisions on capacity and material resources are very interrelated; therefore integrated decision making for both resources is needed. Next we review the literature on maintenance briefly.

1.1.2.1 Literature on Maintenance

There has been considerable research on maintenance policies and practices. We refer the interested readers to (Pierskalla & Voelker 1976), (Sherif & Smith 1981) , (Cho & Parlar 1991), (Wang, 2002), which provide extensive surveys and reviews of the maintenance literature. Also, for a framework of maintenance to classify problems and research, we refer the reader to "The EUT Maintenance Model", a descriptive model developed by (Geraerds 1992), which describes the sub-functions within maintenance and their inter-relations. As mentioned before, we focus on corrective maintenance activities in this thesis.

Despite the sheer volume of studies conducted on corrective maintenance, the number of studies that incorporate the repair capacity in the maintenance systems is limited. These studies can be mainly classified into two groups based on the repair environment, namely machine interference/repairman problem environment and repairable item inventory problem environment.

The machine interference/repairman problem involves a finite population of machines, operating under the supervision of a number of repairmen (or repair facilities or servers), who repair the machine as they break down in a non-pre-emptive manner. (Stecke & Aronson 1985) provide a survey for the performance analysis of the models in this setting. There are also studies that analyze the optimal control for the machine interference problem. The control can be realized either by variable service rates (Crabill 1974), (Winston 1977), or by variable number of repair service facilities (Winston 1978). Further extensions of these models include (Goheen 1977), who assumed Erlang distribution for failure and repair times, (Albright 1980), who included both repair rate and the number of repair service facilities as control variables and (Van Der Duyn Schouten & Wartenhorst 1993), who included general failure and repair service distributions. Another stream of researchers extends the single class problem to multi-class machine-interference problems with heterogeneous machines. (Chandra & Shanthikumar 1983), (Chandra 1986), (Agnihothri 1989) and (Kameda 1982) all analyze different extensions of the classical homogenous single class machine interference problems. Further studies such as (Shawky 1997), (Iravani & Krishnamurthy 2007) incorporated the cross training issues for the repairmen into the multi-class machine-interference problems.

The second group of papers considers the management issues in a repairable-item inventory setting. In these problems, it is considered that a machine is composed of parts and upon a part failure, the failed part is replaced by a spare part, if it is available in the inventory. The control of the inventory levels in single-echelon and multi-echelon systems have been an area of interest for both practitioners and academicians. METRIC (Multi-Echelon Technique for Recoverable Item Control), is the famous approach for the stock allocation problem for repairable items and is developed in 60's. The METRIC approach (Sherbrooke 1967), is a greedy and iterative heuristic that increases the inventory level of a certain item at a certain location in each iteration. This approach spawned much further research. For instance basic METRIC model is extended to the multi-indenture case by MOD-METRIC approach (Muckstadt 1973), and the variance of the pipeline is incorporated to obtain more accurate approximations, which resulted in VARI-METRIC models (Sherbrooke, 1986). All of these studies assume a constant failure rate and an ample repair capacity. Later, (Gross et al. 1983), (Diaz & Fu, 1997), (Perlman et al. 2001) and (Sleptchenko et al. 2003) provide extensions of the existing VARI-METRIC methods by replacing the infinite server queuing model by different, finite capacity systems.

Aside from these METRIC based models; there are studies that deal with modeling of the capacitated service networks via closed/open queuing networks, e.g. (Zijm & Avsar 2003) and other studies are mostly based on the analysis of Markov Process such as (Gupta & Albright 1992) and (Albright & Gupta 1993).

Flexible capacity control in repairable-item inventory models is a rather understudied topic. Many simulation studies are conducted in order to explore the benefits of the use of overtime policies in a repairable-item inventory setting, e.g. (Scudder 1985) and (Scudder & Chua 1987). To the best of our knowledge, (de Haas 1995) is the last study which sheds light to the problem of integration of the flexible manpower and the initial stock decisions in repairable item systems. In all of these simulation studies mentioned above, overtime decisions are not periodic, but can be taken at any point in time.

Another stream that worked on the flexible maintenance workforce planning is the maintenance scheduling literature. Studies on workforce-constrained maintenance scheduling problem developed several meta-heuristic approaches that analyze different aspects of the problem such as conflicting objectives, precedence relations, priority setting, etc. Most of these studies either assumed deterministic/given repair job time requirements (e.g. Yan et al. 2004), or incorporated the randomness by simulation (e.g. Safaei et al. 2010). In (Yang et al. 2003), it has been shown that flexible manpower strategies reduced the operating costs significantly in a test bed that is built on the operating data from a leading Taiwanese airline company.

1.1.3 Capacity Flexibility

Researchers unanimously agree that flexibility is an essential requirement for organizations and systems for a better responsiveness (Bertrand 2003), whereas there is no consensus on the definition of it. The lack of a consensual agreement on the demarcation is due to the fact that the concept of flexibility is conceptually broad, multi-dimensional (Suarez et al. 1995) and polymorphous (Evans 1991). However, volume flexibility, focus of our thesis, is more amenable to definition. It is defined as the ability of an organization to change volume (of output) levels in response to changing socio-economic conditions profitably and with minimal disruptions (Jack & Raturi 2002). There are different drives and sources of volume flexibility.

In this thesis we assume that a change in volume is only possible by a change in the capacity level of the repair shop, therefore we use “capacity flexibility” instead of volume flexibility in the remaining part of the thesis. Empirical studies show that flexible capacity management policies (e.g. flexible staffing, under/over working hours, outsourcing) are commonly used in the manufacturing as well as service industries (Houseman 2001) and (Kalleberg et al. 2003).

For various reasons, most of the time, capacity flexibility can be practiced only periodically. Firstly, a company’s reach to the external capacity pool may be restricted to certain specific times like the start of a day or the start of a week. Secondly, decisions about working times (e.g. working over/under time) are often taken on a periodic basis, in order to abide to labor regulations and to accomplish the timely communication of these working time decisions to the relevant employees. In addition, periodic flexible capacity policies are compatible with the modus operandi of resource planning software systems, most of which also operate on a periodic basis due to decision-information synchronization issues (see e.g. (ORACLE 2000)).

Owing to the reasons listed above, we concentrate on periodic capacity flexibility in this thesis.

1.1.3.1 Literature on Capacity Flexibility/ Capacity Management

In this subsection, we provide a brief overview of the capacity management literature. Most of the capacity management research is studied in production/service environments.

Decisions on capacity investment are first studied in Economics/Econometrics literature as capacity investment problems. See e.g. (Chenery 1952), (Eberly & van Mieghem 1997). These studies take a holistic view on the interactions between the capacity decisions and the performance of the system. A more detailed modeling is necessary for the planning of operations in a production or a service environment.

(Holt et al. 1960) were the first to address the problem of the coordination of production, inventory and capacity decisions, and they develop the aggregate planning model, in which the production, inventory and workforce decisions (such as hiring/firing and over time/under time working hours) are taken for a finite horizon based on forecasted demand over that horizon with the help of linear decision models.

(Pinker 1996), (Milner & Pinker 2001), (Pinker & Larson 2003) develop models with different types of flexible capacity arrangements (such as contingent labor contracting or overtime working hours) in the presence of demand/supply uncertainty over a finite discrete-time horizon. In these studies, different stochastic dynamic programming models are presented in order to obtain the optimal decisions on the capacity levels. Similarly, (Kouvelis & Milner 2002) study the interplay of demand and supply uncertainty in capacity and outsourcing decisions in multi-stage supply chains. These models incorporate several factors of permanent/contingent capacity or outsourcing structures, however the inherent congestion effects of a production/service system are not analyzed.

Later studies extend the problem to integrated capacity and inventory control. (Bradley & Glynn 2002) provided a Brownian motion approximation to study the joint optimal control of the inventory and the capacity in a make-to-stock system. Similarly, (Alp & Tan 2008) use stochastic dynamic programming formulations for the integrated capacity and inventory management problem of a make-to-stock system.

Another relevant research stream is the call center capacity management literature. The rise of the industry in the late 90’s revived the call-center research stream again. Different from maintenance/repair environments, call center environments are very fast moving and the systems are mostly operating under heavy traffic. Therefore, most of the studies in this stream use either heavy traffic or fluid approximations for the workload process in their models. See (Gans et al. 2003) for a general overview, tutorial and a list of prospects for the call-center research.

If a production/service system is modeled as a queuing system, the service rate (or number of servers) of the queue can be interpreted as the capacity level. Mostly, stochastic dynamic programming formulations are utilized in order to determine the optimal service rates of the queuing systems with the help of the uniformization technique (Lippman 1975). (Sennott 1998) provides a comprehensive overview of the usage of stochastic dynamic programming in queuing systems for different control aspects. In most of the queuing control studies that use dynamic programming, the capacity actions are taken based on an event occurrence and the average delay (or equivalently average number of customers) in the system is penalized.

Other approaches than dynamic control often necessitate the performance evaluation of the system first. For instance (Bekker & Boxma, 2007) and (Bekker et al. 2004) first provide performance analysis of several queuing systems with variable service rates. (Zijm & Buitenhek, 1996) discuss a framework for capacity planning and lead time management in manufacturing companies, with an emphasis on the machine shop, where queuing theory results are used to derive approximations for the necessary performance measures.

In spite of its practical relevance, periodic capacity control in queuing systems is not as popular as its continuous counterpart. This can be due to the complexity of deriving the transient queuing behavior that is needed for the analysis of the performance of systems under periodic capacity control. In (Yoo 1996), an unconstrained dynamic programming model is formulated to address the problem for setting staffing levels at a post office’s service window over multiple periods, including the transient behavior of the queue. Afterwards, (Fu et al. 2000) prove the monotonicity of the optimal control by establishing the sub-modularity of the objective function with regard to the initial queue size and the staffing level. In a single processing unit (Buyukkaramikli et al. 2011-a) and parallel processing unit MTO environments, (Buyukkaramikli et al. 2011-b), analyze threshold-type, two-level periodic capacity control policies for single server and multi-server MTO systems, respectively. In both of the studies, it is assumed that capacity actions can be only taken at equidistant points in time and the systems in consideration operate under a lead time performance constraint. In this thesis, we also stick to the assumption that periodicity is a *sine qua non* condition for the realization of the capacity volume flexibility for the repair environments that we study.

1.1.4 Compensating Differentials

Compensating differential is a term used in labor economics literature that denotes the additional amount of income necessary to compensate workers for the non-pecuniary disadvantages (such as risk, unpleasantness or other undesirable attributes) of a particular job. The basic conception of wage differentials dates back to late 18th century, at the beginning of Industrial Revolution (Smith 1776).

The topic is important for both theoretical and empirical research. Theoretically, it can make the legitimate claim to be the fundamental market equilibrium construct in labor economics on the conceptual level (Rosen 1986). Empirically, it contributes to the useful understanding of the determinants of the structure of wages in the market and to make inferences about preferences and technology from observed wage data. The bottom line of the concept is rooted in the utility theory and the theory asserts that workers receive compensating wage premiums when they accept jobs with undesirable nonwage characteristics, holding the worker’s characteristics constant.

As a framework of analysis, compensating wage differentials provide a solid explanation of the wage rate structure of the flexible capacity resources. In our thesis, we assume that a high frequency of decisions over the use of a flexible capacity resource as an undesirable attribute, since the corresponding worker will have less work security and has to be ready and available to be deployed more frequently. In a similar manner, the frequency of task switching is also considered to be undesired, since the level of concentration is disrupted more frequently and a higher working memory is needed for more frequent task switching activities (Rogers & Monsell 1995).

In order to reflect the compensating effects above, we provide several empirically testable functional forms for the wage rate of a unit flexible capacity resource. A common trait of these functions is that they are all decreasing with the period length (or increasing with the changing frequency).

1.1.4.1 Literature on Compensating Differentials

A large corpus of studies does exist on compensating wage differentials, however they are published mainly in economic/econometric journals, aiming to explore the causal relations between the wage rate and the other working conditions/factors. (Rosen 1986) provides an excellent summary of the concept and an overview of the literature.

Different studies focused on different aspects of wage differentials concept, however there are a few related studies that analyze the wage differentials for temporary/on-call and fixed term workers. Among the body of the literature, there are two studies that are particularly interesting to the context of this thesis. In (Hagen 2002), the risk premiums are studied for the temporary workers. In that study, there is no evidence for the wage differential for fixed time contracted workers, but the author commented that the sample size was too small. In (De Graaf-Zijl 2012), the role of uncertainty on the compensation of on-call and fixed term employment contracts is studied by using an analytical framework. In the paper, it is found that compensation differentials does exist for the future uncertainties/unemployment risks, and it is concluded that these are reflected to the temporary/on-call workers' wages additionally.

The impacts of wage differentials are reflected in the operations management literature in a narrow and limited manner. For instance, in many production planning models, the overtime costs are mostly more expensive than the regular-hour costs, see e.g. (Holt et al. 1960), (Bitran et al. 2011) and (Graves 1982). However in all of these studies, the models in concern are deterministic, which leads to non-stochastic demand and service requirements for each job. In addition, in all of these studies, the overtime wage rates are mostly taken constant per unit time. On the other hand, the increasing role of the capacity/workforce agencies in the market and the greater use of contractual agreements make the wage of flexible/temporary resources more responsive and reflexive to different system and usage characteristics. We take this trend into consideration while modeling the additional cost effects of the capacity action frequency in future chapters.

1.1.5 Commoditization

Increasing technology and the effective communication mediums have accelerated the commoditization of products/processes. Commoditization is a process during which non-commodity products become more like a commodity. Commodity is a good for which there is demand, but is supplied without significant functional differentiation across a market.

In most of the popular business publications, commoditization is portrayed as an inevitable tragic end-trap for organizations who cannot innovate, since they cannot generate profit from premium margins obtained from the unique products/ services. Although innovation of a new product/service is often prescribed to avoid the "wrath" of commoditization, the other side of the medallion is most of the time ignored: a ground breaking innovation of a product may require the commoditization of its subcomponents. For instance, the emergence of the smart phones necessitated the commoditization of processors, memory, screen, etc… This perspective frames commoditization as a natural process whose consequences fertilize the ground for the innovation of higher-level, better and more complicated products/ services. This perspective is in line with the insights from innovation theory and complex system theory, where the emergence of complex patterns arise out multiplicity of relatively simple patterns/interactions (Holland 2000).

The commoditization state can be considered as a continuum, which ranges from near zero commoditization at one end to fully commoditization at the other end. Figure 1-1 sketches this continuum of the commoditization state.

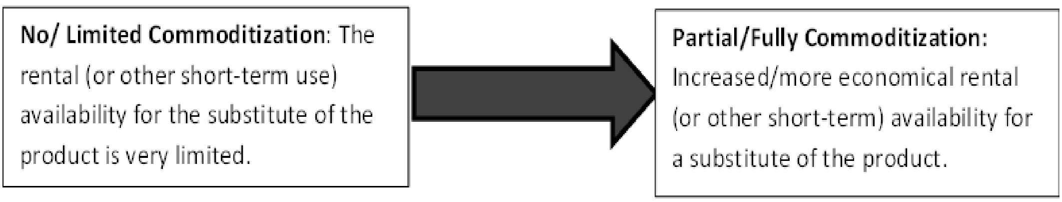

Figure 1-1 The commoditization trend (Holmes 2008)

In this thesis, systems under concern are considered to be either limited/no commoditized or partially/fully commoditized. We believe that as a product becomes more commoditized, the rental/leasing availability of that product, or a substitute, becomes much more common, more widely reachable and economically more attractive. We refer to this process as rentalization, and after the rentalization of a system, the immediate market availability for that system (or a substitute) for short term renting/leasing purposes can be achieved through rental/3rd party supplier channels.

1.1.5.1 Literature on Commoditization

Most of the studies on commoditization are published in popular business magazines. The conversion of the previously non-commodity market into a commodity market, which means declining profits and prices, is not a preferable situation and it is often destined for non-innovator manufacturing/service provider companies. Therefore, bulk of the studies in the literature tries to answer how to avoid or beat the commoditization trend with the best strategic response. Industry-specific studies include (Olson & Sharma 2008) for electronics industry, (Ealey & Troyano-Bermudez 1996) for automotive industry, (McLean 2007) for radiology industry and (Carr 2003) for IT service industry.

(Reimann et al. 2010) conduct an extensive survey in ten industries to better understand the commoditization phenomenon and its role and nature in evolving market competition. (Davenport 2005) sheds lights on the process commoditization and its effects on the business. A recent book, (Holmes 2008), summarizes the foundations of the stages of the commoditization, the impacts of commoditization on business level and individual level, and discusses the best responses of the companies to commoditization in order to survive.

The number of studies that conduct quantitative analysis of the effects of commoditization is limited. (Weil 1996) uses simulation and system dynamics methodology to explain the causal relationships between commoditization dynamics in service and technology-based markets. A similar approach is followed by (Manatayev 2004) for analyzing the commoditization in the third party logistics industry.

To the best of our knowledge, the impacts of the commoditization phenomenon have not been explicitly analyzed at the operational level in the literature. As discussed, after the commoditization, often, the substitute of a product becomes much more common, widely reachable and more economical. In this thesis, parallel to this discussion, we further assume that the rental/leasing (or other short term use) reachability of commoditized systems are quite high, such that the short term hire of a substitute system upon a failure becomes an alternative down-time service strategy for commoditized systems, rather than keeping spare unit stocks. The further effects of the rentalization and the commoditization of the systems on maintenance strategy/operations and their interactions with the capacity decisions will be analyzed more in detail in Chapter 3.

1.2 Problem under study

In this section, we describe the problem under study, which is motivated by the trends and the concepts discussed in the previous subsection. As mentioned earlier, we study the capacity flexibility management problem for a MSP operating in specialized/commoditized system environments, where the systems are prone to failure. Upon a failure, the defective units are sent to the repair shop to get repaired. The MSP is responsible for the availability of the systems so that the operations of the system owners can continue uninterruptedly. Therefore, the MSP is liable for the repair as well as the down-time costs resulting from the system unavailability.

In order to alleviate the down-time costs, in the specialized system environment, MSP holds a spare unit inventory for the most critical subsystem. On the other hand, in the commoditized system environment, MSP makes a long term agreement with a rental store/external 3rd party supplier, and upon a system failure, another substitute system is immediately supplied for a predetermined duration. The predetermined duration of the hiring period is necessary for the substitute system supplier, as it provides a degree of controllability for the rental/leasable asset utilization.

In the specialized system environment, inventory level is the down-time service related decision, however in the commoditized system environment, the hiring duration for the substitute is the down-time service related decision. Note that in the commoditized environment, it is still possible to both keep a spare unit inventory and make an agreement with an external 3rd party supplier, at the same time. In the thesis, we analyze this hybrid strategy as a special case in the later parts (end of Chapter 3) of the thesis.

1.2.1 Flexible Capacity Modes

The MSP aims to minimize its total relevant costs which is the sum of capacity costs, spare unit holding/substitute system hiring costs and down-time costs. The use of capacity flexibility in the integration of capacity and down-time service related decisions forms the leitmotif of this thesis.

We assume that there is a capacity/workforce agency, which can provide the contingent capacity resources to the repair shop, upon a need. Similarly, that capacity agency may have an interest in buying the repair shop's unused in-house capacity, since the agency can assign the idle repair shop capacity to other temporary tasks found in the market and may generate profit out of it. The capacity agency can be an external agency as well as an internal department within the MSP. We focus on periodic capacity flexibility and investigate three different capacity modes in this thesis:

Fixed Capacity Mode: In this mode, all of the capacity is permanent and ready for use in the repair shop. This mode serves as a reference point in order to assess the benefits of other flexible capacity modes. The relevant capacity decision in this mode is the single capacity level of the repair shop.

Periodic Two-Level Capacity Mode: In this mode, we assume two levels of repair shop capacity: permanent level and permanent plus contingent capacity level. The permanent capacity is always deployed in the repair shop, whereas the periodic deployment of the contingent capacity in the repair shop is decided at the start of each period based on the number of defective units waiting to be repaired in the shop. The relevant capacity decisions in this mode are the permanent and contingent capacity levels, the period length and the states (in terms of number of defective units waiting) where the contingent capacity is deployed.

Periodic Sell-back Capacity Mode: A condition for deploying this mode is that the failed units are sent to the repair shop at regular intervals in time. Due to this admission structure, when the repair of all the defective units in the repair shop are completed in a period, it is known that no new defective unit will arrive to the shop at least until the start of the next period, therefore the shop capacity will remain idle at least until the next interval. This allows for a contract, where the repair shop capacity, which is assumed to be multi-skilled and flexible, can be deployed at different tasks during these idle times. The original cost of the multi-skilled repair shop capacity per time unit is higher than that of the permanent capacity that is mentioned in the previous modes, and once the repair shop capacity becomes idle, the capacity is immediately sold at a reduced price back to the capacity agency until the next interval. The underlying reasoning and the motivation of this contract/cost structure will be explained further in Chapter 2 and Chapter 3. Other physical factors arising from the periodic admission structure, such as the pre-admission delay of a defective unit and the clustering of defective system/sub-system arrivals, will be examined in the later chapters, thoroughly. The relations between the capacity/workforce agency and the repair shop can be depicted for the second (Periodic Two-level) and the third (Periodic Sell-back) capacity modes in Figure 1-2a and Figure 1-2b respectively.

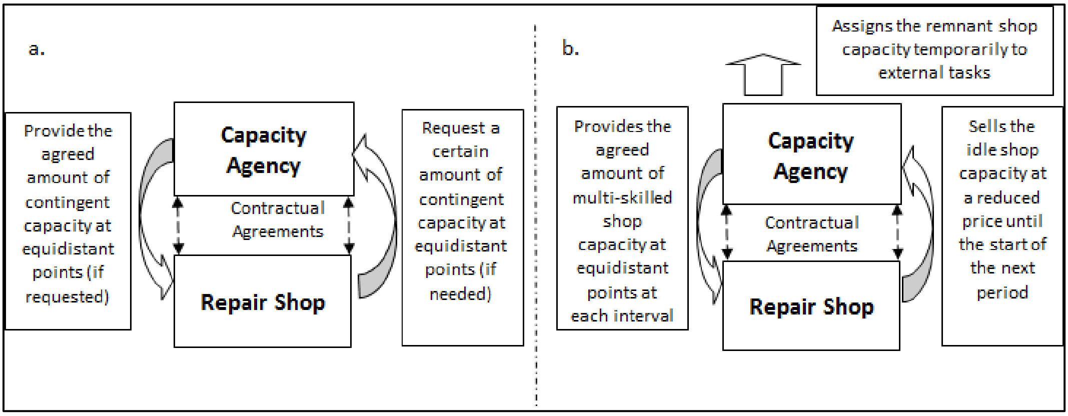

Figure 1-2 The relations between the repair shop and the capacity/workforce agency through contractual agreements in: a) Periodic Two-Level (on the left) b) Periodic Capacity Sell-back Mode (on the right)

Note that in the first capacity policy, the repair shop operates only with permanent capacity and the contractual agreement would enforce the provision of an agreed amount of capacity indefinitely, i.e. for an infinite time, which can be interpreted as the ownership of the capacity resources is taken over by the MSP. In this thesis, we investigate the performance of these three capacity modes for the repair shop, servicing for highly specialized or commoditized systems.

1.2.2 Specialized /Commoditized System Environments

The specialized systems are highly customized (frequently they are designed and built on demand) and not readily available in the market. We assume a specialized system consists of several subsystems and one critical subsystem causes most of the failures, therefore keeping a spare stock for that critical spare subsystem and using a spare subsystem upon failure in order to replace the defective subsystem (due to repair by replacement concept), if available, can reduce the system unavailability and down-time costs drastically. The down-time service related decision for this strategy is the stock level of the spare parts. In Figure 1-3, the actors and their interactions upon a system failure are sketched for specialized system environment.

On the other hand, (partially) commoditized systems that we study in this thesis are less customized and upon a failure, it is easier to find a substitute for the failed system in the market. Key property for the commoditized systems is that they are accessible for short term hiring purposes in the market through rental/other 3rd party supply channels, and we further assume that a 3rd party supplier agrees to provide a substitute system, at a fixed hiring rate, for a pre-determined duration, every time a system fails. We advocate a uniform and deterministic hiring duration for the substitute system due to practical

reasons which will be explained further in Chapter 3. The incurred hiring costs are non-refundable, i.e. if the repair of the defective system is completed before the hiring duration elapses, the hiring cost is still deducted based on the uniform hiring duration,

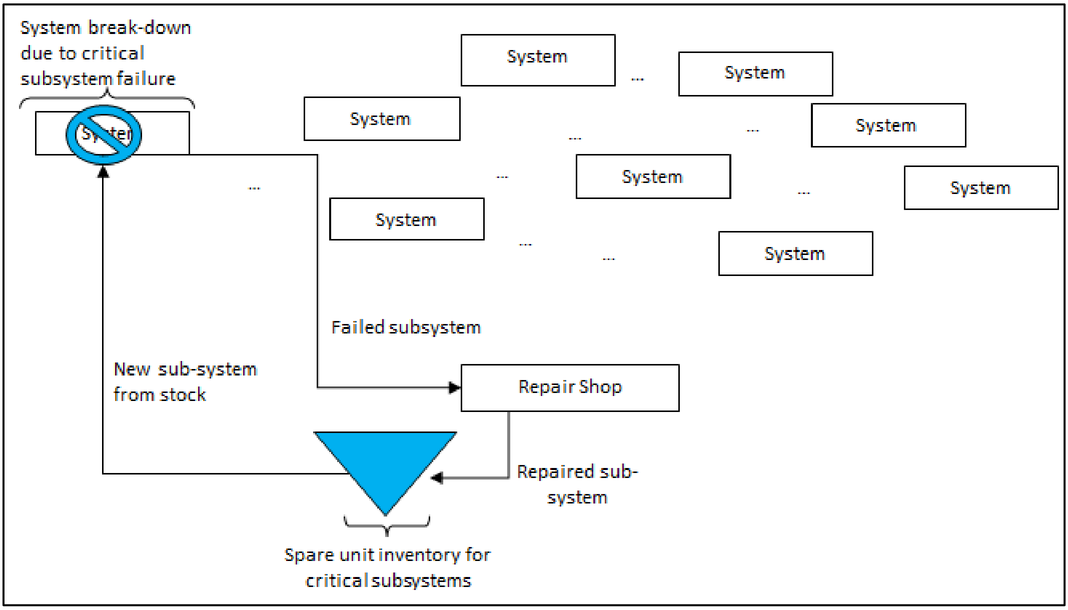

Figure 1-3 The actors and the interactions upon a system failure for specialized system environment.

not usage. On the other hand, if the repair of the defective system took longer than the hiring duration, the down-time cost per unit time is incurred during that non-covered time. Therefore, the down-time service related decision for the commoditized environment is the (uniform) hiring duration of a substitute system, which has to be decided judiciously, taking both the hiring and down-time costs into account. In Figure 1-4, the actors and their interactions upon a system failure are sketched for commoditized system environment.

Note that the special hybrid strategy, where MSP applies both keeping spare unit inventory and hiring substitute upon failure, will be explained further in the end of Chapter 3. Even though the optimal hybrid strategy can be more cost effective, the MSP may still have a tendency to apply the "hire only" strategy. This preference can be explained due to the fact that "hire only" strategy does not require any initial capital investment, unlike keeping a spare unit inventory, which necessitates the purchasing of the units in the stock at the beginning. This necessity implies that the repair shop manager, who is probably a small/medium sized enterprise manager, has to invest a serious amount of money initially for this “keeping spare unit” availability strategy. On the other hand, "hire only" strategy does not require such an initial investment but merely a service fee paid to the 3rd party supplier, due to the long-term agreement of uniform-duration hiring of the substitute system upon a system failure.

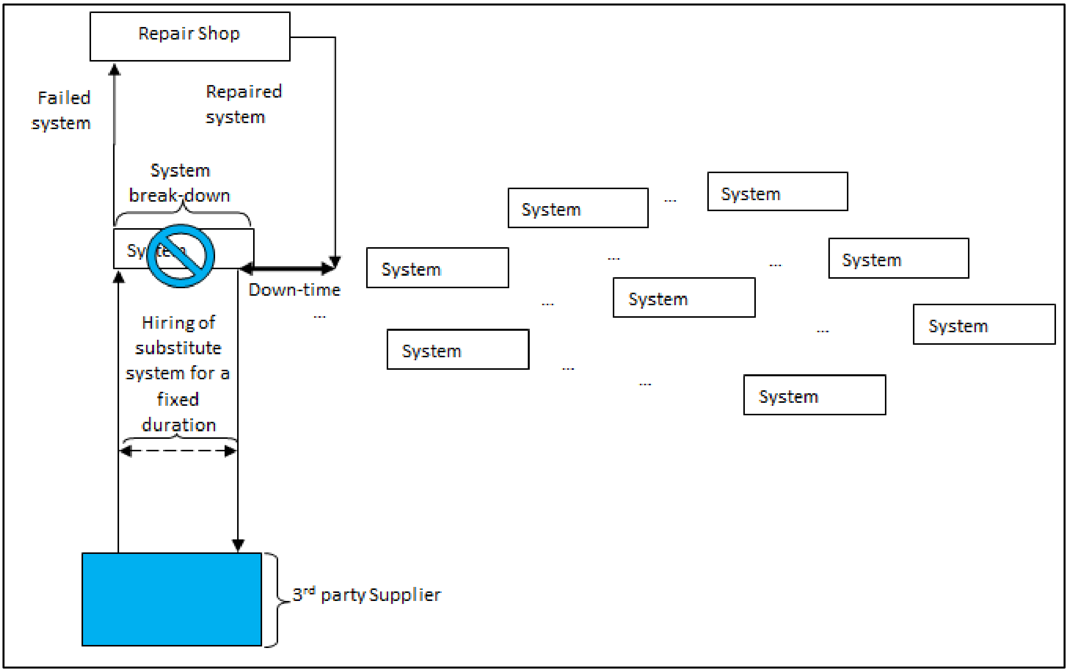

Figure 1-4 The actors and the interactions upon a system failure for (partially) commoditized systems.

In both of the environments, we assume that the MSP serves to numerous systems and the number of total systems is quite high compared to the probability of a system failure in a given unit time, and repairing a defective unit is more cost-effective than scrapping the defective unit and buying a new one. In the next section, the summary, methodologies and the contributions of the thesis will be explained in detail.

1.2.3 Summary, methodologies & Contributions and Outline of thesis

The research presented in this thesis aims at developing decision support models that can integrate the down-time service related and the capacity related decisions of a MSP in two different environments:

1. A specialized system environment, where the substitute of the system/critical subsystem under concern is not available for short term hiring purposes in the market.
2. A (partially) commoditized system environment, where a substitute system/critical subsystem can be hired from a 3rd party supplier upon a system failure.

For each of these environments, three capacity modes, namely fixed, periodic two-level and periodic sell-back capacity modes are investigated.

In the specialized system environment, MSP decides on the stock level for the spare unit inventory next to the capacity related decisions. On the other hand, in the commoditized system environment, rather than keeping a spare unit inventory, MSP signs an agreement with a 3rd party supplier, which guarantees the temporary provision of a substitute system for a predetermined duration upon a system failure. The related decision in this environment is the length of the uniform hiring duration. The MSP tries to integrate this down-time service decision with the capacity-related decisions for each of the three capacity modes.

As mentioned before, the benefits from the capacity flexibility in different modes form the leitmotif of the thesis for both of these (specialized and commoditized) environments. We assume that the capacity flexibility decisions can only be taken at equidistant points in time, and we incorporate the effects of the frequency of these capacity decision points on the operations of the repair shop. Henceforth, the period length, which is the time between two consecutive capacity decision points, arises as a capacity flexibility metric due to the introduced capacity modes. Also the impact of the period length on wage rates of the flexible resources are modeled and explained through the wage differential concept.

We analyze the centralized decision making problem and focus on the cost rate minimization problem of the MSP. We assume that the service fee that the MSP asks for as well as the substitute hiring/rental prices are already given. Therefore, the price determination problem of the service fee or any other decentralized decision making issues are out of the scope.

The objective of this study is to get more insights into the effects of the capacity flexibility possibilities in the operations of MSP firms for specialized and (partially) commoditized systems. In achieving this objective, we raise a number of research questions that will be addressed in different ways in the upcoming chapters. Furthermore, different research methodologies have been applied. We have used analytical stochastic modeling, Markov Decision Process and computer simulation as methodologies in our research, which all provide valuable insights towards understanding the planning and control of capacity management of MSPs. The contributions of this thesis to the literature can be listed as follows:

- In addition to the traditional maintenance problem of specialized system environments, we address the maintenance problem of (partially) commoditized systems and build a maintenance strategy coherent with the increased short-term substitute hiring possibilities resulted from the commoditization and the rentalization of the systems in consideration.
- Different from many other studies, we focus on periodic capacity flexibility, where period length arises both as a decision variable and as a dimension of a system's flexibility measure. Furthermore, we use the wage differential concept from Labor Economics literature to reflect the effects of some capacity related

decisions (such as the frequency) upon the per time unit cost of flexible capacity.

- We introduced novel capacity flexibility policies and substantiated their possible cost savings compared to the fixed capacity policy in both specialized and commoditized environments.
- We integrate the down-time service related decisions with capacity related decisions of a MSP in the presence of three different capacity modes in both of the system environments.

We believe that the framework, design and analysis of the problems addressed as well as the results and the insights obtained in this thesis can help and motivate other researchers/ practitioners to further investigate the cost saving prospects from capacity flexibility in the after sales/maintenance service operations. We also anticipate that the framework described for commoditized systems will be increasingly useful in the future, since the commoditization and rentalization of the systems will be much more widespread due to the increasing information technology and the accelerated mimetic innovations. Therefore all the after-sales service providers have to come up with innovative strategies and compete more on the efficiency of their after-sales operations in order to regain what they lose from the commoditization.

The remainder of the thesis is organized as follows. In Chapter 2, we focus on the use of capacity flexibility in the repair operations of the MSP in the specialized system environment. The capacity related decisions are integrated with the decision on the stock level of the spare unit inventory for all three capacity modes. In Chapter 3 we investigate the same three capacity modes in a (partially) commoditized system environment, where hiring a substitute system for a pre-determined, uniform duration becomes the conventional down-time service upon a failure. In this chapter the decision on the hiring duration is integrated with the other capacity related decisions. In Chapter 4 we provide some preliminary analysis and give the early results on future research topics such as the hybrid strategy where both “keeping stock” and “hire substitute” strategies are followed simultaneously. Finally in Chapter 5, we summarize our results, give the conclusion and discuss the topics covered in this thesis.

2 Specialized System Environment

2.1 Introduction

In this chapter, we focus on the integration of capacity and down-time service related decisions in the specialized, engineer-to-order system environments, in which there are different types of capacity flexibility options available. In this type of environments, each specialized system is designed and built specifically according to its owner's requirements. Defense systems, lithography systems, aircrafts or other advanced/complex, engineered to order capital goods are examples of such specialized systems. Due to the diversity of owners' requirements, each system develops many unique characteristics, which make it hard, if not impossible, to find a substitute for the system upon a failure, as a whole. Other factors that restrain the substitution of a system as a whole are the complexity and the scale of the system.

No matter to what extent each individual system is specialized; these systems are often composed of a number of standard subsystems. The modularity and the commonality of interchangeable parts make the repair by replacement solutions realizable in this maintenance context. We suppose that the same type of subsystems/components are interchangeable between various systems. However, we also assume that the repair processes of different types of subsystems require different technical skills and/or manpower, therefore each sub-system type necessitates either its own repair shop or its own crew in a repair shop.

These assumptions distinguish our approach from some of the other conventional multi-item inventory approaches in the maintenance literature. Although some repair shops can conduct repairs of multiple component types (i.e. (Adan et al. 2009), we observe an ongoing trend of after-sales service differentiation in modular designed system environments. For instance, upon a failure of a system, once the root of the failure is diagnosed, the corresponding failed module is handled distinctly for each type. This proclivity can be seen in the repair/maintenance activities of many specialized and modular system environments such as aviation or defense industry. In (Keizers et al. 2009), it is reported that there were 75 repair shops specialized on the repair of different parts/projects in the Dutch Royal Navy Maintenance and Repair Organization. In a case study conducted in the MRO department of the Canadian Airforce (Nima Safaei et al. 2011), it has been observed that the skilled technicians are divided into many trades (i.e. weapons and electrical armament, airframe mechanical, airframe electrical, propulsion and avionics/electronics), and each trade group is responsible for the overhaul of different types of components/parts. Similarly, the evolutionary pattern of the medical science epitomizes this differentiation/segmentation phenomenon. As the accumulated knowledge on human body and on the diseases mounted, different specializations/branches (e.g. cardiology, neurology, etc...) came into existence and hence resulted in better, more effective and (sometimes) more economical services and treatments. This ongoing trend is expected to continue in the maintenance industry, thanks to the increasing role of the modular design concept of the products and systems.

Parallel to this trend, in this thesis, we suppose that each critical type of subsystem requires a different repair shop for the necessary repair activities. Thus, the integrated capacity related and the down-time service related decisions are taken for each subsystem type separately.

In this chapter, it is assumed that the MSP takes care of the repair and the availability of a critical subsystem (e.g. jet engines, railway locomotives, etc.), which is used in numerous specialized systems (e.g. planes, trains, etc.), installed in a region, in exchange for a service fee. In order to realize the repair process of a critical subsystem, the MSP operates a repair shop, and the overall system availability is improved by spare unit inventory pool for the critical subsystem under concern. This single item repair-to-stock system is modeled using a single inventory/queue formalism, where the processing rate corresponds to the repair shop capacity and the base stock level corresponds to the maximum number of non-defective spare units not in use in the absence of failed systems.

Our objective in this chapter is to minimize the total relevant costs (TRC) of the MSP, which consists of the three components listed below with their abbreviations in parentheses:

1. Capacity related costs of the repair shop (CRC)
2. Down-time costs of a system whose critical subsystem has failed and not replaced with an operating one from the stock (DTC)
3. Holding costs for the critical spare units, both in the stock and in the repair shop (HC)

Given the cost components above, the MSP takes capacity and inventory related decisions simultaneously in order to minimize its TRC. Three capacity modes are investigated in this chapter are: Fixed Capacity Mode (Reference), Two-Level Flexible Capacity Mode and Periodic Capacity Sell-back Mode

As mentioned in Chapter 1, the service provider can make use of periodic capacity flexibility options while integrating its repair shop capacity and spare unit inventory decisions. The reasons behind the periodicity of capacity flexibility were already discussed in Chapter 1.

The flexibility options can be realized through an external agent. We use the umbrella term of “capacity agency” for the contingent capacity supplier(s) in all of the flexibility scenarios. The capacity cost structures of the contractual agreements for each capacity mode will be further explained in this chapter.

The relations between the repair shop, the capacity agency (can be either internal or external), and the specialized systems in this chapter's environment is depicted in Figure 2-1.

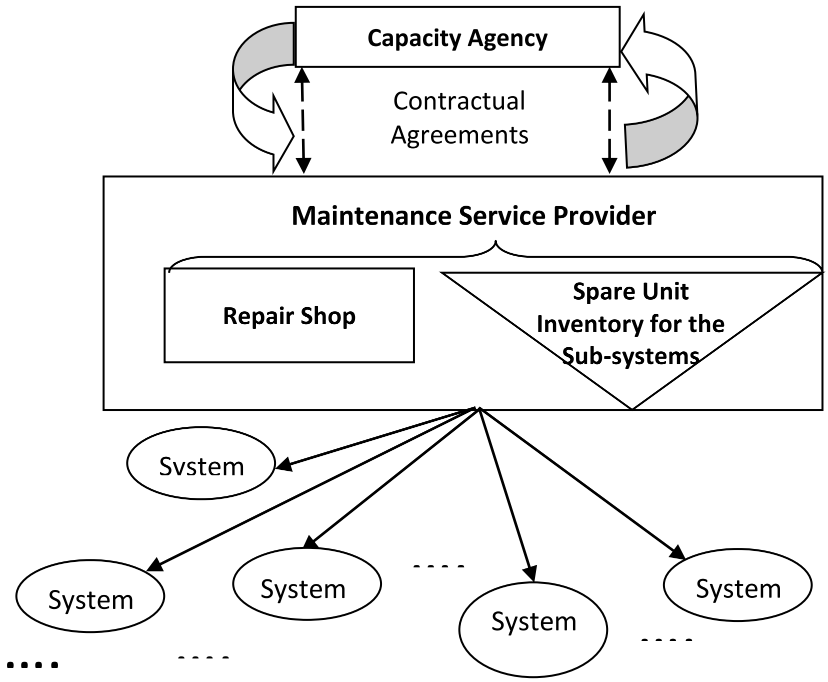

Figure 2-1: The relations between the MSP of a critical subsystem, the capacity agency and the specialized systems through contractual agreements.

We aim to model the maintenance service network for the specialized systems which embodies all the active/passive actors listed above in order to analyze the interplay between the capacity agency and the repair shop, derive the cost performance characteristics and develop a decision support system that integrates the capacity related and the inventory related decisions in order to minimize the TRC of the MSP under different capacity flexibility options. In addition, the developed modeling framework in this chapter enables the researchers/practitioners to foresee how much cost savings can be realized through the use of capacity flexibility compared to the best practice under the fixed capacity setting.

This single-item modeling approach can be simply generalized to multi-item settings by designing different single item repair-to-stock systems for each type of subsystem and by summing up the total costs of each single item model. Under this modeling approach, it is assumed that a separate repair shop unit and a separate spare item stock are operated for each relevant critical subsystem type and that the failures due to the different types of subsystems, their repair process and the relevant capacity/subsystem availability decisions are independent from each other.

The outline of the remainder of this chapter is as follows. In Section 2.2, we provide a brief literature review about the maintenance/repairable item control and capacity management in specialized system environments and list this chapter's contributions. The capacity provision mechanism, the cost structure of a unit of provided capacity and how this is affected by the period length due to the wage differentials are explained in Section 2.3. In Section 2.4, we model and analyze the integrated decision making problem under the fixed capacity mode, which serves as a reference model for the further modes. In Section 2.5 and in Section 2.6 we explain, model and analyze the same problem framework under two-level flexible capacity and capacity sell back modes, respectively. Finally in Section 2.7, we draw overall conclusions over the performance of capacity modes, interpret the differences and finalize this chapter.

2.2 Literature Review

As mentioned in the general literature review presented in the previous chapter, inventory control of the repairable items constitutes one of the strongest streams of the literature in the realm of maintenance. However, the dominant part of the models for the repairable item inventory control are based on the assumption of ample repair capacity, which used to be a benign presumption for most of the military environments. (See (Sherbrooke 1992) and (Muckstadt 2005) for detail).

Several studies generalized this ample supply assumption mostly by incorporating exact queuing network models to the repairable item inventory control problems (See (Gross et al. 1983), (Albright & Soni 1988) and (Albright & Gupta 1993)). A critical aspect of this approach is the inherent computational complexity of the performance evaluation methods of closed queuing networks, which can be prohibitive for multi-item setting of practical problems. Therefore, further studies introduced approximations and other methods for multi-echelon repairable item inventory systems with limited repair facilities (See (Diaz & Fu 1997), (Perlman et al. 2001), (Zijm & Avsar 2003) and (Sleptchenko et al. 2003). The flexible capacity/manpower use in repairable item systems is a rather understudied subject. There are only a few studies, by using simulation, trying to explore the benefits from the use of flexible manpower decisions in multi-echelon/ multi-indenture repairable item systems (See (Scudder & Hausman 1982), (Scudder 1985) and (de Haas 1995)).

As mentioned before, in this chapter, we use a single inventory/queue formalism to model the repair shop operations of the MSP under study. In a different context, similar quantitative formalisms are widely used in manufacturing/production control problems, in the shape of produce-to-stock or make-to-stock systems. (Buzacott & Shanthikumar 1993) and (Altiok 1997) provide good overviews of the stochastic models used in capacity and base stock level decision problems for make-to-stock production systems. A review of the studies that incorporate capacity flexibility in make-to-stock systems is already given in the corresponding literature review section of Chapter 1.

Following from this literature review, the objective of this chapter can be summarized as follows: to integrate the stock level related decisions for spare unit availability with the capacity related decisions of the repair shop, both of which are taken by the MSP for a specific subsystem type, which is critical for many specialized systems installed in a region. In the next section, we explain the capacity provision mechanism, the cost structure of a unit of permanent capacity and a unit of contingent capacity that is delivered from the capacity agency, and how these costs are affected by the period length due to the wage differentials.

2.3 Capacity Provision Mechanism

The capacity agency is a reactive agent in the whole decision making process, and is responsible for the capacity provision mechanism under the periodic two-level and periodic sell-back capacity modes. The capacity provision mechanism has a periodic nature: at equidistant points in time, the capacity agency must be ready to supply an agreed amount of capacity that covers the whole period, that is to say until the next equidistant point. The use of this reserved capacity is decided instantaneously at these equidistant supply points.

In order to be able to supply the required amount of capacity for each period, the capacity agency has to be prepared at the start of each period before the decision is taken. Although the provided capacity is ready to be deployed at the start of each period, it is not guaranteed that it will be used.

In the second (two-level) capacity mode, the permanent capacity is always deployed at the repair shop, and the use of the contingent capacity is decided by the repair shop at the start of each period with regard to the workload situation. If the number of units waiting to be repaired is bigger than a given threshold value, then the provided capacity is deployed and used by the repair shop. Since this decision cannot be known in advance with certainty, this uncertainty on the use of the periodically provided capacity creates an economic factor that causes an opportunity cost, because that capacity could be used somewhere else if it was not reserved for that period.

Similarly, in the third capacity mode (capacity sell-back), the provided capacity is deployed at the repair shop at the start of each period. However, in this capacity mode, additional uncertainty factor is the time during which the provided capacity will be actually deployed at the repair shop. This is due to the fact that the capacity is sold back to the agency to be hired out temporarily for other external tasks as soon as there is no repair waiting in the shop. This uncertainty of the duration of the deployment in the repair shop and frequency of job switching (between the repair shop and the tasks that the capacity agency assigns) create an economic factor that causes an opportunity cost due to the additional skills needed for, and the extra cognitive load generated from task switching as well as the transportation/ transaction costs of the shop capacity.

In the light of the explanations of the opportunity costs for the capacity modes, in the next section, the opportunity cost per time and its relation with the period length will be elucidated.

2.3.1 Opportunity Cost as a Function of Period Length D

Let D denotes the length of the period, which is the time between two equidistant capacity supply points. A longer D mitigates the severity of the lost opportunity effects due to the enforced capacity availability at the start of each period because:

- Longer D gives more room to the capacity agency to benefit from the possibility of using the reserved capacity for other tasks until the start of the next period.
- Longer D implies an improved task security for the provided resources in the second capacity mode and less job switching for the third capacity mode.

These effects are in line with the wage differential theory, a research area in Labor Economics that analyzes the relations between the wage rate and the unpleasantness, risk or other undesirable attributes of a particular job (Rosen 1986).

Let c_p denotes the per time unit cost for a unit of fixed capacity that is deployed at the repair shop indefinitely, e.g. for an infinite period length. This is equivalent to the situation when the ownership of the provided capacity is passed to the repair shop, therefore hereafter c_p is denoted as the permanent unit capacity cost per unit time. Similarly, c_c is the cost that is incurred per unit time for a unit of provided contingent capacity. Due to the opportunity costs resulting from the capacity reservation, c_c is greater than or equal to c_p for finite period lengths and as the period length D goes to infinity, the capacity is provided to the repair shop indefinitely, which can be interpreted that the capacity is owned by the repair shop, thus c_c equates to c_p when $D = \infty$.

The opportunity cost is denoted by $o_c(D, \Delta, \alpha)$, which is always greater than or equal to zero. The opportunity cost $o_c(D, \Delta, \alpha)$ decreases with the period length D in different forms as will be shown in Table 2-1. We assume that c_c is the sum of c_p and $o_c(D, \Delta, \alpha)$. We propose three different functional forms for $o_c(D, \Delta, \alpha)$. Note that other functions (which can be constructed after an empirical investigation) can be also used to model the opportunity costs per unit time, as well. However, we limit ourselves to these three functional forms, namely: linear, inverse proportional and exponential forms, which are quite commonly used in Labor Economics (Rosen 1986).

These proposed functions depend on two additional cost parameters next to the period length: Δ and α. $\Delta > 0$ represents the maximum opportunity cost per time unit due to the availability of the capacity at the start of each period, and $\alpha > 0$ reflects the decreasing rate of the opportunity cost with period length. The proposed functions can be seen in Table 2-1, and for these suggested functional forms of $o_c(D, \Delta, \alpha)$, the effects of Δ and α on c_c are illustrated in Figure 2-2.

Name of the function	$o_c(D, \Delta, \alpha)$
1.Linear	$(\Delta - \alpha D)^+$
2.Inverse Proportional	$\Delta/(1+\alpha D)$
3.Exponential	$\Delta e^{-\alpha D}$

Table 2-1 Three proposed opportunity cost functions.

In the next section, we start with the fixed capacity mode, where the capacity is either owned by the repair shop or provided to the repair shop indefinitely. The minimum cost performance achieved in this mode will act as a reference point to judge the performance of the MSP using other flexibility options in further modes.

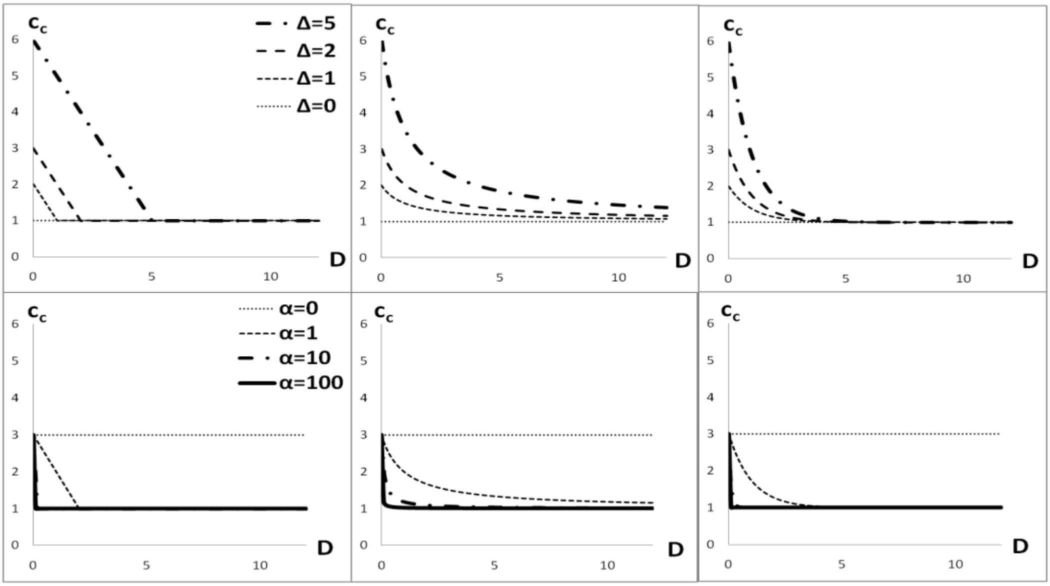

Figure 2-2 The figures on the top depict the behavior of c_c for $\alpha = 1$ and $\Delta = 0, 1, 2, 5$. The figures on the bottom depict the behaviour of c_c for $\Delta = 2$ and $\alpha = 0, 1, 5, 10, 100$. Right to the left: o_c is of the linear, inverse proportional and exponential forms.

2.4 Fixed Capacity Mode

In this section, we analyze the integrated decision making problem of the MSP under the fixed capacity mode. In this mode, all of the capacity is permanent, owned by the repair shop and ready for use all the times. This mode serves as a reference point for the other two capacity modes, necessary to assess the benefits of further flexibility options.

The MSP has to determine the optimal inventory and capacity level decisions in order to minimize its TRC. This section aims at building a modeling framework and a decision support system for the MSP operating under the first capacity mode. Therefore in Subsection 2.4.1 we present the model, assumptions and the problem formulation. In Subsection 2.4.2, the derivation of the total relevant cost per unit time as well as the analytical properties and the optimization procedure are given. Finally in Subsection 2.4.3, we describe the experimental setting and provide the results of the numerical study.

2.4.1 Model, Assumptions and Problem Formulation

We analyze an environment where the MSP operates a repair shop, and is responsible for the repair of the failed critical subsystems of a particular type, from many specialized systems installed in a region. These systems are in the exploitation phase and we assume that the number of the systems that the MSP is responsible for, N, is quite large, whereas the probability of the failure of a given system in a unit time is quite low, which justifies the modeling approach, where the failures come from an infinite population of systems, and the total stream of system failures due to the critical subsystems follow a Poisson process with a constant rate: λ. This is in line with the existing assumptions in the literature, (Sherbrooke 1992) and (Muckstadt 2005), which are already shown to be reasonably justifiable if the total number of the systems is large and the mean time between failures (MTBF in short) due to the critical parts/units is quite long (compared to the down-times).

In this chapter, the systems are supported by a single warehouse where all spare units of the critical subsystem are stocked and there is a single repair shop with a finite capacity where the defective subsystems are repaired and refurbished to “ready-for-use” state, again.

Upon a system failure, the following procedure is ensued. First the reason of the failure is detected. We assume that the detection of the reason of the failure (whether it is caused by the critical subsystem defectiveness or not) is performed by the in-house engineers of the system owners immediately and there is no moral hazard related risks between system owners and the MSP.

Upon the diagnosis, if the failure is due to the critical subsystem, a ready-for-use unit is sent from the stock to the system location, if there is any unit available in the stock. After a new unit has arrived to the system location, the defective subsystem unit can be replaced with the ready-for-use unit and transported to the repair shop in order to get repaired. After the repair process, the defective subsystem unit will be restored to the ready-for-use state again, which we assume that it is *as good as new*. Subsequent to the repair and refurbishing, the subsystem is added to the spare unit inventory and is held in the stock until it is sent to replace another defective subsystem in the future.

If a subsystem failure is followed up with an out-of-stock situation, the demand for the ready for use subsystem will be backordered. We assume that the replacement times (of the defective and ready-for-use critical subsystems) and the transportation times from the repair shop to the customer sites, where the systems are located (or vice versa) are negligible. Each defective subsystem requires an exponentially distributed service time from the repair shop and the repair shop cannot work on more than one repair simultaneously. Therefore the defective units that require repair have to wait for their turn in order to get serviced in the repair shop. A First-Come-First-Served (FCFS) policy is used for the service order of the defective subsystems. We model the repair shop as a single server Markovian queue. The capacity of the repair shop determines the speed of the repair service. Therefore the processing rate μ is considered as the capacity level of the repair shop.

In this thesis, we suppose that all the defective subsystems can be restored to the as good as new condition after the repair. However, in reality, some of the defective units can be in an un-repairable situation, where a new subsystem has to be bought from the original manufacturer (OEM) of the critical subsystem in concern. We exclude these type of situations in our analysis, since the cost burden that they generate are unaffected by our control actions and can be incorporated to the existing analysis easily. For instance, if we assume that a constant ratio of the defective subsystems are not repairable, we can incorporate the existence of un-repairable situations by adding the purchasing costs related to the un-repairable units on top of the other cost components of MSP, and by analyzing the operations of the repair shop queue with the reduced arrival rate.

Due to the low demand characteristics for the repair of the defective units, the spare unit inventory is controlled by a continuous review base-stock policy with a base stock level of S. This policy is commonly used both in academia and industry (see (Feeney & Sherbrooke 1966) and (Muckstadt 2005)).

The capacity cost per unit time is c_p in this fixed capacity mode since all the repair shop capacity is permanent (or supplied indefinitely). We pay h per unit time for each spare unit in the stock/in the repair shop. The down-time costs due to the backorder of the spare subsystems is equal to B per time unit, and we assume that $B > h$. From now on, we use the notation of **C** to denote the capacity policy. In the fixed capacity mode, **C** is a single variable, since the only capacity related decision is the processing rate μ. The inventory related decision is S. The total relevant cost function, TRC, can be represented by **C** and S, and is the sum of capacity related costs $\left(CRC(\mathbf{C})\right)$, down-time costs $\left(DTC(\mathbf{C},S)\right)$ and holding costs $\left(HC(S)\right)$. Given these cost components and the decision variables, the problem of the MSP can be formulated as follows:

$$\begin{gathered}\min_{\mathbf{C},S} TRC(\mathbf{C},S) = CRC(\mathbf{C}) + DTC(\mathbf{C},S) + HC(S) \\ s.t.\;\; S \in N = \{0,1,2,\dots\} \\ \mathbf{C} = \mu > \lambda\end{gathered} \tag{2.1}$$

Given the problem formulation above, in the next subsection, we derive the necessary cost functions used in (2.1), give the analytical properties of $TRC(\mathbf{C},S)$, and present the optimization procedure for the problem.

2.4.2 Derivation and Analysis of the Cost Functions, the Solution Procedure

In this subsection, we first derive and provide the analytical properties of the cost functions used in (2.1). Afterwards, we give the solution procedure for the optimization problem. As it is mentioned previously, $TRC(\mathbf{C}, S)$ consists of three cost components: $CRC(\mathbf{C})$, $DTC(\mathbf{C}, S)$ and $HC(S)$.

For this fixed capacity mode, the capacity related cost per unit time is a linear function of the excess capacity: $\mu - \lambda$, since the baseline capacity level, (λ), is unaffected by the capacity policy. Per time unit cost of the capacity is constant and equal to c_p. Therefore, we have $CRC(\mathbf{C})$ = $c_p(\mu - \lambda)$. The holding cost per unit time is also a linear function of the base stock level S, since we have an additional S number of spare units tied up in the stock/ repair shop and the holding cost rate per unit part is h per time. Hence, we have $HC(S) = hS$.

Per time down-time related cost, $DTC(\mathbf{C}, S)$ is derived from the number of defective units in the repair shop. Let N_d denotes the number of defective units in the repair shop and $P\{N_d = n|\mathbf{C}\}$ denotes the probability that there are n defective units at an arbitrary point of time given that the capacity level is μ. Since we model the repair shop as an $M/M/1$ queue, N_d is identical to the number of customers in a queue with a traffic ratio of $\rho = \lambda/\mu$. Hence, we have the following:

$$P\{N_d = n|\mathbf{C}\} = (1 - \rho)\rho^n, \qquad n = 0,1,2, \ldots \tag{2.2}$$

The expected number of systems that are in down state due to stock out of spare subsystems at an arbitrary point of time can be found from (2.2) as follows:

$$\mathrm{E}((N_d - S)^+|\mathbf{C}) = \sum_{n=S+1}^{\infty} (n - S)(1 - \rho)\rho^n = \frac{\lambda\rho^S}{(\mu - \lambda)} \tag{2.3}$$

For each system that is down due to the failure and the concomitant shortage of the critical subsystem, a cost of B is incurred per unit time. Hence, the average down-time related cost per unit time can be found as follows:

$$DTC(\mathbf{C}, S) = B\big(\mathrm{E}((N_d - S)^+|\mathbf{C})\big) = B\,\frac{\lambda\rho^S}{(\mu-\lambda)}.$$

Hence we have the following:

$$TRC(\mathbf{C}, S) = CRC(\mathbf{C}) + DTC(\mathbf{C}, S) + HC(S) = c_p(\mu - \lambda) + B\,\frac{\lambda\rho^S}{(\mu - \lambda)} + hS$$

(2.4)

As the total relevant cost rate, $TRC(\mathbf{C}, S)$, is derived, next we give two of its analytical properties.

Property 2.1

For a given $\mu > \lambda$, $TRC(\mathbf{C}, S)$ is convex in S.

Proof:

When $\mu > \lambda$ and $S > 0$ we have: $\nabla_S TRC(\mathbf{C}, S) = TRC(\mathbf{C}, S) - TRC(\mathbf{C}, S-1) = h - B\rho^S$. Note that $\nabla_S TRC(\mathbf{C}, S)$ is increasing on its whole domain, and thus $TRC(\mathbf{C}, S)$ is convex.

Property 2.2

For a given $S \geq 0$, $TRC(\mathbf{C}, S)$ is strictly convex in μ.

Proof:

For $S \geq 1$, we have: $\frac{\partial TRC(\mathbf{C},S)}{\partial \mu} = c_p - B\frac{\lambda \rho^S}{(\mu-\lambda)^2}$ and $\frac{\partial^2 TRC(\mathbf{C},S)}{\partial \mu^2} = 2B\frac{\lambda \rho^S}{(\mu-\lambda)^3} + 2B\frac{\lambda^2 \rho^{S-1} S}{(\mu-\lambda)\mu^3} + 2B\frac{\lambda^2 \rho^{S-1} S}{(\mu-\lambda)^2 \mu^2} + B\frac{\lambda^3 \rho^{S-2} S(S-1)}{(\mu-\lambda)\mu^4} > 0$.

Similarly, for $S = 0$, we have: $\frac{\partial TRC(\mathbf{C},S)}{\partial \mu} = c_p - B\frac{\lambda}{(\mu-\lambda)^2}$ and $\frac{\partial^2 TRC(\mathbf{C},S)}{\partial \mu^2} = 2B\frac{\lambda}{(\mu-\lambda)^3} > 0$.

We use the following solution procedure for the optimization problem. Before starting the search procedure, we choose a sufficiently large S_{max} value which is an upper limit for the spare unit base stock level choice. Then we follow the steps given below:

Search Procedure-I

---

1. Let $\mu^*(0) = \lambda + \sqrt{\frac{\lambda B}{c_p}}$ denote the μ that satisfies: $\frac{\partial TRC(\mathbf{C},0)}{\partial \mu} = 0$. Then we have $\mathbf{C}^*(0) = \mu^*(0)$ and $TRC(\mathbf{C}^*(0), 0) = c_p\left(\sqrt{\frac{\lambda B}{c_p}}\right) + \sqrt{c_p(\lambda B)}$.
2. For every $S \in \{1,2, \dots, S_{max}\}$ follow the steps a and b:

    a. Compute $\mu^*(S)$, which is the μ that satisfies: $\frac{\partial TRC(\mathbf{C},S)}{\partial \mu} = 0$.

b. We have $\mathbf{C}^*(S) = \mu^*(S)$ and $TRC(\mathbf{C}^*(S), S) = hS + c_p(\mu^*(S^*) - \lambda) + B\frac{\lambda\left(\frac{\lambda}{\mu^*(S)}\right)^S}{(\mu^*(S) - \lambda)}$

3. After finding $TRC(\mathbf{C}^*(S), S)$ for all $S \in \{1,2,\ldots,S_{max}\}$, we can find S^*, and $\mu^*(S^*)$, which give the global minimum cost rate $TRC^* = TRC(\mathbf{C}^*(S^*), S^*)$ for the fixed capacity mode. S^* can be found as follows:

$$TRC(\mathbf{C}^*(S^*), S^*) = \min_{S \in \{0,1,\ldots S_{max}\}} TRC(\mathbf{C}^*(S), S)$$

In essence, the solution procedure uses the Properties 2.1 and 2.2 of TRC in order to find the single best capacity level for a given base stock level S. After finding the single best capacity level, $\mathbf{C}^*(S)$, for each stock level S up to a confidently large S_{max}, the optimal S^* and $\mathbf{C}^*(S)$ pair can be obtained through a brute force search over different S values. In the next subsection, we present the results of the numerical study that is conducted, where the optimal costs for problem (2.1) are obtained from the Search Procedure-I.

2.4.3 Numerical Study

In this section, we present the results of the numerical study for the fixed capacity mode. First, we describe the experimental design for the computational study. Afterwards, we follow the search procedure and obtain the optimal decision parameters for every instance in the test bed. Finally, the minimum cost performances, the optimal capacity levels and the optimal spare unit stock levels are given in order to generate the managerial insights and form a basis as a reference point to assess the benefits of capacity flexibility in further capacity modes.

2.4.3.1 The Base Case Scenario and the Experimental Design

In our computational study, we take the unit time as a week and normalize the mean arrival rate for the sub-system failures (not from one system but the cumulative failures in the whole environment) $\lambda = 1$ (failures per week). We have a base case scenario, which is described below, and the other 8 scenarios have varying backorder (B) and holding (h) costs per unit time. The parameter values in the base case scenario are based on the following situation:

Suppose that the capital good has a value of €1.000.000, and that the value of the critical subsystem unit is €50.000. The capital good is used in the production process of other products. The economic lifetime of the capital good is assumed to be 10 years, and the cost of the capital good represents 25% of the total costs of the products produced with it (material costs deduced). Further, suppose that the firm sells the products at a price that is 3 times the total production costs (material costs deduced) accumulated during the average machining time used to produce the products (when the capacity of the capital good is used). If the capital good is in use for 16 hours a day, 6 days a week and 50 weeks a year, then the capital good related costs are $\frac{€1.000.000}{(16\times6\times50\times10)} =$ $€20.83$ per hour and the lost revenue due to down-time is: $\left(\frac{€20.83}{0.25}\right) \times\ 3\ \cong\ €250$ per hour. A week (the base time unit) of down-time costs would be $16 \times 6\ \times\ €250\ =$ $€24.000$. For the cost of the workforce capacity of the repair shop, we will use a wage of $€60$ per hour per operator and we assume that a repair of a failed subsystem takes about 80 hours. Then the repair of a failed unit/subsystem on average costs $€4800$. This is much less than the cost price of the subsystem(€50.000), therefore repairing a defective subsystem is a more economical option than scrapping the defective subsystem and buying a new one.

Next, we derive the stock keeping cost per unit time. Assuming a capital rate of 25% per year, 365 days a year and 24 hours a day, stock keeping costs of a spare unit are $\frac{0.25\times50.000}{52}\ \cong\ €\,240$ per week. We scaled the parameter for the cost of workforce per repair to one, and expressed the values for the down-time and stock keeping costs as a multiple of this normalized parameter (for instance: $c_p = \frac{4800}{4800} =\ 1$, $B\ = \frac{24000}{4800} \cong 5$ and stock keeping cost of a sub-system $h\ = \frac{240}{4800} \cong 0.05$).

c_p	h	B/h
1	0.025	50
		100
		200
	0.05	50
		100
		200
	0.25	50
		100
		200

Table 2-2: Values of the analyzed c_p, h and B instances.

Having described the cost setting in the base case scenario, where $c_p = 1$, $h = 0.05$ and $B = 5$, we create the test bed, which consists of a total of 9 scenarios. These different scenarios explore the effects of different $\frac{c_p}{h}$ and $\frac{B}{h}$ ratios around the base case, and we assume that both B and c_p are higher than the holding cost h per unit time. The h ratios range from 0.025 to 0.25 and $\frac{B}{h}$ ratios range from 50 to 200. The values of B and c_p instances that are examined are given in Table 2-2. The base-case scenario that is described above is highlighted.

2.4.3.2 Results

After the Search Procedure-I in the previous subsection, is conducted for each of these nine scenarios, we find the optimal S^*, $\mu^*(S^*)$ and the resulting total relevant costs per unit time. From now on, we use the notation of $TRC_F{}^* = TRC(\mathbf{C}^*, S^*{}_F)$ in order to denote the minimum cost rate achieved under the first (fixed) capacity mode for given cost parameters. Accordingly, $S^*{}_F$ denotes the corresponding optimal stock level for spare units and $\mathbf{C}^* = \mu^*(S^*{}_F)$ refers to the optimal capacity choice. $S^*{}_F$, $\mu^*\ (S^*{}_F)$ and $TRC_F{}^*$ values are tabulated in Table 2-3 for all 9 scenarios:

	$h = 0.025$			$h = 0.05$			$h = 0.25$		
	$S^*{}_F$	$\mu^*\ (S^*{}_F)$	$TRC_F{}^*$	$S^*{}_F$	$\mu^*\ (S^*{}_F)$	$TRC_F{}^*$	$S^*{}_F$	$\mu^*(S^*{}_F)$	$TRC_F{}^*$
$B/h = 50$	12	1.36048	0.74672	9	1.50442	1.07997	5	2.07742	2.62727
$B/h = 100$	14	1.37443	0.80221	10	1.54319	1.16337	5	2.26327	2.84651
$B/h = 200$	15	1.40421	0.85522	11	1.57763	1.24252	6	2.26673	3.05773

Table 2-3 The optimal base stock level $(S^*{}_F)$ and capacity level $\left(\mu^*\ (S^*{}_F)\right)$ decisions and the resulting costs ($TRC_F{}^*$) for the total of 9 scenarios.

From Table 2-3, as it is expected, we can observe that the minimum total relevant cost, $TRC_F{}^*$, increases with B and h. The numerical results suggest that the total cost rate increases with h at a higher rate than it increases with B.

Upon an increase in B, the system responds both with a higher stock level $S^*{}_F$ in order to increase the spare unit availability and with a higher capacity $\mu^*(S^*{}_F)$, in order to provide a faster repair service. These two responses counterbalance the negative effects of the rising down-time costs due to higher B.

On the other hand, upon an increase in h, the MSP has a greater incentive to reduce its stock level to save from holding related costs. The decrease in spare unit stock level $(S^*{}_F)$ is more voluminous for higher backorder costs. As a remedy for the further cost consequences of lower stock levels (which would automatically lead to higher down-time costs), the MSP increases its capacity level so that the MSP can respond faster and it can complete the repairs at a shorter time. The latter effect of increasing capacity dominates the prior effect of spare unit inventory reduction in terms of costs.

Integration of the stock level and capacity level decisions brings drastic savings for the MSP. For instance, in the zero-inventory case, where $S = 0$, the repair shop capacity is the only parameter that the MSP can tune in order to adapt to different operating environments. This leads to very high capacity levels and therefore a lot higher costs. Let $TRC_{F,0}{}^*$ denotes the optimal total relevant costs in the zero-inventory case ($S = 0$). In Table 2-4, we demonstrate the percentage cost savings $100 \times \left(\frac{TRC_{F,0}{}^* - TRC_F{}^*}{TRC_{F,0}{}^*}\right)$ and capacity savings $100 \times \left(\frac{\mu^*(0) - \mu^*(S^*{}_F)}{\mu^*(0)}\right)$ due to the use of the spare unit inventory for the critical sub-system in all 9 scenarios.

	$h = 0.025$		$h = 0.05$		$h = 0.25$	
	$\frac{\tilde{\mu}(0)-\tilde{\mu}(S^*_F)}{\tilde{\mu}(0)}$	$\frac{TRC_{F,0}^*-TRC_F^*}{TRC_{F,0}^*}$	$\frac{\tilde{\mu}(0)-\tilde{\mu}(S^*_F)}{\tilde{\mu}(0)}$	$\frac{TRC_{F,0}^*-TRC_F^*}{TRC_{F,0}^*}$	$\frac{\tilde{\mu}(0)-\tilde{\mu}(S^*_F)}{\tilde{\mu}(0)}$	$\frac{TRC_{F,0}^*-TRC_F^*}{TRC_{F,0}^*}$
$B/h = 50$	36%	67%	42%	66%	54%	63%
$B/h = 100$	47%	75%	52%	74%	62%	72%
$B/h = 200$	57%	81%	62%	80%	72%	78%

Table 2-4 Percentage cost and capacity savings of integrated decision making (inventory and capacity) compared to the zero inventory case in 9 different scenarios.

As it can be seen, incorporating the use of the spare unit inventory in order to ameliorate the down-time, reduces the total relevant costs between 60% to 80% and reduces the capacity levels between 36% to 72%. The percentage savings of costs (due to the spare unit inventory) increase with B and decrease with higher h, which is parallel to the response pattern of S^*_F to different h and B given in Table 2-3. This can be explained due to the fact that the total cost savings are correlated to the difference in the optimal inventory levels between the integrated and “no inventory” settings. Since the inventory level is always “0” in the latter “no inventory” case, the larger S^*_F in the optimal integrated setting, the higher cost savings become.

The breakdown of the total costs under the optimal fixed capacity policy can provide us further managerial insights. Therefore, we investigate how much each of the three components (HC: holding cost for spare units, CRC: capacity related costs, DTC: down-time related costs) has contributed to the total relevant costs under the optimal capacity & stock level decisions for 9 different B/h & h combinations. The percentage contributions of each cost component are shown in the pie charts in Figure 2-3.

As it can be seen from Figure 2-3, the capacity related costs (CRC) constitute the biggest share of the total relevant costs in 7 out of the 9 (h and B/h) combinations. Only when $h = 0.25$, we can see that the holding related costs becomes the biggest cost component when $B/h = 50$ and 200.

Note that each cost result tabulated in Table 2-3 serves us as a reference point to assess the prospects of the flexibility options in the other two capacity modes. For each of the 9 scenarios, in order to consider another flexible capacity policy as an economical alternative, the minimum total relevant costs that can be achieved under that policy must be smaller than the corresponding TRC_F^* value from Table 2-3. In the next section, we set out our analysis to scrutinize the cost saving possibilities in the second, namely periodic two-level, flexible capacity mode

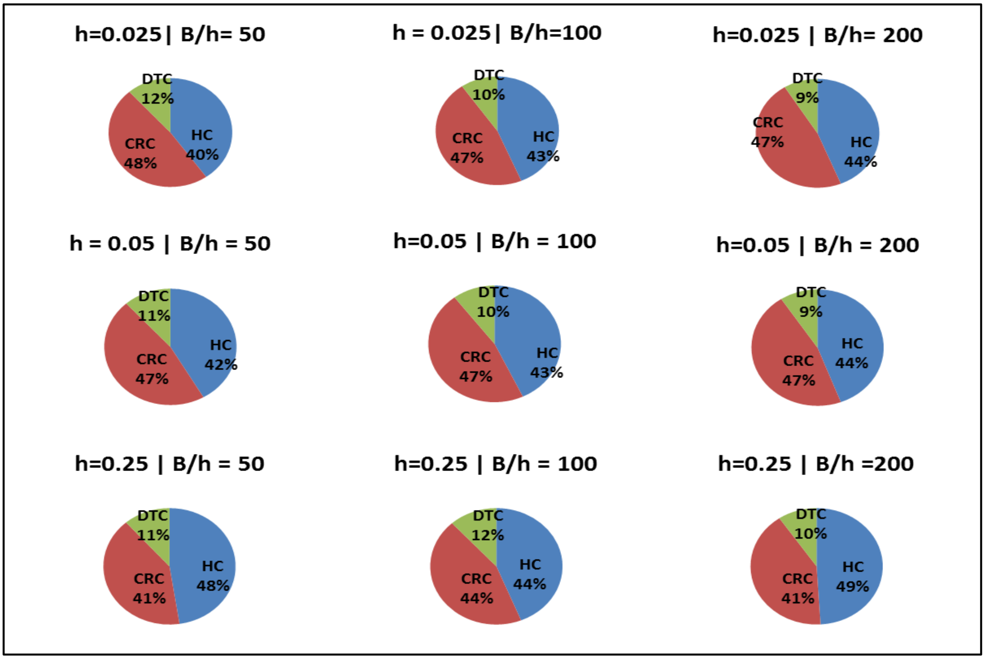

Figure 2-3 The breakdown of total relevant costs (TRC^*) to three cost components (HC, DTC and CRC) under the optimal capacity & stock level decisions ($\mathbf{C}^* = \mu^* \left(S^*{}_F\right), S^*{}_F$) for 9 different ($h$ & B/h) combinations.

2.5 Periodic Two-Level Flexible Capacity Mode

In this section, we analyze the integrated decision making problem of the MSP under the second capacity mode. In this two-level flexible capacity mode, a part of the capacity is permanent (or, in other words, the capacity agency supplies that amount of capacity indefinitely), whereas the other part is contingent, supplied periodically at equidistant points in time at a higher cost rate. The decision on the use of the contingent capacity is given at each equidistant point with regard to the present workload of the repair shop, in terms of the number of defective units in the service or waiting for service.

Similar to the fixed capacity mode, the MSP has to give the optimal inventory and capacity level decisions in order to minimize its TRC. Characterization and the analysis of the cost savings due to the flexible two-level capacity policy in the integrated decision making environment, have our utmost priority in this section. Therefore, we aim at building a modeling framework and a decision support system for the MSP, operating under the periodic flexible two-level capacity mode. In Subsection 2.5.1, we present the model, describe the policy under the periodic two-level flexible capacity mode and introduce the additional decision variables as well as the problem formulation. In Subsection 2.5.2, the derivation and the analysis of the total relevant cost per unit time are given and the relevant MDP formulation is provided. Finally in Subsection 2.5.3, we describe the experimental setting, the search procedure and provide the results of the numerical study with a focus on the sensitivity of cost/policy parameters and the savings under this two level flexible capacity mode compared to the best cost performance under the reference, fixed capacity mode ${TRC_F}^*$.

2.5.1 Model, Two-Level Flexible Capacity Policy and Problem Formulation

The MSP operates in the same environment that is explained in the previous section. Recall that the joint repair shop/ spare unit inventory is modeled as a single-server, queue-to-stock system, where the failures occur following a stationary Poisson demand and each defective unit requires an exponentially distributed amount of dedicated repair service time in the shop in order to regain its *good as new* status.

In this section, we assume that the repair shop can make use of capacity flexibility options due to the periodic two-level flexible capacity mode. In two-level capacity policy, a portion of the total capacity is permanent (i.e. supplied from the capacity agency indefinitely or permanently employed by the repair shop) and the other portion is contingent. The contingent portion is supplied from the capacity agency at equidistant points in time, if it is needed. Suppose D denotes the period length, which is the time between two equidistant points. At the start of each period, the number of defective units in the system is observed and according to this number, the decision on the use of contingent capacity during that period is taken. The dynamics of the effects of wage differentials are reflected on the unit contingent capacity cost rate, which is increasing with the frequency of capacity usage decisions that take place at equidistant points in time.

Under the queue/inventory formalism, the processing rate of the single server queue represents the capacity level of the repair shop. Therefore in the two-level capacity mode, the processing rate of the queue is chosen between two values (a high and a low one) at equidistant points in time according to the number of defective units waiting for repair service.

A periodic, two-level capacity policy, $\mathbf{C}$ = $[D,\ \mu_l,\ \mu_h,\ \vec{\pi}]$, consists of a period length D, a low and a high service rate pair (μ_l, μ_h) and a policy vector, $\vec{\pi}$. The i^{th} row of the policy vector $\vec{\pi}$ denotes the action (0 or 1) that the repair shop will take when there are (i) defective units in the system for $i = 0,1,2, \ldots$.

If $\vec{\pi}(i) = 1$, then the repair shop operates with a processing rate of μ_h , and else, if $\vec{\pi}(i) = 0$, it operates with a processing rate of μ_l , whenever there are i defective units in the environment for $i = 0,1,2, \ldots$.

Under policy $\mathbf{C}$, μ_l can be interpreted as the permanent capacity level and $(\mu_h - \mu_l)$ can be interpreted as the contingent capacity level. For the stability, we assume that $\mu_h > \lambda$.

Under a given policy $\mathbf{C}$ = $[D,\ \mu_l,\ \mu_h,\ \vec{\pi}]$, the number of defective units is observed at the start of each period. If there are m number of defective units and $\vec{\pi}(m) = 0$, then we deploy only the permanent capacity: μ_l, and if $\vec{\pi}(m) = 1$, we deploy permanent + contingent capacity: μ_h. Figure 2-4 illustrates the system under study.

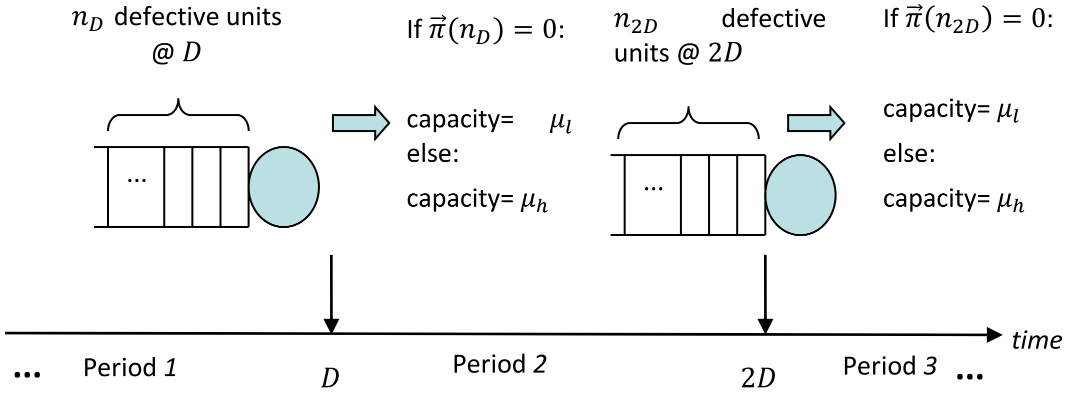

Figure 2-4 Illustration of the system under the second, two level capacity flexible mode.

Suppose the repair shop operates under a stable policy **C** = [D, μ_l, μ_h, $\vec{\pi}$] for an infinite horizon. Let $ACU(\mathbf{C})$ denote the average capacity deployment resulting from the capacity policy **C**. As it is explained in Section 2.3, the per time unit cost for the contingent capacity, c_c, depends on the period length D, and the per time unit cost of the permanent capacity, c_p is constant. Since the baseline capacity level, (λ), is unaffected by the capacity policy, $c_p\lambda$ is deduced from the total capacity related costs. For given c_c and c_p values, the capacity related cost per unit time, $CRC(\mathbf{C})$, can be directly derived from $ACU(\mathbf{C})$ as follows:

$$CRC(\mathbf{C}) = c_p(\mu_l - \lambda) + c_c(ACU(\mathbf{C}) - \mu_l) \tag{2.5}$$

The capacity agency offers a set of possible period lengths, θ, from which the service provider can choose the best period length considering the reflection of wage differentials on c_c. Recall that c_c is also dependent on c_p, and Δ, α coefficients. We pay h per unit time for each spare unit. The down-time costs due to the backorder of the spare units is equal to B per time unit, and we assume that $B > h$. The inventory related decision is S. The capacity related decisions, **C**, are threefold:

1) Length of the period: D

2) The size of the permanent and the contingent capacity levels: (μ_l & $\mu_h - \mu_l$)

3) The policy that determines when the contingent capacity is hired: $\vec{\pi}$ (according to the number of defective units).

The total relevant cost function, TRC, can be represented by **C** and S, and is the sum of capacity related costs $(CRC(\mathbf{C}))$, down-time costs $(DTC(\mathbf{C}, S))$ and holding costs $(HC(S))$. Given these cost components and the decision variables, the problem of the MSP can be formulated as follows:

$$\min_{\mathbf{C},S} TRC(\mathbf{C},S) = CRC(\mathbf{C}) + DTC(\mathbf{C},S) + HC(S)$$
$$s.t. \;\; S \in \{0,1,2,\dots\}$$
$$\mathbf{C} \; = \; [D, \mu_l, \mu_h, \vec{\pi}]$$
$$\mu_h > \lambda, \mu_h > \mu_l, D \in \theta,$$
$$\vec{\pi}(m) \in \{0,1\} \text{ for } \; m = 0,1,2,\dots$$

(2.6)

Given the problem formulation above, in the next subsection, we decompose the problem in (2.6) into smaller sub-problems and re-formulate each of the sub-problem as an infinite horizon average reward Markov Decision Process (MDP). The derivation of the expected immediate reward and the transition probabilities follow after the formulation of the MDP.

2.5.2 Decomposition and the MDP formulation of the Sub-problem.

In this subsection, we first decompose the optimization problem in (2.6). For each possible $D, S,\; \mu_l$ and μ_h, we can write the following sub-problem:

$$\min_{\vec{\pi}} TRC(\mathbf{C},S) = CRC(\mathbf{C}) + DTC(\mathbf{C},S) + HC(S)$$
$$s.t. \;\; \mathbf{C} \; = \; [D, \mu_l, \mu_h, \vec{\pi}]$$
$$\vec{\pi}(m) \in \{0,1\} \text{ for } \; m = 0,1,2,\dots$$

(2.7)

Let $\vec{\pi}^*(D, S, \mu_l, \mu_h)$ denotes the $\vec{\pi}$ that minimizes (2.7) for given $D, S,\; \mu_l$ and μ_h. If $\vec{\pi}^*(D, S, \mu_l, \mu_h)$ is found for all possible $D, S,\; \mu_l$ and μ_h, the optimal solution to problem (2.6) can be found from:

$$\min_{D,\mu_l,\mu_h,S} TRC(\mathbf{C},S) = CRC(\mathbf{C}) + DTC(\mathbf{C},S) + HC(S)$$
$$s.t. \;\; S \in N = \{0,1,2,\dots\}$$
$$\mathbf{C} \; = \; [D, \mu_l, \mu_h, \vec{\pi}^*(D, S, \mu_l, \mu_h)]$$
$$\mu_h > \lambda, \mu_h > \mu_l > 0, D \in \theta$$

(2.8)

Next, for a given base stock level S, period length $D \in \theta$, μ_l and μ_h, we reformulate the sub-problem given in (2.7) as a discrete time, average reward, infinite horizon Markov Decision Process (MDP) in order to find the optimal hiring policy $\vec{\pi}^*(D, S, \mu_l, \mu_h)$ for given S, $D \in \theta$, μ_l and μ_h and other cost parameters (c_p, B, h, Δ and α). In the remainder of this section, we will use only $\vec{\pi}^*$ to denote the optimal policy vector: $\vec{\pi}^*(D, S, \mu_l, \mu_h)$. Let W_π denote state space. In our MDP formulation, the state refers to the number of defective units at the repair shop at the start of each period. Therefore, we have $W = \{0,1,2,\dots\}$. Similarly, let A denote the action space. At each state, the action $a \in A$ determines the action that whether the contingent capacity is deployed in that period. Therefore, we have $A = \{0,1\}$, where "$a = 0$" implies that only permanent capacity (μ_l) is deployed, whereas "$a = 1$" implies that permanent + contingent capacity is deployed (μ_h).

We model the problem under the average cost criteria as follows:

$$g(s) = TRC(\mathbf{C}[D, \mu_l, \mu_h, \vec{\pi}^*], S) = \min_{\pi} \lim_{\substack{T\to\infty \\ \gamma\to 1}} \frac{1}{T} E^{\pi} \left(\sum_{n=1}^{T} \gamma^n r(s_n, a_n) \right) \tag{2.9}$$

Here $g(s)$ indicates the optimal average cost with an initial state $s \in W$, $\vec{\pi}$ is any type of policy, $s_n \in W$ is the state in period n, $0 < \gamma \leq 1$ is the discount factor, $a_n \in A$ is the action in period n and $r(s_n, a_n)$ is the expected period (immediate) costs of taking action a_n in state s_n. In this section, as it will be shown in the subsequent sub-sections, the state information (number of defective parts) s_n, combined with the action taken a_n, is sufficient enough to derive the expected period costs. By sufficiency we mean that any additional information (such as the time information of each of the subsystem failures) does not change the expected period costs.

In this MDP formulation, the optimality equation can be written as below:

$$g(s) = \lim_{\gamma\to 1} \max_{a \in A} \left\{ r(s,a) + \gamma \sum_{s' \in W} p(s'|s,a) g(s') \right\} \tag{2.10}$$

In (2.10) above, $g(s')$ is the optimal discounted value starting in state $s' \in W$ for the infinite horizon problem, $r(s,a)$ is the expected immediate cost of taking action $a \in A$ at state $s \in W$ and $p(s'|s,a)$ is the probability that the state will be s' in the next period given the current state is s and action $a \in A$ is taken.

Theorem 2.1: The underlying MDP in (2.9) is communicating for $\mu_h > 0$.

Proof: Given $D > 0$, $\mu_h > \mu_l > 0$, $\forall s, s' \in W$, $p(s'|s,a)$ is the probability that the state will be s' in the next period given the current state is s and action is $a \in A$. Suppose that in the policy π, $a = 1$ is chosen $\forall s \in W$.

For $s > s'$, we know that $p(s'|s,a)$ involves all possible sample paths. Therefore, $p(s'|s,a)$ is greater than the probability of a specific path, in which no new defective unit has arrived and an exact number of $(s - s')$ defective units are repaired. Hence we have:

$$p(s'|s,a) > P\{\text{no new defective units arrived \& } (s - s') \text{ defective units repaired}\} =$$

$$e^{-\lambda D} \frac{e^{-\mu_h D}(\mu_h D)^{(s-s')}}{(s-s')!} > 0 \text{ if } s' > 0 \text{ and}$$

$$e^{-\lambda D} \sum_{n=s}^{\infty} \frac{e^{-\mu_h D}(\mu_h D)^n}{n!} > 0 \text{ if } s' = 0$$

Similarly, for $s \leq s'$, we know that $p(s'|s,a)$ is greater than the probability of a specific unique path, in which no new defective unit has been repaired and $(s' - s)$ defective units have arrived. Hence we have:

$$p(s'|s,a) > P\{\text{no defective units repaired \& } (s' - s) \text{ defective units arrived}\} =$$

$$e^{-\mu_h D} \frac{e^{-\lambda D}(\lambda D)^{(s'-s)}}{(s'-s)!} > 0 \text{ if } s > 0 \text{ and}$$

$$\frac{e^{-\lambda D}(\lambda D)^{(s'-s)}}{(s'-s)!} > 0 \text{ if } s = 0$$

Hence, it is shown that $\forall s, s' \in W, \exists \pi$ s.t. $p(s'|s,a) > 0$, therefore the MDP in (2.9) is communicating which implies that it is also weakly communicating (Puterman 1994).

Theorem 2.2:

Under average cost criteria, an optimal <u>deterministic</u> policy exists for problem (2.9)

Proof: The reader is referred to Theorem 8.3.2 and 8.5.3 from (Puterman 1994), which show the existence of a deterministic optimal policy in (weakly) communicating MDPs.

If the $r(s,a)$ and $p(s'|s,a)$ values for every $s, s' \in W$ and $a \in A$ are given, the optimal capacity policy $\vec{\pi}^*$ and the minimum cost rate: g, for each S, D, μ_l and μ_h combination can be found from the value iteration and/or the policy iteration algorithms from the literature. (See e.g. (Puterman 1994)). Therefore, for each base stock level S, period length $D \in \theta$, μ_l and μ_h capacity level combinations, we derive: $p(s'|s,a)$ and $r(s,a)$ values for every $s, s' \in W$ and $a \in \{0,1\}$.

Recall that the operations at the repair shop level are modeled as a single server exponential queue. Without any constraints on the waiting room capacity, the formulas used in the MDP formulation would contain infinite sums of Bessel functions, which would make the numerical computations intractable, time-consuming and difficult. However, it is known from the literature (i.e. (Stern 1979)) that the transient and steady state behaviors of a Markovian queue with an infinite waiting room can be approximated with that of the same queue but with a finite waiting room. Hence, we model the system as an $M/M/1/K$ queue with periodically adjustable service rates. The accuracy of the approximation is dependent on the cost factors, policy parameters, traffic ratios and the size of the waiting room K. However, as it will be shown in Section 2.5.3, the analytical results are verified with an extensive simulation study, which suggests that for our numerical test bed, an approximate model with a finite waiting room K around $40-50$ approximates the actual model with an infinite waiting room almost perfectly. The accuracy holds even for very high traffic ratios (when $\rho \approx 1$), which can be explained as follows: under the periodic two-level policy, the repair shop capacity is adapted according to the workload and even though the average deployed capacity, $ACU(\mathbf{C})$, is very close to λ, the number of defective parts waiting never accumulates close to K, because the capacity is already switched to a higher level before the workload reaches to jeopardizing levels. The results of the simulation study is explicated further in the end of this section. Next, we analyze how the transition probabilities are derived.

2.5.2.1 Transition Probabilities:

In this subsection, we elucidate how we derive $p(s'|s,a)$ for a given base stock level S, period length $D \in \theta$, μ_l and μ_h capacity levels. At each period and in each state, we have two possible actions: $a \in \{0,1\}$. When $a = 0$, then the repair shop's capacity level is μ_l and otherwise it is μ_h. With the finite waiting room approximation, we have the state parameters: $s, s' \in \{0,1, \dots K\}$, which denote the total number of defective units waiting to be repaired. Recall that period length is D.

For an $M/M/1/K$ system, let $N_d(t)$ denote the number of defective units present at time t. The service rate is set to μ_l (μ_h) at the start of period number $n = 1,2, \dots$ if the action taken at the start of period n, a_n is equal to 0 (equal to 1). When the service rate is set to μ_l (μ_h), the behavior of $N_d(t)$ in this dynamic system is identical to the behavior of the number of defective units at time t ($t < D$) in a repair shop under fixed capacity regime with a constant service rate of μ_l (μ_h). Therefore, next we analyze the transient behavior of the $N_d(t)$ with a constant service rate.

Let $\mathrm{P}_{ij}(t,\mu)$ indicates the probability that there will be j defective units at time t given that there are i defective units at the beginning, when a constant service rate of μ is used. We will use numerical methods for the computation of the $\mathrm{P}_{ij}(t,\mu)$ values.

For all $i = 0,1, \dots. K$, $\mathrm{P}_{ij}(t,\mu)$ satisfies the following Kolmogorov equations:

if $j = 0$:

$$\mathrm{P}_{i0}(t,\mu) = \mathrm{P}_{i1}(t-dt,\mu)\mu dt + \mathrm{P}_{i0}(t-dt,\mu)(1-\lambda dt)$$

if $0 < j < K$:

$$\mathrm{P}_{ij}(t,\mu) = \mathrm{P}_{i(j-1)}(t-dt,\mu)\lambda dt + \mathrm{P}_{i(j+1)}(t-dt,\mu)\mu dt + \mathrm{P}_{ij}(t-dt,\mu)(1-(\lambda+\mu)dt)$$

if $j = K$:

$$\mathrm{P}_{iK}(t,\mu) = \mathrm{P}_{i(K-1)}(t-dt,\mu)\lambda dt + \mathrm{P}_{iK}(t-dt,\mu)(1-\mu dt)$$

(2.11)

After some algebra we obtain the following set of differential equations for $i = 0,1,\dots.K$:

if $j = 0$:

$$\frac{d\mathrm{P}_{i0}(t,\mu)}{dt} = \mathrm{P}_{i1}(t,\mu)\mu - \mathrm{P}_{i0}(t,\mu)\lambda$$

if $0 < j < K$:

$$\frac{d\mathrm{P}_{ij}(t,\mu)}{dt} = \mathrm{P}_{i(j-1)}(t,\mu)\lambda + \mathrm{P}_{i(j+1)}(t,\mu)\mu - \mathrm{P}_{ij}(t,\mu)\big((\lambda+\mu)\big)$$

if $j = K$:

$$\frac{d\mathrm{P}_{iK}(t,\mu)}{dt} = \mathrm{P}_{i(K-1)}(t,\mu)\lambda - \mathrm{P}_{iK}(t,\mu)\mu$$

(2.12)

The differential equations given in (2.12) can then be represented in matrix multiplication form as follows:

$$\frac{d\mathrm{P}(t,\mu)}{dt} = \mathrm{P}(t,\mu)\mathrm{Q}, \text{ with } \mathrm{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & 0 \\ \mu & (-\mu-\lambda) & \lambda & 0 & \cdots \\ 0 & \mu & (-\mu-\lambda)\,\lambda & & \cdots 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & \mu & -\mu \end{pmatrix}_{(K+1)\times(K+1)}$$

(2.13)

It is well known in the literature (Kulkarni 1995) that the solution for this equation is given by:

$$\mathrm{P}(t,\mu) = e^{\mathrm{Q}t} = \sum_{n=0}^{\infty} \frac{\mathrm{Q}^n}{n!} t^n$$

(2.14)

In order to escape from the infinite summation in (2.14), we will use matrix decomposition techniques (Neuts 1981). Since Q is a tri-diagonal matrix of a birth-death process, we know that it has $K+1$ distinct eigenvalues (Ledermann & Reuter 1954). The maximum of these eigenvalues is equal to zero, and the minimum of them is greater than $(-2(\mu+\lambda))$. Since matrix Q has $K+1$ different real eigenvalues, it has $K+1$ mutually independent eigenvectors, as well.

Let ξ_j and r_j be the j^{th} eigenvalue and its corresponding orthonormal right eigenvector for $0 \leq j \leq K$. Let V denote the diagonal matrix with diagonal entries equal to the eigenvalues (ξ_j 's) of Q and R denote the matrix that consists of right eigenvectors (j^{th}

column of R is equal to the r_j). From matrix eigen-decomposition theory, we can write $\mathrm{Q} = \mathrm{RV}(\mathrm{R})^{-1}$. Hence, by using Taylor series expansion, (2.14) can be rewritten as:

$$\mathrm{P}(t,\mu) = e^{\mathrm{Q}t} = e^{\mathrm{RV}(\mathrm{R})^{-1}t} = \mathrm{R}e^{\mathrm{V}t}(\mathrm{R})^{-1} \tag{2.15}$$

where $e^{\mathrm{V}t}$ is the $(K+1)\times(K+1)$ diagonal matrix, where the j^{th} diagonal element is equal to $e^{\xi_j t}$.

The derivation of the $\mathrm{P}_{ij}(t,\mu)$ holds for an arbitrary t under a system with a constant service rate μ. Since we are interested in the transition probabilities of the discrete-time, infinite-horizon and average cost MDP that is formulated above, we can state the following for given period length $D \in \theta$, low and high capacity levels μ_l and μ_h and for any states (i.e. number of defective units) s & s' $\in \{0,1,\ldots K\}$:

$$p(s'|s,0) = \mathrm{P}_{ss'}(D,\mu_l) \;\&\; \qquad p(s'|s,1) = \mathrm{P}_{ss'}(D,\mu_h) \tag{2.16}$$

Next, we derive the expected immediate periodic cost rate function: $r(s,a)$.

2.5.2.2 Expected Periodic Cost Rate:

Here, we derive the cost rate of a period that results from taking action $a \in \{0,1\}$ in state $s \in \{0,1,\ldots,K\}$. In the MDP literature these types of costs are called immediate costs (rewards) since they are mostly immediate consequences of the state and the action combinations. In our problem, the resulting costs are not the immediate consequence of the actions; however the expected costs can be directly calculated for each state & action combination.

The relevant costs that have to be included are the holding costs, down-time related costs and the capacity related costs. First, note that the holding cost is independent from either the state or the action taken. No matter what the state/action is, there are S spare units in this environment. Therefore hS, holding costs per unit time is incurred in each period independent from the state and the action. The second part, the capacity cost rate per unit time in a period, is only dependent on the action taken at the start of a period. For given μ_l and μ_h, if $(a=0)$, then the capacity related cost rate, $c(a)$, in that period will be equal to $c_p(\mu_l-\lambda)$. On the other hand, if $(a=1)$, then the capacity related cost rate will be equal to $c_p(\mu_l-\lambda)+c_c(\mu_h-\mu_l)$. The last part of TRC is the average down-time related cost rate in a period, and it is dependent on both state $s \in \{0,1,\ldots,K\}$ and the action $a \in \{0,1\}$ taken.

Recall that we pay B per time unit for each backordered defective unit in the environment. For a given base stock level S, given capacity levels (μ_l, μ_h) and given period length D, the expected immediate down-time related cost rate per unit time: $dtc(s,a)$, starting at state s with action a taken, can be calculated as follows:

$$dtc(s,a) = \frac{B}{D}\int_{t=0}^{D}\left(\sum_{n=S}^{K} P\{N_d(t) = n|s,a\}\,(n-S)\right)dt$$

(2.17)

Here $P\{N_d(t) = n|s,a\}$ is the probability that there are n number of defective units in the system after t time units from the start of a period, given the (s,a) state-action combination at the start of that period. In the following, we provide a practical method to calculate (2.17). We can rewrite (2.17) as below:

$$dtc(s,a) = B\sum_{n=0}^{K}((n-S)^{+})\int_{t=0}^{D}\frac{1}{D}P\{N_d(t) = n|s,a\}dt =$$

$$\frac{B}{D}\sum_{n=0}^{K}((n-S)^{+})\left(\mathrm{R}_a\int_{t=0}^{D}e^{\mathrm{V}_a t}dt\;(\mathrm{R}_a)^{-1}\right)_{s,n} = \frac{B}{D}\sum_{n=0}^{K}((n-S)^{+})\left(\mathrm{R}_a e^{\widetilde{\mathrm{V}_a}}(\mathrm{R}_a)^{-1}\right)_{s,n}$$

(2.18)

In (2.18), R_a and V_a matrices are the corresponding eigenvector and eigenvalue matrices of the transition rate matrix of a $M/M/1/K$ queuing system with a service rate of μ_l if action $(a = 0)$ and with a service rate μ_h if action $(a = 1)$ is taken. Therefore, we have $P\{N_d(t) = n|s,a\} = \mathrm{P}_{sn}\,(t,\mu_l)$ for $a = 0$ and $P\{N_d(t) = n|s,a\} = \mathrm{P}_{sn}(t,\mu_h)$ for $a = 1$.

Note that $\widetilde{\mathrm{V}_a}$ is a different $(K+1)\times(K+1)$ sized diagonal matrix than V_a. Suppose $\{\xi_0, \xi_1, \dots \xi_K\}$ are the diagonal entries of the eigenvalue matrix V_a. In such a case, the diagonal entries of $\widetilde{\mathrm{V}}_a$ will be $\left\{\int_{t=0}^{D} e^{\xi_0 t}dt\,, \int_{t=0}^{D} e^{\xi_1 t}dt\,, \dots \int_{t=0}^{D} e^{\xi_K t}dt\right\}$.

Hence if the eigenvalue j, $\xi_j = 0$, then the corresponding j^{th} element of $\widetilde{\mathrm{V}}_a$ will be equal to D; whereas if ξ_j<0 then the corresponding j^{th} element of $\widetilde{\mathrm{V}}_a$ will be equal to $\frac{e^{\xi_j D}-1}{\xi_j}$.

Due to the numerical procedures described above, the derivation of expected period cost rate $r(s,a) = hS + c(a) + dtc(s,a)$ can be found for all states: $s \in \{0,1,\dots K\}$, for all possible actions $a \in \{0,1\}$, for given stock level: S, given period length: D and given capacity levels: μ_l and μ_h.

After the computation of all $r(s,a)$ values and transition probabilities: $p(s'|s,a)$, the optimality equation in (2.10) can be solved with the existing value/policy iteration algorithms in the MDP literature, in order to find the optimal solution to the sub-problem in (2.7). The optimal solution of each sub-problem is going to be used in the minimization of the global problem in (2.8), in which more upper-level decisions such as base stock level S, period length D and capacity levels: μ_l and μ_h are taken. We explain the details of the solution procedure for the problem (2.6) in the next subsection.

2.5.3 Numerical Study

In this section, we use the analysis and the results provided in the previous section in order to assess the performance of the two level flexible capacity mode. In the end, we compare the cost performance of this mode with the minimum cost rate achieved under the fixed capacity mode, $TRC_F{}^*$. The outline of this section is as follows. First, we describe the characteristics and the test bed of the computational study. Second, we present the search procedure in order to find the near-optimal policy parameters for the problem given in (2.6). Afterwards, we assess the potential cost benefits that can be gained in two level flexible mode in comparison to the best policy of the reference, fixed capacity mode. Finally, we discuss the results of the search procedure, examine the sensitivity of and the interactions among the policy and system parameters and the accuracy of the finite waiting room approximation.

In our computational study, we normalize the arrival rate for the subsystem failures in the whole environment, $\lambda = 1$ (failures per time unit) as well as the unit permanent capacity cost per time unit, $c_p = 1$. The other parameters are scaled according to normalized λ and c_p. Similar to Section 2.4.3, we analyze a total of 9 scenarios with 3 different B/h values and 3 different h values, which are already demonstrated in Table 2-2. For each of these 9 scenarios and different (Δ, α) combinations, we execute our solution procedure to find the capacity policy parameters, (D, μ_l, μ_h, k) and S that yield the minimum total costs.

We use a different solution procedure for the optimization problem than the procedure for the fixed capacity mode. However there are many similarities in between these two procedures. For instance, in Search Procedure-II, we also have an arbitrarily large S_{max} value as an upper limit for the spare unit base stock level choice. In addition, the fixed capacity results obtained from the Search-Procedure-I, not only serve as reference points to assess the overall cost performance, but also help us to develop a method for generating meaningful candidate value sets in our search procedure for permanent and permanent plus contingent levels in the two-level flexible capacity mode settings.

Previously, we have discussed that the capacity agency can supply the agreed amount of contingent capacity at a given frequency. This frequency is driven by the period length, which is chosen from θ, the set of candidate period lengths offered. In our thesis we assume $\theta = \{0.5, 1, 1.5, \dots, 4.5, 5\}$. These values are scaled to the normalized inter-arrival time: $\frac{1}{\lambda} = 1$. Thus, a period length of $D = 1$ corresponds to the mean inter-arrival time of system failures due to the part in concern. Next, we give our search procedure followed by the description of its underlying mechanism.

Search Procedure-II

---

0. Follow steps a & b:

a. Let $\mu^*(0) = \lambda + \sqrt{\frac{\lambda B}{c_p}}$ denote the μ that satisfies: $\frac{\partial TRC(\mathbf{C},0)}{\partial \mu} = 0$ from $TRC(\mathbf{C}, S)$ in Equation (2.4) when $S = 0$.

b. For every $S \in \{1,2, \ldots, S_{max}\}$:

   Set $\mu^*(S)$, which is the μ that satisfies: $\frac{\partial TRC(\mathbf{C},S)}{\partial \mu} = 0$ from $TRC(\mathbf{C}, S)$ in Equation (2.4).

1. For every $S \in \{0,1,2, \ldots, S_{max}\}$ follow the steps a & b:

   a. Create the Υ_l and Υ_h sets from $\mu^*(S)$ values as follows:

   $$\Upsilon_l = \left\{\frac{2}{10}\mu^*(S),\ \frac{3}{10}\mu^*(S),\ \frac{4}{10}\mu^*(S),\ \ldots,\ \frac{8}{10}\mu^*(S),\ \frac{9}{10}\mu^*(S)\right\}$$

   $$\Upsilon_h = \left\{\frac{12}{10}\mu^*(S),\ \frac{14}{10}\mu^*(S),\ \frac{16}{10}\mu^*(S),\ \ldots,\ \frac{24}{10}\mu^*(S),\ \frac{26}{10}\mu^*(S)\right\}$$

   b. For every $D \in \theta$, $\mu_l \in \Upsilon_l$ and $\mu_h \in \Upsilon_h$ solve the sub-problem (2.7) and obtain the corresponding $\vec{\pi}^*$.

2. After solving the sub-problem (2.7) for all $D \in \theta$, $\mu_l \in \Upsilon_l$ and $\mu_h \in \Upsilon_h$, we can find the minimum cost for given S, $TRC(\mathbf{C}^*(S), S)$ as follows:

$$TRC(\mathbf{C}^*(S), S) = \min_{D \in \theta, \mu_l \in \Upsilon_l, \mu_h \in \Upsilon_h} (TRC(\mathbf{C}, S): \mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}^*])$$

3. After $TRC(\mathbf{C}^*(S), S)$ for all $S \in \{0, 1, 2, \ldots, S_{max}\}$ are found, we can find S^*, and $\mathbf{C}^*(S^*)$ = [D^*, ${\mu_l}^*$, ${\mu_h}^*$, $\vec{\pi}^*$] values as follows:

$$TRC(\mathbf{C}^*(S^*), S^*) = \min_{S \in \{0,1,2,\ldots,S_{max}\}} TRC(\mathbf{C}^*(S), S)$$

2.5.3.1 Explanation of the Search Procedure-II, Some Key Observations

In the search procedure above, at step 0, we find a reference capacity level for every S: $\mu^*(S)$, which is the capacity level that results in the minimum TRC for a given $S \in \{0,1, \ldots, S_{max}\}$ from the single, fixed capacity model. This reference capacity level is used later to determine the permanent (μ_l) & permanent + contingent (μ_h) capacity levels, which will be used in the two-level flexible capacity policy, $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$.

After the reference point $\mu^*(S)$ is found for each base stock level $S \in \{0,1, \ldots, S_{max}\}$ from the derivative of Equation (2.4), we are ready to construct the sets Υ_l and Υ_h, which contain the low and high capacity levels (μ_l and μ_h), respectively. Note that for a given S and given period length D, μ_l and μ_h should satisfy: $\mu_l \leq \mu^*(S) \leq \mu_h$. Otherwise the resulting total costs under two level capacity: $TRC(\mathbf{C}, S)$ will be already more than the total cost under fixed capacity of $\mu^*(S)$.

Actually, there can be an infinite number of (μ_l, μ_h) possibilities. However, for computational time reasons, we limit ourselves to a total of 64 (μ_l, μ_h) possibilities. For a given stock level S, we have tested 8 equidistantly scattered (with a distance of: $\frac{\mu^*(S)}{10}$) μ_l candidates that are lower, and 8 equidistantly scattered (with a distance of: $\frac{\mu^*(S)}{5}$) μ_h candidates that are higher than $\mu^*(S)$. The reasons for these particular (μ_l, μ_h) choices can be explained as follows:

1) μ_l is bounded by 0, whereas μ_h is not bounded from above, therefore the distance between μ_h candidates is twice higher than the distance between μ_l candidates.
2) Permanent and contingent capacity costs per unit time (c_p and c_c) are a lot higher than the holding cost per unit time h. ($\frac{c_p}{h}$ varies from 4 to 40) This makes the choice of μ_l and μ_h much more sensitive compared to the choice of S.

Therefore, we have $\Upsilon_l = \left\{\frac{2}{10}\mu^*(S),\ \frac{3}{10}\mu^*(S),\ \frac{4}{10}\mu^*(S),\ \ldots,\ \frac{9}{10}\mu^*(S)\right\}$ and $\Upsilon_h = \left\{\frac{12}{10}\mu^*(S),\ \frac{14}{10}\mu^*(S),\ \frac{16}{10}\mu^*(S),\ \ldots,\ \frac{24}{10}\mu^*(S),\ \frac{26}{10}\mu^*(S)\right\}$, which lead to a total of 64 (μ_l, μ_h) pairs for each S, where $\mu_l \in \Upsilon_l$ and $\mu_h \in \Upsilon_h$.

For each base stock level S, period length $D \in \theta$ and (μ_l, μ_h) pair, sub-problem (2.7) is solved in step 2, where the optimal policy $\vec{\pi}^*$ is found with the help of the MDP formulation of the sub-problem in (2.9)

After (2.7) is solved with the MDP formulation of (2.9) for given: $S \leq S_{max}$, $D \in \theta$, $\mu_l \in \Upsilon_l$ and $\mu_h \in \Upsilon_h$, we find the optimal parameters (S^*, D^*, ${\mu_l}^*$ and ${\mu_h}^*$) by brute force search in steps 2 & 3 from the Search Procedure-II. These steps complete the solution procedure for the optimization problem (2.6).

Before starting the discussions, we would like to give one of the key observations concerning the threshold structure of the optimal policy: $\vec{\pi}^*$, which appear ubiquitously in the results of our numerical study.

This observation is in line with the current structural optimality results in the queuing control literature (See (StidhamJr 2002) for an extensive overview), which mostly use event based dynamic programming techniques. The periodicity of the capacity action taking points in our study hinders us to use event based dynamic programming techniques to prove threshold optimality. However the ubiquitous appearance of this threshold optimality occurrence heartens us to postulate this structural property of threshold type policies as a conjecture.

Conjecture 2.1: For any given stock level S, given permanent & permanent + contingent capacity levels ($\mu_l \in \Upsilon_l$ & $\mu_h \in \Upsilon_h$) and given period length $D \in \theta$, the optimal policy: $\vec{\pi}^*$ is always of a threshold type policy and can be characterized with a single, non-negative integer switching point: k^*. Under this optimal policy: $\vec{\pi}^*$, at the start of each period length D, if there are less than k^* defective parts in the repair shop, the repair shop uses only permanent capacity: μ_l $(a = 0)$ otherwise, the repair shop uses both permanent and contingent capacity, $(a = 1)$, and $\mu_h - \mu_l$ is hired on top of the permanent capacity until the next period.

After postulating this conjecture, we continue with the discussion on the results of the numerical study that is conducted, where the optimal costs for (2.6) are obtained from the Search Procedure-II, which is described above.

2.5.3.2 Discussions on the Results of Numerical Study

In this subsection we discuss the results of our numerical study. We first present the total savings in total relevant costs when the best two level flexible capacity policy is employed compared to the best fixed capacity system (${TRC_F}^*$ in short). Afterwards, the discussion on the savings is followed by the sensitivity analyses of the cost and the optimal policy parameters and the discussion on the simulation study which inspects the accuracy of the finite waiting room approximation.

2.5.3.3 Savings of Two-Level Capacity Policies Compared to the Single Level Capacity (${TRC_F}^*$)

In this subsection, we contemplate the potential savings that two level capacity flexibility brings compared to the ${TRC_F}^*$. Total costs can be reduced up to a great extent due to the two-level capacity policy. From our numerical results, we have witnessed that up to 75% savings can be achievable in total costs due to the two-level flexible capacity mode compared to the minimum cost that can be achieved in the single capacity mode ${TRC_F}^*$.

In Table 2-5, we give the maximum percentage savings that two level flexible policies can bring for all 9 different B & c_p scenarios (which are already listed in Table 2-2) with $\Delta = 0, 0.25, 0.5$ and 1 and $\alpha = 0{,}1$ and 2, when c_c is an inversely proportional, exponential and linear function of the period length. Suppose for given cost parameters (B, c_p, Δ and α), and a functional form for c_c , ${TRC_2}^*$ represents the minimum total costs that can be achieved from Search Procedure-II. After ${TRC_2}^*$ is found, the percentage savings in Table 2-5 can be calculated from: $\frac{({TRC_F}^* - {TRC_2}^*)}{{TRC_F}^*}$. The parameter sets, which result in higher percentage savings are highlighted with a darker gray tone.

From Table 2-5, we can observe that the percentage savings seem to decrease with holding cost rate h. This is due to the fact that for higher holding cost rate h, although two level capacity mode can achieve more cost savings, percent wise it is smaller, since the reference cost parameter, ${TRC_F}^*$, is greater and the share of the holding costs (HC) in ${TRC_F}^*$ is bigger. We also observe that the savings increase with the elasticity factor α. The more elastic the wage is with respect to the period length, the cheaper the contingent capacity gets for longer period lengths. Similarly, the maximum opportunity cost Δ, increases the price of the contingent capacity, which leads to a decline in the percentage savings, as well. The down-time related cost rate B seems to increase the percentage savings in general, however there are some contradictory instances where the percentage savings decrease with higher B. The interaction between B and savings in TRC needs further research.

In the table, we can see that the best two-level flexible capacity policy outperforms the best single level capacity for all instances (with different B, h, Δ and α values). From the other extrapolated numerical studies which are not listed on the table, we observe that two level capacity flexibility starts to become costlier compared to the fixed capacity mode, when the cost of the contingent capacity is very high compared to the permanent capacity cost (i.e: $\frac{\Delta}{c_p} > 15$) and when the contingent capacity is rather insensitive to the period length (i.e. $\alpha = 0.5$ and lower). The reason of such an underperformance is due to the fact that in the two level policies, even if the contingent capacity is prohibitively expensive, the use of the contingent capacity cannot be avoided, since the permanent capacity level is too low when deployed alone.

However, in the specialized system environment, the realistic cost parameters for contingent capacity are in the area where the contingent capacity is reasonably priced (i.e. $\frac{\Delta}{c_p} \leq 1$) and therefore using optimal two-level flexibility is always more economical for all the cost-parameter scenarios in the numerical test bed. It can be concluded that the two level capacity policies bring savings compared to ${TRC_F}^*$, at different magnitudes, especially with low max. opportunity costs (Δ) and high elasticity parameters ($\alpha > 0$).

The functional form for the contingent capacity cost structure affects the optimal capacity policy $\mathbf{C}^*$, the inventory decision S^*, and the percentage of the savings if there is a wage differential effect ($\Delta, \alpha > 0$). With the same cost parameters, under optimal policy $\mathbf{C}^*$, linear c_c structure results in greater savings compared to other cost structures (exponential and inversely proportional) in the cost scenarios investigated. The differences in the total cost can be up to 24% between linear and exponential cost structures and can be up to 12% between exponential and inversely proportional cost structures. The differences in percentage savings decrease with higher B. The percentage differences between the savings of two level policy when c_c has exponential form and an inverse proportional form, increase with Δ as well as α. On the other hand, the differences in percentage savings when c_c has linear form and an inversely proportional form increase with Δ, however first increase and then decrease with α. This can be explained due to the fact that when c_c has an exponential structure, it monotonically decreases with D, however when c_c has a linear structure, it stays the same and equal to the permanent capacity cost, c_p, after some D. Therefore the difference between percentage savings can be the biggest, when c_c is equal to c_p for the first time in the linear contingent capacity cost function.

Finally, we explore further how the optimal policy parameters change under the optimal two-level periodic capacity flexible mode compared to the single level capacity mode for different cost parameter settings. In Table 2-6, we show how the optimal two level capacity mode policy parameters (S^*, D^*, k^*, ${\mu_l}^*$ and ${\mu_h}^*$) differ with various (Δ, α) combinations and 4 B & h scenarios.

The data in Table 2-6 illustrate that under the optimal policies pertaining to the two level capacity mode, the cost savings compared to ${TRC_F}^*$ come from both less capacity deployment as well as less spare unit inventory holding costs. It can be seen that for each of the 4 B & h scenario and (Δ, α) combination, S^* under the two level capacity mode is less than or equal to the ${S_F}^*$, which is the optimal stock level under the single-level capacity mode. The differences in spare unit stock levels are higher for lower h and higher B, lower Δ , and higher α parameters. In Table 2-6, it can be seen that the smallest period length $D = 0.5$, is chosen as the optimal period length in most of the instances. However, for lower h values, high Δ and positive elasticity ($\alpha > 0$), higher period lengths ($D > 0.5$) can be optimal, as well. Although two level capacity policies cause savings in capacity related costs (CRC) due to less deployed average capacity, $ACU(\mathbf{C})$, the optimal capacity policy parameters in Table 2-6 (k^*, ${\mu_l}^*$ and ${\mu_h}^*$) can provide further insights.

From Table 2-6, we can observe that in the two level capacity mode, the choice on the optimal permanent capacity ${\mu_l}^*$ is considerably lower in comparison to optimal capacity level in the single capacity mode $\mu^*({S_F}^*)$. Generally, ${\mu_l}^*$ changes parallel to S^*, but apart from S^*, it is also dependent on (Δ, α) parameters, as well. We can see that ${\mu_l}^*$ increases with Δ and decreases with α, which suggests that the repair shop hedges risk by deploying more permanent capacity when the contingent capacity becomes more expensive and its price gets more insensitive with respect to the period length. We do not observe a monotonic relation between the choice of ${\mu_h}^*$ and the other cost parameters. We observed that in the Search Procedure-II, not only the maximum but also the intermediate values of μ_h are chosen in Υ_h, which gives us an enough confidence on the validity of the construction of Υ_h. It is remarkable that the contingent capacity volume always exceeds the optimal capacity level in the single capacity mode $\mu^*({S_F}^*)$, however it is less frequently deployed, especially for high Δ, which leads to higher threshold values k^*.

These inter-relations between the capacity and cost parameters will be further explicated in the next section, where a list of sensitivity analyses are conducted on policy and cost parameters. The cost parameter values used in the next section differ from the core experimental setting explained in Table 2-2, due to the following reasons:

1) In the sensitivity analysis for the specialized system environment, numerical test-bed of the commoditized system setting is used in order to compare the performance of different down-time service strategies (keeping spare unit inventory vs. hiring a substitute from rental/3rd party supplier) under the same capacity mode.

2) In the specialized system experimental setting (Table 2-2), the down-time (B) and capacity costs (c_p, c_c) per unit time are strongly higher than the holding cost per unit time (h) , which can blanket over some of the interactions between capacity and spare unit inventory related decisions. On the other hand, the effects and the sensitivity of the

policy parameters are clearer and can be crystalized in a more concrete way in the numerical test bed of the commoditized setting.

3) In some of the search procedures that are going to be generated in Chapter 3, insights from the sensitivity analysis results from Chapter 2 will be used.

2.5.3.4 Sensitivity Analysis of the Optimal Policy Parameters

In this subsection, we discuss the inter-relations among the cost and optimal policy parameters. We first focus on how the optimal switching point k^* is affected by other cost/policy parameters. Afterwards we investigate the role of the period length on TRC and on the choice of other policy parameters. Finally, we examine how the base stock level S affects TRC and other parameters.

2.5.3.4.1 The Optimal Switching Point: $\boldsymbol{k}^*$

Our main concern in this subsection is how the optimal switching point k^*, changes with different policy and cost parameters. Therefore in Table 2-7, we tabulate how the k^* value from the policy $\vec{\pi}^*$ from problem (2.8) responds to changes in permanent and contingent capacity levels (μ_l, μ_h), period length D and base stock level S when $B = 10$, $c_p = 5$, $\Delta = 5$ and $\alpha = 2$.

In short, the results from Table 2-7 suggest that, ceteris paribus:

- The lower μ_l, the earlier (i.e. smaller number of parts at the repair shop) we switch from low to high capacity (or vice versa).
- The higher μ_h, the later (i.e. larger number of parts at the repair shop) we switch from low to high capacity (or vice versa).
- The larger D, the earlier we switch from low to high capacity.
- The higher S, the later we switch from low to high.

Two Level

h	Δ	B=1.25	inv. Lin	exp	Lin
h =0.025	Δ=1	α=0	36%	36%	36%
		α=1	44%	52%	69%
		α=2	49%	60%	75%
	Δ=0.5	α=0	46%	46%	46%
		α=1	52%	56%	75%
		α=2	56%	63%	75%
	Δ=0.25	α=0	56%	56%	56%
		α=1	60%	62%	75%
		α=2	64%	66%	75%
	Δ=0	all α	75%	75%	75%

h	Δ	B=2.5	inv. Lin	exp	Lin
h =0.025	Δ=1	α=0	39%	39%	39%
		α=1	46%	52%	69%
		α=2	50%	61%	75%
	Δ=0.5	α=0	48%	48%	48%
		α=1	54%	57%	75%
		α=2	57%	64%	75%
	Δ=0.25	α=0	57%	57%	57%
		α=1	61%	62%	75%
		α=2	64%	67%	75%
	Δ=0	all α	75%	75%	75%

h	Δ	B=5	inv. Lin	exp	Lin
h =0.025	Δ=1	α=0	41%	41%	41%
		α=1	47%	53%	70%
		α=2	52%	61%	75%
	Δ=0.5	α=0	50%	50%	50%
		α=1	55%	58%	75%
		α=2	59%	64%	75%
	Δ=0.25	α=0	59%	59%	59%
		α=1	63%	64%	75%
		α=2	65%	68%	75%
	Δ=0	all α	75%	75%	75%

h	Δ	B=2.5	inv. Lin	exp	Lin
h =0.05	Δ=1	α=0	35%	35%	35%
		α=1	41%	44%	68%
		α=2	45%	53%	68%
	Δ=0.5	α=0	45%	45%	45%
		α=1	50%	51%	68%
		α=2	53%	56%	68%
	Δ=0.25	α=0	53%	53%	53%
		α=1	57%	58%	68%
		α=2	60%	62%	68%
	Δ=0	all α	68%	68%	68%

h	Δ	B=5	inv. Lin	exp	Lin
h =0.05	Δ=1	α=0	38%	38%	38%
		α=1	43%	45%	68%
		α=2	47%	54%	68%
	Δ=0.5	α=0	47%	47%	47%
		α=1	52%	53%	68%
		α=2	55%	58%	68%
	Δ=0.25	α=0	55%	55%	55%
		α=1	58%	59%	68%
		α=2	61%	63%	68%
	Δ=0	all α	68%	68%	68%

h	Δ	B=10	inv. Lin	exp	Lin
h =0.05	Δ=1	α=0	40%	40%	40%
		α=1	45%	47%	68%
		α=2	48%	54%	68%
	Δ=0.5	α=0	48%	48%	48%
		α=1	52%	53%	68%
		α=2	56%	58%	68%
	Δ=0.25	α=0	56%	56%	56%
		α=1	59%	60%	68%
		α=2	61%	63%	68%
	Δ=0	all α	68%	68%	68%

h	Δ	B=12.5	inv. Lin	exp	Lin
h =0.25	Δ=1	α=0	30%	30%	30%
		α=1	32%	33%	38%
		α=2	37%	40%	50%
	Δ=0.5	α=0	37%	37%	37%
		α=1	37%	37%	50%
		α=2	43%	45%	50%
	Δ=0.25	α=0	43%	43%	43%
		α=1	43%	43%	50%
		α=2	46%	47%	50%
	Δ=0	all α	50%	50%	50%

h	Δ	B=25	inv. Lin	exp	Lin
h =0.25	Δ=1	α=0	31%	31%	31%
		α=1	34%	35%	39%
		α=2	38%	41%	50%
	Δ=0.5	α=0	38%	38%	38%
		α=1	39%	40%	50%
		α=2	43%	45%	50%
	Δ=0.25	α=0	43%	43%	43%
		α=1	43%	44%	50%
		α=2	47%	48%	50%
	Δ=0	all α	50%	50%	50%

h	Δ	B=50	inv. Lin	exp	Lin
h =0.25	Δ=1	α=0	33%	33%	33%
		α=1	35%	36%	39%
		α=2	38%	41%	49%
	Δ=0.5	α=0	38%	38%	38%
		α=1	41%	42%	49%
		α=2	43%	45%	49%
	Δ=0.25	α=0	43%	43%	43%
		α=1	45%	45%	49%
		α=2	46%	47%	49%
	Δ=0	all α	49%	49%	49%

Table 2-5: The percentage cost savings of two level flexible capacity policy compared to the fixed capacity policy when c_c has an inversely proportional, exponential and linear structure, when $B/h = 50{,}100$ and 200, $h = 0.025{,}0.05$ and 0.25, for $\Delta = 0, 0.25, 0.5$ and 1 & $\alpha = 0, 1$ and 2.

These trends are general, no matter which cost function is chosen (linear, inverse proportional or exponential) for the contingent capacity. These behaviors can be explained as follows: as the permanent capacity (or contingent capacity) gets higher, the repair shop would hire contingent capacity less frequently and at higher workloads. On the other hand, shorter period lengths enable more frequent capacity updates, in other words enable faster recourse actions, which incentivize the repair shop to use contingent capacity at higher workloads. In a similar vein, k^* increases as S increases, although $\mu^*(S)$ is decreasing with S. This is due to the fact that a larger spare part inventory may decrease the DTC (down-time related costs) too much, such that trimming the capacity usage of the repair shop by hiring contingent capacity at higher workloads is cost beneficial.

After discussing the interactions between k^* and the other policy parameters, now we investigate the effects of capacity cost parameters (Δ and α). To illustrate the impacts of Δ and α, we present Table 2-8, where the k^* are tabulated for $B = 10$, $c_p = 5$, $S = 3$ and $D = 2$ under different Δ and α values with 3 different capacity cost structures.

What we observe from Table 2-8 is the following: as the contingent capacity becomes more expensive, the repair shop has more incentive to use only permanent capacity more frequently and at more workloads. Therefore the switching point k^* gets larger. For the same period length D, the cost of the contingent capacity becomes more expensive as Δ increases. On the other hand when the time-elasticity factor, α increases, the contingent capacity gets cheaper, which yields to a more frequent hiring of the contingent capacity, thus a lower k^*.

The cost structure of contingent capacity determines how expensive the capacity related cost rate is and how fast it decreases. From Table 2-7 and Table 2-8, we observe that under exponential cost structure, we have the same or lower optimal switching points compared to the k^* values under 2 other cost structures. However, it can change under different Δ and α values.

Having completed our discussion on how k^* is affected by various parameters, next we summarize our findings on the bidirectional relations between period length, D, TRC and other policy parameters.

B=5, h=0.05			S_F^*		$\mu^*(S_F^*)$		
Single Level			10		1.54		
Two-Level			S^*	D^*	k^*	$\mu^*_l(S^*)$	$\mu^*_h(S^*)$
inv. Linear	Δ=0	all α	6	0.5	4	0.35	3.89
	Δ=0.25	α=0	6	0.5	5	0.53	4.60
		α=1	6	0.5	5	0.53	4.60
		α=2	6	0.5	4	0.35	3.89
	Δ=0.5	α=0	7	0.5	6	0.85	4.41
		α=1	6	0.5	5	0.71	4.60
		α=2	6	0.5	5	0.53	4.60
	Δ=1	α=0	8	0.5	7	0.98	4.25
		α=1	7	0.5	6	0.85	4.07
		α=2	7	0.5	6	0.85	4.41

B=25, h=0.25			S_F^*		$\mu^*(S_F^*)$		
Single Level			5		2.26		
Two-Level			S^*	D^*	k^*	$\mu^*_l(S^*)$	$\mu^*_h(S^*)$
inv. Linear	Δ=0	all α	4	0.5	3	0.50	5.97
	Δ=0.25	α=0	4	0.5	3	0.50	5.97
		α=1	4	0.5	3	0.50	5.97
		α=2	4	0.5	3	0.50	5.97
	Δ=0.5	α=0	4	0.5	3	0.75	5.47
		α=1	4	0.5	3	0.75	5.47
		α=2	4	0.5	3	0.50	5.97
	Δ=1	α=0	5	0.5	4	0.91	5.43
		α=1	5	0.5	4	0.91	4.53
		α=2	4	0.5	3	0.75	5.47

B=5, h=0.025			S_F^*		$\mu^*(S_F^*)$		
Single Level			15		1.40		
Two-Level			S^*	D^*	k^*	$\mu^*_l(S^*)$	$\mu^*_h(S^*)$
inv. Linear	Δ=0	all α	7	0.5	5	0.34	4.41
	Δ=0.25	α=0	8	0.5	6	0.65	4.25
		α=1	7	0.5	5	0.51	4.07
		α=2	7	0.5	5	0.51	4.07
	Δ=0.5	α=0	8	0.5	7	0.82	4.25
		α=1	9	0.5	7	0.79	4.12
		α=2	8	0.5	6	0.65	4.25
	Δ=1	α=0	11	0.5	10	1.06	3.92
		α=1	11	1	9	0.90	3.62
		α=2	10	1	8	0.77	4.01

B=10, h=0.05			S_F^*		$\mu^*(S_F^*)$		
Single Level			11		1.58		
Two-Level			S^*	D^*	k^*	$\mu^*_l(S^*)$	$\mu^*_h(S^*)$
inv. Linear	Δ=0	all α	6	0.5	4	0.38	4.56
	Δ=0.25	α=0	7	0.5	5	0.54	4.70
		α=1	7	0.5	5	0.54	4.34
		α=2	6	0.5	4	0.38	4.56
	Δ=0.5	α=0	7	0.5	6	0.90	4.70
		α=1	7	0.5	5	0.72	4.34
		α=2	7	0.5	5	0.54	4.70
	Δ=1	α=0	8	0.5	7	1.04	4.51
		α=1	7	0.5	6	0.90	4.34
		α=2	7	0.5	6	0.90	4.70

Table 2-6: The optimal two-level capacity mode policy parameters (S^*, D^*, k^*, ${\mu_l}^*$ and ${\mu_h}^*$) under different 4 B & h scenarios (**1:** $B = 5, h = 0.05$ **2:** $B = 25, h = 0.25$ **3:** $B = 5, h = 0.025$ and **4:** $B = 10, h = 0.25$) and various (Δ, α) combinations when $c_p = 1$.

$S = 3$, $D = 1$	$\mu_l =$	$0.2\mu^*$	$0.3\mu^*$	$0.4\mu^*$	$0.5\mu^*$	$0.6\mu^*$	$0.7\mu^*$	$0.8\mu^*$	$0.9\mu^*$
$\mu_h = 1.1\mu^*$	inv.	2	2	2	2	2	3	3	3
	lin.	2	2	2	2	3	3	3	3
	exp.	2	2	2	2	2	2	3	3
$\mu_h = 1.2\mu^*$	inv.	2	2	2	2	3	3	3	3
	lin.	2	2	2	3	3	3	3	3
	exp.	2	2	2	2	3	3	3	3

$S = 3$, $D = 2$	$\mu_l =$	$0.2\mu^*$	$0.3\mu^*$	$0.4\mu^*$	$0.5\mu^*$	$0.6\mu^*$	$0.7\mu^*$	$0.8\mu^*$	$0.9\mu^*$
$\mu_h = 1.1\mu^*$	inv.	2	2	2	2	2	2	3	3
	lin.	2	2	2	2	2	2	3	3
	exp.	1	2	2	2	2	2	2	3
$\mu_h = 1.2\mu^*$	inv.	2	2	2	2	3	3	3	3
	lin.	2	2	2	2	3	3	3	3
	exp.	2	2	2	2	2	3	3	3

$S = 3$, $D = 3$	$\mu_l =$	$0.2\mu^*$	$0.3\mu^*$	$0.4\mu^*$	$0.5\mu^*$	$0.6\mu^*$	$0.7\mu^*$	$0.8\mu^*$	$0.9\mu^*$
$\mu_h = 1.1\mu^*$	inv.	1	1	2	2	2	2	3	3
	lin.	1	1	1	2	2	2	2	3
	exp.	1	1	1	2	2	2	2	3
$\mu_h = 1.2\mu^*$	inv.	1	2	2	2	2	3	3	3
	lin.	1	1	2	2	2	2	3	3
	exp.	1	1	2	2	2	2	3	3

$S = 6$, $D = 2$	$\mu_l =$	$0.2\mu^*$	$0.3\mu^*$	$0.4\mu^*$	$0.5\mu^*$	$0.6\mu^*$	$0.7\mu^*$	$0.8\mu^*$	$0.9\mu^*$
$\mu_h = 1.2\mu^*$	inv.	2	3	3	3	3	4	4	4
	lin.	2	3	3	3	3	4	4	4
	exp.	2	2	3	3	3	3	4	4
$\mu_h = 1.2\mu^*$	inv.	3	3	3	3	4	4	4	5
	lin.	3	3	3	3	4	4	4	5
	exp.	3	3	3	3	4	4	4	5

Table 2-7: The optimal switching point (k^*) derived from (2.9) when $B = 10$, $c_p = 5$, $\Delta = 5$ and $\alpha = 2$ for different S, D, μ_l and μ_h values & 3 different capacity cost structures (taken from Table 2-1).

$\Delta = 5$, $\alpha = 0$	$\mu_l =$	$0.2\ \mu^*$	$0.3\ \mu^*$	$0.4\ \mu^*$	$0.5\ \mu^*$	$0.6\ \mu^*$	$0.7\ \mu^*$	$0.8\ \mu^*$	$0.9\ \mu^*$
$\mu_h = 1.1\ \mu^*$	inv.	2	2	2	2	2	3	3	3
	lin.	2	2	2	2	2	3	3	3
	exp.	2	2	2	2	2	3	3	3
$\mu_h = 1.2\ \mu^*$	inv.	2	2	3	3	3	3	4	4
	lin.	2	2	3	3	3	3	4	4
	exp.	2	2	3	3	3	3	4	4

$\Delta = 5$, $\alpha = 1$	$\mu_l =$	$0.2\ \mu^*$	$0.3\ \mu^*$	$0.4\ \mu^*$	$0.5\ \mu^*$	$0.6\ \mu^*$	$0.7\ \mu^*$	$0.8\ \mu^*$	$0.9\ \mu^*$
$\mu_h = 1.1\ \mu^*$	inv.	2	2	2	2	2	3	3	3
	lin.	2	2	2	2	3	3	3	3
	exp.	1	2	2	2	2	2	3	3
$\mu_h = 1.2\ \mu^*$	inv.	2	2	2	2	3	3	3	3
	lin.	2	2	2	3	3	3	3	4
	exp.	2	2	2	2	2	3	3	3

$\Delta = 5$, $\alpha = 2$	$\mu_l =$	$0.2\ \mu^*$	$0.3\ \mu^*$	$0.4\ \mu^*$	$0.5\ \mu^*$	$0.6\ \mu^*$	$0.7\ \mu^*$	$0.8\ \mu^*$	$0.9\ \mu^*$
$\mu_h = 1.1\ \mu^*$	inv.	2	2	2	2	2	2	3	3
	lin.	2	2	2	2	2	2	3	3
	exp.	1	2	2	2	2	2	2	3
$\mu_h = 1.2\ \mu^*$	inv.	2	2	2	2	3	3	3	3
	lin.	2	2	2	2	3	3	3	3
	exp.	2	2	2	2	2	3	3	3

$\Delta = 10$, $\alpha = 2$	$\mu_l =$	$0.2\ \mu^*$	$0.3\ \mu^*$	$0.4\ \mu^*$	$0.5\ \mu^*$	$0.6\ \mu^*$	$0.7\ \mu^*$	$0.8\ \mu^*$	$0.9\ \mu^*$
$\mu_h = 1.2\ \mu^*$	inv.	2	2	2	2	2	3	3	3
	lin.	2	2	2	3	3	3	3	4
	exp.	1	2	2	2	2	2	2	3
$\mu_h = 1.2\ \mu^*$	inv.	2	3	3	3	3	4	4	4
	lin.	2	3	3	3	3	4	4	4
	exp.	2	2	3	3	3	3	4	4

Table 2-8: The optimal switching point (k^*) derived from (2.9) when for $B = 10$, $c_p = 5$, $S = 3$ and $D = 2$.

2.5.3.4.2 Period Length: D

The period length D plays a central role in this capacity mode, as it induces the level of the adaptability of the repair shop capacity to the workload and it determines the per time unit cost of the contingent capacity due to the capacity provision mechanism. Accordingly, in this part we investigate the sensitivity of the total costs and the optimal policy parameters to the period length D under different contingent cost functions with 3 different structures and Δ and α parameters.

In order to pursue an investigation on the effects of period length per se, we worked on a scenario when the capacity agency supplies the agreed amount of capacity only at a given frequency (i.e. $\theta = \{D\}$, a set that consists of a single period length) and we run the search procedure described above in order to optimize other policy parameters (${\mu_l}^*$, ${\mu_h}^*$ and k^*). We follow these steps for increasing values of period length by the order of: $D = 0.5, 1, 1.5, \ldots 5$ with different Δ and α values. Therefore note that the TRC^* notation in this part is used for the optimal total costs <u>for a given period length</u>. In Figure 2-5, the behavior of the optimal TRC^* at increasing period lengths, under 2 level capacity policy is illustrated for $B = 10, c_p = 5, \alpha = 2$ and $\Delta = 0, 5$ and 10, when c_c function has exponential (left) and inverse proportional (right) structure.

In Figure 2-5, it can be observed that in the two level capacity mode, all optimal TRC^* values at given period lengths D from 0.5 to 5 for the chosen Δ values ($0, 5$ and 10) engender smaller total cost realizations compared to the optimal cost in the single capacity mode, ${TRC_F}^*$. Of course this occurrence can take a different turn for even higher values of Δ coupled with a low elasticity α, or even longer period lengths ($D > 5$). Therefore, any meaningful and viable two-level capacity mode alternative at a given period length D should yield a total cost value within a region that is bounded by ${TRC_F}^*$ from above.

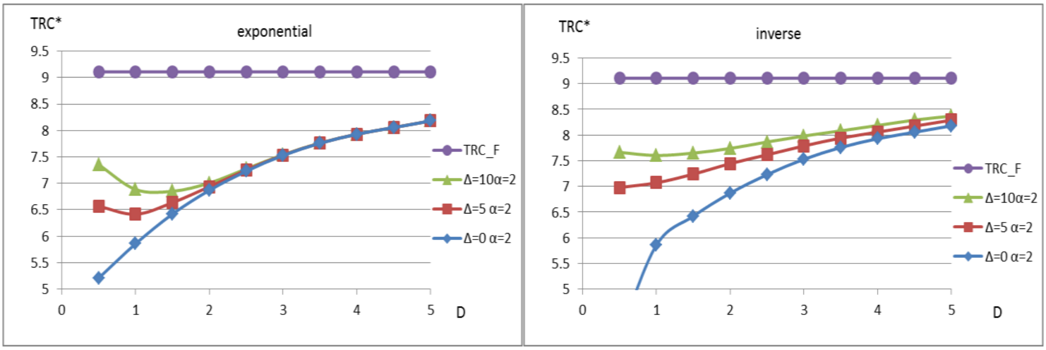

Figure 2-5 The behavior of the optimal TRC^* at increasing values of D, under 2 level capacity policy for $B = 10$, $c_p = 5, \alpha = 2$ and $\Delta = 0, 5$ and 10, when c_c has exponential (left) and inversely proportional (right) structure.

Furthermore, we can observe from Figure 2-5 that the minimum total costs with positive Δ are higher than the minimum total costs when $\Delta = 0$. This is self-evident, since the per time unit cost for contingent capacity, c_c, is the cheapest and equal to c_p when $\Delta = 0$, no matter what structure that the c_c function has. In such a case choosing the shortest possible period length will be the optimum, since more frequent updating possibilities would increase the responsiveness of the repair shop capacity to its workload, and thus a more economic use of its resources without any added cost factors. On the other hand, for $\Delta > 0$, due to the wage differential reflections of the contingent capacity costs, TRC^* gets higher at short period lengths, which stimulates the choice of a longer period. That's the reason why the gaps between TRC^* with different Δ values are widest when $D = 0.5$, which is the shortest period length that is analyzed. On the contrary, TRC^* for different Δ overlap each other at longer period lengths, since the effects of the wage differentials slim down to a negligible extent after some D. This overlapping take place at an earlier stage (after $D = 2$) when c_c has an exponential structure compared to the case when it has an inverse proportional structure (after $D = 4$). Actually, it can be observed that for each $\Delta > 0$, the TRC^* values that pertain to the inverse proportional c_c functional structure resemble the magnified and horizontally stretched version of a piece (between $D = 1$ and $D = 2$) from the corresponding TRC^* values pertaining to the exponential c_c functional structure. This magnifying effect can be explained by the fact that the logarithm of the exponential form is quite similar to the inverse proportional form, and when we use the same (Δ, α) parameters for both of the functional forms, the elasticity of the capacity costs to the period length is smaller in the inversely proportional case.

Despite the slight resemblance of the exponential and inverse proportional cost structures for the contingent capacity, TRC^* can behave quite differently with different period lengths when c_c has a linear structure. In order to comprehend the dynamics of the interplays between the period length, total costs and cost parameters at a deeper level, we present Figure 2-6, which illustrates TRC^* values under 2 level capacity mode at increasing period lengths (D from 0.5 to 5), when c_c has a linear structure for $B = 10$ and $c_p = 5$. The figure on the left emphasizes the effects of different Δ values (0, 5 and 10) when $\alpha = 2$ and the figure on the right emphasizes the role of the elasticity, by illustrating TRC^* with $\alpha = 0,2,5$ and ∞ when $\Delta = 5$.

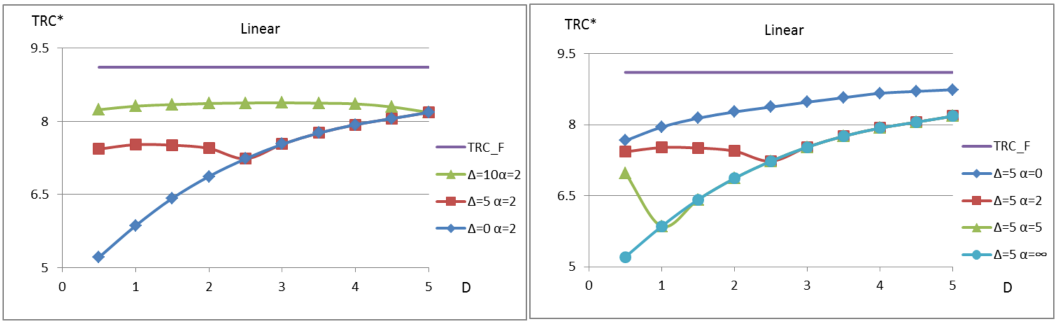

Figure 2-6 The behavior of the optimal TRC^* under 2 level capacity mode at increasing values of period length D, for $B = 10$, $c_p = 5$ and when c_c function has the linear structure. Left: $\alpha = 2$ and $\Delta = 0, 5$ and 10. Right: Δ= 5 and $\alpha = 0,2,5$ and ∞ .

In the linear cost structure, c_c equates to c_p for period lengths bigger than or equal to Δ/α values. For $D > \Delta/\alpha$, all TRC^* values with $\Delta > 0$ coincide with the TRC^* values with $\Delta = 0$. In comparison to those of the inverse linear and exponential cost structures, TRC^* curve pertaining to the linear cost structure is much more flat, except for the carving around Δ/α. This carving is important because it mostly determines the optimal period length for $\Delta > 0$. From Figure 2-6, we can observe that the optimal period length is the smallest possible D for $\Delta = 0$ and it is equal to: Δ/α for $\Delta > 0$ (unless $\Delta/\alpha < 5$). The other characteristics of TRC^* with linear cost structure are similar to those of the previously mentioned TRC^* values with exponential and inverse proportional structures. After discussing the role of Δ on the behavior of TRC^*, next we can scrutinize the effects of different time elasticity factors: α on the behaviour of TRC^*.

The time elasticity of per time unit contingent capacity cost is the other important factor that determines the behavior of the minimum capacity costs in response to the period length D. What we can first conclude from the right figure in Figure 2-6 is the following: for a given $\Delta > 0$, TRC^* values decrease with increasing α, because for a given period length D, the contingent capacity gets cheaper for increasing elasticity, when every other cost parameters remain unchanged. Two extreme values that elasticity factor α can take are 0 and ∞ respectively. If $\alpha = \infty$, the contingent capacity is perfectly elastic, and at the start of a period if it is communicated to the capacity agency that the contingent capacity is not needed for that period, the capacity agency can immediately find another substitute task until the next period which is equivalent in terms of financial returns. This would vanish the cost burden due to the lost opportunities and wage differentials, therefore the optimal total costs would behave as TRC^* with $\Delta = 0$.

On the other hand, when $\alpha = 0$, the provider cannot assign the contingent capacity for any other task, hence the wage differentials remain the same and the contingent capacity costs are inelastic and not affected by the period length. Given a certain $\Delta > 0$, any intermediate α value between 0 and ∞ will result in an optimal total cost that is located in this band, bounded by TRC^* curve with $\alpha = \infty$ from below and TRC^* curve with $\alpha = 0$ from above. It can be further observed that, in all capacity cost structures, with $\Delta > 0$ and an intermediate α value, TRC^* gets closer to the above mentioned upper bound for shorter period lengths (as D goes to 0) and on the contrary, it gets closer to the lower bound (even completely overlaps in the linear cost structure case) for longer period lengths D.

For completely elastic ($\alpha = \infty$) and completely inelastic ($\alpha = 0$) contingent capacity cost structures, we observe that TRC^* values display a monotonically increasing behaviour with the period length D. However, for the other mid-values of α, TRC^* displays a U-shaped structure; where both Δ and α parameters are critical in determining the steepness of the curve. Therefore the optimal period length choice (D^*) is affected by these cost parameters to a great extent. In order to understand the nature of the dynamics between the optimal choice of the period length D^* and the contingent capacity cost parameters (Δ, α), we present Table 2-9, where the optimal period length choices are tabulated under different contingent capacity cost structures for $\Delta = 0, 5$ and 10 when $\alpha = 2$ (above) and for $\alpha = 0, 2, 5$ and ∞ when $\Delta = 5$ (below).

D^*	$\Delta = 0$ $\alpha = 2$	$\Delta = 5$ $\alpha = 2$	$\Delta = 10$ $\alpha = 2$	
Inverse	0.5	0.5	1	
Linear	0.5	2.5	5	
Exp	0.5	1	1.5	
D^*	$\Delta = 5$ $\alpha = 0$	$\Delta = 5$ $\alpha = 2$	$\Delta = 5$ $\alpha = 5$	$\Delta = 5$ $\alpha = \infty$
Inverse	0.5	0.5	0.5	0.5
Linear	0.5	2.5	1	0.5
Exp	0.5	1	0.5	0.5

Table 2-9 The optimal period length: D^* when $B = 10$ and $c_p = 5$ for inverse, linear and exponential cost structures when $\Delta = 0$,5 and 10 for $\alpha = 2$ (above) and when $\alpha = 0, 2, 5$ and ∞ for $\Delta = 5$ (below).

From the first tabular in Table 2-9, we can see that, when α remains constant, D^* increases with Δ. This increase is most evident for linear contingent capacity cost structure and least evident for inverse linear contingent capacity cost structure. On the other hand, for the response of D^* to an increase in α, we can draw different conclusions. Generally speaking, we can say that D^* first increases and then decreases with α, when Δ is positive. In the second tabular in Table 2-9, this behavior can be obviously seen for linear and exponential cost structures, whereas D^* is insensitive to an increase in α for the inverse linear cost structure. However, different cost parameter selections (with higher values of Δ) result in the suggested trend for the effects in response to α, for the inverse proportional cost structure, as well.

After finishing the analysis of the bi-directional relations between the period length D and TRC^*, next we discuss the effects of the base stock level decision: S.

2.5.3.4.3 Base Stock Level S

In this part, we discuss how the base stock level decision affects the optimal total costs and other policy parameters and vice versa. Note that in this part, TRC^* corresponds to the minimum costs that can be achieved for a given base stock level S. In order to conduct the analysis of the effects of the base-stock level per se, we fix a base-stock level S, and find the optimal capacity policy parameters for that base stock level S according to the Search Procedure-II described above. Three figures in Figure 2-7 illustrate how TRC^* changes with increasing values for base-stock levels for $B = 10$, $c_p = 5$ and for different Δ values ($0, 5$ and 10) and exponential, inverse proportional

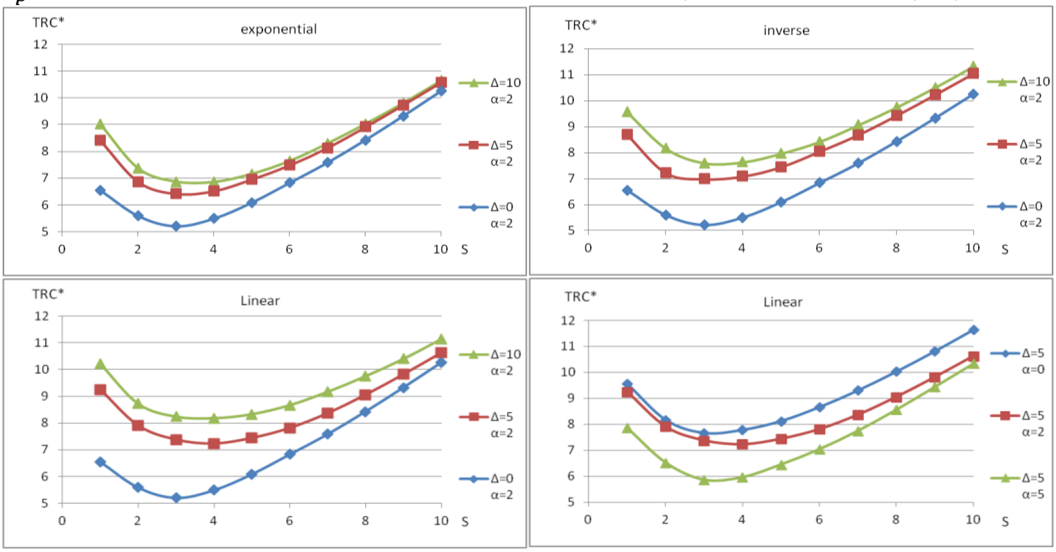

Figure 2-7 TRC^* for increasing base-stock levels S with Δ= $0, 5, 10$, $\alpha = 2$ Clockwise direction: exponential, inverse proportional and linear contingent capacity cost structures with $\alpha = 0, 2, 5$ and $\Delta = 5$ in linear contingent capacity cost structure.

and linear contingent capacity cost structures when ($\alpha = 2$). The last figure in Figure 2-7 shows the behavior or TRC^* with increasing base stock levels but for different α values, $(0, 2, 5)$ for linear cost structure, for given Δ= 5.

The data in Figure 2-7 suggest that TRC^* is a convex function of S. It is already discussed that total costs are convex with base stock level for a given policy **C**. It seems that optimizing the parameters of capacity policy at each base-stock level S, does not distort this behavior of TRC^*. As expected, increasing contingent capacity costs (higher Δ or lower α) increase TRC^* values, however at a moderated extent due to the freedom of optimizing of other policy parameters. We observe that increase in Δ at lower Δ values affect TRC^* more than increase in Δ at higher Δ values. Since the TRC^* function is quite flat around its minimum, variations in Δ or α hardly changes (at maximum 1 unit) the optimal base-stock level decision. In Table 2-10, we tabulate how the optimal base-stock level changes under different capacity cost structures when $B = 10, c_p = 5$, for $\Delta = 0,5$ and 10 for $\alpha = 2$(above) and for $\alpha = 0, 2,$ and 5 when $\Delta = 5$ (below).

From the first tabular in Table 2-10, similar to the response of D^*, we can see that S^* increases with Δ, when α remains constant. This increase is most evident for linear contingent capacity cost structure. On the other hand, the response of S^* to an increase in α for a given Δ is different. From the second tabular in Table 2-10, we can say that S^* first increases and then decreases with α, when Δ is given and positive. The choices on the base stock level and on the period length are inter-connected.

S^*	$\Delta = 0\ \alpha = 2$	$\Delta = 5\ \alpha = 2$	$\Delta = 10\ \alpha = 2$	
inverse	*3*	*3*	*3*	
linear	*3*	*4*	*4*	
Exp	*3*	*3*	*4*	
S^*	$\Delta = 10\ \alpha = 0$	$\Delta = 10\ \alpha = 2$	$\Delta = 10\ \alpha = 5$	$\Delta = 10\ \alpha = \infty$
Inverse	*3*	*3*	*3*	*3*
Linear	*3*	*4*	*4*	*3*
Exp.	*3*	*4*	*3*	*3*

Table 2-10 The optimal base-stock level: S^* under different contingent capacity structures when $B = 10$ and $c_p = 5$ for $\Delta = 0, 5$ and 10 for $\alpha = 2$ (above) and for $\alpha = 0, 2, 5$ and ∞ when $\Delta = 10$ (below).

Figure 2-8 can be helpful to understand the interplay between period length D and the base stock level S.

The figure on the left demonstrates that the optimal period length, D^*, increases for higher values of stock levels, whereas the in figure on the right, we can see that the optimal stock level S^* increases with period length D. The data in

Figure 2-8 suggests that in the presence of wage differentials ($\Delta, \alpha > 0$), spare unit availability and capacity update frequency are complementary to some extent. As the frequency of capacity updates is higher, less spare units in the stock are needed, whereas if we have more spare units waiting in the stock, it can be compensated with less frequent capacity updates with cheaper contingent capacity costs.

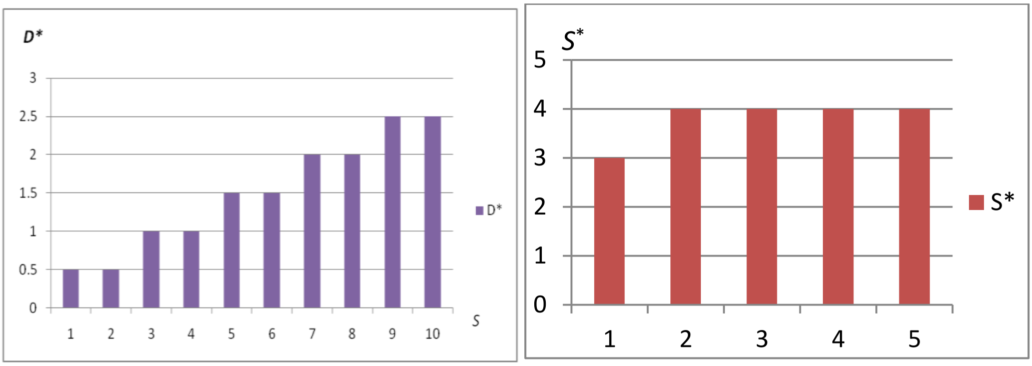

Figure 2-8 Left: optimal base-stock levels for increasing period lengths. Right: the optimal period length for increasing base-stock levels when $B = 10$, $c_p = 5$ for $\Delta = 5$ and $\alpha = 2$.

This concludes our sensitivity analysis section. In the next subsection, the accuracy of the finite waiting room approximation in the analytical method is examined with an extensive simulation study.

2.5.3.5 Accuracy of Finite Waiting Room Approximation

In this subsection, we examine the accuracy of our finite waiting room approximation by comparing the total cost rate of the proposed analytical model (having a finite waiting room of 40) with the cost rate obtained by simulating the real environment having a repair shop that has an infinite waiting room. In our simulations, we used a run length of $2{,}000{,}000{,}000$ defective unit arrivals (when $\lambda = 1$) in a single replication, where the average total cost rate TRC is calculated under a policy: $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$.

We investigated a total of 243 different scenarios with different B, c_p, Δ and α and resulting optimal policy parameters from Search Procedure-II. The percentage error, $\%err$, of using the approximation for TRC in a scenario can be found as:

$$\%err = 100 \times \frac{(TRC_{sim} - TRC_{app})}{TRC_{sim}} \tag{2.19}$$

Here, TRC_{sim} is the total relevant costs obtained from the simulation and TRC_{app} is the total relevant costs obtained from the analytical model. Table 2-11 summarizes the accuracy of the approximations.

average %|*err*|	min %*err*	median %*err*	max %*err*
0.13%	-0.42%	-0.02%	0.38%

Table 2-11 Accuracy of the approximation for the TRC values

In Table 2-11, the absolute value, minimum, median and the maximum for the percentage errors are listed, respectively. From the table, we can see that the approximation can mimic the performance of the original, infinite waiting room environment almost perfectly, which demonstrates the accuracy of our method.

This simulation study finalizes this section for the second capacity mode. In the next section, Section 2.6, we set out to analyze the third (and last) capacity mode. The conclusive remarks on the second and third capacity modes will be provided in the overall conclusions section, Section 2.7.

2.6 Periodic Capacity Sell-back Mode

In this section, we analyze the integrated decision making problem of the MSP under the third, sell-back capacity mode. In this capacity mode, the failed units are sent to the repair shop at regular intervals in time. Due to this admission structure, when the repair of all the failed units in the shop are completed in a period, it is known that there will not be any job left at least until the start of the next period. This synchronization of the arrivals allows for a contract, where the capacity agency supplies a fixed amount of capacity at regular intervals in time, covering for the whole interval duration. However, if all the repairs in the shop are completed before the end of an interval, the capacity can be temporarily sold back to the capacity agency, at a reduced price, until the next interval. The capacity agency can then deploy the sold capacity temporarily for external/other tasks. In this setting, all capacity used in the repair shop can be interpreted as contingent and the idle time notion will be cancelled.

In this mode, at the start of each period, the agreed capacity is ready for the use at the repair shop during that period. However, when the capacity becomes idle, it is immediately assigned to other tasks through the capacity agent. Frequent task switching, searching/ assigning alternative tasks for the idle capacity, and the risk of not finding an appropriate ad-hoc assignment for the idle repair shop capacity, all create an economic factor that leads to an opportunity cost, which decreases with the period length. Due to this opportunity cost, unit capacity cost in this mode, c_c, is higher than the original permanent capacity price c_p.

On the other hand, in this capacity mode, when all the repairs are completed, i.e. when the repair shop becomes idle, the agency buys back the capacity temporarily until the start of the next period at a reduced price, which is lower than c_p. The other effects arising from the periodic admission structure, such as the down-time related costs due to the pre-admission delay of a failed unit and the burstiness due to the clustering of failed unit arrivals have indirect effects on the total relevant costs, and are analyzed thoroughly in the later sections of this chapter.

Similar to the previous capacity modes, MSP has to give the optimal inventory and capacity level decisions in order to minimize its TRC. The relevant capacity decisions in this mode are the period length D and the reserved capacity level μ. It is of paramount importance to characterize and analyze the cost effects due to the sell-back option enabled in this third mode. Therefore we aim at building a modeling framework and a decision support system for the service provider operating under the periodic capacity sell-back mode. In Subsection 2.6.1 we describe the dynamics of the periodic admission and the capacity sell-back option, introduce the additional decision variables as well as the problem formulation. In Subsection 2.6.2, the derivation and the analysis of the total relevant cost per unit time are given. Finally in Subsection 2.6.3, we describe the experimental setting, the search procedure and provide the results of the numerical study with a particular focus on the savings under the capacity sell-back mode compared to the best cost performance under the reference, fixed capacity mode ${TRC_F}^*$. This part is followed by an elaborate sensitivity analysis of the cost/policy parameters and a simulation study for the accuracy check of the analytical approximations used.

2.6.1 Periodic Admission and Capacity Sell-back Structure and Problem Formulation

The MSP operates in the same environment that is explained in the previous sections. Recall that the joint repair shop/ spare unit inventory is modeled as a single-server, queuing to stock system, where the failures occur following a stationary Poisson demand and each defective subsystem requires an exponentially distributed amount of dedicated repair service time in the shop in order to regain its *good as new* status.

In this section, there are two major differences in the repair shop compared to the reference, single level capacity model that is described in Section 2.4. The first deviation is in the admission structure of the failed units to the repair shop, and the second deviation is in the capacity sell-back option.

In this third mode, upon a subsystem failure in one of the systems, a new ready-for-use unit from the stock is sent to the system, if there is any available in the stock. However, that failed subsystem is not sent to the repair shop immediately, but the shipment is postponed until the start of the next regular interval. The length of each interval, D, is an important decision parameter. For a given interval length D, periodic admission points to the repair shop are introduced at equidistant times: $D, 2D, 3D, \ldots$

In such a case, all the subsystem units that failed within an interval are sent to the repair shop at the end of that interval simultaneously and the operations at the repair shop level resulting from the periodic admission structure can be modeled as a $D^X/M/1$ queue, where the processing rate of the single server queue represents the capacity level of the repair shop, and X is a discrete random variable which represents the number of parts failed within an interval. Due to the Poisson arrival stream for the part failures, X also follows a Poisson distribution with mean λD. Note that as D goes to 0, this model reduces to the reference $M/M/1$ model. The analysis and the characteristics of this gated, $D^X/M/1$ type of queues are given in Subsection 2.6.2.

A periodic, capacity sell-back policy, $\mathbf{C} = [D, \mu]$, consists of a period length D, and a processing rate of μ. Under policy $\mathbf{C}$, the capacity agency supplies the agreed capacity level μ indefinitely, at a price of c_c, and the failed units are sent to the repair shop at regular intervals of length D. Similar to the previous capacity mode, c_c is higher than or equal to c_p, decreases with D and goes to c_p as D goes to infinity. If all the failed units in the shop are repaired before the end of an interval, or if there is no unit to be repaired at the start of an interval, the shop capacity is sold back to the capacity agency at a reduced price of $c_p R$, for $0 < R < 1$. After the capacity is sold back, it is temporarily re-assigned to other task(s) until the start of the next interval. The operations at the repair shop level and the interactions with the capacity agency are illustrated in Figure 2-9.

From Figure 2-9, it can be seen that all the subsystem units that failed in the first period are admitted to the repair shop at time D. After all the repairs are completed during the second period, the capacity is immediately sold back to the capacity agency, where it is re-assigned to other tasks until $2D$. At $2D$, upon the admission of the newly failed units, the capacity is returned back to the repair shop in order to work on the repair of the failed units again.

Due to this admission structure, when the repair of all the failed units in the shop are completed in a period, it is known that there will not be any job left in the repair shop at least until the start of the next period. This partial certainty on the idle times of the repair shop capacity allows for a contract, where the capacity agency buys the repair shop capacity at a reduced price during its idle times.

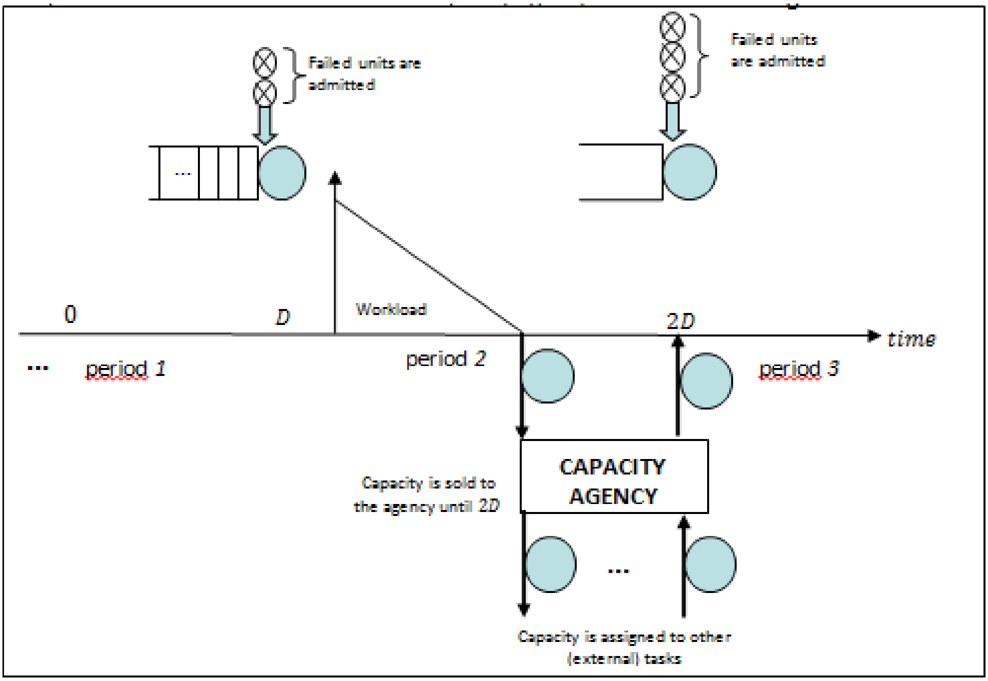

Figure 2-9 Illustration of the system under the third: capacity sell-back mode.

2.6.1.1 Updated Capacity Provision/Buying Mechanism

Under this contract, the capacity is provided to the repair shop indefinitely, however frequent job switching and searching for ad-hoc assignments for the idle capacity create an economic factor that lead to an opportunity cost $o_c(D, \Delta, \alpha)$. This opportunity cost decreases with the period length and is dependent on two other cost parameters: Δ and α. The cost structure of $o_c(D,\Delta,\alpha)$ is similar to the structures described in Table 2-1 and Figure 2-2. Due to the $o_c(D,\Delta,\alpha)$, the unit capacity cost rate for the repair shop c_c, is greater than or equal to the c_p. When all the part repairs in the shop are completed before the end of an interval, the capacity is temporarily sold to the capacity agency until the next interval. The sell-back price is lower than c_p: c_pR, where $0 < R < 1$, because we assume that there is a risk of not finding an appropriate ad-hoc task and even in the presence of a temporary task, that temporary task can be less profitable than the core repair shop activity. Having discussed the contract particular to this capacity mode, next we elaborate the problem formulation for the MSP.

2.6.1.2 Problem formulation

Suppose the repair shop operates under a stable (i.e. $\mu > \lambda$) policy, $\mathbf{C} = [D,\mu]$, for an infinite horizon. For a given per time unit permanent and contingent capacity costs, c_p and c_c, and a given capacity sell-back cost reduction rate R (for $0 < R < 1$), the average capacity related costs, $CRC(\mathbf{C})$ can be directly calculated as follows:

$$CRC(\mathbf{C}) = c_c\,\mu - c_p\lambda - \left(c_pR\right)(\mu-\lambda) = \left(c_c - c_pR\right)\mu - c_p\lambda(1-R) \tag{2.20}$$

In (2.20), we excluded the costs pertaining to the baseline costs ($c_p\lambda$) from the amount that is paid to the capacity agency ($c_c\,\mu$), for deploying μ level of capacity. The repair shop capacity is sold to the capacity agency during the idle times. Therefore the repair shop gains a revenue of: $(\mu-\lambda)c_p\,R$ per time unit.

We assume that the capacity agency offers a set of possible period lengths, θ, from which the service provider can choose the best period length considering the pros and cons of periodic admission and capacity sell-back options. We pay h per unit time for each spare unit in the stock / in the repair shop. The down-time costs due to the backorder of the spare parts is equal to B per time unit, and we assume that $B > h$. The inventory related decision is S. The capacity related decisions, $\mathbf{C}$, are twofold:

1) Length of the period: D

2) The size of the repair shop capacity level: μ

The total relevant cost function, TRC, can be represented by $\mathbf{C}$ and S, and is the sum of capacity related costs ($CRC(\mathbf{C})$), down-time costs ($DTC(\mathbf{C},S)$) and holding costs ($HC(S)$). Given these cost components and the decision variables, the problem of the MSP can be formulated as follows:

$$\min_{\mathbf{C},S} TRC(\mathbf{C},S) = CRC(\mathbf{C}) + DTC(\mathbf{C},S) + HC(S)$$

$$s.t.\;\; S \in N = \{0,1,2,\dots\}$$

$$\mathbf{C} = [D,\mu]$$

$$\mu > \lambda,\, D \in \theta,$$

(2.21)

Motivated by the problem formulation above, in the next subsection, we derive the necessary cost functions used in (2.21) and give some of the analytical properties of the components of $TRC(\mathbf{C}, S)$.

2.6.2 Derivation & Analysis of the Cost Functions

In this subsection, we derive the components of the total relevant costs per unit time. The capacity related costs per unit time, $CRC(\mathbf{C})$, for given cost parameters (c_p , c_c and R) and policy parameters (μ and D) are already derived in (2.20). Also, the derivation of the holding costs for a given stock level S, $HC(S) = hS$, remains unchanged compared to the previous capacity modes.

Finally, in order to derive the last remaining cost component, which is the down-time related costs per unit time, $DTC(\mathbf{C}, S)$, we need to delve into the detailed modeling of the operations at the repair shop level.

We know from the previous sections that $DTC(\mathbf{C}, S)$ can be derived from the expected number of backordered repair demands, $\mathrm{E}((N_d - S)^+ | \mathbf{C})$, where N_d is the random variable that corresponds to the number of defective units at an arbitrary point of time, both in the repair shop waiting for (or in the) service and outside the repair shop, waiting to be admitted.

We analyze the performance characteristics of the $D^X/M/1$ queue in order to obtain $\mathrm{E}((N_d - S)^+ | \mathbf{C})$. For the sake of consistency with the queuing terminology, from now on "defective units" and "customers" are being used interchangeably.

2.6.2.1 Analysis of the $D^X/M/1$ Queue

In the $D^X/M/1$ model, at any time point t, the number of customers, $N_d(t)$, is the sum of the number of customers in the queue (including the one in the service), $N_q(t)$, and the number of customers outside the queue, $N_o(t)$, waiting to be admitted. The capacity mode related parameters may affect $N_q(t)$, whereas the latter part, $N_o(t)$, which follows a discrete Poisson distribution, is independent of the capacity level choice μ, but is dependent on the failure arrival rate λ and the period length D.

At the start of each period, all customers outside the queue are admitted into the queue based on their arrival order, therefore we have $N_d(nD) = N_q(nD)$ and $N_o(nD) = 0$ for $n = 1,2, \ldots$ To have a better understanding of the interrelations between $N_o(t)$, $N_q(t)$ and $N_d(t)$, their behaviors are illustrated on a sample path for $0 \le t \le 4D$ in Figure 2-10.

From Figure 2-10, it can be seen that $N_o(t)$ increases at each failure arrival and resets itself to zero at the start of each period (at times: $D, 2D, 3D$ and $4D$). At these period start points, the failed subsystem units are admitted to the repair shop simultaneously, which results in the abrupt increases in $N_q(t)$. Similarly, at each repair completion,

$N_q(t)$ decreases by one. $N_d(t)$ is the sum of these two variables and therefore increases with a failure arrival and decreases with a repair completion.

Recall that $DTC\ (\mathbf{C}, S) = B\big(\mathrm{E}((N_d - S)^+|\mathbf{C})\big)$, and that $\mathrm{E}((N_d - S)^+|\mathbf{C})$ can be derived from the time average probabilities of the number of defective units (i.e. $P\{N_d = i\}$ for $i = 0,1,2\ldots$), which are obtained in the following two steps:

The first step is the derivation of the steady state probability vector, $\mathrm{v}(\mathbf{C})$, for the number of customers at the start of a period in the queue. This requires the analysis of the behavior of $N_d(nD)$ for $n = 0,1,2\ldots$

After the derivation of $\mathrm{v}(\mathbf{C})$, we can proceed to the second step, which is the generation of the time average probabilities for the number of defective units in the environment, consisting of the defective units both in and out of the repair shop.

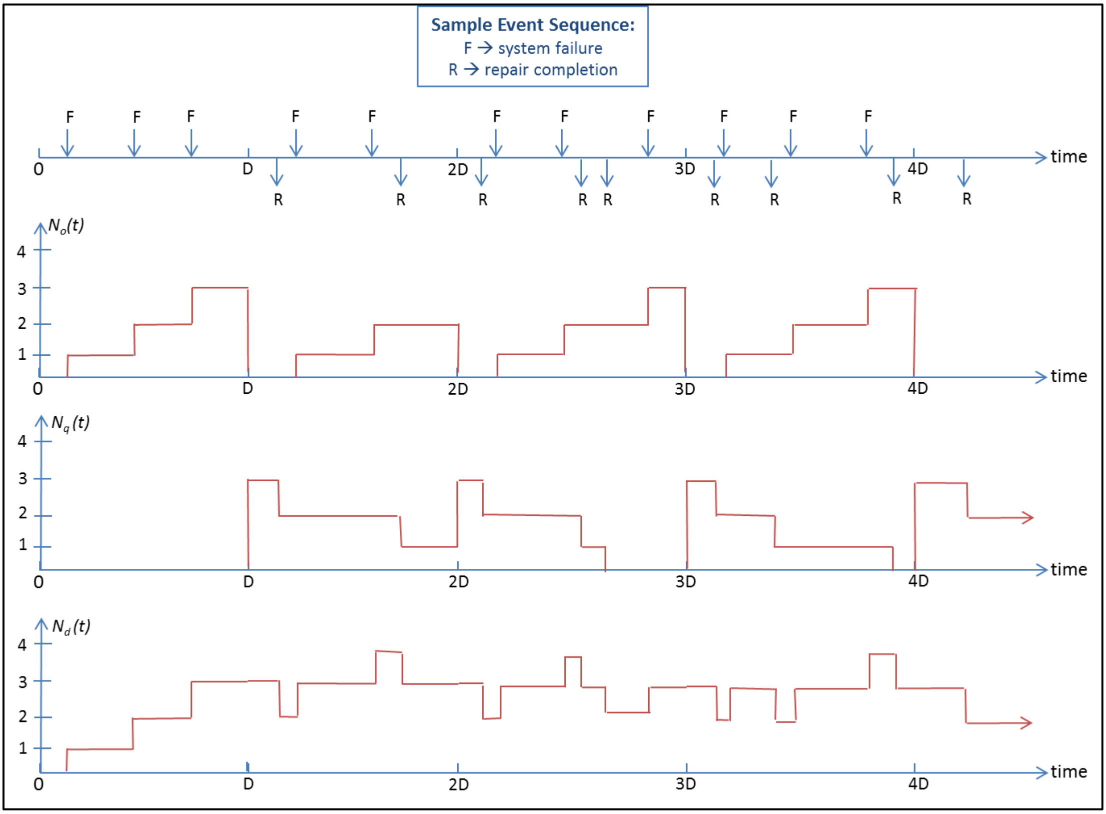

Figure 2-10 The illustration of the behavior of $N_o(t)$, $N_q(t)$ and $N_d(t)$ for the given sample event sequence.

2.6.2.2 *Steady State Probabilities for the Number of Customers at the Start of a Period*

Suppose $N_o(nD^-)$ denotes the number of customers outside the queue at time nD, but just before the customer admissions. As mentioned above, for all $n = 0, 1, 2\ \ldots,$

$N_o(nD^-)$ follows a Poisson distribution with mean: λD. As can be seen from Figure 2-10, these customers are admitted into the queue immediately after time nD, and therefore we have $N_o(nD) = 0$.

In a similar manner, suppose that $C^n(t)$ is a random variable, which represents the number of service completions in "t" time units after the start of the n^{th} period, provided that there is an infinite number customers waiting for service at time nD. Due to this assumption, $C^n(t)$ simply follows a Poisson distribution, having a mean of μt for any $t > 0$.

In the light of the discussions above, we can write the following recursive relation for $N_d(nD)$ for any $n = 0, 1, 2 \ldots$:

$$N_d(nD) = N_q(nD^-) + N_o(nD^-) = \left(N_d\big((n-1)D\big) - C^n(D)\right)^+ + N_o(nD^-) \tag{2.22}$$

Here $(a)^+ = a$ if $a > 0$, and $(a)^+ = 0$ otherwise. In essence, (2.22) demonstrates the Lindley-type (Lindley 2008) recursive relation between $N_d(nD)$ and $N_d\big((n-1)D\big)$, where $N_o(nD^-)$ and $C^n(D)$ are independent from each other and independent from $N_d\big((n-1)D\big)$, therefore, $N_d(nD)$ has the Markov property for $n = 0, 1, 2 \ldots$

We assume that $\mathrm{P}(D, \mathbf{C})$ is the transition probability matrix for the $N_d(nD)$ process under the capacity policy: $\mathbf{C} = [D, \mu]$, when $n = 0,1, \ldots$ With a period length D, $\mathrm{P}_{ij}(D, \mathbf{C})$ denotes the probability that there will be j defective units at the start of the next period, given that there are i defective units at the start of the current period.

For a given capacity policy $\mathbf{C} = [D, \mu]$, $\mathrm{P}_{ij}(D, \mathbf{C})$ can be derived from the recursive behavior in (2.22) explicitly for all $i, j \geq 0$ as below:

$$\mathrm{P}_{ij}(D, \mathbf{C}) = \sum_{a=max(0,j-i)}^{j} P\left\{\left(i - C^n(D)\right)^+ = (j-a)\right\} P\{N_o(nD^-) = a\} \tag{2.23}$$

Here, for all $i = 0, 1, 2, \ldots$, we can write:

$P\left\{\left(i - C^n(D)\right)^+ = j - a\right\} = \frac{e^{-\mu D}(\mu D)^{(i-(j-a))}}{(i-(j-a))!}$ for $0 < j - a \leq i$ and

$$P\left\{\left(i - C^n(D)\right)^+ = 0\right\} = \sum_{k=i}^{\infty} \frac{e^{-\mu D}(\mu D)^k}{k!}$$

The rationale behind (2.23) is to sum up the probability of all possible combination of events (repair completions and new unit failures) where the number of defective units is "i" at the start of this period, and becomes "j" at the start of the next period.

After $\mathrm{P}_{ij}(D,\mathbf{C})$ is found from (2.23) for all $i \,\&\, j \geq 0$, we can construct $\mathrm{P}(D,\mathbf{C})$, the transition probability matrix of the $N_d(nD)$ process for $n = 0,1,2\ldots$ under policy $\mathbf{C} = [D,\mu]$. Let $\mathrm{v}(\mathbf{C})$ be the steady state vector of the probabilities for the number of defective units at the start of a period under policy $\mathbf{C} = [D,\mu]$. $\mathrm{v}(\mathbf{C})$ can be easily derived from $\mathrm{P}(D,\mathbf{C})$from the following equations:

$$\mathrm{v}(\mathbf{C}) = \mathrm{v}(\mathbf{C})\mathrm{P}(D,\mathbf{C})$$
$$\textstyle\sum_{i=0}^{\infty} \mathrm{v}_i(\mathbf{C}) = 1$$

(2.24)

After deriving the $\mathrm{v}(\mathbf{C})$ vector, we can proceed to the next step, which is to generate the time average probabilities for the number of defective units in the environment both in and outside the repair shop.

2.6.2.3 Time Average Probabilities for the Number of Defective Units in the Environment

Next, we derive the time average probabilities for the number of customers, both in and outside the queue. In order to derive the time average probabilities, we first need the time-dependent probabilities for the number of customers, not only at the equidistant admission points, but also for all of the other intermediate time points in between.

Therefore, let $t \in [0,D)$ denote the time elapsed until the start of the current period. Then we have $N_d(nD+t)$, which corresponds to the random variable that indicates the number of customers in the whole environment, t time units after the start of the n^{th} interval. As discussed previously, $N_d(nD+t)$ consists of two independent random variables: $N_q(nD+t)$, which is the number of customers in the queue (or in the service) and $N_o(nD+t)$, which is the number of the customers outside the queue, waiting to be admitted.

$N_q(nD+t)$ can be generated from $N_d(nD)$ following the same reasoning of (2.23), since it is the remnant number of customers from $N_d(nD)$, t time units after the n^{th} admission point. With the help of the previously defined $C^n(t)$, the probability of having i customers remnant in the queue, t time units after the n^{th} admission, $P\{N_q(nD+t) = i\}$ can be formulated as follows:

$$P\{N_q(nD+t) = i\} = P\{(N_d(nD) - C^n(t))^+ = i\} =$$
$$\sum_{k=i}^{\infty} P\{(N_d(nD) - C^n(t))^+ = i | N_d(nD) = k\} P\{N_d(nD) = k\} =$$

$$\sum_{k=i}^{\infty} P\{(k - C^n(t))^+ = i\} P\{N_d(nD) = k\}$$

(2.25)

Suppose we have the following limiting probabilities for any $i = 0,1,2, \ldots$:

$$P\{N_o{}^t = i\} = \lim_{n\to\infty} P\{N_o(nD+t) = i\}, \text{ and}$$

$$P\{N_q{}^t = i\} = \lim_{n\to\infty} P\{N_q(nD+t) = i\} = \lim_{n\to\infty} P\{(N_d(nD) - C^n(t))^+ = i\} =$$

$$\lim_{n\to\infty} \sum_{k=i}^{\infty} P\{(k - C^n(t))^+ = i\} P\{N_d(nD) = k\} =$$

$$\sum_{k=i}^{\infty} \lim_{n\to\infty} \left(P\{(k - C^n(t))^+ = i\} P\{N_d(nD) = k\}\right) = \sum_{k=i}^{\infty} P\{(k - C(t))^+ = i\} \mathrm{v}_k(\mathbf{C})$$

(2.26)

In (2.25) and (2.26), $P\left\{(k - C^n(t))^+ = i\right\}$ defines the probability that "i" number of defective units are left at time "$nD + t$", given that there were "k" defective units at the start of the n^{th} period, for $k \geq i$.

From these definitions, we can write the following for $t < D$, under $\mathbf{C} = [D, \mu]$:

$$P\left\{(k - C^n(t))^+ = i\right\} = \frac{e^{-\mu t}(\mu t)^{(k-i)}}{(k-i)!} \text{ for } 0 < i \leq k \text{ and}$$

$$P\left\{(k - C^n(t))^+ = 0\right\} = \sum_{j=k}^{\infty} \frac{e^{-\mu t}(\mu t)^j}{j!}$$

In the light of discussions above, with the help of the newly derived $P\{N_q{}^t = i\}$ and $P\{N_o{}^t = i\}$ values, we can obtain $P\{N_d{}^t = k|\mathbf{C}\}$, which is the probability that there will be "k" defective units after "t" time units from the start of an arbitrary interval, under policy $\mathbf{C} = [D, \mu]$, as follows:

$$P\{N_d{}^t = k|\mathbf{C}\} = \sum_{i=0}^{k} P\{N_o{}^t = k - i\} P\{N_q{}^t = i\}$$

(2.27)

From (2.27), we can proceed on deriving the time average of the instances, when there are "k" defective units as below:

$$P\{N_d = k|\mathbf{C}\} = \frac{\int_{t=0}^{D} P\{N_d{}^t = k|\mathbf{C}\}dt}{D} = \int_{t=0}^{D} \frac{1}{D}\sum_{i=0}^{k} P\{N_o{}^t = k - i\}P\{N_q{}^t = i\}\ dt$$
$$= \int_{t=0}^{D} \frac{1}{D}\sum_{i=0}^{k} \frac{e^{-\lambda t}(\lambda t)^{k-i}}{(k-i)!}\sum_{j=i}^{\infty} P\left\{\left(j - C(t)\right)^{+} = i\right\} \mathrm{v}_j(\mathbf{C})\,dt$$
$$= \frac{1}{D}\sum_{i=0}^{k}\sum_{j=i}^{\infty} \mathrm{v}_j(\mathbf{C}) \int_{t=0}^{D} \frac{e^{-\lambda t}(\lambda t)^{k-i}}{(k-i)!} P\left\{\left(j - C(t)\right)^{+} = i\right\} dt$$

(2.28)

Note that the integral given in (2.28) is easily computable since they can reduce to some lower incomplete gamma functions after some algebra. After $P\{N_d = k|\mathbf{C}\}$ is obtained for all $k = 0,1,2,\ldots$, we can obtain the average down-time related costs per time, $DTC(\mathbf{C}, S) = B\left(\mathrm{E}((N_d - S)^{+}|\mathbf{C})\right)$ below.

Recall that the $\mathrm{E}((N_d - S)^{+}|\mathbf{C})$ is the expectation of the number of backordered units at an arbitrary point of time given the inventory level S and the capacity policy $\mathbf{C} = [D, \mu]$. We can write $\mathrm{E}((N_d - S)^{+}|\mathbf{C})$ and $TRC(\mathbf{C}, S)$ as below:

$$\mathrm{E}((N_d - S)^{+}|\mathbf{C}) = \mathrm{E}(N_d|\mathbf{C}) - S + \mathrm{E}((S - N_d)^{+}|\mathbf{C})$$
$$= \sum_{k=0}^{\infty}(k)P\{N_d = k|\mathbf{C}\} - S + \sum_{k=0}^{S}(S - k)P\{N_d = k|\mathbf{C}\}$$

$$TRC(\mathbf{C}, S) = CRC(\mathbf{C}) + DTC(\mathbf{C}, S) + HC(S)$$
$$= hS + \left(c_c - c_p R\right)\mu - c_p\lambda(1 - R) + B\left(\mathrm{E}((N_d - S)^{+}|\mathbf{C})\right)$$

(2.29)

Next, we examine some of the properties of the $TRC(\mathbf{C}, S)$ function. These properties are useful since their implications can be exploited during the numerical search procedure that will be given in the next section.

2.6.2.4 Analytical Properties of Total Relevant Cost Function

---

Theorem 2.3: For a given stock level S, and period length D, $TRC(\mathbf{C}, S) = hS + \left(c_c - c_p R\right)\mu - c_p\lambda(1 - R) + B\left(\mathrm{E}((N_d - S)^{+}|\mathbf{C})\right)$ is convex in capacity level μ.

Proof: Proof of the theorem is given in the Appendix (Chapter 6)

Theorem 2.4: For a given capacity policy $\mathbf{C} = [D, \mu]$, $TRC(\mathbf{C}, S)$ is unimodal with S.

Proof: Let $\Delta TRC(\mathbf{C}, S) = TRC(\mathbf{C}, S) - TRC(\mathbf{C}, S-1)$ for $S = 1, 2, \ldots$ Then we have:

$$\Delta TRC(\mathbf{C}, S) = h + B \sum_{n=S}^{\infty} (n - S)\ P\{N_d = n|\mathbf{C}\} - B \sum_{n=S-1}^{\infty} (n - S + 1)\ P\{N_d = n|\mathbf{C}\}$$
$$= h - B \times P\{N_d \geq S|\mathbf{C}\} \tag{2.30}$$

As can be seen from (2.30), $\Delta TRC(\mathbf{C}, S)$ increases with S, and changes sign once from negative to positive. Therefore, the optimal stock level S^* for a given capacity policy $\mathbf{C}$ will be the largest S, where $\Delta TRC(\mathbf{C}, S)$ is still negative.

These properties will help us to search for the optimal policy parameters more efficiently. In the succeeding section, we provide the design of our test bed, details and the results of the search procedure applied on the instances from the designed test bed and finally a critical discussion over the results.

2.6.3 Numerical Study

In this section, we use the analysis and the results provided in the previous section in order to assess the performance of the third, capacity sell-back mode. We compare the cost performance of this mode with the minimum cost rate achieved under the fixed capacity mode, $TRC_F{}^*$. The outline of this section is as follows. First, we describe the characteristics and the test bed of the computational study. Second, we present the search procedure in order to find the optimal policy parameters for the problem given in (2.21). Finally, we discuss the results of the search procedure, starting from assessing the potential cost benefits that can be gained in the capacity sell-back mode in comparison to the best policy of the reference, fixed capacity mode. Afterwards, we examine the interactions among the policy and system parameters.

2.6.3.1 Numerical Test Bed and the Search Procedure

In our computational study, we normalize the arrival rate for the subsystem failures in the whole environment, $\lambda = 1$ (failures per time unit) as well as the unit permanent capacity cost per time unit, $c_p = 1$. The other parameters are scaled according to the normalized λ and c_p. Similar to Sections 2.4.3 and 2.5.3, we analyze a total of 9 scenarios with 3 different B/h values and 3 different h values, which are already given in Table 2-2. For each of these 9 scenarios and different $\Delta = (0, 0.5, 1)$, $\alpha = (0, 1, 2)$ as well as $R = (0.1, 0.2, \ldots 1)$ combinations, we execute our solution procedure to find the capacity policy parameters, (D, μ) and S that yield the minimum total costs.

We developed a new solution procedure, Search Procedure-III, for the optimization problem in (2.21). Similar to the previous search procedures, we also have an arbitrarily large S_{max} value as an upper limit for the spare unit inventory level choice. The periodic admission determines the frequency of defective subsystem unit admissions as well as the frequency of the possible capacity sell-back actions. This frequency is determined by the period length, which is chosen from θ, the set of candidate period lengths offered. In our thesis we assume $\theta = \{0.5, 1, 1.5, \dots, 4.5, 5\}$, which are scaled according to the normalized inter-arrival time $\frac{1}{\lambda} = 1$.

Note that all the derivations/equations in Section 2.6.2 pertain to infinite state variables, which may create hindrances during the computational study. Many performance measures of similar batch arriving queuing models to $D^X/M/1$ are shown to possess the geometric tail property (Tijms 1994). In the numerical results, it is observed that the steady state probabilities of the defective units follow the geometric tail property ubiquitously. Therefore, we truncate the infinite state system of equations to largely enough finite (with a large waiting room of K) ones. Simulation studies that we conduct, which will be discussed at the end of this subsection, suggest that the analytical finite system approximation mimics the performance of the infinite state system almost perfectly.

Having mentioned this, subsequently we give our search procedure followed by the description of its underlying mechanism.

Search Procedure-III

---

0. For each inventory Level $S \in \{0,1,2,\dots,S_{max}\}$ follow the steps from a to b:

   a. Given S, for each period length $D \in \theta$ follow the steps I & ii:

      i. Find the capacity level $\mu^*(D,S)$={ $\mu > \lambda : \frac{d\, TRC(\mathbf{C},S)}{d\mu} = 0\}$, where $TRC(\mathbf{C},S)$ is from (2.29)

      ii. Suppose $\mathbf{C} = [D, \mu^*(D,S)]$. Then calculate:

$$TRC(\mathbf{C},S) = hS + \left(c_c - c_p R\right) \times \mu^*(D,S) - c_p\lambda(1-R) + B\left(\mathrm{E}((N_d - S)^+ | \mathbf{C})\right)$$

   b. After following step a for all $D \in \theta$, we can find the minimum cost for a given S, $TRC(\mathbf{C}^*(S),S)$ as follows:

$$TRC(\mathbf{C}^*(S),S) = \min_{D\in\theta}(TRC(\mathbf{C},S) :\ \mathbf{C} = [D, \mu^*(D,S)]) \tag{2.31}$$

1. After $TRC(\mathbf{C}^*(S), S)$ for all $S \in \{0,1,2,\dots,S_{max}\}$ are found, we can find S^*, and $\mathbf{C}^*(S^*) = [D^*, \mu^*(D^*, S^*)]$ values, which would give the global minimum cost rate $TRC^* = TRC(\mathbf{C}^*(S^*), S^*)$ for the capacity sell-back mode. S^* can be found from brute force search as follows:

$$TRC(\mathbf{C}^*(S^*), S^*) = \min_{S \in \{0,1,2,\dots,S_{max}\}} TRC(\mathbf{C}^*(S), S)$$

2.6.3.2 Explanation of the Search Procedure-III

In the search procedure above, at step 0, we start with a certain stock level $S \in \{0,1,\dots,S_{max}\}$ and a certain period length $D \in \theta$.

Afterwards, in sub-step $0.a.i$, for the chosen S and D parameters, we find the capacity level that minimizes the total relevant costs $TRC(\mathbf{C}, S)$ for a given S and a given D. Due to the convexity of $TRC(\mathbf{C}, S)$ with respect to μ, the optimal capacity level, $\mu^*(D, S)$, for given S and D can be found from μ, which yields the $\frac{d\,TRC(\mathbf{C},S)}{d\mu} = 0$ result for the $TRC(\mathbf{C}, S)$ from Equation (2.29).

Subsequent to finding the optimal capacity level $\mu^*(D, S)$, for a given S and D, then we proceed to calculate $TRC(\mathbf{C}, S)$ for that S and capacity policy $\mathbf{C} = [D, \mu^*(D, S)]$.

These steps in 0-a are repeated for all $D \in \theta$. Then, we can find the optimal capacity policy, $\mathbf{C}^*(S)$, for a given stock level S from (2.31), by employing a brute-force search over possible D values in θ. Finally, in the last step, step 1, the global optimal policy parameters: (S^*, D^*, and μ^*) are searched over all possible stock level candidates $S \in \{0,1,2,\dots,S_{max}\}$, which finalizes the search procedure.

Next, we present and discuss the results of the numerical study that was conducted, where the optimal costs for problem (2.21) are obtained from Search Procedure-III.

We first present the total savings in total relevant costs when the best periodic sell-back capacity policy is employed compared to the best fixed capacity system (${TRC_F}^*$ in short). Afterwards, the sensitivity analysis results of the cost and the optimal policy parameters will be discussed in detail, and finally the accuracy of the finite waiting room approximation will be discussed through the results of the simulation study.

2.6.3.3 Possible Savings of Capacity Sell-back Policies Compared to the Single Level Capacity (${TRC_F}^*$)

In this subsection, we contemplate the potential savings that the capacity sell-back policy can bring compared to the ${TRC_F}^*$. Total costs can be reduced up to a great extent due to the capacity sell-back policy. From our numerical results, we have witnessed that up to 92% savings can be achievable in total costs due to the capacity sell-back mode compared to the minimum cost that can be achieved in the single capacity mode ${TRC_F}^*$. However, in some of the cost parameter instances, for instance, when the maximum opportunity cost is high and insensitive and there is no (or very limited) sell-back opportunity, the capacity sell-back policy may lead to drastic increases (up to

174%), rather than savings in costs, in comparison to $TRC_F{}^*$. Therefore the implementation of the third capacity mode and the feasibility of it have to be scrutinized additionally, taking the cost parameters into account, due to the potential serious consequences.

In Table 2-12, we give the maximum percentage savings that capacity sell-back flexible policies can bring for all 9 different B & h_r scenarios (which are already listed in Table 2-2) with $\Delta = 0, 0.25, 0.5$ and 1 and $\alpha = 0$,1 and 2, when $c_p = 1$ and c_c is an inversely proportional, exponential and linear function of the period length respectively, for $R = 0, 0.1, 0.5, 0.9$ and 1. Let $TRC_3{}^*$ represent the minimum total costs that can be achieved from Search Procedure-III, for given cost parameters (B, h_r, R, Δ and α), and a functional form for c_c. After $TRC_3{}^*$ is found, the percentage savings in Table 2-12 can be calculated from: $\frac{(TRC_F{}^* - TRC_3{}^*)}{TRC_F{}^*}$. The cells are color coded according to their percentage saving values. If $TRC_3{}^*$ is higher than $TRC_F{}^*$, the corresponding cell is white, on the other hand, if $TRC_3{}^*$ results in savings, the corresponding cell is shaded in gray, where the higher percentage savings have darker gray tones.

From Table 2-12, we can observe that the periodic capacity sell-back mode is outperformed to the fixed capacity mode, especially for low sell back rates (e.g. $R = 0, 0.1$), high opportunity cost factors (e.g. $\Delta = 1$) and low time elasticity (e.g. $\alpha = 0$).

The percentage savings/losses seem to decrease with holding cost rate h. This can due to the fact that for higher holding cost rate h, although the absolute change in costs due to the sell-back capacity mode gets bigger, percent wise it gets smaller, since the reference cost parameter, $TRC_F{}^*$, is greater and the share of the holding costs (HC) in $TRC_F{}^*$ is bigger for larger h. Parallel to our observations before, the percentage savings increase/percentage losses decrease with the elasticity factor α. This can be explained as follows: the more elastic the contingent capacity cost gets (with respect to the period length D), the cheaper contingent capacity becomes, which leads to additional savings or alleviation of the losses. On the other hand, the maximum opportunity cost, Δ, has an adversary effect, since higher Δ causes the contingent capacity to be more expensive, which leads to an increase in total costs , $TRC_3{}^*$.

From Table 2-12, the exact effect of the down-time related cost rate, B, is unclear. For some instances, when all the cost parameters stay the same except for B, we can see that an increase in B can accompany both an increase as well as a decrease in percentage savings. The dynamics between the backorder cost rate B and percentage savings $\frac{(TRC_F{}^* - TRC_3{}^*)}{TRC_F{}^*}$ seem to be indirect and may be related on many other factors, therefore further research is needed on this interaction.

The possibility of selling the capacity back to the provider makes the TRC less sensitive with respect to B and h. This generally holds true, as the flexible capacity policies make the systems more robust, not only to the changes/wrong estimates for the cost parameters, but also to the changes/ wrong estimates for the demand rate, as well. This, we will mention more specifically, later in Chapter 4.

The functional form of c_c seems to play an important role in the cost performance of the capacity sell-back mode when $\Delta > 0$ and $\alpha > 0$, because when there is no maximum opportunity cost, or when the capacity cost is inelastic to the period length, all three functional forms yield to the same c_c . On the other hand, when both $\Delta > 0$ and $\alpha > 0$, for given B, h, Δ & α , the linear functional form appears to be the form that results in the biggest savings (or smallest loss) and the inverse proportional form appears to be the form that results in the smallest savings (or biggest loss).

From Table 2-12, it is remarkable that in all of the 9 B & h combinations and all of the Δ & α settings, the sell-back rate R is the primary determinant of the cost savings (losses). For each B, h, Δ & α quartet, when $R = 0$, the $TRC^*{}_3$ is always surpassed by $TRC^*{}_F$ (the % difference can be up to -174%), and as R increases, the gap between $TRC^*{}_3$ and $TRC^*{}_F$ decreases. When $\alpha > 0$, there is a threshold R value, after which $TRC^*{}_3$ starts to outperform $TRC^*{}_F$, and the cost savings of the capacity sell-back policy increases after that threshold R value. This threshold R value is lower, when the capacity price , c_c is lower, i.e. when Δ is small, when α is high and when c_c has a linear cost structure.

Finally, we explore further how the optimal policy parameters change under the optimal capacity sell-back mode compared to the single level capacity mode for different cost parameter settings. In Table 2-13 we show how the optimal periodic sell-back capacity mode policy parameters (S^*, D^* and μ^*) differ with various (Δ, α) combinations and 4 different B & h scenarios. The $S^*$and μ^* values from the periodic sell-back capacity mode, which are higher than the reference optimal ${S_F}^*$ and $\mu^*({S_F}^*)$ values from the optimal single-level capacity policies are highlighted.

The data in Table 2-13 illustrate that, under the optimal policies pertaining to the periodic sell-back capacity mode, the spare unit inventory holding costs can be higher or lower than the holding costs under the optimal single-level capacity policy, depending on the cost parameters. This is different in two-level capacity mode, which always engenders lower (or the same) S^* values compared to the ${S_F}^*$ in the single-level capacity mode.

The optimal stock levels under the sell-back capacity mode tend to decrease with higher holding cost rate h (parallel to other capacity modes) and with higher sell-back rate R values. The latter tendency of decreasing S values can be attributed to the fact that higher sell-back rate R motivates the MSP to buy higher capacity, which necessitates less spare units. However, no matter how high the capacity level is, the need for spare units never disappears. Even in one of the instances, when ($\Delta = 0$) and the sell-back rate R is 1, the optimal capacity level appears to be ∞, however optimal stock level is still greater than zero, despite the fact that repairs are conducted instantaneously.

The reason for the need for the spare unit, even in the presence of infinite capacity, can be explained as follows. In this hypothetical scenario, MSP would choose to buy as much capacity as possible, since they can sell it back to the agency with no penalty. However, even in this case, system requires additional spare units, in order to cover for the availability during the delays due to the periodic admission structure (pre-admission delays).

Inversely Proportinal

h	B=1.25		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.025	Δ=1	α=0	-174%	-171%	-156%	-140%	-135%
		α=1	-44%	-40%	-20%	7%	16%
		α=2	-31%	-26%	-5%	24%	35%
	Δ=0.5	α=0	-89%	-85%	-68%	-47%	-40%
		α=1	-29%	-25%	-4%	26%	37%
		α=2	-21%	-16%	5%	36%	50%
	Δ=0.25	α=0	-46%	-41%	-22%	4%	13%
		α=1	-19%	-15%	7%	38%	52%
		α=2	-14%	-9%	12%	45%	60%
	Δ=0	all α	-1%	4%	26%	63%	93%

h	B=2.5		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.025	Δ=1	α=0	-164%	-161%	-147%	-130%	-126%
		α=1	-42%	-37%	-18%	9%	18%
		α=2	-29%	-25%	-4%	25%	37%
	Δ=0.5	α=0	-84%	-80%	-63%	-42%	-36%
		α=1	-28%	-23%	-2%	27%	39%
		α=2	-20%	-15%	6%	37%	50%
	Δ=0.25	α=0	-43%	-39%	-20%	6%	15%
		α=1	-18%	-14%	8%	39%	52%
		α=2	-13%	-9%	13%	46%	61%
	Δ=0	all α	-1%	4%	26%	63%	92%

h	B=5		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.025	Δ=1	α=0	-157%	-153%	-139%	-122%	-118%
		α=1	-40%	-35%	-16%	11%	20%
		α=2	-28%	-23%	-3%	27%	38%
	Δ=0.5	α=0	-80%	-76%	-59%	-38%	-32%
		α=1	-26%	-22%	-1%	28%	40%
		α=2	-19%	-14%	7%	38%	51%
	Δ=0.25	α=0	-41%	-37%	-18%	8%	16%
		α=1	-17%	-13%	8%	40%	53%
		α=2	-13%	-8%	14%	46%	61%
	Δ=0	all α	-1%	4%	26%	63%	92%

h	B=2.5		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.05	Δ=1	α=0	-132%	-129%	-115%	-99%	-94%
		α=1	-43%	-39%	-20%	5%	13%
		α=2	-31%	-27%	-7%	20%	30%
	Δ=0.5	α=0	-69%	-65%	-48%	-27%	-21%
		α=1	-29%	-25%	-5%	23%	33%
		α=2	-21%	-17%	4%	33%	45%
	Δ=0.25	α=0	-36%	-31%	-12%	13%	21%
		α=1	-19%	-14%	6%	36%	48%
		α=2	-14%	-9%	12%	43%	56%
	Δ=0	all α	-2%	3%	25%	62%	90%

h	B=5		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.05	Δ=1	α=0	-125%	-122%	-108%	-92%	-87%
		α=1	-41%	-37%	-18%	7%	15%
		α=2	-29%	-25%	-6%	22%	31%
	Δ=0.5	α=0	-65%	-61%	-44%	-24%	-18%
		α=1	-27%	-23%	-3%	24%	34%
		α=2	-20%	-16%	5%	34%	46%
	Δ=0.25	α=0	-34%	-30%	-11%	14%	23%
		α=1	-18%	-13%	7%	37%	48%
		α=2	-13%	-9%	12%	43%	57%
	Δ=0	all α	-2%	3%	25%	61%	90%

h	B=10		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.05	Δ=1	α=0	-119%	-116%	-102%	-86%	-82%
		α=1	-39%	-35%	-16%	9%	17%
		α=2	-28%	-24%	-4%	23%	32%
	Δ=0.5	α=0	-62%	-58%	-42%	-21%	-15%
		α=1	-26%	-22%	-2%	25%	35%
		α=2	-19%	-15%	6%	35%	46%
	Δ=0.25	α=0	-32%	-28%	-9%	16%	24%
		α=1	-17%	-13%	8%	37%	49%
		α=2	-13%	-8%	13%	43%	57%
	Δ=0	all α	-1%	3%	25%	61%	89%

h	B=12.5		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.25	Δ=1	α=0	-77%	-74%	-61%	-46%	-42%
		α=1	-44%	-40%	-25%	-5%	1%
		α=2	-33%	-29%	-12%	10%	17%
	Δ=0.5	α=0	-42%	-38%	-23%	-4%	2%
		α=1	-28%	-24%	-7%	15%	23%
		α=2	-22%	-18%	0%	25%	34%
	Δ=0.25	α=0	-24%	-20%	-2%	21%	29%
		α=1	-17%	-13%	5%	31%	40%
		α=2	-14%	-10%	9%	36%	46%
	Δ=0	all α	-4%	1%	21%	54%	79%

h	B=25		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.25	Δ=1	α=0	-74%	-71%	-58%	-43%	-39%
		α=1	-42%	-39%	-23%	-4%	3%
		α=2	-32%	-28%	-11%	10%	17%
	Δ=0.5	α=0	-40%	-37%	-21%	-2%	3%
		α=1	-27%	-23%	-6%	16%	23%
		α=2	-21%	-17%	1%	25%	34%
	Δ=0.25	α=0	-22%	-19%	-1%	22%	30%
		α=1	-16%	-12%	6%	31%	40%
		α=2	-13%	-9%	10%	36%	46%
	Δ=0	all α	-4%	0%	21%	54%	79%

h	B=50		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.25	Δ=1	α=0	-71%	-68%	-55%	-41%	-36%
		α=1	-40%	-37%	-21%	-2%	4%
		α=2	-30%	-26%	-10%	11%	18%
	Δ=0.5	α=0	-39%	-35%	-20%	-1%	5%
		α=1	-26%	-22%	-5%	17%	24%
		α=2	-20%	-16%	2%	26%	34%
	Δ=0.25	α=0	-22%	-18%	-1%	22%	30%
		α=1	-16%	-12%	6%	31%	40%
		α=2	-13%	-9%	9%	36%	46%
	Δ=0	all α	-4%	1%	21%	53%	77%

Exponential

	B=1.25		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.025	Δ=1	α=0	-174%	-171%	-156%	-140%	-135%
		α=1	-14%	-10%	12%	45%	63%
		α=2	-8%	-3%	19%	54%	75%
	Δ=0.5	α=0	-89%	-85%	-68%	-47%	-40%
		α=1	-12%	-8%	14%	48%	66%
		α=2	-7%	-2%	20%	56%	77%
	Δ=0.25	α=0	-46%	-41%	-22%	4%	13%
		α=1	-10%	-5%	17%	51%	70%
		α=2	-6%	-1%	21%	57%	78%
	Δ=0	all α	-1%	4%	26%	63%	93%

	B=2.5		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.025	Δ=1	α=0	-164%	-161%	-147%	-130%	-126%
		α=1	-14%	-9%	13%	46%	64%
		α=2	-7%	-3%	20%	54%	75%
	Δ=0.5	α=0	-84%	-80%	-63%	-42%	-36%
		α=1	-12%	-7%	15%	49%	67%
		α=2	-6%	-2%	21%	56%	77%
	Δ=0.25	α=0	-43%	-39%	-20%	6%	15%
		α=1	-10%	-5%	17%	51%	69%
		α=2	-6%	-1%	22%	57%	78%
	Δ=0	all α	-1%	4%	26%	63%	92%

	B=5		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.025	Δ=1	α=0	-157%	-153%	-139%	-122%	-118%
		α=1	-13%	-8%	13%	46%	64%
		α=2	-7%	-2%	20%	55%	75%
	Δ=0.5	α=0	-80%	-76%	-59%	-38%	-32%
		α=1	-11%	-6%	15%	49%	67%
		α=2	-6%	-1%	21%	56%	76%
	Δ=0.25	α=0	-41%	-37%	-18%	8%	16%
		α=1	-9%	-4%	17%	51%	70%
		α=2	-5%	-1%	22%	57%	78%
	Δ=0	all α	-1%	4%	26%	63%	92%

	B=2.5		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.05	Δ=1	α=0	-132%	-129%	-115%	-99%	-94%
		α=1	-18%	-14%	7%	38%	54%
		α=2	-10%	-5%	16%	49%	68%
	Δ=0.5	α=0	-69%	-65%	-48%	-27%	-21%
		α=1	-15%	-11%	10%	42%	58%
		α=2	-9%	-4%	18%	51%	70%
	Δ=0.25	α=0	-36%	-31%	-12%	13%	21%
		α=1	-12%	-8%	14%	46%	62%
		α=2	-7%	-2%	19%	53%	72%
	Δ=0	all α	-2%	3%	25%	62%	90%

	B=5		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.05	Δ=1	α=0	-125%	-122%	-108%	-92%	-87%
		α=1	-17%	-13%	8%	39%	54%
		α=2	-9%	-5%	17%	50%	68%
	Δ=0.5	α=0	-65%	-61%	-44%	-24%	-18%
		α=1	-14%	-10%	11%	43%	58%
		α=2	-8%	-4%	18%	51%	70%
	Δ=0.25	α=0	-34%	-30%	-11%	14%	23%
		α=1	-11%	-7%	14%	46%	62%
		α=2	-7%	-2%	20%	53%	72%
	Δ=0	all α	-2%	3%	25%	61%	90%

	B=10		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.05	Δ=1	α=0	-119%	-116%	-102%	-86%	-82%
		α=1	-16%	-12%	9%	39%	54%
		α=2	-9%	-4%	17%	49%	67%
	Δ=0.5	α=0	-62%	-58%	-42%	-21%	-15%
		α=1	-14%	-9%	12%	43%	58%
		α=2	-8%	-3%	18%	51%	70%
	Δ=0.25	α=0	-32%	-28%	-9%	16%	24%
		α=1	-11%	-6%	15%	47%	62%
		α=2	-6%	-2%	20%	53%	72%
	Δ=0	all α	-1%	3%	25%	61%	89%

	B=12.5		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.25	Δ=1	α=0	-77%	-74%	-61%	-46%	-42%
		α=1	-29%	-25%	-8%	16%	25%
		α=2	-18%	-13%	5%	32%	45%
	Δ=0.5	α=0	-42%	-38%	-23%	-4%	2%
		α=1	-23%	-18%	0%	25%	35%
		α=2	-14%	-10%	9%	37%	50%
	Δ=0.25	α=0	-24%	-20%	-2%	21%	29%
		α=1	-16%	-12%	7%	34%	44%
		α=2	-11%	-7%	13%	42%	56%
	Δ=0	all α	-4%	1%	21%	54%	79%

	B=25		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.25	Δ=1	α=0	-74%	-71%	-58%	-43%	-39%
		α=1	-28%	-24%	-7%	17%	26%
		α=2	-17%	-13%	6%	33%	45%
	Δ=0.5	α=0	-40%	-37%	-21%	-2%	3%
		α=1	-21%	-17%	1%	26%	35%
		α=2	-14%	-9%	9%	37%	50%
	Δ=0.25	α=0	-22%	-19%	-1%	22%	30%
		α=1	-15%	-11%	7%	34%	45%
		α=2	-11%	-7%	13%	42%	55%
	Δ=0	all α	-4%	0%	21%	54%	79%

	B=50		R=0	R=0.1	R=0.5	R=0.9	R=1
h = 0.25	Δ=1	α=0	-71%	-68%	-55%	-41%	-36%
		α=1	-27%	-23%	-6%	17%	26%
		α=2	-16%	-12%	6%	33%	45%
	Δ=0.5	α=0	-39%	-35%	-20%	-1%	5%
		α=1	-20%	-17%	1%	26%	35%
		α=2	-13%	-9%	10%	37%	50%
	Δ=0.25	α=0	-22%	-18%	-1%	22%	30%
		α=1	-14%	-10%	8%	34%	44%
		α=2	-10%	-6%	13%	41%	55%
	Δ=0	all α	-4%	1%	21%	53%	77%

Linear

h =	B=1.25		R=0	R=0.1	R=0.5	R=0.9	R=1
0.025	Δ=1	α=0	-174%	-171%	-156%	-140%	-135%
		α=1	-2%	3%	25%	62%	89%
		α=2	-1%	4%	26%	63%	93%
	Δ=0.5	α=0	-89%	-85%	-68%	-47%	-40%
		α=1	-1%	4%	26%	63%	93%
		α=2	-1%	4%	26%	63%	93%
	Δ=0.25	α=0	-46%	-41%	-22%	4%	13%
		α=1	-1%	4%	26%	63%	93%
		α=2	-1%	4%	26%	63%	93%
	Δ=0	all α	-1%	4%	26%	63%	93%

h =	B=2.5		R=0	R=0.1	R=0.5	R=0.9	R=1
0.025	Δ=1	α=0	-164%	-161%	-147%	-130%	-126%
		α=1	-2%	3%	25%	62%	89%
		α=2	-1%	4%	26%	63%	92%
	Δ=0.5	α=0	-84%	-80%	-63%	-42%	-36%
		α=1	-1%	4%	26%	63%	92%
		α=2	-1%	4%	26%	63%	92%
	Δ=0.25	α=0	-43%	-39%	-20%	6%	15%
		α=1	-1%	4%	26%	63%	92%
		α=2	-1%	4%	26%	63%	92%
	Δ=0	all α	-1%	4%	26%	63%	92%

h =	B=5		R=0	R=0.1	R=0.5	R=0.9	R=1
0.025	Δ=1	α=0	-157%	-153%	-139%	-122%	-118%
		α=1	-2%	3%	25%	61%	88%
		α=2	-1%	4%	26%	63%	92%
	Δ=0.5	α=0	-80%	-76%	-59%	-38%	-32%
		α=1	-1%	4%	26%	63%	92%
		α=2	-1%	4%	26%	63%	92%
	Δ=0.25	α=0	-41%	-37%	-18%	8%	16%
		α=1	-1%	4%	26%	63%	92%
		α=2	-1%	4%	26%	63%	92%
	Δ=0	all α	-1%	4%	26%	63%	92%

h =	B=2.5		R=0	R=0.1	R=0.5	R=0.9	R=1
0.05	Δ=1	α=0	-132%	-129%	-115%	-99%	-94%
		α=1	-3%	1%	23%	59%	85%
		α=2	-2%	3%	25%	62%	90%
	Δ=0.5	α=0	-69%	-65%	-48%	-27%	-21%
		α=1	-2%	3%	25%	62%	90%
		α=2	-2%	3%	25%	62%	90%
	Δ=0.25	α=0	-36%	-31%	-12%	13%	21%
		α=1	-2%	3%	25%	62%	90%
		α=2	-2%	3%	25%	62%	90%
	Δ=0	all α	-2%	3%	25%	62%	90%

h =	B=5		R=0	R=0.1	R=0.5	R=0.9	R=1
0.05	Δ=1	α=0	-125%	-122%	-108%	-92%	-87%
		α=1	-3%	1%	23%	59%	85%
		α=2	-2%	3%	25%	61%	90%
	Δ=0.5	α=0	-65%	-61%	-44%	-24%	-18%
		α=1	-2%	3%	25%	61%	90%
		α=2	-2%	3%	25%	61%	90%
	Δ=0.25	α=0	-34%	-30%	-11%	14%	23%
		α=1	-2%	3%	25%	61%	90%
		α=2	-2%	3%	25%	61%	90%
	Δ=0	all α	-2%	3%	25%	61%	90%

h =	B=10		R=0	R=0.1	R=0.5	R=0.9	R=1
0.05	Δ=1	α=0	-119%	-116%	-102%	-86%	-82%
		α=1	-3%	2%	23%	59%	84%
		α=2	-1%	3%	25%	61%	89%
	Δ=0.5	α=0	-62%	-58%	-42%	-21%	-15%
		α=1	-1%	3%	25%	61%	89%
		α=2	-1%	3%	25%	61%	89%
	Δ=0.25	α=0	-32%	-28%	-9%	16%	24%
		α=1	-1%	3%	25%	61%	89%
		α=2	-1%	3%	25%	61%	89%
	Δ=0	all α	-1%	3%	25%	61%	89%

h =	B=12.5		R=0	R=0.1	R=0.5	R=0.9	R=1
0.25	Δ=1	α=0	-77%	-74%	-61%	-46%	-42%
		α=1	-9%	-5%	16%	46%	69%
		α=2	-4%	1%	21%	54%	79%
	Δ=0.5	α=0	-42%	-38%	-23%	-4%	2%
		α=1	-4%	1%	21%	54%	79%
		α=2	-4%	1%	21%	54%	79%
	Δ=0.25	α=0	-24%	-20%	-2%	21%	29%
		α=1	-4%	1%	21%	54%	79%
		α=2	-4%	1%	21%	54%	79%
	Δ=0	all α	-4%	1%	21%	54%	79%

h =	B=25		R=0	R=0.1	R=0.5	R=0.9	R=1
0.25	Δ=1	α=0	-74%	-71%	-58%	-43%	-39%
		α=1	-8%	-4%	16%	47%	69%
		α=2	-4%	0%	21%	54%	79%
	Δ=0.5	α=0	-40%	-37%	-21%	-2%	3%
		α=1	-4%	0%	21%	54%	79%
		α=2	-4%	0%	21%	54%	79%
	Δ=0.25	α=0	-22%	-19%	-1%	22%	30%
		α=1	-4%	0%	21%	54%	79%
		α=2	-4%	0%	21%	54%	79%
	Δ=0	all α	-4%	0%	21%	54%	79%

h =	B=50		R=0	R=0.1	R=0.5	R=0.9	R=1
0.25	Δ=1	α=0	-71%	-68%	-55%	-41%	-36%
		α=1	-8%	-4%	16%	46%	67%
		α=2	-4%	1%	21%	53%	77%
	Δ=0.5	α=0	-39%	-35%	-20%	-1%	5%
		α=1	-4%	1%	21%	53%	77%
		α=2	-4%	1%	21%	53%	77%
	Δ=0.25	α=0	-22%	-18%	-1%	22%	30%
		α=1	-4%	1%	21%	53%	77%
		α=2	-4%	1%	21%	53%	77%
	Δ=0	all α	-4%	1%	21%	53%	77%

Table 2-12 The % cost savings of the capacity sell-back mode compared to the fixed capacity policy, when $h = 0.025, 0.05, 0.25$ and $B/h = 50, 100, 200$ for $\Delta = 0, 0.25, 0.5, 1$, $\alpha = 0, 1, 2$ when $R = 0, 0.1, 0.5, 0.9, 1$ when c_c has inverse linear, exponential and linear structure.

In Table 2-13, it can be seen that the smallest possible period length, $D = 0.5$, is chosen as the optimal period length D^*, especially for the instances, when there is no opportunity cost ($\Delta = 0$), or when c_c is inelastic ($\alpha = 0$). What we can further observe that the optimal period length D^* tends to increase with Δ, and first increase and then decrease with α. These interactions of D^* under the periodic sell-back mode are parallel to the interactions of D^* under two-level capacity mode. Additionally, we can also see that D^* decreases with higher h and increases with higher R, but it is rather insensitive to the changes in B only.

In the two level capacity mode, the percentage cost savings with respect to $TRC^*{}_F$ result from both lower stock levels, S^*, and lower capacity deployment ($ACU(\mathbf{C})$) in comparison to the single-level capacity mode (Table 2-6). However, in Table 2-13, it is remarkable that under the periodic sell-back capacity mode, at least one of the optimal stock level S^*, or the optimal capacity μ^* is higher than the optimal reference values under the single level capacity mode. Although μ^* from the third capacity mode can be a lot higher than the $\mu^*(S^*{}_F)$ from the single-level capacity mode, one should always keep it in mind that, in the periodic sell-back capacity mode, more capacity does not always lead to higher capacity related costs (CRC), since the excess idle capacity can be sold back to the agency and due to this sell-back opportunity, it may be more profitable to deploy higher capacity compared to the $\mu^*(S^*{}_F)$ from the fixed capacity mode.

These inter-relations between the capacity and cost parameters will be further explicated in the next section, where a list of sensitivity analyses are conducted on policy and cost parameters. After the sensitivity analysis subsection, we check the accuracy of using the finite waiting room approximation for the analysis of the repair shop operations, by comparing the results from the analytical computation with the results from the simulation.

2.6.3.4 Sensitivity Analysis of the Cost and Optimal Policy Parameters

In this subsection, we discuss the interrelations among the cost and optimal policy parameters. We first focus on how total relevant cost per time, TRC, responds to changes in capacity level μ under different cost/policy parameters. Afterwards, we investigate the effect of the period length D on TRC as well as on the choice of the other policy parameters. Finally, we examine how the base stock level decision S influences the TRC and other parameters.

2.6.3.4.1 The Capacity Level: $\boldsymbol{\mu}$

Recall that in the previous section, we have proved the convexity of TRC with respect to μ. In this sub-section, we first elaborate further on the behavior of TRC with increasing capacity levels, μ. Afterwards, we discuss how the optimal capacity level μ^* changes with different cost/policy parameters.

B=5, h=0.05

Single Level			S_F^*			$\mu^*(S_F^*)$					
			10			1.54					
			R=0			R=0.5			R=1		
Periodic Sell-back			S*	D*	μ*	S*	D*	μ*	S*	D*	μ*
inv. proportional	Δ=0	all α	11	0.5	1.52	8	0.5	1.76	2	0.5	inf
	Δ=0.25	α=0	12	0.5	1.47	9	0.5	1.64	6	0.5	2.10
		α=1	12	2.0	1.52	10	2.0	1.69	7	2.5	2.74
		α=2	12	1.5	1.50	9	1.5	1.73	6	2.0	3.10
	Δ=0.5	α=0	13	0.5	1.42	11	0.5	1.52	8	0.5	1.76
		α=1	14	3.0	1.48	11	3.0	1.67	9	3.5	2.34
		α=2	13	2.5	1.50	11	2.5	1.66	7	2.5	2.68
	Δ=1	α=0	14	0.5	1.38	13	0.5	1.42	11	0.5	1.52
		α=1	15	4.5	1.49	13	4.5	1.63	11	5.0	2.15
		α=2	14	3.5	1.49	12	3.5	1.64	9	3.5	2.30

B=25, h=0.25

Single Level			S_F^*			$\mu^*(S_F^*)$					
			5			2.26					
			R=0			R=0.5			R=1		
Periodic Sell-back			S*	D*	μ*	S*	D*	μ*	S*	D*	μ*
inv. proportional	Δ=0	all α	6	0.5	2.16	5	0.5	2.58	2	0.5	inf
	Δ=0.25	α=0	6	0.5	2.10	5	0.5	2.44	4	0.5	3.28
		α=1	6	0.5	2.12	5	0.5	2.47	4	1.0	4.17
		α=2	6	0.5	2.13	5	0.5	2.50	3	0.5	4.63
	Δ=0.5	α=0	7	0.5	1.94	6	0.5	2.16	5	0.5	2.58
		α=1	7	1.0	2.02	6	1.0	2.32	5	1.5	3.47
		α=2	7	1.0	2.04	6	1.0	2.35	4	1.0	3.92
	Δ=1	α=0	7	0.5	1.88	7	0.5	1.94	6	0.5	2.16
		α=1	8	2.0	2.00	7	1.5	2.17	6	2.0	2.94
		α=2	7	1.5	2.09	7	1.5	2.22	5	1.5	3.32

B=5, h=0.025

Single Level			S_F^*			$\mu^*(S_F^*)$					
			15			1.40					
			R=0			R=0.5			R=1		
Periodic Sell-back			S*	D*	μ*	S*	D*	μ*	S*	D*	μ*
inv. proportional	Δ=0	all α	16	0.5	1.39	12	0.5	1.55	2	0.5	inf
	Δ=0.25	α=0	17	0.5	1.36	14	0.5	1.46	9	0.5	1.79
		α=1	18	3.0	1.39	15	3.0	1.51	9	3.0	2.40
		α=2	18	2.5	1.38	14	2.5	1.53	8	2.5	2.65
	Δ=0.5	α=0	19	0.5	1.32	16	0.5	1.39	12	0.5	1.55
		α=1	20	4.5	1.37	17	4.5	1.49	12	5.0	2.18
		α=2	19	3.5	1.38	15	3.5	1.53	10	4.0	2.48
	Δ=1	α=0	20	0.5	1.29	19	0.5	1.32	16	0.5	1.39
		α=1	20	5.0	1.38	18	5.0	1.47	13	5.0	1.90
		α=2	20	4.5	1.37	17	5.0	1.50	12	5.0	2.15

B=10, h=0.05

Single Level			S_F^*			$\mu^*(S_F^*)$					
			11			1.58					
			R=0			R=0.5			R=1		
Periodic Sell-back			S*	D*	μ*	S*	D*	μ*	S*	D*	μ*
inv. proportional	Δ=0	all α	12	0.5	1.55	9	0.5	1.79	2	0.5	inf
	Δ=0.25	α=0	13	0.5	1.50	10	0.5	1.67	7	0.5	2.12
		α=1	13	2.0	1.55	11	2.0	1.72	7	2.0	2.79
		α=2	13	1.5	1.54	10	1.5	1.76	7	2.0	3.02
	Δ=0.5	α=0	14	0.5	1.46	12	0.5	1.55	9	0.5	1.79
		α=1	15	3.0	1.51	12	3.0	1.71	9	3.0	2.40
		α=2	14	2.5	1.53	12	2.5	1.69	8	2.5	2.65
	Δ=1	α=0	16	0.5	1.39	14	0.5	1.46	12	0.5	1.55
		α=1	17	4.5	1.49	14	4.5	1.67	12	5.0	2.18
		α=2	15	3.5	1.53	13	3.5	1.68	10	4.0	2.48

Table 2-13: The optimal periodic sell-back capacity mode policy parameters (S^*, D^* and μ^*) under different 4 B & h scenarios (**1:** $B = 5, h = 0.05$ **2:** $B = 25, h = 0.25$ **3:** $B = 5, h = 0.025$ and **4:** $B = 10, h = 0.25$) and various (Δ, α) combinations when $c_p = 1$ and $R = 0, 0.5$ and 1.

How TRC responds to changes in the capacity level is greatly dependent on other policy/cost parameters. In order to have a better understanding of the effects of the capacity level on total costs, we plot the TRC as a function of μ in Figure 2-11. In this figure, we analyze the response of TRC to increasing values of μ, under the fixed capacity mode (TRC_F) as well as under the capacity sell-back mode with different sell-back reduction rates ($R = 0, 0.5$ and 1) when: Left: $B = 5$, $h = 0.05$, $c_p = 1$, $\Delta = 0$, $D = 0.5$ and $S = 2$ Right: $B = 5$, $h = 0.05$, $c_p = 1$, $\Delta = 0.5$, $\alpha = 2$, $D = 2.5$ and $S = 13$.

From Figure 2-11, we can observe a general tendency of the response of TRC to increasing values of μ. For a constant stock level S and a constant period length D, if the capacity sell-back cost reduction rate R is strictly less than 1, TRC first decreases and then increases with the capacity level μ. This is in line with the convexity result, which we have proven in the previous section. Note that the difference between the fixed policy line (TRC_F) and the line when there is no capacity sellback option ($R = 0$) shows the additional costs due to the delays/burstiness in the periodic admission. It can be also observed that, in case there is no opportunity cost for the flexible workforce ($\Delta = 0$), the TRC gradually transmutes itself to a decreasing function from a U-shaped one as R goes to 1. However if ($\Delta > 0$), TRC is still a U-shaped function of μ, when $R = 1$.

In the first extreme case, when $R = 1$, in the optimal solution, the repair shop operates in an hypothetical environment, where the capacity agency is always able to assign any amount of capacity to another task during the idle times in each period, and these alternative tasks are at least as profitable as the core activity of the repair shop, so that the provider agrees to pay back the full amount of c_p in order to employ the repair shop capacity for the alternative tasks during its idle times. In such a case, when there is no opportunity cost ($\Delta = 0$), TRC decreases with the capacity level μ, and for any period length D and any stock level S, $\mu = \infty$ would be optimal.

The other extreme is when $R = 0$, i.e. when the repair shop capacity is not sold back to the provider during the idle times. In such a case, the fixed capacity mode (TRC_F) outperforms the sell-back mode for every μ, since the periodicity of the admission structure causes unnecessary repair delays as well as burstiness in the defective unit arrivals to the repair shop. A financial return is only possible for $R > 0$, when the provider agrees to employ the repair shop capacity for another task during the idle times in exchange for an amount of $c_p R$ per time unit. That financial return, for each R, implies a break-even capacity level, after which the capacity sell-back mode outperforms the fixed capacity mode. That break-even capacity level gets higher for lower R values. We also observe that the effects of the sell-back reduction rate R on TRC are much more visible for higher levels of capacity. This can be explained by the fact that a higher capacity level results in more idle time; therefore the differences in CRC due to the selling the capacity back to the agency between different R values materialize more when the capacity level μ is higher.

A higher R value not only reduces costs but also stimulates the repair shop to install more capacity due to the financial regains from capacity sell-back possibilities during idle times. Note that even if there was no financial gains from capacity sell-back (i.e. $R = 0$), the repair shop would already require a higher capacity than the fixed capacity mode in order to neutralize the negative effects on performance of the periodic admissions. These effects can be observed in Table 2-13, where the best capacity levels are tabulated under the single capacity mode as well as in the periodic sell-back capacity mode with increasing values of R for a given period length ($D = 0.5$), when there is no opportunity costs ($\Delta = 0$) and when ($S = 2, h = 1, B = 10, c_p = 1$), ($S = 2, h = 1, B = 10, c_p = 5$) and ($S = 5, h = 1, B = 10, c_p = 10$) respectively.

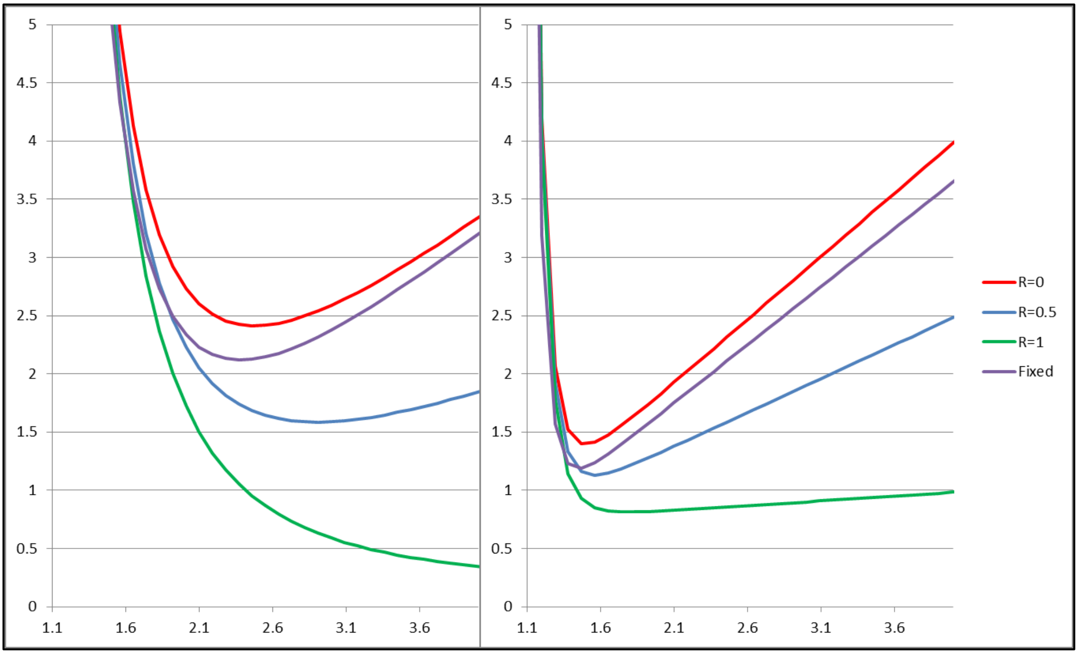

Figure 2-11: : TRC as a function of μ under fixed capacity mode (TRC_F) and under capacity sell-back mode when: Left: $B = 5$, $h = 0.05$, $c_p = 1$, $\Delta = 0$, $D = 0.5$ and $S = 2$ Right: $B = 5$, $h = 0.05$, $c_p = 1$, $\Delta = 0.5$, $\alpha = 2$, $D = 2.5$ and $S = 13$ and c_c has an inverse proportional form.

From Table 2-13, we can observe that under the capacity sell-back mode, the capacity choice $\mu^*({S^*}_F)$ gets higher as R increases, and when $\Delta = 0$ or when a D and α combination cancels the opportunity costs and make c_c equal to c_p, (for finite period lengths, only possible in linear cost structure), $\mu^*({S^*}_F)$ finally goes to infinity for full sell/back rate, i.e. $R = 1$.

The pure influence of deploying periodic admissions becomes clear when we compare the optimal capacity under the single capacity mode with the optimal capacity under the sell-back mode for $R = 0$. The latter is always higher, especially for lower c_p and lower S, both of which magnify the repair shop's capacity response to the periodicity of defective part admissions. A higher sell-back rate ($R > 0$) amplifies this capacity response further; and in order to identify the pure gains from sell-back due to the higher $R > 0$, μ^* under $R > 0$ should be compared with the μ^* under $R = 0$.

	$\mu^*(S^*_F)$	Capacity sell-back mode(when $D = 0.5$)				
		$R = 0$	$R = 0.25$	$R = 0.5$	$R = 0.75$	$R = 1$
$S = 2, h = 1, B = 10, c_p = 1$	2.74	2.86	3.06	3.46	4.06	∞
$S = 2, h = 1, B = 10, c_p = 5$	1.86	1.90	1.96	2.16	2.46	∞
$S = 5, h = 1, B = 10, c_p = 10$	1.55	1.56	1.60	1.66	1.86	∞

Table 2-14 Optimal μ for the single capacity mode and the periodic sell-back capacity mode with increasing R for given period length ($D = 0.5$), when $\Delta = 0$ and when 1-($S = 2, h = 1, B = 10, c_p = 1$), 2-($S = 2, h = 1, B = 10, c_p = 5$) and 3-($S = 5, h = 1, B = 10, c_p = 10$), respectively.

Similar to the second capacity mode, period length D plays a very central role in the sell-back capacity mode, as well. Therefore, in the following subsection, we summarize our findings about the effects of the period length D on TRC, μ^* and S^* and we discuss how the optimal period length D^* changes with respect to other cost parameters.

2.6.3.4.2 Period Length D

Similar to the second capacity mode, the period length D plays a very central role also in this capacity mode, since it determines how frequent the defective parts are admitted to the repair shop as well as the opportunity costs incurred to the contingent capacity price.

Firstly, note that the capacity sell-back mode with periodic admissions transmutes itself to the continuous admission, single capacity mode when the period length D goes to 0 for $\Delta = 0$ and $R = 0$. If $R = 0$, the single capacity mode outperforms the periodic sell-back mode, since there is no regain from selling the capacity back to the agency during the idle times. Furthermore, a longer period length D can only be better for $\Delta > 0$ and $\alpha > 0$. Otherwise, when there is no opportunity cost, or if the opportunity cost is inelastic, MSP would choose the smallest possible period length, since there is no additional benefits of a longer intervals for the periodic admission and longer D amplifies burstiness effect in the repair shop.

In order to understand the effects of the period length on the total cost rate TRC as well as on the other policy parameters (μ and S), we present Table 2-15, where the TRC, μ^* and S^* values under different scenarios are given for increasing values of period length D when $R = 0$, $\Delta = 0, B = 10$ and $c_p = 5$. In Table 2-15, under the first column, we show the response of TRC to increasing values of D when $S = 3$ and $\mu = 1.861$, which are the optimal stock, S^*_F and capacity level, $\mu^*(S^*_F)$ under the single capacity mode.

Afterwards, the capacity level is relaxed and it is optimized at each period length for $S = 3$. These optimal $\mu^*(S)$ levels and the resulting TRC values are given in the 2^{nd} and the 3^{rd} columns. Finally, the stock level decision is also relaxed and both stock and capacity levels are optimized for each period length D. The resulting optimal S^* & μ^* levels and the corresponding TRC values for each D are given in 4^{th}, 5^{th} and 6^{th} columns respectively.

	$S = 3,$ $\mu = 1.86$ $TRC_F = 9.107$	$S = 3,$ μ optimized		S & μ optimized		
D	TRC	μ^*	TRC	S^*	μ^*	TRC
0.5	9.595	1.911	9.583	4	1.759	9.494
1	10.327	1.961	10.267	4	1.859	9.932
1.5	11.349	2.011	11.197	5	1.759	10.411
2	12.682	2.061	12.397	5	1.809	10.931
2.5	14.331	2.111	13.873	6	1.759	11.424
3	16.280	2.161	15.617	6	1.809	11.958
3.5	18.507	2.261	17.605	7	1.759	12.475
4	20.981	2.311	19.808	7	1.759	13.032
4.5	23.671	2.361	22.200	8	1.759	13.542
5	26.547	2.461	24.748	8	1.759	14.108

Table 2-15: Under $R = 0, \Delta = 0, B = 10, h = 1$ and $c_p = 5$: Column 1: TRC for increasing values of D when $S = 3$ and $\mu = 1.861$. Columns 2 & 3: μ^* and TRC values for increasing values of D when $S = 3$. Columns 4 & 5 & 6: S^* , μ^* and TRC values for increasing values of D.`

The data in the first column confirms our expectations about the effects of the periodicity and the period length on total costs. The TRC values with $D > 0$ are higher than TRC with $D = 0$ ($a.k.a$ TRC_F), especially for long period lengths (e.g. when $D = 3.5$, TRC is more or less twice as higher than TRC_F). Under the 2nd and 3rd columns, the capacity level μ is optimized to minimize TRC for each period length, while keeping the stock level unchanged at $S = 3$. We observe that the optimized capacity level, μ^* increases with the period length; however we also observe that this optimization brings about a very minor cost reduction compared to the TRC with constant μ and S. By contrast, if the capacity level is optimized jointly with the stock level for each period length D, the reduction in the total costs is a lot higher, which can be seen from the last column. When we have a closer look on how S^* and μ^* change with different period lengths, we can conclude that it is primarily the stock level decision that responds to an increase in the period length, and decreases the TRC to a great extent; whereas the capacity level does not necessarily increase with D, but plays a secondary, fine-tuning role. One important reason of this behavior is due to the fact that the capacity cost is 5 times higher than the cost of holding one more unit part in the stock.

Next, we analyze how the cost parameters R, Δ and α affect the behaviour of TRC under different period lengths, D. For each period length D, the capacity and stock levels are simultaneously optimized in order to minimize TRC with different sell-back reduction rates and contingent capacity costs. The response of the optimal TRC to increasing period lengths when $B = 5$ and $h = 0.05$ for different R, Δ and α values are illustrated in Figure 2-12.

From the upper left chart in Figure 2-12, we can see that TRC first decrease and then increase with different D values. Note that the $TRC,$ as a function of $D,$ with different R values are almost parallel to each other. When $\Delta = 0$, each $R > 0$ that is sufficiently large, begets a break-even point for the period length. The single capacity mode outperforms the sell-back mode operating under a period that is longer than this break-even point. The break-even point for the period length gets bigger for larger values of R.

The right upper chart in Figure 2-12 illustrates the response of the TRC values to increasing period lengths under different maximum opportunity costs ($\Delta = 0, 0.5$ and 1) when $R = 1$ and $\alpha = 1$, when c_c has an inverse proportional form. If $\Delta = 0$, then the optimal period length is always the smallest possible one, which is 0.5 in our numerical study. On the other hand, a positive transaction cost $\Delta > 0$, increases the TRC values significantly, especially for short period lengths, and changes the structure of TRC in response to D, from an increasing pattern to a U-shaped pattern (Note that when $\Delta = 1$, TRC increases for $D > 5$).

However, it is remarkable that the increasing parts of the U-shaped TRC curves for $\Delta > 0$ are almost flat, which suggests that operating with higher than optimal period length does not increase the costs dramatically since the capacity sell-back and the wage differential mechanisms counterbalance the negative effects of higher period lengths. This change in the cost response structure stimulates the repair shop to choose a period length longer than 0.5 due to the possible savings from capacity costs. Note that the gaps between the curves of TRC with different Δ are the highest for small D values, since the c_c variations start to diminish for longer D if $\alpha > 0$ due to the wage differentials.

The lower chart in Figure 2-12 illustrates the behavior of TRC under different elasticity factors. It can be seen that when $\alpha = 0$, TRC increases with D, therefore $D = 0.5$ is the optimal period length for this case. However for positive elasticity, TRC first decreases with D, which shifts the optimal period length further ($D > 0.5$). These are parallel to the observations in the previous, two-level capacity mode. However, we observe that the differences between the TRC curves for $\alpha = 1$ and $\alpha = 2$ are almost identical, and the gap in between them diminishes for longer D. We also observe that the TRC response to longer D values in Figure 2-12 are a lot more robust compared to the two-level mode, which is due to the fact the very high sell-back rate R hedges the adverse effects of longer period lengths in Figure 2-12.

The optimal period length choice (D^*) seems to be very sensitive to the cost parameters. In order to understand the nature of the dynamics between the optimal choice of the period length D^*, and the capacity cost parameters (Δ, α and R) we present Table 2-16, where the optimal period length choices are tabulated for Top: $\Delta = 0{,}0.5$ and 1 when $\alpha = 1$ and $R = 0.5$; Between: $\Delta = 1$ when $\alpha = 0, 1, 2$ and ∞ and $R = 0.5$; Below: $\Delta = 0.5$ when $\alpha = 1$ and $R = 0, 0.5$ and 1, when c_c has an inverse proportional, linear and exponential structure respectively.

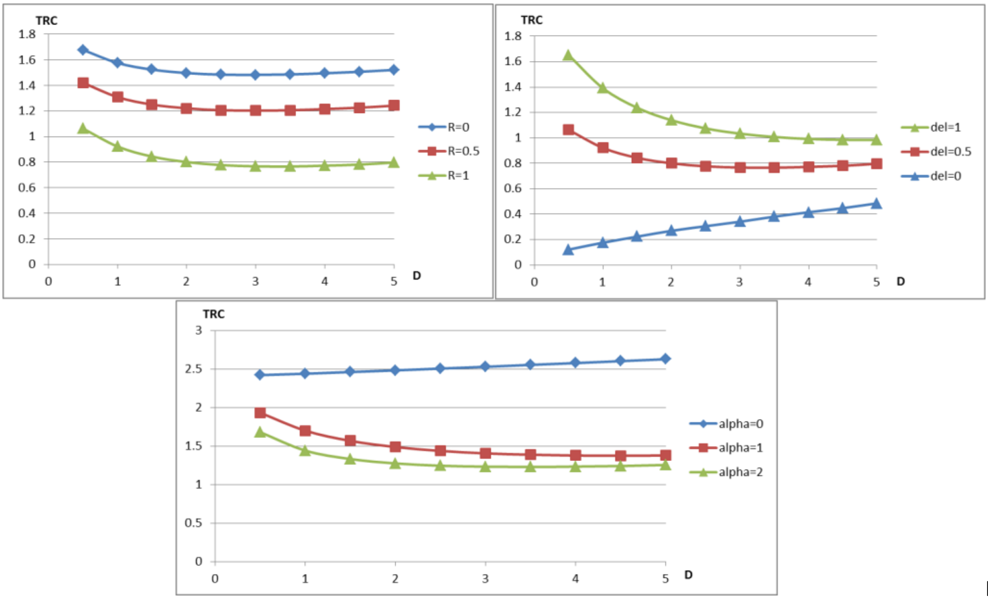

Figure 2-12: TRC as a function of D, with optimized S^* and μ^* for each $D \in \theta = \{0.5, 1, \ldots, 5\}$ when $B = 5$ and $h = 0.05$ Upper left: $\Delta = 0.5$ & $\alpha = 1$ & $R = 0, 0.5$ and 1. Upper right: $R = 1$ & $\alpha = 1$& $\Delta = 0, 0.5$ and 1. Lower: $R = 0.5$ & $\Delta = 1$&$\alpha = 0, 1,2$; c_c: inverse proportional.

D^*	$\Delta = 0, \alpha = 1, R = 0.5$	$\Delta = 0.5, \alpha = 1, R = 0.5$	$\Delta = 1, \alpha = 1, R = 0.5$	
Inverse	0.5	3	4.5	
Linear	0.5	0.5	1	
Exp	0.5	3	3.5	
D^*	$\Delta = 1, \alpha = 0, R = 0.5$	$\Delta = 1, \alpha = 1, R = 0.5$	$\Delta = 1, \alpha = 2, R = 0.5$	$\Delta = 1, \alpha = \infty, R = 0.5$
Inverse	0.5	4.5	3.5	0.5
Linear	0.5	1	0.5	0.5
Exp	0.5	3.5	2	0.5
D^*	$\Delta = 0.5, \alpha = 1, R = 0$	$\Delta = 0.5, \alpha = 1, R = 0.5$	$\Delta = 0.5, \alpha = 1, R = 1$	
Inverse	3	3	3.5	
Linear	0.5	0.5	0.5	
Exp	3	3	3	

Table 2-16 The optimal period length: D^* when $B = 5; h = 0.05$ for Top: $\Delta = 0$,0.5 and 1 when $\alpha = 1$; $R = 0.5$ Between: $\Delta = 1$ when $\alpha = 0, 1, 2$ and ∞ and R = 0.5 Below: $\Delta = 0.5$ when $\alpha = 1$ and $R = 0, 0.5$ and 1; for c_c: inverse proportional, linear and exponential structure respectively.

From Table 2-16, it can be seen that D^* is at its lowest when c_c has a linear structure and at its highest when c_c has an inverse proportional structure. It can be further observed that the responses of D^* to increasing Δ and α in the sell-back mode resemble the D^* responses in the two-level capacity mode. In both of these capacity modes, D^* increases with higher opportunity costs, and it first increases and then decreases with higher elasticity factor. In the tabular below, we can see that D^* is rather insensitive to the changes in R. It can be interpreted that its mode of effect on the capacity level, μ^*, works more or less the same for different period lengths. Only when c_c has an inverse proportional cost structure, optimal period length D^* is changed to the closest larger period length when $R = 1$, which can be explained due to the fact that the sell-back possibility with high return incentivizes to buy more capacity, which hedges the risks of the adverse effects of higher D.

After finishing the analysis of the bi-directional relations between the period length D and TRC^*, next we discuss the effects of the base stock level decision: S.

2.6.3.4.3 Base Stock Level S

In this part, we discuss how the base stock level decision affects the optimal total costs and other policy parameters and vice versa.

Our first observation is the following: a higher stock level S makes the total costs more insensitive to changes in other policy parameters. This is due to the fact that a higher stock level curtails the effects of a change in capacity level and/or the period length. We can distinguish this curtailing effect easily in Figure 2-13, where on the left chart the response of TRC to different μ is illustrated for $S = 2$ and $S = 10$ when $c_p = 1, D = 0.5, \Delta = 0, B = 5$, $h = 0.05$ and $R = 0.5$ and on the right chart the response of TRC to different D is shown for for $S = 2$ and $S = 10$ when $\mu = 2$, $\Delta = 0, R = 0.5, B = 5$ $h = 0.05$ and $c_p = 1$.

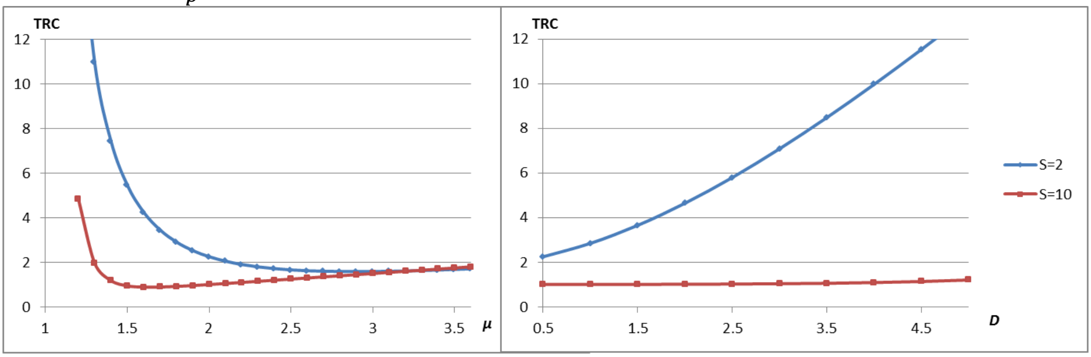

Figure 2-13: On the left: response of TRC to different μ for $S = 2$ and $S = 10$ when $c_p = 1, D = 0.5, \Delta = 0, B = 5$, $h = 0.05$ and $R = 0.5$. On the right: The response of TRC to different D for $S = 2$ and $S = 10$ when $\mu = 2$, $\Delta = 0, R = 0.5, B = 5$ $h = 0.05$ and $c_p = 1$.

Next, we want to analyze how TRC behaves with different stock levels when the capacity policy parameters D and μ are optimized at each stock level S. Figure 2-14 illustrate how TRC changes with increasing values of base-stock levels for $B = 5$, $c_p = 1$, $h = 0.05$ and for different R values ($0, 0.5$ and 1) when ($\Delta = 0.5$, $\alpha = 1$) on the left chart and for different Δ values ($0, 0.5, 1$) when ($R = 0.5$, $\alpha = 1$) on the right. Note that in both of the charts, c_c has an inverse proportional structure.

The data in Figure 2-14 suggest that TRC is a convex function of the stock level when D & μ are optimized at each S. It is already discussed that when all the parameters of a capacity policy $\mathbf{C}$ is given, total costs are convex with the base stock level. It seems that optimizing the parameters of capacity policy at each base-stock level S, does not distort this convexity property.

As expected, lower sell-back rates, R, and higher capacity max. opportunity costs, Δ, increase the TRC values, especially for lower values of S. We also observe that a change in one of the cost parameters can alter the optimal stock level decision significantly. It is evident from Figure 2-14 that the adversary changes in the cost parameters (i.e. lower sell-back rate or higher opportunity costs for contingent capacity) make the MSP hold more spare parts in the stock.

This tendency can be more clearly seen in Table 2-17, where the optimal base-stock levels, S^*, under different capacity cost parameters (R & Δ &α) are tabulated when $B = 5$, $c_p = 1$ and $h = 0.05$, when c_c has an inverse proportional, linear and exponential structure respectively.

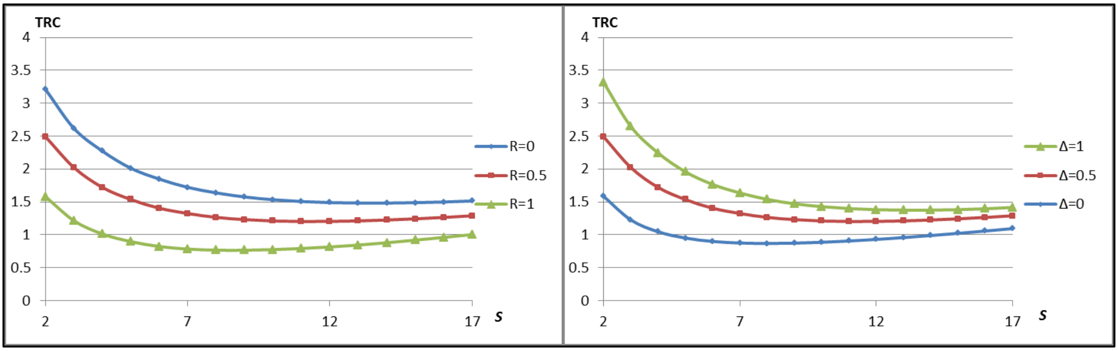

Figure 2-14: The response of TRC for increasing values of base-stock levels S for $B = 5$, $c_p = 1$, $h = 0.05$ and for different R values $(0, 0.5$ and $1)$ when ($\Delta = 0.5$, $\alpha = 1$) on the left chart and for different Δ values $(0, 0.5, 1)$ when ($R = 0.5$, $\alpha = 1$) on the right.

From Table 2-17, we can see that S^* is at its lowest when c_c has a linear structure and at its highest when c_c has an inverse proportional structure. S^* increases with higher opportunity costs and lower elasticity factor. These responses can be explained as follows: a higher contingent capacity cost gives impetus to a higher period length and less installed capacity, both of which necessitate a higher stock level to compensate the negative effects of the periodic admissions and insufficient capacity. Both higher Δ and lower α engenders a more expensive capacity cost.

On the other hand, the response of S^* to an increase in R for given Δ&α is different. With a higher R, the repair shop is encouraged to install more capacity, which attenuates the need for additional spare units, therefore the stock level S^* faces a decrease. This decrease is more evident when c_c has a linear structure, because when $R = 1$, optimal capacity level is infinite, but the MSP still holds an inventory for the availability of the spare units due to the pre-admission delay.

S^*	$\Delta=0, \alpha=1, R=0.5$	$\Delta=0.5, \alpha=1, R=0.5$	$\Delta=1, \alpha=1, R=0.5$	
Inverse	8	11	13	
Linear	8	8	9	
Exp	8	11	12	
S^*	$\Delta=1, \alpha=0, R=0.5$	$\Delta=1, \alpha=1, R=0.5$	$\Delta=1, \alpha=2, R=0.5$	$\Delta=1, \alpha=\infty, R=0.5$
Inverse	13	13	12	8
Linear	13	9	8	8
Exp	13	12	10	8
S^*	$\Delta=0.5, \alpha=1, R=0$	$\Delta=0.5, \alpha=1, R=0.5$	$\Delta=0.5, \alpha=1, R=1$	
Inverse	14	11	9	
Linear	11	8	2	
Exp	13	11	7	

Table 2-17: The optimal stock level: S^* when $B=5$ and $h=0.05$ for Top: $\Delta=0, 0.5$ and 1 when $\alpha=1$ and $R=0.5$ Between: $\Delta=1$ when $\alpha=0,1,2,\infty$ and $R=0.5$ Below: $\Delta=0.5$ when $\alpha=1$ and $R=0,0.5,1$; when c_c: inverse proportional, linear and exponential structure respectively.

This concludes our sensitivity analysis. In the next subsection, the accuracy of the finite waiting room approximation in the analytical method is examined with an extensive simulation study.

2.6.3.5 Accuracy of Finite Waiting Room Approximation

In this subsection, we examine the accuracy of our finite waiting room approximation by comparing the total cost rate of the proposed analytical model (having a finite waiting room of 100) with the cost rate obtained by simulating the real environment having a repair shop that has an infinite waiting room.

In our simulations, we used a run length of 5,000,000,000 defective part arrivals (when $\lambda=1$) in a single replication, where the average total cost rate TRC is calculated under a policy: $\mathbf{C} = [D, \mu]$.

We investigated a total of 360 different scenarios with different B, h, R, Δ, α and resulting policy parameters. The percentage error, $\%err$, of using the approximation for TRC in a scenario can be found as:

$$\%err = 100 \times \frac{\left(TRC_{sim} - TRC_{app}\right)}{TRC_{sim}} \tag{2.32}$$

Here, TRC_{sim} is the total relevant costs obtained from the simulation and TRC_{app} is the total relevant costs obtained from the analytical model. Table 2-18 summarizes the accuracy of the approximations.

In Table 2-18, the absolute value, minimum, median and the maximum for the percentage errors are listed, respectively. From the table, we can see that the approximation can mimic the performance of the original, infinite waiting room environment almost perfectly, which evinces the accuracy of our method.

average $\%|err|$	**min** $\%err$	**median** $\%err$	**max** $\%err$
0.05%	-0.14%	0.03%	0.25%

Table 2-18 Accuracy of the approximation for the TRC values

This simulation study finalizes this section for the second capacity mode. In the next section, Section 2.7, we discuss and summarize our findings in a nutshell and provide our conclusive remarks on the applicability of these capacity modes in the specialized environment and build the links to the commoditized environment which we will cover in Chapter 3.

2.7 Concluding Remarks

2.7.1 Summary

In this chapter, we studied the integrated inventory and capacity problem of a MSP who is running a repair shop and is responsible for the availability of different specialized systems in an environment, each of which contains a common critical subsystem that is prone to failure. In order to decrease the down-time costs, the repair shop keeps an inventory for the critical subsystem units, such that a failed critical subsystem unit can be replaced with a spare one immediately, if it is available. In this specialized system operating environment, the MSP is offered the possibility to close a contract with a capacity agency that allows him to make use of different sorts of capacity modes for the repair shop.

In the first capacity mode, the repair shop capacity is deployed once and its level is fixed. We use the optimal cost performance of this fixed capacity mode as a reference to assess the cost performance of the other two capacity modes.

In the second capacity mode, two-level flexible mode, the MSP is offered the possibility to close a contract with a capacity agency that allows him to periodically hire a pre-specified amount of capacity during the next period. By closing such a contract, the MSP creates a capacity volume flexibility in the repair shop. The tactical decision problem the MSP faces is setting the spare unit inventory level, the repair shop permanent capacity level, and the terms of the contract with the capacity agency (which includes the decisions on the contingent capacity level, period length and the workload in which the contingent capacity is hired).

In the third capacity mode, the failed subsystems are admitted to the repair shop periodically, which creates a level of certainty on the duration of the idle time of the repair shop capacity (from the start of an idleness to the end of a period). Due to this partial certainty, the MSP is offered the possibility to close a contract with a capacity agency that allows him to sell the idle capacity back to the provider at a reduced price until the start of the next period. By closing such a contract, the MSP is able to sell back his idle capacity to the agency, and the capacity agency assigns the idle capacity to other external tasks until the start of the next period. The tactical decision problem this MSP faces then, is setting the spare unit inventory level and the terms of the contract with the capacity agency (which includes decisions on the capacity level, and the period length).

In both second and third capacity modes, the period length plays a very central role. It is a flexibility metric in both of the capacity modes. A shorter period length enables a better tailoring of the repair shop capacity to the workload. However, in both of two-level and sell-back modes, the unit time cost of a unit of contingent capacity provided from the capacity agency is higher for shorter period lengths, due to the opportunity costs caused by the differential effects of the wages of the contingent capacity.

In the second (two-level) capacity mode, the use of the contingent capacity is decided by the repair shop at the start of each period with regard to the workload situation. Since this decision cannot be known in advance with certainty, this uncertainty on the use of the periodically provided capacity creates an economic factor that causes an opportunity cost, because that capacity could be used somewhere else if it was not reserved for that period. In the third capacity mode (capacity sell-back), the provided capacity is deployed at the repair shop at the start of each period. However, in this capacity mode, additional uncertainty factor is the time during which the provided capacity will stay at the repair shop, because the capacity will be sold back to the agency in order to be assigned to another external task, as soon as it becomes idle. This uncertainty on the duration of the deployment in the repair shop and frequency of job switching (between the repair shop and the other external tasks that the capacity agency assigns) create an economic factor that causes an opportunity cost due to the required additional skills as well as the extra cognitive load generated from task switching as well as the transportation/ transaction costs of the shop capacity. The opportunity costs are assumed to decrease with period length duration in both of the capacity modes.

We developed a decision support system for both two-level and sell-back capacity modes which integrates the down-time and capacity decisions of the MSP in order to minimize its total relevant costs. We compared the savings and the optimal policy parameters of both of these capacity modes with the optimal fixed capacity mode results first, and with each other afterwards.

For both two-level and periodic sell-back capacity modes, we analyze the performance of the MSP and develop computational approaches based on the decomposition of the overall problem in a number of sub problems that can be solved as infinite period Markov decision problems or convex optimization problems. Moreover, we performed a computational study to investigate the possible benefits of closing a contract with a capacity agency given certain values for the cost parameters (down-time costs, inventory holding costs and permanent/ contingent capacity costs) for both of the flexible capacity modes.

2.7.2 Results

2.7.2.1 Cost Savings of the Flexible (2 Level and Periodic Sell-back) Capacity Modes

In the two-level capacity mode, the results show that maximum savings (with respect to the costs under the optimal fixed capacity policy) can range from 50% to 75%. These maximum savings occur when there is no opportunity cost for the contingent capacity. The data further show that these maximum savings are quite sensitive to the holding cost rate: the higher the holding costs, the lower the percentage savings, and they are quite insensitive to the down-time costs, at least for the range of down-time costs studied.

A substantial part of the relative cost savings is maintained in case of nonzero capacity flexibility costs. As could be expected, lower cost savings, (as low as 30%), are obtained if maximum opportunity costs for the contingent capacity is high (i.e. as high as the permanent capacity costs) and is inelastic to the period length. However, for most of the nonzero opportunity cost realizations, high cost savings, in the range around 50%, can be obtained, especially if the opportunity cost has a linear cost structure and a high elasticity to period length. Most of the time, the percentage savings are the least for inverse proportional cost structure in the specialized environment.

In the periodic sell-back capacity mode, the cost-savings of the optimal periodic sell-back capacity policy seem to be much more volatile than the two-level capacity policy. The numerical study indicates that the maximum savings can vary from negative values (i.e. higher costs than fixed capacity mode) to saving values of 93%. In this mode, all of the capacity is considered as contingent, since it is assigned to other tasks as soon as the idle time starts, and these maximum savings occur when there is no opportunity cost for the contingent capacity. However, it is remarkable that even if there is no opportunity cost for the capacity, the periodic sell-back capacity mode can still perform worse than the fixed capacity mode, especially for low sell-back rates (when the sell-back rate is less than 0.1). This underperformance can be attributed to the burstiness caused by the periodic admission. Similar to the two-level capacity mode, the data from the numerical results further illustrate that the maximum savings are rather insensitive to the down-time costs, and quite sensitive to the holding cost rate. It seems that higher holding cost rates lead to lower percentage savings for the periodic sell-back capacity mode, as well.

Our decision system optimizes the decision variables in the flexible capacity contract, being the period length, and lower- and upper capacity level, to minimize the sum of down-time costs, capacity costs, and inventory costs. Our results show that there seems to exist a complex, non-intuitive relationship between the optimal control values and the cost parameters of the decision problem. If flexible capacity costs are independent of the choice of the period length, which is the case when there is no max. opportunity cost, or when there is perfect or none period length elasticity period, the system chooses the shortest period length offered and sets the capacity levels accordingly. If flexible capacity cost decrease with period length, we see that the system chooses a larger period length and operates at a lower permanent capacity level, resulting in less frequent but larger changes in capacity. However, this relationship is not strict. We observe that if flexible capacity costs strongly depend on period length, the system further decreases the permanent capacity level, but chooses a somewhat smaller period length, resulting in more frequent and still larger changes in capacity. Thus the choice of the period length of the flexible capacity contract is obvious and should be based on a careful study of its effect on total costs. Further research of this complex interaction is needed.

A last remarkable observation from our study is that, the stronger the flexible capacity costs go down with period length, the smaller the optimal base stock level tends to be. This suggests that cheaper flexible capacity results in lower stock keeping costs.

2.7.2.2 Intra-Environment, Inter-Mode Comparisons

In this section, we wrap up our findings and worked on the integrated down-time service and capacity management problem of a MSP operating in a specialized system environment under different capacity modes. First, we observe that, under the fixed capacity mode, keeping a stock of the critical subsystem and integrating the stock level decision with the capacity level decision reduces the total relevant costs of the MSP substantially, compared to the situation in which the MSP does not hold a spare unit stock but optimizes its capacity level only. The intervention of the stock level can decrease the total relevant costs up to 80%. Capacity relevant costs compose the biggest slice of the total relevant costs, therefore the cost saving prospects of the flexible capacity modes are explored. Among the investigation of the three capacity modes, it is witnessed that the optimal performance of the fixed capacity mode is surpassed by the optimal two-level capacity policy in all of the cost parameter realizations in the studied test-bed. Beyond the regular test-bed, some stress testing scenarios are conducted, and it is observed that the two-level capacity modes are exceeded by the fixed level capacity mode only after when the maximum opportunity costs become extraordinarily higher in comparison to the permanent capacity cost (e.g. 10 times or more expensive). We observe that the cost savings of the two- level capacity mode(with respect to the fixed capacity mode) in the total costs derive from both lower stock levels and lower capacity deployment.

When we have a closer look at the cost performance of the periodic sell-back capacity mode, the first characteristic that we have detected is its hazard, hazard of being the least economical capacity mode among the listed capacity modes. This cost under-performance gets especially critical when there is no or limited sell-back prospects, low period length elasticity and high opportunity costs. However, as the capacity sell-back rates increase and the opportunity costs start to lessen, the periodic sell-back capacity mode begins to save costs and it comes to be the most economical capacity mode under the full sell-back rate option when there is no opportunity cost.

When making a pair-wise comparison between the two-level and periodic sell-back capacity modes, one should pay attention on the comparability between the maximum opportunity costs in two different modes. The factors that are driving the maximum opportunity costs in two level capacity mode are different than those in the periodic sell-back capacity mode. We can posit the task insecurity of the contingent capacity as the leading compensation factor in the two level capacity mode, whereas the frequent task switching would be the main compensation factor in the periodic sell-back capacity mode. Based on these two maximum opportunity costs, if the elasticity and the other cost parameters are the same, the MSP is recommended to choose the two level capacity mode when the maximum opportunity cost in the periodic sell-back capacity is higher. When these two opportunity costs are the same, the MSP is again recommended to choose two-level capacity policy if the sell-back rate is less than 0.9. There is a threshold between 0.9 and 1, after which the periodic sell-back capacity mode beats the other two modes. If the opportunity cost of the periodic sell-back capacity mode starts to become smaller than the opportunity cost of the two-level mode, the afore-mentioned threshold value for the sell-back rate gets smaller.

In both two-level and periodic sell-back capacity modes, for given cost parameters, the total relevant costs are the minimum when the contingent capacity costs follow the linear cost structure and maximum when the contingent capacity follow the inversely proportional cost structure.

In the periodic sell-back capacity mode, there can be an implication of savings in the shipment costs due to the periodic admission of the defective subsystem units to the repair shop. This structure in the periodic sell-back capacity mode may create milk-run occasions, which facilitate the collection and produce pooling possibilities in the transportation of the defective subsystems. Other modes do not have this saving prospect since the defective subsystems are admitted continuously. Therefore, the shipment costs are necessary to incorporate in intra-environment inter-mode comparisons, especially if the shipment costs are high or is contracted to a 3PL.

3 Commoditized System Environment

3.1 Introduction

Due to the increasing technology and the effective communication mediums, the commoditization of products/processes has been accelerated. Commoditization is a process during which a non-commodity product becomes more like a commodity, for which there is demand, but is supplied without significant qualitative differentiation across a market.

In this chapter, we focus on the integration of capacity and maintenance related decisions in the (partly) commoditized system environment, where there are different types of capacity flexibility options available. The systems in this environment fall more to the right half of the commoditization continuum represented in Figure 1-1. We are especially interested in systems that are sort of *capital-goods*, which are utilized during the realization of the production and service activities for the system owners' businesses. Trucks, cranes, printers, photocopy machines, forklifts, refrigerated cargoes, computer systems, heavy duty vehicles, cooling towers, some common medical equipment (i.e. anesthesia machines, ultrasound devices, etc.) can be examples of such (partly) commoditized systems. Note that rather than being commoditized as a whole system, in some situations, it is possible that only some of the components/parts from a system can be commoditized. For instance, although a specific military aircraft as a whole cannot be considered as a commoditized system, some components of that aircraft such as the engine and the gas turbine can be considered as commoditized subsystems, since there are numerous manufacturers/suppliers of these types of components/subsystems, and the lease market of aircraft parts is quite vibrant. The modeling framework developed for the individual commoditized systems, which is described in Section 3.4, can be replicated for the commoditized components/parts, as well.

There are a number of characteristics that are particular to commoditized systems. First, a (partly) commoditized system is much less specialized than the systems under concern in Chapter 2, which probably indicates that it has been designed and built in order to meet the specifications that are shared by a much more populous number of potential end-users. Second, in the commoditized system environment, the technological and the financial barriers for a competing manufacturer to design and produce a similar system is significantly lower compared to those in the specialized system environment. This lower barrier enables the small-medium sized enterprises' (SME's) entry to the maintenance market of commoditized systems. This barrier for SME's is a lot higher for the maintenance market of specialized systems, in which semi-autonomous business units from the OEMs have much more competitive advantage. As a result, the maintenance market of the commoditized systems (forklifts, trucks, cranes, etc.) is more fragmented than the maintenance market of the non-commoditized systems. The characteristics listed above have important consequences, particularly in the realm of system availability. For instance, a higher number of potential end-users and the presence of numerous alternative manufacturers both contribute to the increase in demand for both long and short-term use of the commoditized systems. This leads to an increase in the short term supply/operating lease options, which enhances the short-term availability options for the commoditized system under concern. In our research, we assume that the short-term availability is realized through rental/other 3rd party supply channels in the market. These 3rd party suppliers can immediately provide a substitute of a (partly) commoditized system/commoditized component with identical functionality via short-term supply agreements (via rental/leasing).

In this chapter, we consider that a MSP takes care of the repair and the availability of many commoditized systems installed in a region, in exchange for a service fee. In order to realize the repair process, the MSP operates a repair shop. In order to improve the overall system availability, rather than keeping a stock of commoditized systems, the MSP makes use of a new down-time service strategy, namely, *hire upon failure* strategy, which is begotten from the improved short-term availability of the commoditized systems.

In this strategy, it is assumed that a short term supply agreement is signed between the MSP and one of the 3rd party supply channels, where the supply channel agrees to provide a substitute system, upon a system failure, for a fixed duration, at a constant price rate. These sorts of arrangements are common for industrial equipment via operating lease agreements, where the equipment is acquired on a short-term basis. Some of the advantages of such a short-term hiring strategy include: less initial investment compared to the keeping spare stocks for the substitute systems (requires buying the substitute systems first), improved cash flow and tax-deductibility, since the hiring costs for the substitute are considered as an operating costs. This is a very attractive option for the players in the fragmented commoditized maintenance market, since most of the MSP's in the commoditized market are also SME's, and the initial investment to build its own stock of the spare systems, in order to sustain the availability during the down-times, can be non-affordable for a SME, due to the scale and the cash-flow limitations.

The capacity of the repair shop determines the speed of the repair process of a failed system. Similar to Chapter 3, our objective in this chapter is to minimize the total relevant costs (TRC) of the MSP, which consists of the three components listed below with their abbreviations in parentheses:

1. Capacity related costs of the repair shop (CRC)

2. Down-time costs of a system that is failed and not supplied with a substitute system from the supply channel anymore since the repair of that system is still not complete when the hiring duration is expired (DTC)

3. Hiring costs for the substitute commoditized system (HC)

Given the cost components above, the MSP takes the capacity and down-time service related decisions simultaneously in order to minimize its TRC.

Similar to Chapter 2, the MSP can make use of periodic capacity flexibility options while integrating its repair shop capacity and down-time service related decisions. The reasons behind the periodicity of capacity flexibility were already discussed in Chapter 1. Also the structure of flexible capacity options and the actors involved were already described in detail in Chapter 2.

In Figure 3-1, we depict the relations between the repair shop, the capacity agency (can be either internal or external), rental channel and the commoditized systems that are prone to failure.

In the light of the discussions above, three capacity modes are investigated in this chapter. These modes are:

1. Fixed Capacity Mode (Reference)
2. Two-Level Flexible Capacity Mode
3. Periodic Sell-back Capacity Mode

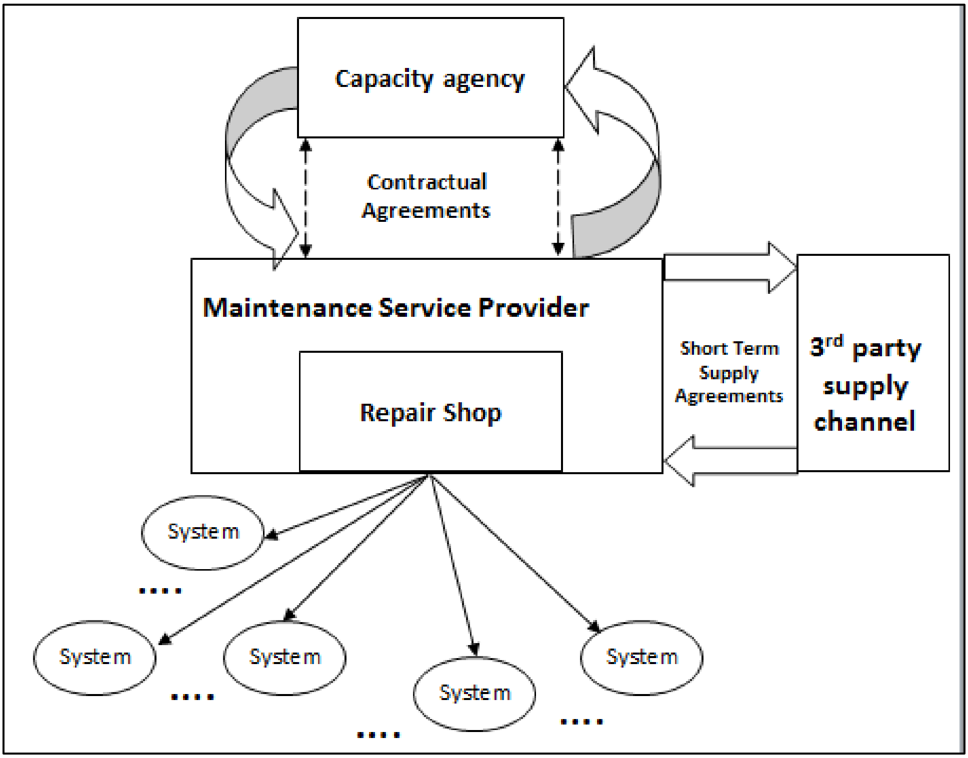

Figure 3-1: The relations between the maintenance service provider, the capacity agency, the rental channel and the commoditized systems through contractual agreements.

In each of the three modes above, the capacity related decisions and the capacity cost structures are different. We aim to model the maintenance service network for commoditized systems, which embodies all the active/passive actors listed above in order to analyze the interplays between the 3rd party supply channel, the capacity agency and the repair shop. With the analysis, we can derive the cost performance characteristics and develop a decision support system that integrates the capacity and the down-time service related decisions in order to minimize TRC of the MSP under different capacity flexibility options. In addition, the developed modeling framework in this chapter enables the researchers/practitioners to foresee how much cost savings can be realized through the use of capacity flexibility compared to the best practice under the fixed capacity setting.

The outline of this chapter is as follows. In Section 3.2, we provide a brief literature review about the rental environment models in the operations management literature. The details of the structure of the short term supply/rental agreement and its justifications are explained in Section 3.3. In Section 3.4, we model and analyze the integrated decision making problem under the fixed capacity mode, which serves as a reference model for the further modes. In Section 3.5 and in Section 3.6 we explain, model and analyze the same problem framework under two-level flexible capacity and periodic sell back capacity modes, respectively. Finally in Section 3.7, we draw overall conclusions over the performance of capacity modes, interpret the differences and finalize this chapter.

3.2 Literature Review

In Subsection 1.1.5, the studies on commoditization literature have already been given. Recall that we have enclosed that subsection, by reasoning how the commoditization process may lead to the *rentalization* process, where the rental availability of a system (or a substitute) becomes more common, more widely reachable and more economic. The consequences of the rentalization process have mainly shaped the reasoning behind our model. Therefore, in this section, we review the relevant literature on rental systems and their implications on the operations. Note that we exclude studies on leasing, and focus on short-term rental in our literature search.

The early studies for rental systems include (Tainiter 1964), which uses a queuing framework to model rental situations by showing the correspondence between service systems and queuing loss systems. Later, (Whisler 1967) formulates a discrete-time, finite-horizon, dynamic programming model to find a cost optimal policy for the rental equipment inventory.

Most of the latter studies focused on the operations of rental stores in specific sectors. These specific sectors include vehicle/car rental and video rental stores.

In the video rental industry, we mostly observe a vertically separated structure (including a rental store and a DVD/cassette supplier) and a non-stationary and temporal demand pattern, where the rental demand for a new cassette/DVD is very high initially but then declines significantly over time. Due to this demand behavior, the major problem of the video rental systems is to deal with the left over cassettes after the peak season has elapsed. Therefore many studies in the literature focused on contractual agreements between the rental store and the DVD/cassette supplier. It has been shown several times that price/revenue sharing contracts are coordinating and create a win-win situation for both of the parties in various settings (see e.g. (Gerchak et al. 2006), (van der Veen & Venugopal 2005), (Dana Jr & Spier 2001)).

Different from video rental stores, in vehicle/car rental systems, most of the time, the size of the rental fleet size appears to be the biggest decision problem. (Parikh 1977) uses a queuing system for modeling rail car rental systems by using an $M/G/c$ queue with backorders to determine the steady-state distribution of the number of rail cars in use and the optimal fleet size and structure under a service-level constraint. (Turnquist & Jordan 1986) address the fleet-sizing problem for a container network by assuming a deterministic demand but allowing for stochastic transportation times. (Du & Hall 1997) approach the fleet-sizing problem from an inventory theory perspective by analytically deriving the optimal fleet sizes and a method of balancing empty and loaded flows in a transportation network. (Love 1985) develops a source/sink type of inventory control for vehicle rental systems, where the rented vehicle could be returned to another location than the rental store. In (Geraghty & E. Johnson 1997), an implemented revenue management project in National Car Rental is described, where the revenue of National Car Rental is increased drastically. In (Pachon et al. 2003) a tactical planning model for the vehicle fleet management is developed, where the daily tactical plan is further decomposed into a fleet deployment and a transportation model. In (Savin et al. 2005) the capacity management of a rental system is studied through a loss system with two customer classes, and where the optimal capacity rationing decision is studied under Poisson arrival and exponentially distributed service times assumption. Later, (Papier & Thonemann 2008) develop and solve analytical models for fleet planning of a rail car rental system. In a later study, (Papier & Thonemann 2009), the use capacity rationing in a stochastic rental system in the presence of advanced demand information (ADI) is investigated by using stochastic dynamic programming techniques.

Besides the studies mentioned above, there is a stream of literature on closed-loop supply chains, where the quantitative formalisms that are used to model re-manufacturable product supply chain can be also applicable to model the rental systems. We refer the interested reader to (Fleischmann 1997), for a comprehensive literature review.

We have found only a few papers that analyze generic, sector non-specific rental situations. Among these, we are particularly interested in (Tang & Deo 2008), where the problem on determining the optimal rental price and duration is studied under retail competition. Motivated from many observations in the industry, they assume pre-specified, certain (fixed) rental duration to be decided upon. In this chapter, we also follow that approach and assume that the rental duration is fixed and to be determined in the rental agreement. We do not consider competition among different rental stores, but rather focus on a setting where the rental duration is a part of the integrated decision making process for the MSP. The further details of the structure of the rental agreement and its justifications are explained in the next chapter.

3.3 Short Term/Rental Substitute Provision Mechanism

In this section we detail out the short term rental substitute provision mechanism. Note that in this section, where we analyze the maintenance network of the commoditized systems, the MSP's agreement with a 3rd party/rental supplier takes the place of the strategy of keeping a spare stock for systems/subsystems. Similar to the capacity agency, it is assumed that the 3rd party supplier is a reactive agent in the decision making process. Upon a part/ system failure, the 3rd party supplier provides a substitute system for a certain duration for a fixed hiring rate.

In the industry, we see that a lot of companies provide integrated solutions by integrating rental/lease & repair services for their customers (for example see www.crepa.nl or www.dremed.com). In practice, many of the rental suppliers may allow the customer to specify the rental duration at each rental instance, according to their needs. This is for sure a more convenient solution for a one-time customer.

However, due to a number of reasons listed below, a rental supplier may choose to rent its products for a fixed, specified duration, even possibly at a reduced price, to some of its contracted, long-term customers. Firstly, fixed rental duration increases the utilization of the rental units by effectuating a better control of the return processes. Secondly, the long term contracted customer may require a timely service, in particular, no occurrence of stock-outs/delays, which may trigger the rental retailer to invoke upon some rationing policies within its rental unit stock. In such a case, a fixed rental duration would eliminate the rental duration variance for the corresponding units, and would eventually decrease the necessary number of allocated rental units for the contracted customer as well as the number of possible emergent measures in case of stock-out instances. Thirdly, if the transportation tasks of the rental units belong also to the rental supplier, fixed rental duration provides certainty about the returning time, therefore facilitates the planning of collecting items and may lead to savings in transportation costs. Finally, a uniform rental duration is also attractive for the MSP, because in that situation, the rental duration is decided in the beginning, once and for all, which would drastically reduce the decision making burden on the MSP.

Owing to the reasons above, we assume that the rental supplier agrees to provide a substitute upon each system failure for a fixed duration, L , at a fixed hiring cost rate h_r . In leasing environments, it is often witnessed that the leasing rate decreases with the leasing duration. However, in our thesis, we assume that L is considerably short compared to the capital leasing durations and therefore we assume that h_r is constant no matter how short/long the hiring duration L is. The short-term supply agreement brings liabilities for both parts. Upon a system failure, the 3rd party supplier has to provide a substitute system to the MSP immediately. This brings the responsibility of the substitute system availability upon a failure to the 3rd party supplier, which may necessitate to keep more rental units in the store and also to take other emergent actions upon stock-out instances.

Upon the receipt of the substitute, the MSP transfers the rental substitute to the failed system's location immediately. On the other hand, after L time units, the MSP must return back the hired substitute system, irrespective of whether the repair of the corresponding failed system is completed or not. Therefore, the MSP has to consider an additional safety time while deciding on the uniform hiring duration L.

The role of the hiring duration L for the system availability in the commoditized environment is analogous to that of the spare part stock level S in the specialized system environment. In the former environment, as soon as a system fails, the failed system is sent to the repair shop and a substitute system hired from the 3rd party supplier immediately resumes the system owners' operations. The hired substitute remains in use for another L time units. If the repair of the failed system is not finished after L time units, the system owners' operations are halted until the repair completion. Therefore, after a failure, in this strategy, it is only possible to have a down-time after the rental duration is elapsed. On the other hand, in the specialized system environment, upon a system failure (due to the critical subsystem failure), if there is a spare unit in the stock, it is immediately sent to the failed system location to replace the failed critical unit so that the operations of the system owner can continue. If there is no spare unit available in the stock, the demand for this unit is backordered, and the operations are halted until the repair of that backordered unit is completed and sent to the failed system's location. Therefore keeping a spare unit inventory doubtlessly improves the system availability; however the randomness of the system availability upon a failure instance still remains. On the contrary, by having a fixed supply agreement with a 3rd party supplier, upon a system failure we have guaranteed system availability for another L time units after the failure.

Having discussed the rental provision mechanism, in the next section, we start with analyzing the fixed capacity mode in the (partly) commoditized environment. In the next section, the capacity is provided to the repair shop indefinitely and the minimum cost performance achieved will act as a reference point to judge the performance of the service providers using flexibility options in further modes.

3.4 Fixed Capacity Mode

In this section, we analyze the integrated decision making problem of the service provider under the fixed capacity mode, where all of the capacity is permanent and ready for use in the repair shop. Recall that different from Chapter 2, the MSP does not hold a stock of spare parts, but rather hires a substitute for the failed system upon a failure for a fixed hiring duration. This mode serves as a reference point for the other two capacity modes, necessary to assess the benefits of further flexibility options.

The MSP has to determine the optimal hiring duration and capacity level decisions in order to minimize its TRC. This section aims at building a modelling framework and a decision support system for the service provider operating under the first capacity mode. Therefore in Subsection 3.4.1 we present the model, assumptions and the problem formulation. In Subsection 3.4.2, the derivation of the total relevant cost per unit time as well as the analytical properties and the optimization procedure are given. Finally in Subsection 3.4.3, we describe the experimental setting and provide the results of the numerical study.

3.4.1 Model, Assumptions and Problem Formulation

We analyze an environment, where a service provider operates a repair shop, and is responsible for the repair of the failed systems, from a population that involves numerous (partly) commoditized systems installed in a region. Similar to the specialized system environment, in this chapter we also assume that the systems are in the exploitation phase, therefore the failure occurrences are stationary and we also further assume that the number of the systems that the service provider is responsible for, N, is quite large, which justify the modeling approach, in which the failures come from an infinite population of systems and follow a Poisson process with a constant rate λ.

In this chapter, the service provider does not keep a spare item stock for critical systems, but makes an agreement with a rental/ 3rd party supplier. Upon a system failure, the following procedure is applied. Immediately after the failure, a substitute system is sent from the rental supplier to the failed system's location and the failed system is shipped to the repair shop. After L units of time, the rented substitute system is returned back to the 3rd party supplier. If the repair of the failed system is finished before L units of time, the operations of the system owners can continue without any interruption. On the other hand, if the repair of the failed system is not finished, the system owners' operations are halted until the repair of the failed system is completed.

We assume that the replacement and the transportation times from the repair shop/rental store to the customer sites where the systems are located (or vice versa) are negligible. Each defective system requires an exponentially distributed service time from the repair shop and the defective systems that require repair have to wait for their turn in order to get serviced in the repair shop. We model the repair shop as a single server Markovian queue. The capacity of the repair shop determines the speed of the repair service. Therefore the processing rate μ is considered as the capacity level of the repair shop.

In this thesis, we suppose that all the defective systems can be restored to the as good as new condition after the repair.

The capacity cost per unit time is c_p in this fixed capacity mode, since all the repair shop capacity is permanent. We pay h_r per unit time for each substitute system in use. The down-time costs due to the halted operations of the system owner is equal to B per time unit, and we assume that $B > h_r$. From now on, we use the notation of $\mathbf{C}$ to denote the capacity policy. In the fixed capacity mode, $\mathbf{C}$ is a single variable, since the only capacity related decision is the processing rate μ. The rental duration related decision is L. The total relevant cost function, TRC, can be represented by $\mathbf{C}$ and L, and is the sum of capacity related costs $\left(CRC(\mathbf{C})\right)$, down-time costs $\left(DTC(\mathbf{C}, L)\right)$ and hiring related costs $\left(HC(L)\right)$. Given these cost components and the decision variables, the problem of the MSP can be formulated as follows:

$$\min_{\mathbf{C},L} TRC(\mathbf{C}, L) = CRC(\mathbf{C}) + DTC(\mathbf{C}, L) + HC(L)$$
$$s.t. \;\; L > 0$$
$$\mathbf{C} = \mu > \lambda \tag{3.1}$$

Given the problem formulation above, in the next subsection, we derive the necessary cost functions used in (3.1), give the analytical properties of $TRC(\mathbf{C}, L)$, and present the optimization procedure for the problem.

3.4.2 Derivation and Analysis of the Cost Functions, the Solution Procedure

In this subsection, we first derive and provide the analytical properties of the cost functions used in (3.1). Afterwards, we give the optimal decision variables that minimize total relevant costs. As it is mentioned previously, $TRC(\mathbf{C}, L)$ consists of three cost components: $CRC(\mathbf{C})$, $DTC(\mathbf{C}, L)$ and $HC(L)$.

The capacity related cost per unit time is a linear function of μ, since the capacity policy is solely the processing rate and per time unit cost of it is constant and equal to c_p. Similar to the previous analyses, we exclude the baseline capacity costs: $c_p\lambda$. Therefore, we have $CRC(\mathbf{C})$ = $c_p(\mu - \lambda)$. Similarly, the rental hiring cost per unit time is also a linear function of the hiring duration L, since upon each failure, a substitute system is hired from the rental supplier for L units of time. Hence, we have: $HC(L) = \lambda \, L \, h_r$.

Per time down-time related cost, $DTC(\mathbf{C}, L)$ is closely related to the sojourn time distribution of a defective system in the repair shop. Let S_d denotes the sojourn time of a defective system in the repair shop and $P\{S_d > x|\mathbf{C}\}$ denotes the probability that the defective system will spend more than x time units in the repair shop under the capacity policy $\mathbf{C} = \{\mu\}$, when $\mu > \lambda$. Since we model the repair shop as an $M/M/1$ queue, S_d is exponentially distributed with a rate of $\mu - \lambda$. Hence, we have the following:

$$P\{S_d > x|\mathbf{C}\} = e^{-(\mu-\lambda)x}, \qquad x > 0 \tag{3.2}$$

Let $f_{S_d}(x|\mathbf{C}) = \frac{\partial(1-P\{S_d>x|\mathbf{C}\})}{\partial x}$. The expected down-time that the owner of a failed system will face due to the elapse of rental duration L, given capacity μ can be found from (3.2) as follows:

$$\mathrm{E}((S_d - L)^{+}|\mathbf{C}) = \int_{x=L}^{\infty} (x - L)\, f_{S_d}(x|\mathbf{C})\mathrm{dx} = \int_{x=L}^{\infty} (x - L)\, (\mu - \lambda)e^{-(\mu-\lambda)x}dx$$
$$= \frac{e^{-(\mu-\lambda)L}}{(\mu - \lambda)} \tag{3.3}$$

Each system that is down due to the failure of a system in concern is incurred B per unit time. Hence, the down-time related cost per unit time can be found as follows:

$$DTC(\mathbf{C}, L) = \lambda B\big(\mathrm{E}((S_d - L)^{+}|\mathbf{C})\big) = \lambda B \frac{e^{-(\mu-\lambda)L}}{(\mu - \lambda)}$$

Hence we have the following:

$$TRC(\mathbf{C}, L) = CRC(\mathbf{C}) + DTC(\mathbf{C}, L) + HC(L) = c_p(\mu - \lambda) + \lambda\left(h_r L + B\frac{e^{-(\mu-\lambda)L}}{(\mu - \lambda)}\right) \tag{3.4}$$

As the total relevant cost rate, $TRC(\mathbf{C}, L)$, is derived, next we give some of its analytical properties.

Property 3.1

For a given $\mu > \lambda$, $TRC(\mathbf{C}, L)$ is strictly convex in L.

Proof:

If $\mu > \lambda$, we have: $\frac{\partial TRC(\mathbf{C},L)}{\partial L} = \lambda h_r - \lambda B e^{-(\mu-\lambda)L}$ and $\frac{\partial^2 TRC(\mathbf{C},L)}{\partial L^2} = \lambda B(\mu - \lambda)e^{-(\mu-\lambda)L} > 0$

Property 3.2

For a given $L \geq 0$, $TRC(\mathbf{C}, L)$ is strictly convex in μ.

Proof:

For $L \geq 0$, we have: $\frac{\partial TRC(\mathbf{C},L)}{\partial \mu} = c_p - \lambda B\left(\frac{e^{-(\mu-\lambda)L}}{(\mu-\lambda)^2} + \frac{Le^{-(\mu-\lambda)L}}{(\mu-\lambda)}\right)$ and $\frac{\partial^2 TRC(\mathbf{C},L)}{\partial \mu^2} = \lambda B\left(\frac{2\times e^{-(\mu-\lambda)L}}{(\mu-\lambda)^3} + \frac{2Le^{-(\mu-\lambda)L}}{(\mu-\lambda)^2} + \frac{L^2 e^{-(\mu-\lambda)L}}{(\mu-\lambda)}\right) > 0.$

Property 3.3

$TRC(\mathbf{C},L)$ is jointly convex in both L and μ.

Proof:

We know that in order to prove the joint convexity of $TRC(\mathbf{C},L)$, we need to show that both of the eigenvalues of the Hessian matrix of $TRC(\mathbf{C},L)$, $\mathrm{H}\big(TRC(\mathbf{C},L)\big)$ are non-negative. We have:

$$\mathrm{H}\big(TRC(\mathbf{C},L)\big)=\begin{bmatrix}\frac{\partial^2 TRC(\mathbf{C},L)}{\partial L^2} & \frac{\partial^2 TRC(\mathbf{C},L)}{\partial L\partial\mu}\\ \frac{\partial^2 TRC(\mathbf{C},L)}{\partial\mu\partial L} & \frac{\partial^2 TRC(\mathbf{C},L)}{\partial\mu^2}\end{bmatrix}=$$

$$\lambda B\begin{bmatrix}(\mu-\lambda)e^{-(\mu-\lambda)L} & Le^{-(\mu-\lambda)L}\\ Le^{-(\mu-\lambda)L} & \left(\frac{2e^{-(\mu-\lambda)L}}{(\mu-\lambda)^3}+\frac{2Le^{-(\mu-\lambda)L}}{(\mu-\lambda)^2}+\frac{L^2e^{-(\mu-\lambda)L}}{(\mu-\lambda)}\right)\end{bmatrix}$$

We know that if $\frac{\partial^2 TRC(\mathbf{C},L)}{\partial L^2}\times\frac{\partial^2 TRC(\mathbf{C},L)}{\partial\mu^2}\geq\frac{\partial^2 TRC(\mathbf{C},L)}{\partial L\partial\mu}\times\frac{\partial^2 TRC(\mathbf{C},L)}{\partial\mu\partial L}$, then $\mathrm{H}\big(TRC(\mathbf{C},L)\big)$ is a positive semi-definite matrix (Horn & Johnson 1985). In our case, the above inequality holds, therefore $H\big(TRC(\mathbf{C},L)\big)$ is positive semi-definite, therefore both of the eigenvalues are non-negative.

Property 3.4

For a given population failure rate λ, and cost parameters B, h_r and c_p, we have the optimal rental hiring duration $L^*=\frac{-\ln\left(\frac{h_r}{B}\right)}{\sqrt{\frac{\lambda h_r\left(1-\ln\left(\frac{h_r}{B}\right)\right)}{c_p}}}$ and optimal capacity level $\mu^*=\lambda+\sqrt{\frac{\lambda h_r\left(1-\ln\left(\frac{h_r}{B}\right)\right)}{c_p}}$.

Proof:

For a given capacity level μ, we can find the optimal hiring duration for that given capacity level, $L^*(\mu)=\frac{-\ln\left(\frac{h_r}{B}\right)}{(\mu-\lambda)}$ which is the L that satisfies: $\left.\frac{\partial TRC(\mathbf{C},L)}{\partial L}\right|_{\mathbf{C}=\mu}=0$.

Similarly we obtain $\frac{\partial TRC(\mathbf{C},L)}{\partial\mu}=c_p-\lambda B\left(\frac{e^{-(\mu-\lambda)L}}{(\mu-\lambda)^2}+\frac{Le^{-(\mu-\lambda)L}}{(\mu-\lambda)}\right)$. If we plug $L^*(\mu)$ into this function and solve the capacity level μ that satisfies $\frac{\partial TRC(\mathbf{C},L)}{\partial\mu}=0$, then we obtain $\mu^*=\lambda+\sqrt{\frac{\lambda h_r\left(1-\ln\left(\frac{h_r}{B}\right)\right)}{c_p}}$ as well as $L^*=L^*(\mu^*)=\frac{-\ln\left(\frac{h_r}{B}\right)}{\sqrt{\frac{\lambda\times h_r\times\left(1-\ln\left(\frac{h_r}{B}\right)\right)}{c_p}}}$.

In the next subsection, we present the results of the numerical study that is conducted, where the optimal costs for problem (3.1) are obtained from the analytically derived optimal parameters L^* and μ^* that are introduced above.

3.4.3 Numerical Study

In this section, we present the results of the numerical study for the fixed capacity mode. First, we describe the experimental design for the computational study. Afterwards, we obtain the optimal decision parameters for every instance in the test bed. Finally, the minimum cost performances, the optimal capacity levels and the optimal hiring durations are given in order to generate managerial insights and form a basis as a reference point to assess the benefits of capacity flexibility in further capacity modes.

3.4.3.1 The Base Case Scenario and the Experimental Design

In our computational study, we take the unit time as a week and normalize the mean arrival rate for the system failures (not from one system but the cumulative failures in the whole environment) $\lambda = 1$ (failures per week). We have a base case scenario, which is described below, and the other 8 scenarios have varying backorder (B) and hiring (h_r) costs per unit time. The parameter values in the base case scenario are founded on the following situation:

Suppose that the capital good has a value of €250.000. This amount is lower than the value in the previous specialized system environment. The capital good is used in the production process of other products. The economic lifetime of the capital good is assumed to be 10 years, and the cost of the capital good represents 20% of the total costs of the products produced with it (material costs deduced). Further, suppose that the firm sells the products at a price that is 3 times the total production costs (material costs deduced) accumulated during the average machining time used to produce the products (when the capacity of the capital good is used). If the capital good is in use for 250 days per year and 8 hours per day, then the capital good related costs are $\frac{€250.000}{(250\times8\times10)} = €12.5$ per hour and the lost revenue due to down-time is: $\left(\frac{€12.5}{0.2}\right)\times 3 \cong €187.5$ per hour. A week (the base time unit) of down-time costs would be $40 \times €187.5 = €7500$. For the cost of the workforce capacity of the repair shop, we will use a wage of €60 per hour per operator and we assume that a repair of a failed subsystem takes about 60 hours. Then the repair of a failed unit/subsystem on average costs €3600. This is much less than the cost price of the system(€250.000), therefore repairing a defective system is a more economical option than scrapping the defective system and buying a new one.

Next, we derive the hiring cost per unit time. We assume that the 3rd party supplier adds on a 50% premium on top of the capital good related costs. Therefore, the hiring cost per week is equal to: $\frac{1.5\times250.000}{52\times10} \cong €\ 721.15$ per week.

We scaled the parameter for the cost of workforce per repair to one, and expressed the values for the down-time and stock keeping costs as a multiple of this normalized parameter (for instance: $c_p = \frac{3600}{3600} \cong 1$, $B = \frac{7500}{3600} \cong 2$ and hiring the substitute system for a week, $h_r = \frac{721.15}{3600} = 0.2$).

c_p	h_r	B/h_r
1	0.1	5
		10
		20
	0.2	5
		10
		20
	1	5
		10
		20

Table 3-1: Values of the analyzed c_p, h_r and B instances.

Having described the cost setting in the base case scenario, where $c_p = 1$, $h_r = 0.2$, and $B/h_r = 10$, we create the test bed, which consists of a total of 9 scenarios. These different scenarios explore the effects of different $h_r$and B/h_r ratios around the base case, and we assume that both B and c_p are higher than the hiring cost h_r per unit time. The h_r ratios range from 0.1 to 1 and B/h_r ratios range from 5 to 20. The values of B and c_p instances that are examined are given in Table 3-1. The base-case scenario that is described above is highlighted.

3.4.3.2 Results

From the analysis and the properties given in the previous subsection, we can find the optimal policy parameters, μ^* and L^* for each of these nine scenarios and the resulting total relevant costs per unit time. From now on, we use the notation of ${TRC_{Fr}}^*$, ${L_{Fr}}^*$ and ${\mu_{Fr}}^*$ in order to denote the minimum cost rate, optimal hiring duration and the optimal capacity under the first (fixed) capacity mode in the commoditized environment where the rental availability of a substitute system is present. In Table 3-2, ${L_{Fr}}^*$, ${\mu_{Fr}}^*$ and ${TRC_{Fr}}^*$ values are tabulated for all 9 scenarios, when we have $c_p = \lambda = 1$:

	$h_r = 0.1$			$h_r = 0.2$			$h_r = 1$		
	$L_{Fr}{}^*$	$\mu_{Fr}{}^*$	$TRC_{Fr}{}^*$	$L_{Fr}{}^*$	$\mu_{Fr}{}^*$	$TRC_{Fr}{}^*$	$L_{Fr}{}^*$	$\mu_{Fr}{}^*$	$TRC_{Fr}{}^*$
$B/h_r = 5$	3.15	1.511	1.022	2.23	1.722	1.445	1.00	2.615	3.231
$B/h_r = 10$	4.01	1.575	1.149	2.83	1.813	1.625	1.27	2.817	3.635
$B/h_r = 20$	4.74	1.632	1.264	3.35	1.894	1.788	1.50	2.999	3.998

Table 3-2: The optimal rental hiring duration ($L_{Fr}{}^*$) and the capacity level ($\mu_{Fr}{}^*$) decisions and the resulting costs ($TRC_{Fr}{}^*$) for the total of 9 scenarios, when $c_p = \lambda = 1$.

From Table 3-2, as it is expected from the analytical representation, we can observe that the minimum total relevant cost, $TRC_{Fr}{}^*$, increases with B and h_r . Upon an increase in B, the system responds with a higher rental duration $L_{Fr}{}^*$, in order to increase the system availability ; and with a higher capacity $\mu_{Fr}{}^*$, in order to provide a faster repair service to balance the negative effects of the increasing down-time costs.

On the other hand, upon an increase in h_r, we can see that the MSP has a greater incentive to reduce its hiring duration to save from the hiring related costs. In order to remedy for the further cost consequences of shorter rental hiring durations (which would automatically lead to higher down-time costs), the service provider increases its capacity level.

Integration of the hiring duration and capacity level decisions brings drastic savings for the MSP. For instance, in the no-hiring case, where $L = 0$, the repair shop capacity is the only parameter that the MSP can tune in order to adapt to different operating environments. This leads to inflated capacity levels and therefore very much higher costs. Let $TRC_{Fr,0}{}^*$ denotes the optimal total relevant costs in the no-hiring case ($L = 0$). In Table 3-3 , we demonstrate the percentage cost savings $100 \times \left(\frac{TRC_{Fr,0}{}^* - TRC_{Fr}{}^*}{TRC_{Fr,0}{}^*}\right)$ and savings in capacity level $100 \times \left(\frac{\mu^*(0) - \mu^*(S^*{}_{Fr})}{\mu^*(0)}\right)$ due to the integration of the rental hiring option to the capacity decision in all 9 scenarios.

	$h_r = 0.1$		$h_r = 0.2$		$h_r = 1$	
	$\frac{\mu^*(0)-\mu^*{}_{Fr}}{\mu^*(0)}$	$\frac{TRC_{Fr,0}{}^*-TRC_{Fr}{}^*}{TRC_{Fr,0}{}^*}$	$\frac{\mu^*(0)-\mu^*{}_{Fr}}{\mu^*(0)}$	$\frac{TRC_{Fr,0}{}^*-TRC_{Fr}{}^*}{TRC_{Fr,0}{}^*}$	$\frac{\mu^*(0)-\mu^*{}_{Fr}}{\mu^*(0)}$	$\frac{TRC_{Fr,0}{}^*-TRC_{Fr}{}^*}{TRC_{Fr,0}{}^*}$
$B/h_r = 5$	11%	28%	14%	28%	19%	28%
$B/h_r = 10$	21%	43%	25%	43%	32%	43%
$B/h_r = 20$	32%	55%	37%	55%	45%	55%

Table 3-3 Percentage cost and capacity savings of integrated decision making (hiring duration and capacity) compared to the no hiring case in 9 different scenarios.

As it can be seen, incorporating the use of the hiring option in order to ameliorate the down-time, reduces the total relevant costs between 28% to 55% and reduces the capacity levels between 19% to 45%. The percentage savings of the costs (due to the hiring of a substitute in the integrated decision making framework) are lower than the savings in the specialized system environment, which were displayed in Table 2-4 This change in savings arise from the different cost parameters used in the test bed for the commoditized system environment, as the c_p/h ratio in the specialized system environment is higher than the c_p/h_r ratio in the commoditized setting.

It is remarkable that the percentage cost savings of integrated decision making is insensitive to h_r when B/h_r is kept constant. This can be explained by the fact that when B/h_r is kept constant, both $TRC_{Fr}{}^*$ and $TRC_{Fr,0}{}^*$ increase with h_r multiplicatively at the same rate, and during the calculation of the percentage difference, the effect of c_p cancels out. Similarly, the increase of the savings with B under the same h_r can be explained due to the fact that the $\left(TRC_{Fr,0}{}^* - TRC_{Fr}{}^*\right)$ increases faster with B compared with $TRC_{Fr,0}{}^*$ alone.

The breakdown of the total costs under the optimal fixed capacity policy can provide us further managerial insights. Therefore, we investigate how much each of the three components (HC: hiring rental cost for a substitute system, CRC: capacity related costs, DTC: down-time related costs) has contributed to the total relevant costs under the optimal capacity & stock level decisions for 9 different B/h_r & h_r combinations. The percentage contributions of each cost component can be seen from the pie charts given in Figure 3-2.

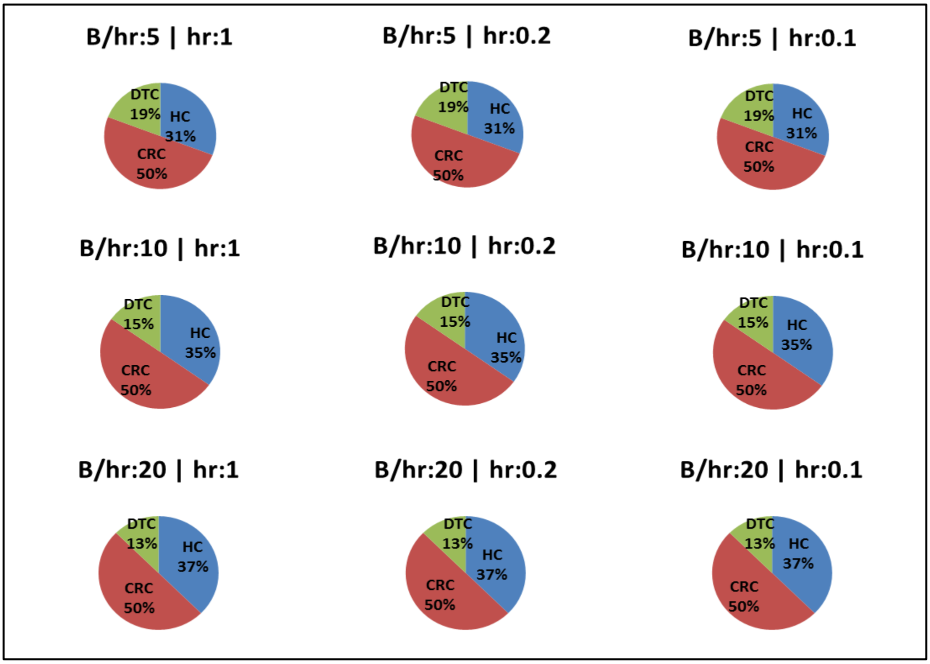

Figure 3-2: The breakdown of total relevant costs ($TRC_{Fr}{}^*$) to three cost components (HC, DTC and CRC) under the optimal capacity & rental hiring duration (**C** = $\mu_{Fr}{}^*$, $L_{Fr}{}^*$) decisions for 9 different (B/h_r & h_r) combinations.

It is remarkable that under the fixed capacity mode in the commoditized system, the ratio of capacity related costs to TRC is always 50% under the optimal capacity and the optimal rental hiring duration decisions. In addition, we also observe that the cost breakdown of the total costs does not change with different h_r, when B/h_r remains constant.

For higher B values, we observe that the $DTC\%$ in TRC decreases and $HC\%$ in TRC increases. We can also deduce these general behaviors from the $TRC(\mathbf{C}, L)$ formula in (3.4) when we plug the optimal $L^* = \frac{-\ln\left(\frac{h_r}{B}\right)}{\sqrt{\frac{\lambda h_r\left(1-\ln\left(\frac{h_r}{B}\right)\right)}{c_p}}}$ and optimal $\mu^* = \lambda + \sqrt{\frac{\lambda h_r\left(1-\ln\left(\frac{h_r}{B}\right)\right)}{c_p}}$ from Property 3.4, we obtain $TRC(\mathbf{C}^*, L^*) = \sqrt{(\lambda h_r c_p)}\left(\sqrt{1-\ln\left(\frac{h_r}{B}\right)} + \frac{1}{\sqrt{1-\ln\left(\frac{h_r}{B}\right)}} + \frac{-\ln\left(\frac{h_r}{B}\right)}{\sqrt{1-\ln\left(\frac{h_r}{B}\right)}}\right)$.

In this equation above, the first term in the parentheses corresponds to the CRC, the second to the DTC and the last term to the HC.

Each cost result tabulated in Table 3-2 serves us as a reference point to assess the prospects of the flexibility options in the two other capacity modes. Similar to the analysis in Chapter 2, for each of the 9 scenarios, in order to consider another flexible capacity policy as an alternative, the total relevant costs resulting from that policy must be smaller than the corresponding ${TRC_{Fr}}^*$ value that is tabulated in Table 3-2. In the next section, we set out our analysis to scrutinize the cost saving possibilities in the second, namely two-level, flexible capacity mode.

3.5 Two-Level Flexible Capacity Mode

In this section, we analyze the integrated decision making problem of the MSP under the second capacity mode. This two-level flexible capacity mode is same as the capacity mode described in Section 2.5, where a part of the capacity is permanent (or, in other words, the capacity agency supplies that amount of capacity indefinitely), whereas the other part is contingent, supplied on demand periodically at equidistant points in time at a higher cost rate. The decision on the use of the contingent capacity is given at each equidistant point with regard to the present workload of the repair shop, in terms of the number of defective systems in the service or waiting for service.

Similar to the fixed capacity mode, the MSP has to decide on the optimal rental hiring duration and the capacity level in order to minimize its TRC. Characterization and the analysis of the cost savings of flexibility due to the integration of the capacity and maintenance related decisions under the duality of the capacity provision, have our utmost priority in this section. Therefore we aim at building a modeling framework and a decision support system for the service provider operating under the two-level flexible capacity mode.

Since the capacity mode is identical to the mode described in Section 2.5, we skip re-introducing the capacity mode, but briefly mention about the model and the decision problem of the service provider. In Subsection 3.5.2, the derivation and the analysis of the total relevant cost per unit time as well as the sojourn time distribution in the repair shop are provided. Finally in Subsection 3.5.3, we describe the experimental setting, the search procedure and provide the results of the numerical study with a focus on the sensitivity of cost/policy parameters and the savings under this two level flexible capacity mode compared to the best cost performance under the reference, fixed capacity mode ${TRC_{Fr}}^*$.

3.5.1 Model, Two-Level Flexible Capacity Policy and Problem Formulation

The MSP operates in the same environment that is explained in the previous section(s). Recall that the repair shop is modeled as a single-server queue where the failures occur following a stationary Poisson demand and each defective system requires an exponentially distributed amount of dedicated repair service time in the shop in order to regain its *good as new* status.

In this section, we assume that the repair shop can make use of capacity flexibility options due to the two-level flexible capacity mode. Recall that a periodic, two-level, capacity policy, **C** = [D, μ_l, μ_h, $\vec{\pi}$], consists of a period length D, a low and a high service rate pair (μ_l, μ_h) and a policy vector, $\vec{\pi}$, which is the probability vector that consist of the probability values for the repair shop deploying the higher service rate at each state, which is the number of defective parts in the system. The details of this capacity mode and the policy , **C** = [D, μ_l, μ_h, $\vec{\pi}$], is described previously in Section 2.5.1.

Suppose the repair shop operates under a stable policy **C** = [D, μ_l, μ_h, $\vec{\pi}$] for an infinite horizon.

The capacity agency offers a set of possible period lengths, θ, from which the MSP can choose the best period length considering the reflection of wage differentials on c_c. Recall that c_c is also dependent on c_p, Δ and α coefficients. We pay h_r per unit time for the substitute system during the rental hiring duration. The down-time costs due to halt of the operations is equal to B per time unit, and we assume that $B > h_r$. The hiring related decision is the rental hiring duration L. The capacity related decisions, **C**, are threefold:

1) Length of the period: D

2) The size of the permanent and the contingent capacity levels: (μ_l & $\mu_h - \mu_l$)

3) The policy that determines when the contingent capacity is hired: $\vec{\pi}$ (based on the number of defective systems).

The total relevant cost function, TRC, can be represented by **C** and L, and is the sum of capacity related costs $\left(CRC(\mathbf{C})\right)$, down-time costs $\left(DTC(\mathbf{C}, L)\right)$ and hiring related costs

$\left(HC(L)\right)$. Given these cost components and the decision variables, the problem of the MSP can be formulated as follows:

$$\min_{\mathbf{C},L} TRC(\mathbf{C},L) = CRC(\mathbf{C}) + DTC(\mathbf{C},L) + HC(L)$$
$$s.t.\ \ L \geq 0$$
$$\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$$
$$\mu_h > \lambda, \mu_h > \mu_l, D \in \theta,$$
$$0 \leq \vec{\pi}(m) \leq 1 \text{ for } m = 0,1,2, \dots \tag{3.5}$$

Given the problem formulation above, in the next subsection, we analyze the problem in (3.5) and analyze each of the sub-problems. The derivation of the performance characteristics necessary for the analysis of the sub-problems are also given in the subsection 3.5.2.

3.5.2 Analysis and the Derivation of Necessary Functions.

In this subsection, we first analyze the optimization problem in (3.5). Note that in this problem, we need to explain a few more additional steps since the optimization problem (3.5) differs greatly from the problem (2.6) in Section 2.5.

Firstly, time requirement for solving problem (3.5) is a lot more demanding than time requirement for solving problem (2.6). The continuous nature of the rental duration enables us to set any positive value of L, and cancels out the sclerotic behavior of the stock level, which can take only integer values. Although this brings about an added flexibility, it also increases the computational burden tremendously, in such a way that it is numerically infeasible to search for the real optimal decision variables (especially, in terms of rental duration) globally.

In addition, the problem (3.5) cannot be decomposed into and formulated as conventional discrete time, infinite horizon Markov Decision Process problems in its current format, because the state information, which is the number of defective systems in the repair shop at the start of a period, is not detailed enough to derive the expectation of the immediate periodic reward (cost). This is due to the fact that the contribution of the down-time related costs to the immediate periodic reward necessitates not only the number of defective systems in the repair shop at the start of a period, but also the failure time records of each defective system, because the down-time costs are incurred after L time units following each failure. Since the failure times of the systems are not recorded, only the number of defective systems can be used in the capacity decision making process. Henceforth, rather than tackling the optimal policy structure problem for the MSP, we propose a prescriptive approach for the periodic capacity policy structure in (3.5). In this section, we focus on two-level threshold type policies and assess the possible savings of two-level capacity mode in the commoditized environment with regard to the total costs under the optimal fixed capacity policy, ${TRC_{Fr}}^*$.

A two level threshold-type policy includes a switching point k, which implies that the contingent capacity is deployed ($\vec{\pi}(m) = 1$) in a period, only if there are at least k number of defective systems in the repair shop at the start of that period ($m \geq k$). Note that k can also be a non-integer number. In such a case, we have a randomized action taking in the capacity policy as follows: at the start of a period, if there are more (less) than $\lfloor k \rfloor$ defective systems in the repair shop, the contingent capacity is (not) deployed and if there are exactly $\lfloor k \rfloor$ number of defective systems, the contingent capacity is deployed with a probability of: $1 - (k - \lfloor k \rfloor)$.

In the light of discussions above, we can reformulate (3.5) as follows:

$$\min_{D,\mu_l,\mu_h,L,k} TRC(\mathbf{C}, L) = CRC(\mathbf{C}) + DTC(\mathbf{C}, L) + HC(L)$$

$$s.t.\ \ L, k \geq 0$$

$$\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$$

$$\mu_h > \lambda,\ \mu_h > \mu_l,\ D \in \theta$$

$$\vec{\pi}(\lfloor k \rfloor) = 1 - (k - \lfloor k \rfloor)$$

$$\vec{\pi}(m) = 0 \text{ if } m < \lfloor k \rfloor \ \&\ \vec{\pi}(m) = 1 \text{ if } m > \lfloor k \rfloor \text{ for } m \in Z^+ - \{\lfloor k \rfloor\}$$

(3.6)

It is known that the optimal rental hiring duration L^* has to satisfy the following first order condition:

$$\frac{\partial TRC(\mathbf{C}, L)}{\partial L} = \lambda\left(h_r + B \times P\{S_d \geq L | \ \mathbf{C}\}\right) = 0 \rightarrow\ P\{S_d \geq L | \ \mathbf{C}\} = \frac{h_r}{B}$$

(3.7)

Therefore, if $P\{S_d \geq L | \ \mathbf{C}\} = \frac{h_r}{B}$ is added to (3.6) as a constraint, there will not be any implications for the optimal solution, since the optimal hiring duration would already satisfy $P\{S_d \geq L | \ \mathbf{C}\} = \frac{h_r}{B}$. On the other hand, it may have important consequences in terms of decomposability. Note that rather than investigating all possible computations, we investigate a finite number of rental duration (L) and period length (D) candidates as well as a limited number of permanent & contingent capacity level combinations (μ_l, μ_h) due to the computational burden. Suppose $\vec{\pi}^*$ denotes the threshold type policy with a switching point k^*, where $P\{S_d \geq L | \ \mathbf{C}\} = \frac{h_r}{B}$ when $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$, for given $L > 0$, $D \in \theta$ and (μ_l, μ_h).

Note that in all of our numerical results, for given L, D, μ_l and μ_h, we observe that $P\{S_d \geq L | \ \mathbf{C}\}$ is increasing with k. We can observe this behavior in

Figure 3-3, where the $P\{S_d \geq L | \ \mathbf{C}\}$ values are given for increasing values of k levels, for $\lambda = 1$, $\mu_l = 0.24342$, $\mu_h = 1.7039$, $L = 5$ and $D = 2$.

Therefore, if we have $P\{S_d \geq L|\, \mu_h\} < \frac{h_r}{B}$ (i.e. $k = 0$) and $P\{S_d \geq L|\, \mu_l\} > \frac{h_r}{B}$ (i.e. $k = \infty$), then we can find a k^* such that $P\{S_d \geq L|\, \mathbf{C}\} = \frac{h_r}{B}$, since we can also choose non-integer switching points. After k^* is found for all $L > 0$, $D \in \theta$ and (μ_l, μ_h) candidates, the overall optimal policy parameters can be found from (3.8):

$$\min_{D,\mu_l,\mu_h,L} TRC(\mathbf{C}, L) = CRC(\mathbf{C}) + DTC(\mathbf{C}, L) + HC(L)$$
$$s.t. \quad L \geq 0$$
$$\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}^*]$$
$$\mu_h > \lambda, \mu_h > \mu_l, D \in \theta \tag{3.8}$$

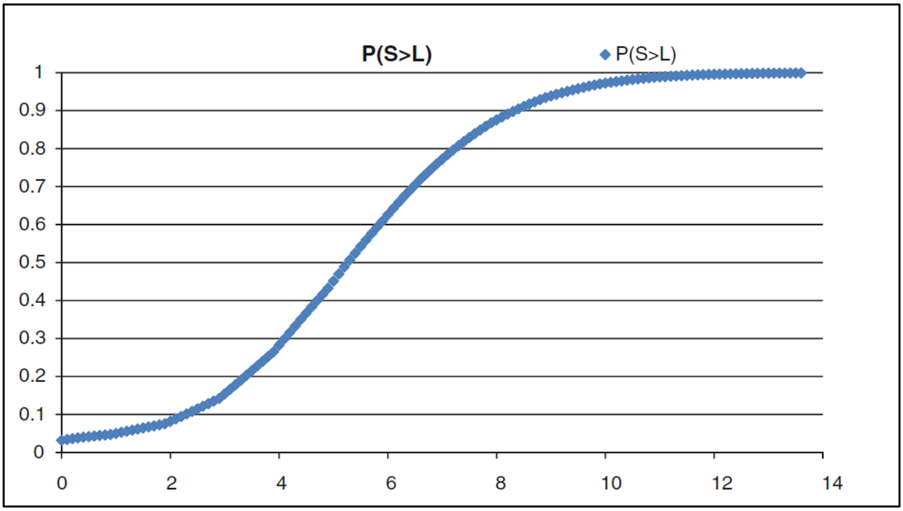

Figure 3-3: The figure depicts how $P\{S_d \geq L|\, \mathbf{C}\}$ values are increasing for increasing values of k levels, when $\lambda = 1, \mu_l = 0.24342, \mu_h = 1.7039, L = 5$ and $D = 2$ are given.

Next we derive the components of total relevant costs $TRC(\mathbf{C}, L)$ as well as the sojourn time distribution $P\{S_d \geq L|\, \mathbf{C}\}$ for given L and capacity policy: $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$.

3.5.2.1 Derivation of Cost Components

Here, we will derive the cost components of the total relevant costs $TRC(\mathbf{C}, L)$ for given L and $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$. First, the rental hiring related costs are the same as the previous, fixed capacity mode, which is $HC(L) = \lambda h_r L$. Next, we will analyze the capacity related costs, $CRC(\mathbf{C})$.

Recall that in Section 2.5, the transition probability matrix $\mathrm{P}(t, \mu)$ was derived for $t \leq D$, where $\mathrm{P}_{ij}(t, \mu)$ denotes the probability of having j defective systems in the repair shop after t time units, given that there were i defective systems at the beginning. From $\mathrm{P}(t,\mu_l)$ and $\mathrm{P}(t,\mu_h)$ matrices, we can obtain $\mathrm{P}(D,\mathbf{C})$, which is the transition probability matrix for the number of defective systems in the repair shop at the start of a period, under a threshold type capacity policy, $\mathbf{C}$ = [D, μ_l, μ_h, $\vec{\pi}$], with a switching point: k.

Suppose $\vec{\mathrm{P}}_n(D,\mathbf{C})$ denotes the n^{th} row of the $\mathrm{P}(D,\mathbf{C})$.Then we can write:

$$\vec{\mathrm{P}}_n(D,\mathbf{C}) = \vec{\mathrm{P}}_n(D,\mu_l) \ \text{ if } n < \lfloor k \rfloor$$
$$\vec{\mathrm{P}}_n(D,\mathbf{C}) = \vec{\mathrm{P}}_n(D,\mu_h) \ \text{ if } n > \lfloor k \rfloor$$
$$\vec{\mathrm{P}}_n(D,\mathbf{C}) = (k - \lfloor k \rfloor) \times \vec{\mathrm{P}}_n(D,\mu_l) + (1 - k + \lfloor k \rfloor) \times \vec{\mathrm{P}}_n(D,\mu_h) \ \text{ if } n = \lfloor k \rfloor \tag{3.9}$$

Having constructed $\mathrm{P}(D,\mathbf{C})$ completely, one can obtain the steady state probabilities for the number of defective systems in the repair shop at the start of a period, $\mathrm{v}(D,\mathbf{C})$, from the equalities below:

$$\mathrm{v}(D,\mathbf{C}) = \mathrm{v}(D,\mathbf{C}) \times \ \mathrm{P}(D,\mathbf{C}), \textstyle\sum_{i=0}^{K} \mathrm{v}_i(D,\mathbf{C}) = 1 \tag{3.10}$$

From $\mathrm{v}(D,\mathbf{C})$, we can obtain the average capacity deployment $ACU(\mathbf{C})$, which will be used in deriving the capacity related costs $CRC(\mathbf{C})$, resulting from capacity policy $\mathbf{C}$.

$$ACU(\mathbf{C}) = \sum_{i=0}^{\lfloor k \rfloor - 1} \mathrm{v}_i(D,\mathbf{C})\mu_l + \mathrm{v}_{\lfloor k \rfloor}(D,\mathbf{C})\big((k - \lfloor k \rfloor)\mu_l + (1 - k + \lfloor k \rfloor)\mu_h\big) + \sum_{i=\lfloor k \rfloor + 1}^{K} \mathrm{v}_i(D,\mathbf{C})\mu_h \tag{3.11}$$

Recall that in the two level capacity mode, there is an external capacity agency which provides the contingent capacity periodically at a higher price than permanent capacity (i.e. $c_c > c_p$), where c_c is sensitive to the period length D due to the (Δ, α) coefficients. Similar to the previous capacity modes, we exclude the costs related to the baseline capacity level: $c_p\lambda$. For given c_c and c_p values, the capacity related cost per unit time, $CRC(\mathbf{C})$, can be directly derived from $ACU(\mathbf{C})$ as follows:

$$CRC(\mathbf{C}) = c_p(\mu_l - \lambda) \ + \ (ACU(\mathbf{C}) \ - \mu_l)c_c \tag{3.12}$$

Next, we focus on deriving the sojourn time distribution of a defective system in the repair shop, $P\{S_d \geq L |\ \mathbf{C}\}$, which is needed both in the derivation of $DTC(\mathbf{C}, L)$, as well as it is used in the constraint: $P\{S_d \geq L |\ \mathbf{C}\} = \frac{h_r}{B}$.

3.5.2.2 Sojourn Time Distribution

In order to derive an explicit formula for the sojourn time distribution of a defective system, we first define an extended Markov Process, $\big(X(t), Y(t)\big)$, where $X(t)$ denotes the number of defective systems in the repair shop at time t, and $Y(t)$ denotes the position of a tagged defective system in the queue (including the one that is under service) at time t.

Recall that the repair shop is modeled as a $M/M/1/K$ queue with a sufficiently large K. Suppose G denotes the set of all states of the stochastic process $\left(X(t), Y(t)\right)$. Then we have $|G| = 1 + 2 + 3 \ldots + K + (K + 1) - 1 = \frac{(K+3)\times K}{2}$, which is the cardinality of set G.

Since the repair shop employs a $FCFS$ priority rule, we always have $0 \leq Y(t) \leq X(t) \leq K$, since the number indicating the position of a tagged defective system cannot exceed the total number of defective systems in the system. Also, the $FCFS$ rule provides that $Y(t)$ is non-increasing in t, because the position of the tagged defective system decreases one by one as the repairs of the defective systems before the tagged system are completed. When $Y(t) = 0$, the tagged system's repair is literally finished. Therefore, any $(j, 0)$ is an absorbing state of the $\left(X(t), Y(t)\right)$ process for $0 \leq j \leq K - 1$. Note that upon a failure at time t, if there are already $n - 1$ defective systems in the queue, then we have $\left(X(t), Y(t)\right) = (n, n)$ for all $n > 0$. Due to the capacity policy $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$, the service rate of the queue can change at the start of each period. Therefore, we first characterize the transient behavior of $\left(X(t), Y(t)\right)$ under a constant service rate of μ, which will help us to analyze the $\left(X(t), Y(t)\right)$ process under $\mathbf{C}$. The state diagram of the $\left(X(t), Y(t)\right)$ under a constant service rate μ is given in Figure 3-4.

Note that states $\mathrm{s} = (s_1, 0)$ are absorbing states of this process. After constructing Q for an arbitrary μ, we can analyze the transient probability behaviour of the $(X(t)\,, Y(t))$ process under constant service rates $\mu = \mu_l$ and $\mu = \mu_h$.

Let $\mathrm{U}^l(t)$ and $\mathrm{U}^h(t)$ denote the transition probability matrices of the $(X(t)\,, Y(t))$ process in t time units, under constant service rates $\mu = \mu_l$ and $\mu = \mu_h$, respectively. These matrices contain ${\mathrm{U}_{\mathrm{r,s}}}^l(t)$ $\left({\mathrm{U}_{\mathrm{r,s}}}^h(t)\right)$ values, which denote the probability that the system will be in state s, $(X(t) = s_1, Y(t) = s_2)$, given that it was in state r in the beginning: $(X(0) = r_1, Y(0) = r_2)$, when the service rate is μ_l (μ_h) throughout time t.

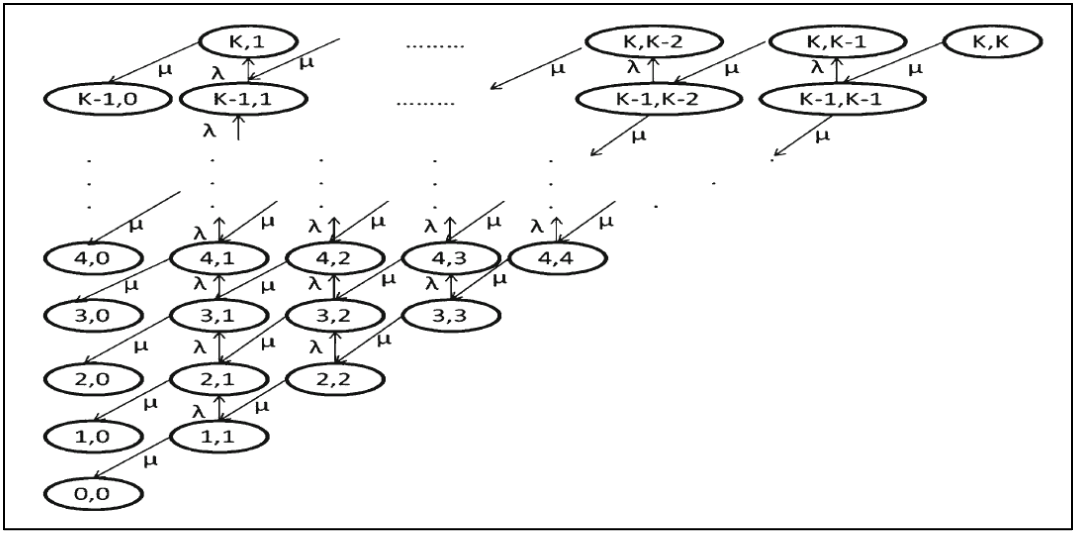

Figure 3-4: The state diagram of the $(X(t), Y(t))$ process under a constant service rate μ.

$Q_{r,s} = \lambda$ if $s_1 = r_1 + 1$ for $r_1 = 1, 2, \dots, K-1$ and $r_2 = s_2$ for $0 < r_2 \le r_1$;
$Q_{r,s} = \mu$ if $s_1 = r_1 - 1$ and $s_2 = r_2 - 1$ for $0 < r_1 \le K$, $r_2 \le r_1$;
$Q_{r,s} = -\sum_{r \neq m} Q_{r,m}$ if $s = r$ for all non-absorbing state r;
$Q_{r,s} = 0$ for all other (r, s) pairs.
(3.13)

We can find $U^l(t)$ $(U^h(t))$ matrices from the transition rate matrix Q, with the help of the uniformization technique. The idea behind the uniformization technique is to make the stay time in each state exponential with the same mean. For a given service rate μ, the leaving rate $l_r = \sum_{r \neq m} Q_{r,m}$ is found for each state $r \in G$. Then we take the uniformized event occurrence rate, which is $\psi = \max_r l_r$. After the uniformization procedure, the time between two transitions in the modified process is exponentially distributed with rate ψ, and given that a transition occurs, the probability that the transition is from state r to s, under a constant service rate μ, is denoted as $\widetilde{P}_{r,s}$ and it satisfies:

$\widetilde{P}_{r,s} = \frac{Q_{r,s}}{\psi} + \delta_{r,s}$, where $\delta_{r,s} = 1$ if $r = s$ and $\delta_{r,s} = 0$ otherwise, for every $r, s \in G$.

Property 3.5: Let $\widetilde{P}$ denotes the uniformized $|G| \times |G|$ transition probability matrix that consists of $\widetilde{P}_{r,s}$ values, which are defined above. Then we can write: $U^l(t) = \sum_{k=0}^{\infty} e^{-\psi t} \frac{(\psi t)^k}{k!} (\widetilde{P})^n$ when $\mu = \mu_l$ and $U^h(t) = \sum_{k=0}^{\infty} e^{-\psi t} \frac{(\psi t)^k}{k!} (\widetilde{P})^n$ when $\mu = \mu_h$

Proof: The proof and the reasoning behind can be found in (Kulkarni 1995). Also, several numerical approximation techniques are available in the literature (e.g. (Kulkarni 1995)) in order to escape from the infinite sum.

After finishing the analysis under the constant service rate, next we focus on the $(X(t)\,,Y(t))$ process under policy **C**, at the start of each period $(t\colon t = nD\ for\ n = 0,1,2,\ldots)$. Note that the $(X(nD)\,,Y(nD))$ process has the following property for $n = 1,2,\ldots$

Property 3.6: For any period length D, $(X(nD)\,,Y(nD))$ under capacity policy $\mathbf{C} = [D,\mu_l,\mu_h,\vec{\pi}]$ satisfies the Markovian property.

Now, let $\mathrm{A}(D,\mathbf{C})$ be the $|G| \times |G|$ transition probability matrix of the $(X(nD)\,,Y(nD))$ process under capacity policy $\mathbf{C} = [D,\mu_l,\mu_h,\vec{\pi}]$ with period length D and for positive integer n values. This matrix consists of $\mathrm{A}_{\mathrm{r,s}}(D,\mathbf{C})$ values, which denote the probability that the system will be in state s=(s_1,s_2) at the end of a period (which means there will be s_1 defective systems in the repair shop and the tagged system's position will be s_2), given that it was at state $\mathrm{r} = (r_1,r_2)$ at the start of that period (which means there are r_1 systems in the repair shop and the tagged system's position was r_2).

$$\mathrm{A}_{\mathrm{r,s}}(D,\mathbf{C}) = P\{X\big((n+1)D\big),Y\big((n+1)D\big) = (s_1,s_2)|\big(X(nD),Y(nD)\big) = (r_1,r_2)\}$$

$$\text{for } \mathrm{r} = (r_1,r_2) \text{ and } \mathrm{s} = (s_1,s_2) \text{ where } \mathrm{r},\mathrm{s} \in G.$$

Since **C** is a two-level threshold type of policy with a switching point (not necessarily an integer) of k, from the definition of the $\mathrm{U}^l(t)$ and $\mathrm{U}^h(t)$ matrices, we have $\mathrm{A}_{\mathrm{r,s}}(D,\mathbf{C}) = \mathrm{U}_{\mathrm{r,s}}{}^l(D)$ if $r_1 < \lfloor k \rfloor$ and $\mathrm{A}_{\mathrm{r,s}}(D,\mathbf{C}) = \mathrm{U}_{\mathrm{r,s}}{}^h(D)$ if $r_1 > \lfloor k \rfloor$ and $\mathrm{A}_{\mathrm{r,s}}(D,\mathbf{C}) = (k - \lfloor k \rfloor)\mathrm{U}_{\mathrm{r,s}}{}^l(D) + (1 - k + \lfloor k \rfloor)\,\mathrm{U}_{\mathrm{r,s}}{}^h(D)$ for all $\mathrm{r} = (\lfloor k \rfloor, r_2), \mathrm{s} = (s_1,s_2) \in G$.

Recall that S_d is the sojourn time of a defective system in the repair shop. After analyzing the transient behavior of $(X(t),Y(t))$ process under capacity policy $\mathbf{C} = [D,\mu_l,\mu_h,\vec{\pi}]$, we can start deriving $P\{S_d \geq x|\ \mathbf{C}\}$ for an arbitrary $x > 0$. In order to obtain $P\{S_d \geq x|\ \mathbf{C}\}$, we investigate the possible trajectories that the $(X(t),Y(t))$ process follows in x time units after an arbitrary failure.

Property 3.7: For any value of x, the trajectory of the $(X(t),Y(t))$ process for $t < x$ follows one of the following four patterns:

1)x spreads over $\lceil x/D \rceil$ periods and initial service rate is μ_l

2) x spreads over $\lceil x/D \rceil$ periods and initial service rate is μ_h

3) x spreads over $\lceil x/D \rceil + 1$ periods and initial service rate is μ_l

4) x spreads over $\lceil x/D \rceil + 1$ periods and initial service rate is μ_h.

Proof: The sketch of our proof is as follows. Suppose it is known that there has been a failure in a period. For the sake of the convenience, let the beginning point denote the start of the period that the failure realizes. If the number of defective systems in the repair shop at the beginning point is less than $\lfloor k \rfloor$, the initial service rate is equal to μ_l, otherwise, if it is more than $\lfloor k \rfloor$, it is equal to μ_h. When there are exactly $\lfloor k \rfloor$ number of defective systems, the service rate is μ_l with probability $(k - \lfloor k \rfloor)$ and μ_h with probability $(1 - k + \lfloor k \rfloor)$.

Let t denote the time between the failure of a system and the end of the first period after the failure. From the Conditional Distribution of the Arrival Times (Ross 1983), it is known that t is uniformly distributed over $(0, D)$. If a failure occurs between $(0, \lceil x/D \rceil D - x)$ in a period, then the duration x spreads over $\lceil x/D \rceil$ consecutive periods, whereas if the failure occurs between $(\lceil x/D \rceil D - x, D)$, then x spreads over $\lceil x/D \rceil + 1$ periods.

For each of these four trajectory patterns, the probability vector for the number of defective systems in the repair shop, upon a new failure is generated. If upon a failure, there are already j defective systems $(j < K)$ in the repair shop, the extended state of the system, $(X(D-t), Y(D-t))$, upon that failure will be $(j+1, j+1)$ and the transient behaviour is traced throughout x from that time point of failure.

Note that upon a failure, if there are already K other defective systems in the repair shop, that failed system is not accepted due to the modeling assumption of the $M/M/1/K$ system. Let R denote the probability of such a hypothetical event. R can be calculated from:

$$R = \frac{1}{D}\left(\left(\sum_{i=0}^{\lfloor k \rfloor - 1} \mathrm{v}_i(D, \mathbf{C}) + (k - \lfloor k \rfloor) \times \mathrm{v}_{\lfloor k \rfloor}(D, \mathbf{C})\right) \times \int_{t=x}^{D} \mathrm{P}_{iK}(D - t, \mu_l)\, dt \right.$$
$$\left. + \left((1 - k + \lfloor k \rfloor) \times \mathrm{v}_{\lfloor k \rfloor}(D, \mathbf{C}) + \sum_{i=\lfloor k \rfloor + 1}^{K} \mathrm{v}_i(D, \mathbf{C})\right) \times \int_{t=x}^{D} \mathrm{P}_{iK}(D - t, \mu_h)\, dt \right) \qquad (3.14)$$

While deriving $P\{S_d \geq x |\, \mathbf{C}\}$, we condition the probability that a failed system is accepted to the shop for repair. Arbitrarily small values of the non-accepting probability R can be obtained by assuming large enough K in the $M/M/1/K$ model. The accuracy of the finite waiting room will be further assessed by a simulation study later in this section. The mathematical formulation of the above sketch is given in Property 3.8 below:

Property 3.8: $P\{S_d \geq x|\ \mathbf{C}\}$ for $0 \leq x < D$ when $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$, a two-level threshold type policy with a switching point k, can be written as: $\frac{1}{D(1-R)} \sum_{i=h,l} \sum_{j=1,2} P\{S_d \geq x|\ (\mathbf{C}\,\&(i,j))\}$

Where:

1) $$P\{S_d \geq x|\ (\mathbf{C}\,\&(l,1))\} = \sum_{i=0}^{\lfloor k\rfloor-1} \mathrm{v}_i(D,\mathbf{C}) \sum_{j=0}^{K-1} \int_{t=x}^{D} \mathrm{P}_{ij}(D-t,\mu_l)\, \overline{F_l}^{(j+1)}(x)dt$$

$$+\left((k-\lfloor k\rfloor)\mathrm{v}_{\lfloor k\rfloor}(D,\mathbf{C})\right) \sum_{j=0}^{K-1} \int_{t=x}^{D} \mathrm{P}_{\lfloor k\rfloor j}(D-t,\mu_l)\, \overline{F_l}^{(j+1)}(x)dt$$

2) $$P\{S_d \geq x|\ (\mathbf{C}\,\&(l,2))\}$$
$$= \sum_{i=0}^{\lfloor k\rfloor-1} \mathrm{v}_i(D,\mathbf{C}) \sum_{j=0}^{K-1} \int_{t=0}^{x} \mathrm{P}_{ij}(D-t,\mu_l) \vec{\mathrm{U}}_{(j+1,j+1)}{}^{l}(t)\, \overrightarrow{\overline{\mathrm{F}}(x-t)}^{Tr} dt$$

$$+\left((k-\lfloor k\rfloor)\mathrm{v}_{\lfloor k\rfloor}(D,\mathbf{C})\right) \sum_{j=0}^{K-1} \int_{t=0}^{x} \mathrm{P}_{\lfloor k\rfloor j}(D-t,\mu_l) \vec{\mathrm{U}}_{(j+1,j+1)}{}^{l}(t)\, \overrightarrow{\overline{\mathrm{F}}(x-t)}^{Tr} dt$$

3) $$P\{S_d \geq x|\ (\mathbf{C}\,\&(h,1))\} = \sum_{i=\lfloor k\rfloor+1}^{K} \mathrm{v}_i(D,\mathbf{C}) \sum_{j=0}^{K-1} \int_{t=x}^{D} \mathrm{P}_{ij}(D-t,\mu_h)\, \overline{F_h}^{(j+1)}(x)dt$$

$$+\ \left((1-k+\lfloor k\rfloor)\mathrm{v}_{\lfloor k\rfloor}(D,\mathbf{C})\right) \sum_{j=0}^{K-1} \int_{t=x}^{D} \mathrm{P}_{\lfloor k\rfloor j}(D-t,\mu_h)\, \overline{F_h}^{(j+1)}(x)dt$$

4) $$P\{S_d \geq x|\ (\mathbf{C}\,\&(h,2))\}$$
$$= \sum_{i=\lfloor k\rfloor+1}^{K} \mathrm{v}_i(D,\mathbf{C}) \sum_{j=0}^{K-1} \int_{t=0}^{x} \mathrm{P}_{ij}(D-t,\mu_h) \vec{\mathrm{U}}_{(j+1,j+1)}{}^{h}(t)\, \overrightarrow{\overline{\mathrm{F}}(x-t)}^{Tr} dt$$

$$+\left((1-k+\lfloor k\rfloor)\mathrm{v}_{\lfloor k\rfloor}(D,\mathbf{C})\right) \sum_{j=0}^{K-1} \int_{t=0}^{x} \mathrm{P}_{\lfloor k\rfloor j}(D-t,\mu_h) \vec{\mathrm{U}}_{(j+1,j+1)}{}^{h}(t)\, \overrightarrow{\overline{\mathrm{F}}(x-t)}^{Tr} dt$$

(3.15)

Property 3.9: In a similar manner, $P\{S_d \geq x|\ \mathbf{C}\}$ for $(n-1)D \leq x < nD$ if $n \geq 2$ when $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$, a two-level threshold type policy with a switching point k, can be written as:

$$\frac{1}{D(1-R)} \sum_{i=h,l} \sum_{j=n,n+1} P\{S_d \geq x|\ (\mathbf{C}\ \&(i,j))\}$$

Where:

1) $P\left\{S_d \geq x \middle| \left(\mathbf{C}\ \&(l,n)\right)\right\} =$

$$\sum_{i=0}^{\lfloor k\rfloor -1} \mathrm{v}_i(D,\mathbf{C}) \sum_{j=0}^{K-1} \int_{t=x-(n-1)D}^{D} \mathrm{P}_{ij}(D-t,\mu_l)\vec{\mathrm{U}}_{(j+1,j+1)}{}^{l}(t)\left(\mathrm{A}(D,\mathbf{C})\right)^{n-2}\overrightarrow{\bar{\bar{\mathrm{F}}}(x-(n-2)D-t)}^{Tr}dt$$

$+$

$$\left((k-\lfloor k\rfloor)\mathrm{v}_{\lfloor k\rfloor}(D,\mathbf{C})\right)$$

$$\sum_{j=0}^{K-1} \int_{t=x-(n-1)D}^{D} \mathrm{P}_{\lfloor k\rfloor j}(D-t,\mu_l)\vec{\mathrm{U}}_{(j+1,j+1)}{}^{l}(t)\ \left(\mathrm{A}(D,\mathbf{C})\right)^{n-2}\overrightarrow{\bar{\bar{\mathrm{F}}}(x-(n-2)D-t)}^{Tr}dt$$

2) $P\left\{S_d \geq x \middle| \left(\mathbf{C}\ \&(l,n+1)\right)\right\} =$

$$\sum_{i=0}^{\lfloor k\rfloor -1} \mathrm{v}_i(D,\mathbf{C}) \sum_{j=0}^{K-1} \int_{t=0}^{x-(n-1)D} \mathrm{P}_{ij}(D-t,\mu_l)\vec{\mathrm{U}}_{(j+1,j+1)}{}^{l}(t)\left(\mathrm{A}(D,\mathbf{C})\right)^{n-1}\overrightarrow{\bar{\bar{\mathrm{F}}}(x-(n-1)D-t)}^{Tr}dt +$$

$$\left((k-\lfloor k\rfloor)\mathrm{v}_{\lfloor k\rfloor}(D,\mathbf{C})\right)\times$$

$$\sum_{j=0}^{K-1} \int_{t=0}^{x-(n-1)D} \mathrm{P}_{\lfloor k\rfloor j}(D-t,\mu_l)\vec{\mathrm{U}}_{(j+1,j+1)}{}^{l}(t)\ \left(\mathrm{A}(D,\mathbf{C})\right)^{n-1}\overrightarrow{\bar{\bar{\mathrm{F}}}(x-(n-1)D-t)}^{Tr}dt$$

3) $P\left\{S_d \geq x \middle| \left(\mathbf{C}\ \&(h,n)\right)\right\} =$

$$\sum_{i=\lfloor k\rfloor+1}^{K} \mathrm{v}_i(D,\mathbf{C}) \sum_{j=0}^{K-1} \int_{t=x-(n-1)D}^{D} \mathrm{P}_{ij}(D-t,\mu_h)\quad \vec{\mathrm{U}}_{(j+1,j+1)}{}^{h}(t)\left(\mathrm{A}(D,\mathbf{C})\right)^{n-2}\overrightarrow{\bar{\bar{\mathrm{F}}}(x-(n-2)D-t)}^{Tr}dt$$

$$+\left((1-k+\lfloor k\rfloor)\mathrm{v}_{\lfloor k\rfloor}(D,\mathbf{C})\right)\times$$

$$\sum_{j=0}^{K-1} \int_{t=x-(n-1)D}^{D} \mathrm{P}_{\lfloor k\rfloor j}(D-t,\mu_h)\quad \vec{\mathrm{U}}_{(j+1,j+1)}{}^{h}(t)\ \left(\mathrm{A}(D,\mathbf{C})\right)^{n-2}\overrightarrow{\bar{\bar{\mathrm{F}}}(x-(n-2)D-t)}^{Tr}dt$$

4) $P\left\{S_d \geq x \middle| \left(\mathbf{C}\ \&(h,n+1)\right)\right\} =$

$$\sum_{i=\lfloor k\rfloor+1}^{K} \mathrm{v}_i(D,\mathbf{C}) \sum_{j=0}^{K-1} \int_{t=0}^{x-(n-1)D} \mathrm{P}_{ij}(D-t,\mu_h)\vec{\mathrm{U}}_{(j+1,j+1)}{}^{h}(t)\left(\mathrm{A}(D,\mathbf{C})\right)^{n-1}\overrightarrow{\bar{\bar{\mathrm{F}}}(x-(n-1)D-t)}^{Tr}dt$$

$+$

$$\left((1-k+\lfloor k\rfloor)\mathrm{v}_{\lfloor k\rfloor}(D,\mathbf{C})\right)\times$$

$$\sum_{j=0}^{K-1} \int_{t=0}^{x-(n-1)D} \mathrm{P}_{\lfloor k\rfloor j}(D-t,\mu_h)\vec{\mathrm{U}}_{(j+1,j+1)}{}^{h}(t)\ \left(\mathrm{A}(D,\mathbf{C})\right)^{n-1}\overrightarrow{\bar{\bar{\mathrm{F}}}(x-(n-1)D-t)}^{Tr}dt$$

(3.16)

Above, in (3.15) and (3.16), a^{Tr} is the transpose of any vector a,

$$\vec{\mathrm{U}}_{(j+1,j+1)}{}^{l}(t) = \left(\mathrm{U}_{(j+1,j+1),s^1}{}^{l}(t), \mathrm{U}_{(j+1,j+1),s^2}{}^{l}(t), \dots, \mathrm{U}_{(j+1,j+1),s^{|G|}}{}^{l}(t)\right)$$

$$\vec{\mathrm{U}}_{(j+1,j+1)}{}^{h}(t) = \left(\mathrm{U}_{(j+1,j+1),s^1}{}^{h}(t), \mathrm{U}_{(j+1,j+1),s^2}{}^{h}(t), \dots, \mathrm{U}_{(j+1,j+1),s^{|G|}}{}^{h}(t)\right)$$

$$\overrightarrow{\overline{\mathrm{F}}(x)} = \left(I(s^1{}_1, s^1{}_2, k, x), I(s^2{}_1, s^2{}_2, k, x), \dots, I\left(s^{|G|}{}_1, s^{|G|}{}_2, k, x\right)\right)$$

Where $I(i,j,k,x) = \bar{F}_l{}^j(x) = P\{B_1 > x\}$, for $B_1 \sim Erlang(j, \mu_l)$ when $i < \lfloor k \rfloor$ and $I(i,j,k,x) = \bar{F}_h{}^j(x) = P\{B_2 > x\}$, for $B_2 \sim Erlang(j, \mu_h)$ when $i > \lfloor k \rfloor$ and finally $I(\lfloor k \rfloor, j, k, x) = (k - \lfloor k \rfloor)\bar{F}_l{}^j(x) + (1 - k + \lfloor k \rfloor)\bar{F}_h{}^j(x)$.

Note that the n^{th} element of $\vec{\mathrm{U}}_{(j+1,j+1)}{}^{l}(t)$ $\left(\vec{\mathrm{U}}_{(j+1,j+1)}{}^{h}(t)\right)$ above denotes the probability that after t time units from the failure, the number of defective systems will be $s^n{}_1$, $(X(D) = s^n{}_1)$, and the position of the system that has failed at time $D - t$ will be $s^n{}_2$ after t time units, $(Y(D) = s^n{}_2)$, for $\mu = \mu_l$ $(\mu = \mu_h)$.

Here, the formula in (3.16) generalizes the formula given in 3.15 for $S_d > D$ by keeping the track of the extended state of a defective system in the shop throughout a period with the help of the $\mathrm{A}(D, \mathbf{C})$ matrix.

As we have derived the sojourn time distribution of a defective system in the repair shop, we can find $P\{S_d \geq L | \mathbf{C}\}$ for any $L > 0$ and any policy $\mathbf{C}$.

Let $f_{S_d}(x|\mathbf{C}) = \frac{\partial(1 - P\{S_d \geq x | \mathbf{C}\})}{\partial x}$. We can obtain $f_{S_d}(x|\mathbf{C})$ by redefining $I(i,j,k,x) = f_l{}^j(x)$ when $i < \lfloor k \rfloor$ and $I(i,j,k,x) = f_h{}^j(x)$ when $i > \lfloor k \rfloor$ and $I(\lfloor k \rfloor, j, k, x) = (k - \lfloor k \rfloor) f_l{}^j(x) + (1 - k + \lfloor k \rfloor) f_h{}^j(x)$, where $f_l{}^j(x)$ is the density of $B_1 \sim Erlang(j, \mu_l)$ and $f_h{}^j(x)$ is the density of $B_2 \sim Erlang(j, \mu_h)$ at x.

In the light of discussions above, we can derive $DTC(\mathbf{C}, L)$, since we have:

$$DTC(\mathbf{C}, L) = \lambda B\left(\mathrm{E}((S_d - L)^+ | \mathbf{C})\right) = \lambda B \int_{x=L}^{\infty} (x - L)\, f_{S_d}(x|\mathbf{C}) \mathrm{d}x$$

As we have derived all the necessary performance measures, in the next section we can discuss about the procedure and present the results of the numerical study that is conducted.

3.5.3 Numerical Study

In this section, we use the analysis and the results provided in the previous section in order to assess the performance of the two level flexible capacity mode under rental availability. In the end, we compare the cost performance of this mode with the minimum cost rate achieved under the fixed capacity mode with rental availability, $TRC_{Fr}{}^*$. The outline of this section is as follows.

First, we describe the characteristics and the test bed of the computational study. Second, we present the search procedure in order to find the near-optimal policy parameters for the problem given in (3.6). Finally, we discuss the results of the search procedure, examine the sensitivity of and the interactions among the policy and system parameters and finally assess the potential cost benefits that can be gained in two level flexible mode in comparison to the total costs under the optimal policy under the fixed capacity mode with rental availability.

In our computational study, we normalize the arrival rate for the system failures, $\lambda = 1$ (failures per time unit) as well as the unit cost per time for the unit permanent capacity costs, $c_p = 1$. The other parameters are scaled according to these normalized λ and c_p parameters. Similar to 3.4.3, we analyze a total of 9 scenarios with three different B/h_r and three different h_r values, which are already given in Table 3-1. For each of these 9 scenarios and different (Δ, α) combinations, we execute our solution procedure to find the capacity policy parameters, (D, μ_l, μ_h, k) and the rental duration L that yield the minimum total costs.

The capacity agency can supply the agreed amount of contingent capacity at a given frequency. The frequency is determined by the period length, which is chosen from θ, the set of candidate period lengths offered. In our thesis we assume $\theta = \{0.5, 1, \ldots, 4.5, 5\}$, which are scaled to the normalized inter-arrival time: $\frac{1}{\lambda} = 1$. Correspondingly, a period length of $D = 2$ corresponds to one capacity update opportunity in two inter-arrival times of system failures on the average.

We develop a solution procedure, which is specifically designed for the optimization problem in (3.6). The fixed capacity results $({L_{Fr}}^*, {\mu_{Fr}}^*)$ obtained from Subsection 3.4.3 not only serve as a reference point to assess the overall cost performance (savings of two level capacity flexibility with respect to ${TRC_{Fr}}^*$), but also help us to generate meaningful candidate parameter sets in our search procedure for the permanent and permanent plus contingent capacity levels as well as a meaningful set for rental hiring durations.

Unlike the fixed capacity mode, we cannot find the optimal hiring duration and the optimal capacity policy parameters that minimize the total costs in problem (3.6) analytically. What we can do for this capacity mode is to derive the total relevant costs $TRC(\mathbf{C}, L)$ for a given capacity policy $\mathbf{C}$ and a given hiring duration L. Time requirements for computing the total cost performance under this capacity mode, especially for the down-time related costs, $DTC(\mathbf{C}, L)$ is quite demanding. While computing analytically $P\{S_d \geq x | \mathbf{C}\}$ for a given x is quite efficient, it becomes computationally non-economical to calculate $TRC(\mathbf{C}, L)$ analytically, since it requires the numerical integration of $P\{S_d \geq x | \mathbf{C}\}$ over x, from L to infinity.

Therefore, we used analytical computational approach to find the threshold policies which satisfy $P\{S_d \geq L | \mathbf{C}\} = h_r/B$. After finding the candidate policies for the optimal solution, the total costs under these candidate policies are obtained from discrete event simulation.

Recall that for the search procedures in the specialized system environment, we have specified a S_{max} value, which is taken large enough to make sure that the optimal base stock level, S^*, is smaller than S_{max} for all the problem instances. Due to the normalized failure arrival rate ($\lambda = 1$), S_{max} values would be affordable for an enumerative search of S^* in $\{0,1,2, \ldots, S_{max}\}$. On the other hand, in the commoditized environment, rental hiring duration L is a continuous decision variable. In addition, the computational requirements in this section enforces us to focus on a limited number of hiring duration possibilities. Therefore, an affordable set for candidate hiring durations has to be generated before performing the search procedure developed to solve (3.6).

3.5.3.1 Generation of Candidate Sets for Hiring Durations

Next, we describe the generation of the set of candidate hiring durations, $\psi_{(B,h_r)}$, for each of the 9 different (B, h_r) combinations given in Table 3-1. These $\psi_{(B,h_r)}$ sets are going to be used in the search procedure to find the optimal parameters for (3.6). For a given (B, h_r) pair, the set $\psi_{(B,h_r)}$ consists of equidistant points (we took a distance of 0.5 in our thesis), that are symmetrically dispersed around ${L_{Fr}}^*$, which is the optimal hiring duration under the same (B, h_r) pair, in the fixed capacity mode.

The half-length of the set $\psi_{(B,h_r)}$ is driven by the sensitivity of S^* with regard to different Δ and α capacity cost parameters under the same (B, h_r) combination in the two level capacity mode for the specialized system environment. The sensitivity of S^* to different Δ and α values (Δ from 0 to 10 and α from 0 to 5) can be collated from the results given in Section 2.5. The reason that we use the sensitivity of S^* while creating the candidate set for L, is due the fact that both spare stock in the specialized system environment and the rental hiring duration in the commoditized system environment serve to the same purpose, which is to alleviate the down-time related costs by improving the overall system availability. In addition, in fixed capacity mode results, we have observed similarities between the effects of (B, c_p, c_c) on the optimal stock level (S^*) and effects of (B, c_p, c_c) on the optimal hiring duration (L^*) decisions. In addition, the interrelations between the capacity policy $\mathbf{C}$ and the hiring duration L resemble those between $\mathbf{C}$ and S, only the response of the stock level decision is more sclerotic than the hiring duration decision since S is a positive integer by definition. Therefore, for each of the nine (B, h_r) combination, we designated the $\psi_{(B,h_r)}$ set in such a way that, it contains the L^* in the middle and it contains a hiring duration candidate that is at least 1 unit less than the smallest S^* value (${S^*}_{min}$) as well as a hiring duration that is at least 1 unit more than the largest S^* value (${S^*}_{max}$) that are encountered for these (B, h_r) values, under the two level capacity policy of the specialized system environment, with different Δ and α capacity cost parameters. In Table 3-4, for each of the 9 (B, h_r) pairs, we give the ${L_{Fr}}^*$, ${S^*}_{min}$ and ${S^*}_{max}$ values as well as the $\psi_{(B,h_r)}$ sets, which include the candidate hiring durations that will be explored in the following search method. The ${L_{Fr}}^*$ for each (B, h_r) is the central and bold member of the $\psi_{(B,h_r)}$ set.

		$h_r = 1$	$h_r = 0.2$	$h_r = 0.1$
$B/h_r = 5$	$\psi_{(B,h_r)}$	(0, 0.5, **1**, 1.5, 2)	(0, 0.23, 0.73, 1.23, 1.73, **2.23**, 2.73, 3.23, 3.73, 4.23, 4.73)	(1.15, 1.65, 2.15, 2.65, **3.15**, 3.65, 4.15, 4.65, 5.15)
	(S^*_{min}, S^*_{max})	(1, 1)	(2, 3)	(3, 4)
$B/h_r = 10$	$\psi_{(B,h_r)}$	(0, 0.27, 0.77, **1.27**, 1.77, 2.27, 2.77, 3.27)	(0.33, 0.83, 1.33, 1.83, 2.33, **2.83**, 3.33, 3.83, 4.33, 4.83, 5.33)	(2.01, 2.51, 3.01, 3.51, **4.01**, 4.51, 5.01, 5.51, 6.01)
	(S^*_{min}, S^*_{max})	(2, 2)	(3, 4)	(3, 5)
$B/h_r = 20$	$\psi_{(B,h_r)}$	(0, 0.5, 1, **1.5**, 2, 2.5, 3)	(0.35, 0.85, 1.35, 1.85, 2.35, 2.85, **3.35**, 3.85, 4.35, 4.85, 5.35, 5.85, 6.35)	(2.24, 2.74, 3.24, 3.74, 4.24, **4.74**, 5.24, 5.74, 6.24, 6.74, 7.24)
	(S^*_{min}, S^*_{max})	(2, 2)	(4, 5)	(4, 6)

Table 3-4 The candidate hiring duration sets $\psi_{(B,h_r)}$ with ${L_{Fr}}^*$, in the middle with bold font and (S^*_{min}, S^*_{max}) values encountered in Section 2.5 with the same (B, h_r) pair under different $(\varDelta, \alpha)$ values and c_c :linear, inverse linear and exponential; for $B/h_r = 5, 10, 20$; $h_r = 0.1, 0.2, 1$.

As we have determined the candidate hiring duration sets, $\psi_{(B,h_r)}$, next we give our search procedure followed by the description of its underlying mechanism.

Search Procedure-IV

---

0. Follow steps a & b:
   a. For every $L \in \psi_{(B,h_r)}$:

      Set $\mu^*(L) = \lambda - \frac{\ln(\frac{h_r}{B})}{L}$, which is the μ that satisfies: $P(S_d \geq L |\, \mu) = \frac{h_r}{B}$ under single capacity mode.

   b. Choose a period length $D \in \theta$.
   c. Create the $\varUpsilon_l$ and $\varUpsilon_h$ sets from $\mu^*(L)$ values as follows:

      $$\varUpsilon_l = \left\{\frac{2}{10}\mu^*(L),\ \frac{3}{10}\mu^*(L),\ \frac{4}{10}\mu^*(L),\ \ldots\ ,\ \frac{8}{10}\mu^*(L),\ \frac{9}{10}\mu^*(L)\right\}$$
      $$\varUpsilon_h = \left\{\frac{11}{10}\mu^*(L),\ \frac{12}{10}\mu^*(L)\right\}$$

   d. For every $D \in \theta$, $\mu_l \in \varUpsilon_l$ and $\mu_h \in \varUpsilon_h$ find the k^* that satisfies the $P\{S_d \geq x |\, \mathbf{C}\} = \frac{h_r}{B}$, where $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}^*]$, and $\vec{\pi}^*$ is the threshold policy with a threshold point of k^*.
1. After finding k^* (and therefore $\vec{\pi}^*$) for all $D \in \theta$, $\mu_l \in \varUpsilon_l$ and $\mu_h \in \varUpsilon_h$, we can find the minimum cost for given L, $TRC(\mathbf{C}^*(L), L)$ as follows:

$$TRC(\mathbf{C}^*(L), L) = \min_{D \in \theta, \mu_l \in \Upsilon_l, \mu_h \in \Upsilon_h} (TRC(\mathbf{C}, L): \mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}^*])$$

2. After $TRC(\mathbf{C}^*(L), L)$ for all $L \in \psi_{(B,h_r)}$ are found, we can find L^*, and $\mathbf{C}^*(L^*)$ = $[D^*, \mu_l{}^*, \mu_h{}^*, \vec{\pi}^*]$ values, which would give the global minimum cost rate for problem (3.6): $TRC^* = TRC(\mathbf{C}^*(L^*), L)$ for this two level flexible capacity mode.

   L^* can be found from brute force search as follows:

$$TRC(\mathbf{C}^*(L^*), L^*) = \min_{L \in \psi_{(B,h_r)}} TRC(\mathbf{C}^*(L), L)$$

3.5.3.2 Explanation of the Search Algorithm

In the search procedure above, at step 0, we find a reference capacity level for every L: $\mu^*(L)$, which is the capacity level that results in the minimum TRC for a given $L \in \psi_{(B,h_r)}$ in the single, fixed capacity model. This reference capacity level is used later to determine the permanent (μ_l) & permanent + contingent (μ_h) capacity levels, which will be used in the two-level flexible capacity policy, $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$. For a given L, the reference $\mu^*(L)$ is found from $\mu^*(L) = \lambda - \frac{\ln(\frac{h_r}{B})}{L}$, which is the capacity level μ that satisfies: $P\{S_d \geq L |\, \mu\} = \frac{h_r}{B}$.

After the reference point $\mu^*(L)$ is found for each candidate hiring duration $L \in \psi_{(B,c_p)}$, we are ready to construct the sets Υ_l and Υ_h, which contain the low and high capacity levels (μ_l and μ_h), respectively. Note that for a given L and period length D, μ_l and μ_h should satisfy: $\mu_l \leq \mu^*(L) \leq \mu_h$. Otherwise the resulting $P\{S_d \geq x |\, \mathbf{C}\}$ due to the capacity policy $\mathbf{C}$ will never be equal to $\frac{h_r}{B}$.

From a previous related study, (Buyukkaramikli et al. 2011-a), we have observed that the choice on the permanent capacity level plays a more important role in determining the savings from average capacity deployment. Therefore, in our thesis, we put more emphasis on the decision on the permanent capacity level during the search procedure. For that reason, for a given hiring duration L, among a number of equidistantly scattered capacity alternatives around the reference $\mu^*(L)$ value, 80% of them are taken smaller and the remaining 20% are taken larger than $\mu^*(L)$. Actually, there can be infinite number of (μ_l , μ_h) possibilities. However, for computational time reasons, we limit ourselves to a total of 10 equidistantly scattered capacity alternatives, where $\Upsilon_l = \left\{\frac{2}{10}\mu^*(L),\ \frac{3}{10}\mu^*(L),\ \frac{4}{10}\mu^*(L),\ \dots,\ \frac{8}{10}\mu^*(L),\ \frac{9}{10}\mu^*(L)\right\}$ and $\Upsilon_h = \left\{\frac{11}{10}\mu^*(S),\ \frac{12}{10}\mu^*(S)\right\}$, which lead to a total of 16 (μ_l, μ_h) pairs for each L, where $\mu_l \in \Upsilon_l$ and $\mu_h \in \Upsilon_h$.

For each hiring duration L, period length $D \in \theta$ and (μ_l, μ_h) pair, we find the k^*, which satisfies $P\{S_d \geq x | \mathbf{C}\} = \frac{h_r}{B}$, where $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}^*]$, and $\vec{\pi}^*$ is the threshold policy with a threshold point of k^*. After k^* is found for every: $L \in \psi_{(B,h_r)}$, $D \in \theta$, $\mu_l \in \Upsilon_l$ and $\mu_h \in \Upsilon_h$, we find the optimal parameters (L^*, D^*, ${\mu_l}^*$ and ${\mu_h}^*$) by brute force search in steps 1 & 2. This search completes the solution procedure for our problem.

Next, we present and discuss the results of the numerical study that is conducted, where the optimal costs for problem (3.6) are obtained from the search procedure that is described above. We first present the savings in total relevant costs when the best two level flexible capacity policy is employed compared to the best fixed capacity system (${TRC_{Fr}}^*$ in short). Afterwards we conduct a sensitivity analysis of the system and the optimal policy parameters. After discussing the interrelations, we provide the results of the simulation study where the accuracy of the finite waiting room approximation was assessed.

3.5.3.3 Savings Compared to the TRC under the Optimal Single Level Capacity (${TRC_{Fr}}^*$)

Total costs can be reduced up to a great extent due to the two-level capacity flexibility. From our numerical results, we have witnessed that up to a 56% savings can be achievable in total costs due to the two-level flexible capacity mode compared to the minimum cost that can be achieved in the single capacity mode: ${TRC_{Fr}}^*$.

In Table 3-5, we give the maximum percentage savings that two level flexible policies can bring for all *9* different $\frac{B}{h_r}$ & h_r scenarios with $c_p = 1$, $\Delta = 0, 0.25, 0.5$ and 1 and $\alpha = 0$,1 and 2, when c_c has the linear, exponential and inverse proportional functional form. Suppose for given cost parameters (B, h_r, Δ and α), and a functional form for c_c, ${TRC_{2r}}^*$ denotes the minimum total costs that can be achieved under all two level capacity policies in the commoditized environment that Search Procedure-IV goes through. After ${TRC_{2r}}^*$ is found, the corresponding entry in Table 3-5 can be calculated from: $({TRC_{Fr}}^* - {TRC_{2r}}^*)/ {TRC_{Fr}}^*$.

We first observe that, although the same permanent capacity cost, $c_p = 1$ is used for both environments, the percentage savings in the commoditized environment is lower compared to the savings of the 2-level policy in the specialized setting, which was given in Table 2-5. This can be explained by the fact that the cost parameters related to the operating systems differ greatly for the commoditized setting from the specialized setting. For instance the unit time holding cost per a critical subsystem is lower than the unit time hiring cost of a substitute system and down-time costs are higher in specialized system environment. Higher B/h and c_p/h ratios (compared to B/h_r and c_p/h_r) make the role of the flexible capacity policies in the specialized system much more critical than in the commoditized system.

From Table 3-5, we can observe that the percentage savings seem to decrease with the hiring cost rate h_r. This is due to the fact that the reference cost parameter ${TRC_{Fr}}^*$ increases with h_r faster than the cost savings $({TRC_{Fr}}^* - {TRC_{2r}}^*)$ increase with h_r, since an inflated h_r means that a higher portion of the total costs is hiring costs (HC), and therefore less sensitive to the changes in capacity policy than the other cost parts (DTC and CRC).

Some of the trends we have seen in Table 2-5 are also observable here, too. For instance, the savings in total costs increase with the period length elasticity α and decrease with maximum opportunity cost factor Δ. Similar to the specialized system environment, under the same capacity cost parameters, linear cost structure for the contingent capacity results in the most savings, and the inverse proportional results in the least. Down-time cost factor seems to increase the percentage savings, however contradictory instances, where a higher B leads to lower savings, exist, which necessitates a further enquiry.

Finally, we explore further how the optimal policy parameters change under the optimal two-level periodic capacity flexible mode compared to the single level capacity mode for different cost parameter settings. In Table 3-6, we show how the optimal two level capacity mode policy parameters (S^*, D^*, k^*, ${\mu_l}^*$ and ${\mu_h}^*$) differ with various (Δ, α) combinations and 4 B & h_r scenarios.

The data in Table 3-6 illustrate that under the optimal policies pertaining to the two level capacity mode, the cost savings compared to ${TRC_{Fr}}^*$ come from both less capacity deployment as well as less hiring costs due to the shorter hiring durations. It can be seen that for each of the 4 B & h_r scenario and (Δ, α) combination, L^* under the two level capacity mode is less than or equal to the ${L_F}^*$, which is the optimal hiring duration under the single-level capacity mode. The differences in hiring durations are generally higher for lower h_r and higher B/h_r, lower Δ , and higher α parameters. In Table 3-6, it can be seen that the smallest period length $D = 0.5$, is chosen as the optimal period length in most of the instances. However, for lower h values, high Δ and positive elasticity ($\alpha > 0$), higher period lengths ($D > 0.5$) can be optimal, as well. Although two level capacity policies cause savings in capacity related costs (CRC) due to less deployed average capacity, $ACU(\mathbf{C})$, the optimal capacity policy parameters in Table 3-6 (k^*, ${\mu_l}^*$ and ${\mu_h}^*$) can provide further insights.

Two Level

	B=0.5		Inv Pro	exp	Lin
h_r =0.1	Δ=1	α=0	14%	14%	14%
		α=1	22%	28%	47%
		α=2	27%	38%	54%
	Δ=0.5	α=0	24%	24%	24%
		α=1	30%	34%	54%
		α=2	35%	42%	54%
	Δ=0.25	α=0	35%	35%	35%
		α=1	41%	42%	54%
		α=2	44%	47%	54%
	Δ=0	all α	54%	54%	54%

	B=1		Inv Pro	exp	Lin
h_r =0.1	Δ=1	α=0	18%	18%	18%
		α=1	25%	31%	49%
		α=2	30%	40%	55%
	Δ=0.5	α=0	28%	28%	28%
		α=1	34%	36%	55%
		α=2	39%	44%	55%
	Δ=0.25	α=0	39%	39%	39%
		α=1	44%	45%	55%
		α=2	47%	49%	55%
	Δ=0	all α	55%	55%	55%

	B=2		Inv Pro	exp	Lin
h_r =0.1	Δ=1	α=0	21%	21%	21%
		α=1	28%	33%	50%
		α=2	33%	41%	56%
	Δ=0.5	α=0	31%	31%	31%
		α=1	36%	38%	56%
		α=2	41%	45%	56%
	Δ=0.25	α=0	41%	41%	41%
		α=1	45%	46%	56%
		α=2	48%	50%	56%
	Δ=0	all α	56%	56%	56%

	B=1		Inv Pro	exp	Lin
h_r =0.2	Δ=1	α=0	13%	13%	13%
		α=1	17%	22%	37%
		α=2	23%	30%	45%
	Δ=0.5	α=0	22%	22%	22%
		α=1	28%	30%	45%
		α=2	32%	34%	45%
	Δ=0.25	α=0	32%	32%	32%
		α=1	35%	36%	45%
		α=2	38%	40%	45%
	Δ=0	all α	45%	45%	45%

	B=2		Inv Pro	exp	Lin
h_r =0.2	Δ=1	α=0	17%	17%	17%
		α=1	39%	25%	22%
		α=2	47%	33%	26%
	Δ=0.5	α=0	26%	26%	26%
		α=1	47%	32%	31%
		α=2	47%	37%	34%
	Δ=0.25	α=0	34%	34%	34%
		α=1	47%	39%	38%
		α=2	47%	42%	40%
	Δ=0	all α	47%	47%	47%

	B=4		Inv Pro	exp	Lin
h_r =0.2	Δ=1	α=0	19%	19%	19%
		α=1	24%	27%	40%
		α=2	29%	35%	47%
	Δ=0.5	α=0	29%	29%	29%
		α=1	33%	34%	47%
		α=2	36%	39%	47%
	Δ=0.25	α=0	36%	36%	36%
		α=1	40%	40%	47%
		α=2	42%	43%	47%
	Δ=0	all α	47%	47%	47%

	B=5		Inv Pro	exp	Lin
h_r =1	Δ=1	α=0	10%	10%	10%
		α=1	13%	15%	18%
		α=2	18%	23%	35%
	Δ=0.5	α=0	18%	18%	18%
		α=1	24%	25%	35%
		α=2	27%	29%	35%
	Δ=0.25	α=0	27%	27%	27%
		α=1	30%	30%	35%
		α=2	31%	32%	35%
	Δ=0	all α	35%	35%	35%

	B=10		Inv Pro	exp	Lin
h_r =1	Δ=1	α=0	11%	11%	11%
		α=1	15%	17%	21%
		α=2	20%	24%	35%
	Δ=0.5	α=0	20%	20%	20%
		α=1	25%	26%	35%
		α=2	28%	30%	35%
	Δ=0.25	α=0	28%	28%	28%
		α=1	30%	31%	35%
		α=2	31%	32%	35%
	Δ=0	all α	35%	35%	35%

	B=20		Inv Pro	exp	Lin
h_r =1	Δ=1	α=0	14%	14%	14%
		α=1	17%	19%	24%
		α=2	23%	27%	40%
	Δ=0.5	α=0	23%	23%	23%
		α=1	28%	29%	40%
		α=2	31%	34%	40%
	Δ=0.25	α=0	31%	31%	31%
		α=1	34%	35%	40%
		α=2	35%	37%	40%
	Δ=0	all α	40%	40%	40%

Table 3-5 : The % cost savings of two level flexible capacity policy compared to the fixed capacity policy when c_c : inversely proportional, exponential and linear structure, when $B/h_r = 5{,}10,\ 20,\ \ h_r = 0.1{,}0.2{,}1,$ for $\Delta =\ 0, 0.25, 0.5$ and 1 & $\alpha = 0, 1{,}2$ in the commoditized system environment.

From Table 3-6, we can observe that in the two level capacity mode, the choice on the optimal permanent capacity $\mu_l{}^*$ is considerably lower in comparison to optimal capacity level in the single capacity mode $\mu_{Fr}{}^*$. We can see that $\mu_l{}^*$ increases with Δ and decreases with α, which suggests that the repair shop hedges risk by deploying more permanent capacity when the contingent capacity becomes more expensive and its price gets more insensitive with respect to the period length. We do not observe a monotonic relation between the choice of $\mu_h{}^*$ and the other cost parameters.

The threshold value k^* can take non-integer values as well. k^* tends to increase with Δ and decrease with α, which is parallel to the risk-hedging behavior of the repair shop when capacity is more expensive. These inter-relations between the capacity and cost parameters will be further explicated in the next section, where a list of sensitivity analyses are conducted on policy and cost parameters.

B =2, h_r =0.2	Single Level			L_F*			μ_F*	
				2.83				1.81
	Two-Level			L*	D*	k*	μ*$_l$	μ*$_h$
		Δ=0	all α	2.33	0.5	1.6	0.40	2.39
	inv. Proportional	Δ=0.25	α=0	2.83	0.5	2.1	0.54	2.18
			α=1	2.33	0.5	1.6	0.40	2.39
			α=2	2.33	0.5	1.6	0.40	2.39
		Δ=0.5	α=0	2.33	0.5	2.1	0.99	2.39
			α=1	2.33	0.5	1.9	0.80	2.39
			α=2	2.83	0.5	2.1	0.54	2.18
		Δ=1	α=0	2.83	0.5	3	1.27	2.18
			α=1	2.33	0.5	2.1	0.99	2.39
			α=2	2.33	0.5	2.1	0.99	2.39

B =1, h_r =0.1	Single Level			L_F*			μ_F*	
				4.01				1.57
	Two-Level			L*	D*	k*	μ*$_l$	μ*$_h$
		Δ=0	all α	1.27	0.5	1.15	0.56	3.38
	inv. Proportional	Δ=0.25	α=0	1.27	0.5	1.15	0.56	3.38
			α=1	1.27	0.5	1.15	0.56	3.38
			α=2	1.27	0.5	1.15	0.56	3.38
		Δ=0.5	α=0	1.27	0.5	1.15	0.56	3.38
			α=1	1.27	0.5	1.15	0.56	3.38
			α=2	1.27	0.5	1.15	0.56	3.38
		Δ=1	α=0	1.27	0.5	1.79	1.97	3.38
			α=1	1.27	0.5	1.21	0.85	3.38
			α=2	1.27	0.5	1.15	0.56	3.38

B =2, h_r =0.1	Single Level			L_F*			μ_F*	
				4.74				1.63
	Two-Level			L*	D*	k*	μ*$_l$	μ*$_h$
		Δ=0	all α	3.23	0.5	1.9	0.39	2.31
	inv. Proportional	Δ=0.25	α=0	3.23	0.5	2.1	0.58	2.31
			α=1	3.23	0.5	1.9	0.39	2.31
			α=2	3.23	0.5	1.9	0.39	2.31
		Δ=0.5	α=0	4.23	0.5	3.1	0.85	2.05
			α=1	3.23	0.5	2.1	0.58	2.31
			α=2	3.23	0.5	2.1	0.58	2.31
		Δ=1	α=0	4.23	0.5	3.9	1.20	2.05
			α=1	4.23	1.0	3	0.85	2.05
			α=2	4.23	1.0	3	0.85	2.05

B=4, h_r=0.2	Single Level			L_F*			μ_F*	
				3.35				1.58
	Two-Level			L*	D*	k*	μ*$_l$	μ*$_h$
		Δ=0	all α	2.84	0.5	1.77	0.62	2.46
	inv. Proportional	Δ=0.25	α=0	2.84	0.5	1.91	0.62	2.46
			α=1	2.84	0.5	1.77	0.41	2.46
			α=2	2.84	0.5	1.77	0.41	2.46
		Δ=0.5	α=0	2.84	0.5	2.1	0.82	2.46
			α=1	2.84	0.5	2.1	0.82	2.46
			α=2	2.84	0.5	1.91	0.62	2.46
		Δ=1	α=0	3.34	0.5	2.89	1.14	2.28
			α=1	2.84	0.5	2.32	1.03	2.46
			α=2	2.84	0.5	2.1	0.82	2.46

Table 3-6 : The optimal two-level capacity mode policy parameters (L^*, D^* and μ^*) under different 4 B & h_r scenarios (**1:** $B = 2, h_r = 0.2$ **2:** $B = 1, h_r = 0.1$ **3:** $B = 2, h_r = 0.1$ and **4:** $B = 4, h_r = 0.2$) and various (Δ, α) combinations when $c_p = 1$.

3.5.3.4 Sensitivity Analysis of the Optimal Policy Parameters

In this subsection, we discuss the inter-relations among the cost and optimal policy parameters. We first focus on how the optimal switching point k^* is affected by other cost/policy parameters. Afterwards we investigate the role of the period length on TRC and on the choice of other policy parameters. Finally, we examine how the hiring duration L affects TRC and other parameters.

3.5.3.4.1 The Optimal Switching Point: k^*

In this subsection, our main concern is how the optimal switching point k^*, changes with different policy and cost parameters. Recall that k^* is the switching point that satisfies $P\{S_d \geq L | \mathbf{C}\} = \frac{h_r}{B}$, where $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}^*]$, and $\vec{\pi}^*$ is the threshold policy with k^*. Therefore unlike the specialized system environment, k^* in this section is the same for all contingent cost function structures (e.g. linear, inverse linear and exponential) and for all capacity cost parameters Δ and α. In Table 3-7, we tabulate how k^* value responds to changes in permanent and contingent capacity levels (μ_l, μ_h), period length D and hiring duration L when $B = 10$, $c_p = 5$.

$L = 2.83, D = 1$									$L = 2.83, D = 2$								
μ_h/μ_l	$0.2\ \mu^*$	$0.3\ \mu^*$	$0.4\ \mu^*$	$0.5\ \mu^*$	$0.6\ \mu^*$	$0.7\ \mu^*$	$0.8\ \mu^*$	$0.9\ \mu^*$	μ_h/μ_l	$0.2\ \mu^*$	$0.3\ \mu^*$	$0.4\ \mu^*$	$0.5\ \mu^*$	$0.6\ \mu^*$	$0.7\ \mu^*$	$0.8\ \mu^*$	$0.9\ \mu^*$
$1.1\ \mu^*$	1.23	1.33	1.47	1.66	1.94	2.23	2.69	3.57	$1.1\ \mu^*$	0.59	0.76	0.99	1.15	1.39	1.78	2.32	3.29
$1.2\ \mu^*$	1.7	1.88	2.07	2.27	2.53	2.9	3.44	4.4	$1.2\ \mu^*$	1.16	1.31	1.53	1.83	2.15	2.53	3.14	4.16
$L = 2.83, D = 3$									$L = 4.83, D = 2$								
μ_h/μ_l	$0.2\ \mu^*$	$0.3\ \mu^*$	$0.4\ \mu^*$	$0.5\ \mu^*$	$0.6\ \mu^*$	$0.7\ \mu^*$	$0.8\ \mu^*$	$0.9\ \mu^*$	μ_h/μ_l	$0.2\ \mu^*$	$0.3\ \mu^*$	$0.4\ \mu^*$	$0.5\ \mu^*$	$0.6\ \mu^*$	$0.7\ \mu^*$	$0.8\ \mu^*$	$0.9\ \mu^*$
$1.1\ \mu^*$	0.24	0.33	0.46	0.64	0.91	1.27	1.89	2.95	$1.1\ \mu^*$	1.75	2.02	2.2	2.45	2.82	3.27	3.97	5.21
$1.2\ \mu^*$	0.47	0.65	0.91	1.2	1.57	2.09	2.74	3.87	$1.2\ \mu^*$	2.54	2.8	3.08	3.36	3.75	4.27	5.01	6.34

Table 3-7 The optimal switching point (k^*) when $B = 10$, $c_p = 5$ for different L, D, μ_l and μ_h.

In short, the results from Table 3-7 suggest that, ceteris paribus:

- The lower μ_l, the earlier (i.e. smaller number of defective systems at the repair shop) we switch from low to high capacity (or vice versa).
- The higher μ_h, the later (i.e. larger number of defective systems at the repair shop) we switch from low to high capacity (or vice versa).
- The larger D, the earlier we switch from low to high capacity.
- The longer L, the later we switch from low to high.

These trends are general, no matter which cost function is chosen (linear, inverse or exponential) for the contingent capacity or what the cost parameters Δ and α are. These behaviors can be explained as follows: as the permanent capacity (or contingent capacity) gets higher, the repair shop would hire contingent capacity less frequently and at higher workloads. On the other hand, shorter period lengths enable more frequent capacity updates, in other words enable faster recourse actions, which incentivize the repair shop to take more risks by using contingent capacity at higher workloads. In a similar vein, k^* increases as L increases, although $\mu^*(L)$ is decreasing with L. This is due to the fact that a higher rental hiring duration may decrease the DTC (down-time related costs) too much, such that trimming the capacity usage of the repair shop by hiring contingent capacity at higher workloads is cost beneficial.

Having completed our discussion on how k^* is affected by various parameters, next we summarize our findings on the bidirectional relations between period length, D, TRC and other policy parameters.

3.5.3.4.2 Period Length D

The period length D plays a central role in this capacity mode, as it induces the level of the adaptability of the repair shop capacity to the workload and it determines the per time unit cost of the contingent capacity due to the capacity provision mechanism. Accordingly, in this part we investigate the sensitivity of the total costs and the optimal policy parameters to the period length D under different contingent cost functions with 3 different structures and Δ and α parameters.

In order to pursue an investigation on the effects of period length per se, we worked on a scenario when the capacity agency supplies the agreed amount of capacity only at a given frequency (i.e. $\theta = \{D\}$, a set that consists of a single period length) and we run the search procedure described above in order to optimize other policy parameters (L^*, ${\mu_l}^*$, ${\mu_h}^*$ and k^*). We follow these steps for increasing values of period length by the order of: $D = 0.5, 1, 1.5, \ldots 5$ with different Δ and α values. Therefore note that the TRC^* notation in this part is used for the optimal total costs for a given period length. In Figure 3-5 the behavior of the optimal TRC^* at increasing period lengths, under two level capacity policy is illustrated for $B = 10, c_p = 5, \alpha\ = 2$ and $\Delta =\ 0, 5$ and 10, when c_c function has exponential (left) and inverse linear (right) structure.

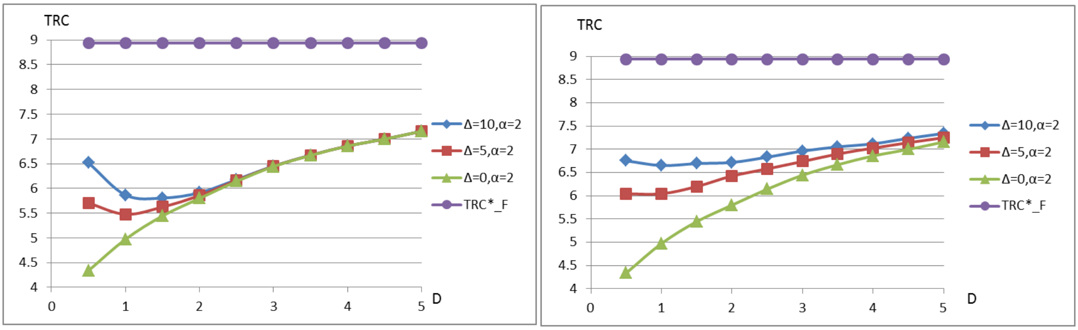

Figure 3-5: The behavior of the optimal TRC^* at increasing values of D, under 2 level capacity policy for $B = 10, c_p = 5, \alpha\ = 2$ and $\Delta =\ 0, 5$ and 10, when c_c has exponential (left) and inverse (right) structure.

In Figure 3-5, it can be observed that in the two level capacity mode, all optimal TRC^* values at given period lengths D from 0.5 to 5 for the chosen Δ values $(0, 5$ and $10)$ engender smaller total cost realizations compared to the optimal cost in the single capacity mode, ${TRC_F}^*$. Of course this can be reverse for even higher values of Δ coupled with a low elasticity α, or even longer period lengths. Therefore, any two-level capacity mode alternative at a given period length D should yield a total cost value less than ${TRC_F}^*$.

Furthermore, we can observe from Figure 3-5 that the minimum total costs with positive Δ are higher than the minimum total costs when $\Delta =\ 0$. This is self-evident, since the per time unit cost for contingent capacity, c_c, is the cheapest and equal to c_p when $\Delta =\ 0$, no matter what structure that the c_c function has. In such a case choosing the shortest possible period length will be the optimum, since more frequent updating possibilities would increase the responsiveness of the repair shop capacity to its workload, and thus a more economic use of its resources without any added cost factors. On the other hand, for $\Delta > 0$, due to the wage differential reflections of the contingent capacity costs, TRC^* gets higher at short period lengths, which stimulates the choice of a longer period. That's the reason why the gaps between TRC^* with different Δ values are widest when $D = 0.5$, which is the shortest period length that is analyzed. On the contrary, TRC^* for different Δ overlap each other at longer period lengths, since the effects of the wage differentials slim down to a negligible extent after some D.

TRC^* can behave quite differently with different period lengths when c_c has a linear structure. In order to comprehend the dynamics of the interplays between the period length, total costs and cost parameters at a deeper level, we present Figure 3-6, which illustrates TRC^* values under 2 level capacity mode at increasing period lengths (D from 0.5 to 5), when c_c has a linear structure for $B = 10$ and $c_p = 5$. The figure on the left emphasizes the effects of different Δ values ($0, 5$ and 10) when $\alpha = 2$ and the figure on the right emphasizes the role of the elasticity, by illustrating TRC^* with $\alpha = 0,2,5$ and ∞ when $\Delta = 5$.

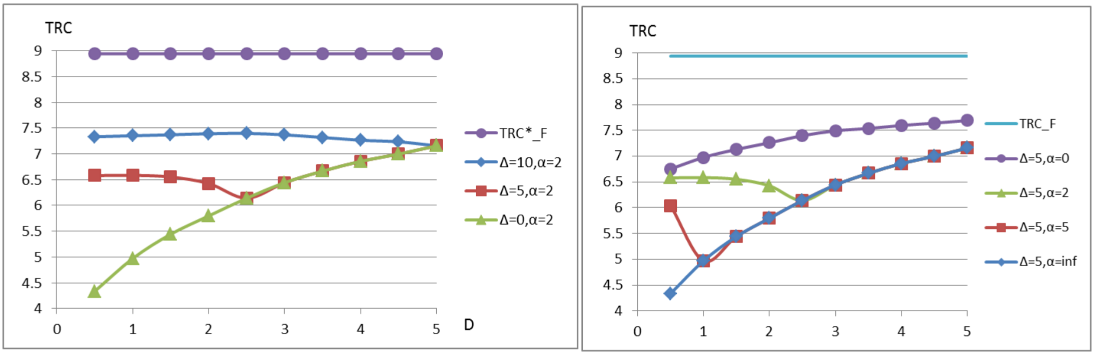

Figure 3-6 : The behavior of the optimal TRC^* under 2 level capacity mode at increasing values of period length D, for $B = 10$, $c_p = 5$ and when c_c function has the linear structure. Left: $\alpha = 2$ and $\Delta = 0, 5$ and 10. Right: $\Delta = 5$ and $\alpha = 0,2,5$ and ∞ .

In the linear cost structure, c_c equates to c_p for period lengths bigger than or equal to Δ/α values. For $D > \Delta/\alpha$, all TRC^* values with $\Delta > 0$ coincide with the TRC^* values with $\Delta = 0$. TRC^* curve pertaining to the linear cost structure has a carving around Δ/α, which is important because it mostly determines the optimal period length for $\Delta > 0$. From Figure 3-6, we can observe that the optimal period length is the smallest possible D for $\Delta = 0$ and it is equal to: Δ/α for $\Delta > 0$, unless $\Delta/\alpha < 5$. The other characteristics of TRC^* with linear cost structure are similar to those of the previously mentioned TRC^* values with exponential and inverse linear structures. After discussing the role of Δ on the behavior of TRC^*, next we can scrutinize the effects of different time elasticity factors: α on the behaviour of TRC^*.

The time elasticity of per time unit contingent capacity cost is the other important factor that determines the behavior of the minimum capacity costs in response to the period length D. What we can first conclude from the right figure in Figure 3-6 is the following:

for a given $\Delta > 0$, TRC^* values decrease with increasing α, because for a given period length D, the contingent capacity gets cheaper for increasing elasticity, when every other cost parameters remain unchanged. Two extreme values that elasticity factor α can take are 0 and ∞ respectively. If $\alpha = \infty$, the contingent capacity is perfectly elastic, and at the start of a period if it is communicated to the provider that the contingent capacity is not needed for that period, the capacity agency can immediately find another substitute task until the next period which is equivalent in terms of financial returns. This would vanish the cost burden due to the lost opportunities and wage differentials, therefore the optimal total costs behave as TRC^* with $\Delta = 0$. On the other hand, when $\alpha = 0$, the provider cannot assign the contingent capacity for another task, hence the wage differentials remain the same and the contingent capacity costs are inelastic and not affected by the period length. Given a certain $\Delta > 0$, any intermediate α value will result in an optimal total cost that is located in this band, bounded by TRC^* curve with $\alpha = \infty$ from below and TRC^* curve with $\alpha = 0$ from above. It can be further observed that, in all capacity cost structures, with $\Delta > 0$ and an intermediate α value, TRC^* gets closer to the above mentioned upper bound for shorter period lengths (as D goes to 0) and on the contrary, it gets closer to the lower bound (even completely overlaps in the linear cost structure case) for longer period lengths D.

D^*	$\Delta = 0$ $\alpha = 2$	$\Delta = 5$ $\alpha = 2$	$\Delta = 10$ $\alpha = 2$	
inverse	0.5	0.5	1	
linear	0.5	2.5	5	
exp	0.5	1	1.5	
D^*	$\Delta = 5$ $\alpha = 0$	$\Delta = 5$ $\alpha = 2$	$\Delta = 5$ $\alpha = 5$	$\Delta = 5$ $\alpha = \infty$
Inverse	0.5	0.5	0.5	0.5
Linear	0.5	2.5	1	0.5
Exp	0.5	1	0.5	0.5

Table 3-8: The optimal period length: D^* when $B = 10$ and $c_p = 5$ for inverse, linear and exponential cost structures when $\Delta = 0,5$ and 10 for $\alpha = 2$(above) and when $\alpha = 0, 2, 5$ and ∞ for $\Delta = 5$ (below).

For completely elastic ($\alpha = \infty$) and completely inelastic ($\alpha = 0$) contingent capacity cost structures, we observe that TRC^* values display a monotonically increasing behaviour with the period length D. However, for the other mid-values of α, TRC^* displays a U-shaped structure; where both Δ and α parameters are critical in determining the steepness of the curve. Therefore the optimal period length choice (D^*) is affected by these cost parameters to a great extent. In order to understand the nature of the dynamics between the optimal choice of the period length D^* and the contingent capacity cost parameters (Δ, α), we present Table 3-8, where the optimal period length choices are tabulated under different contingent capacity cost structures for $\Delta = 0, 5$ and 10 when $\alpha = 2$ (above) and for $\alpha = 0, 2, 5$ and ∞ when $\Delta = 5$ (below).

From the first tabular in Table 3-8, we can see that, when α remains constant, D^* increases with Δ. This increase is most evident for linear contingent capacity cost structure and least evident for inverse linear contingent capacity cost structure. On the other hand, for the response of D^* to an increase in α, we can draw different conclusions. Generally speaking, we can say that D^* first increases and then decreases with α, when Δ is positive. In the second tabular in Table 3-8, this behavior can be obviously seen for linear and exponential cost structures, whereas D^* is insensitive to an increase in α for the inverse linear cost structure. However, different cost parameter selections (with higher values of Δ) result in the suggested effects in response to α, for the inverse linear cost structure, as well. Note that D^* values in Table 3-8 are the same as the D^* values in Table 2-9, which suggest that the availability strategy (whether the availability upon system breakdown is from a spare part stock or a rental agreement) does not affect the response of period length choice to different capacity cost parameters, significantly.

After finishing the analysis of the bi-directional relations between the period length D and TRC^*, next we discuss the effects of the rental hiring duration decision: L.

3.5.3.4.3 Rental Hiring Duration Length L

In this part, we discuss how the rental hiring duration length affects the optimal total costs and other policy parameters and vice versa. Note that in this part, TRC^* corresponds to the minimum costs that can be achieved for a given rental hiring duration L. In order to conduct the analysis of the effects of the rental hiring duration per se, we fix a rental hiring duration length L, and find the optimal capacity policy parameters for that L according to the Search Procedure-IV described above. The first three figures in

Figure 3-7 illustrate how TRC^* changes with increasing values for hiring durations for $B = 10$, $c_p = 5$ and for different Δ values $(0, 5$ and $10)$ and exponential, inverse linear and linear contingent capacity cost structures when $(\alpha = 2)$. The last figure in

Figure 3-7 shows the behavior of TRC^* with increasing L but for different α values, $(0, 2, 5)$ for linear cost structure, for given $\Delta = 5$.

The data in

Figure 3-7 suggest that TRC^* resembles a convex function of L. It is already discussed that total costs are convex with L for a given policy $\mathbf{C}$. It seems that optimizing the parameters of capacity policy at each hiring duration length L, does not distort this behaviour of TRC^*. As expected, increasing contingent capacity costs (higher Δ or lower α) increase TRC^* values, however, the increase is moderated due to the optimization of other policy parameters. We observe that increase in Δ at lower Δ values affect TRC^* more than increase in Δ at higher Δ values. In Table 4.8, we tabulate how the optimal rental hiring duration changes under different capacity cost structures when $B = 10, c_p = 5$, for $\Delta = 0{,}5$ and 10 for $\alpha = 2$(above) and for $\alpha = 0, 2,$ and 5 when $\Delta = 5$ (below).

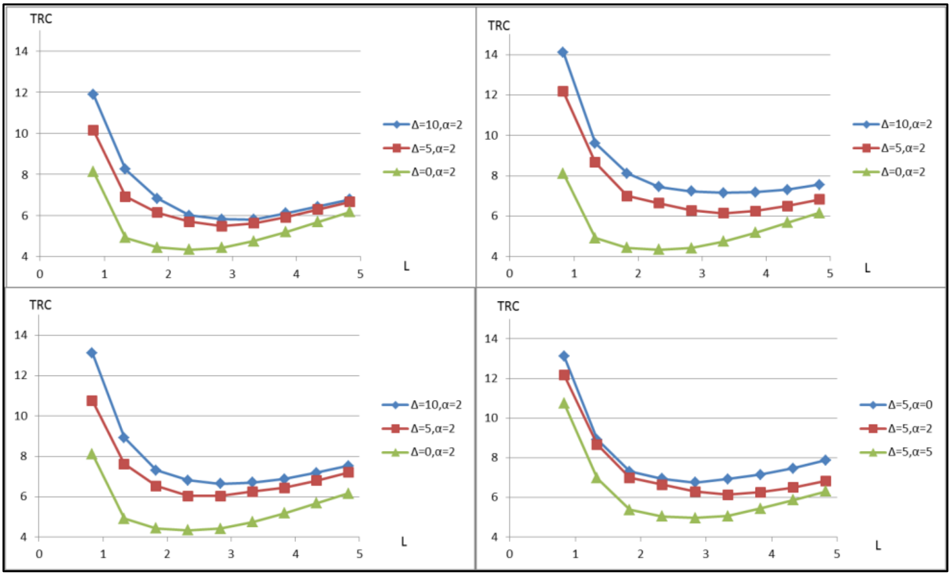

Figure 3-7 : Optimal total cost TRC^* for increasing hiring duration lengths L with $\Delta = 0, 5, 10$, $\alpha = 2$ and for clockwise direction: exponential, inverse linear and linear contingent capacity cost structure and with $\alpha = 0, 2, 5$ and $\Delta = 5$ in linear contingent capacity cost structure.

L^*	$\Delta = 0, \alpha = 2$	$\Delta = 5, \alpha = 2$	$\Delta = 10, \alpha = 2$	
Inverse	*2.33*	*2.33*	*2.83*	
Linear	*2.33*	*3.33*	*3.33*	
Exp	*2.33*	*2.83*	*3.33*	
L^*	$\Delta = 10, \alpha = 0$	$\Delta = 10, \alpha = 2$	$\Delta = 10, \alpha = 5$	$\Delta = 10, \alpha = \infty$
Inverse	*2.83*	*2.83*	*2.83*	*2.33*
Linear	*2.83*	*3.33*	*3.33*	*2.33*
Exp.	*2.83*	*3.33*	*2.83*	*2.33*

Table 3-9: The optimal base-stock level: S^* under different contingent capacity structures when $B = 10$ and $c_p = 5$ for $\Delta = 0, 5$ and 10 for $\alpha = 2$(above) and for $\alpha = 0, 2, 5$ and ∞ when $\Delta = 10$ (below).

From the first tabular in Table 3-9, we can see that L^* increases with Δ, when α remains constant. On the other hand, the response of L^* to an increase in α for a given Δ is different. From the second tabular in Table 3-9, we can say that L first increases and then decreases with α, when Δ is given and positive. Note that the L^* values in Table 3-9, are always less than the S^* values for the two level capacity mode with the same (Δ, α) combinations in Table 2-10.

The choices on the hiring duration and on the period length are inter-connected. Figure 3-8 can be helpful to understand the interplay between period length D and the hiring duration L.

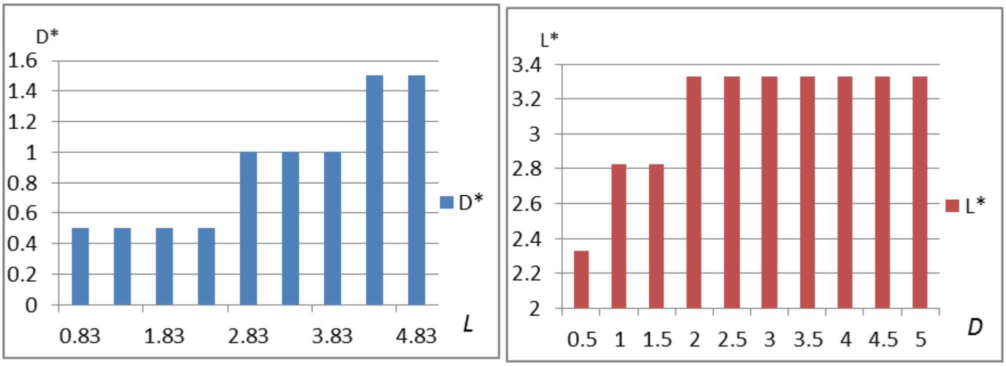

Figure 3-8: Right: the optimal period length for increasing hiring durations. Left: optimal hiring duration lengths for increasing period length when $B = 10$, $c_p = 5$ for Δ= 5 and $\alpha = 2$.

The figure on the left demonstrates that the optimal period length, D^*, increases for higher values of hiring duration length L, and similarly the in figure on the right, we can see that the optimal hiring duration L^* increases with period length D. The data in Figure 3-8 suggest that in the presence of wage differentials ($\Delta, \alpha > 0$), availability due to hiring duration length and capacity update frequency are complementary to some extent. As the frequency of capacity updates is higher, a shorter rental hiring duration is sufficient, whereas if we have a high rental hiring duration, the system would try to cut costs by less frequently capacity updating with cheaper contingent capacity costs.

This concludes our sensitivity analysis section. In the next subsection, we check the accuracy of using the finite waiting room approximation for the repair shop.

3.5.3.5 Accuracy of Finite Waiting Room Approximation

In this subsection, we examine the accuracy of our finite waiting room approximation by comparing the Average Capacity Deployed ($ACU(\mathbf{C})$) and the $P\{S_d \geq L| \mathbf{C}\}$ values from the analytical model (having a finite waiting room of 40) with the ($ACU(\mathbf{C})$) and the $P\{S_d \geq L| \mathbf{C}\}$ values obtained by simulating the real environment having a repair shop that has an infinite waiting room.

In our simulations, we used a run length of 2,000,000,000 defective part arrivals (when $\lambda = 1$) in a single replication, where the average total cost rate TRC is calculated under a policy: $\mathbf{C} = [D, \mu_l, \mu_h, \vec{\pi}]$.

We investigated a total of 9922 different scenarios with different B, c_p, L, D, Δ and α and resulting policy parameters. The percentage error, $\%err$, of using the analytical approximation for ACU and $P\{S_d \geq L| \mathbf{C}\}$in a scenario can be found as:

$$\%err(ACU) = 100 \times \frac{(ACU_{sim} - ACU_{app})}{ACU_{sim}}$$

$$\%err(P\{S_d \geq L| \mathbf{C}\}) = 100 \times \frac{\left(P\{S_d \geq L| \mathbf{C}\}_{sim} - P\{S_d \geq L| \mathbf{C}\}_{app}\right)}{P\{S_d \geq L| \mathbf{C}\}_{sim}} \tag{3.17}$$

Here, ACU_{sim} ($P\{S_d \geq L|\, \mathbf{C}\,\}_{sim}$) is the average deployed capacity (probability that the time until repair is higher than the rental hiring duration length) obtained from the simulation and ACU_{app} ($P\{S_d \geq L|\, \mathbf{C}\,\}_{app}$) are the same statistics derived from the analytical approximation. Table 3-10 summarizes the accuracy of the approximations.

	average %|*err*|	min *%err*	median *%err*	max *%err*
ACU	0.017%	-0.13%	-0.002%	0.13%
$P\{S_d \geq L\|\, \mathbf{C}\,\}$	0.29%	-1.96%	-0.01%	1.96%

Table 3-10 : Accuracy of the approximation for the ACU and $P\{S_d \geq L|\, \mathbf{C}\,\}$ values

In Table 3-10, the absolute value, minimum, median and the maximum for the percentage errors are listed, respectively. From the table, we can see that the approximation can mimic the performance of the original, infinite waiting room environment almost perfectly, which demonstrates the accuracy of our method. We can observe that the percentage of the errors in $P\{S_d \geq L|\, \mathbf{C}\,\}$ are higher than those of ACU. This is partly due to the fact that most of the $P\{S_d \geq L|\, \mathbf{C}\,\}$ values are less than 0.2, therefore although the real differences are quite small, the percentage errors seem to be conflated. Furthermore the event of S_d being larger than L is not as frequent as the capacity deployment points, which happen at the start of each period, which may require a longer simulation length for more accurate results.

This simulation study finalizes this section for the second capacity mode. In the next section, Section 3.6, we set out to analyze the third (and last) capacity mode. The conclusive remarks on second capacity mode will be provided in the overall conclusions section, Section 3.7.

3.6 Periodic Capacity Sell-back Mode

In this section, we analyze the integrated decision making problem of the service provider under the third capacity mode. This capacity mode is same as the capacity mode described in Section 3.6, in which the failed parts are sent to the repair shop at regular intervals in time. Due to this admission structure, when the repair of all the failed parts in the shop are completed in a period, it is known that there will not be any job left at least until the start of the next period. This synchronization of arrivals allows for a contract, where the capacity agency supplies a fixed amount of capacity at regular intervals in time, covering for the whole interval duration. However, if all the repairs in the shop are completed before the end of an interval, the capacity can be temporarily sold back to the capacity agency, at a reduced price, until the next interval.

In this mode, we assume that the provider supplies the capacity at the original permanent capacity price c_p. On the other hand, when all the repairs are completed, i.e. when the repair shop becomes idle, the provider buys back the capacity temporarily until the start of the next period at a reduced price, which is lower than c_p. There is also a fixed set-up cost, incurred at the start of each period due to the preparation of the capacity as well as the additional transactions due to the temporary re-assignment of the sold-back capacity.

Similar to the previous capacity modes in the commoditized system environment, we assume that the rental availability of a substitute system upon failure is possible through rental providers. Distinctive from the previous modes, in this mode, a rental hiring duration consists of two parts. The first part, the pre-admission hiring part, starts immediately from the system failure and includes the time until the start of the next interval, when that failed system is admitted to the repair shop. The second part, post-admission hiring part, starts from the admission into the repair shop and targets to alleviate the down-time costs during the time that the defective system spends in the repair shop. The pre-admission hiring duration is variable and non-controllable, whereas the post-admission hiring duration is fixed and to be determined by the MSP. The rental provider agrees to provide a substitute system for both pre-admission and post-admission hiring durations, but charges different prices for them. Under these circumstances, the MSP has to give the optimal post-admission rental hiring duration and the capacity level decisions in order to minimize its TRC. The relevant capacity decisions in this mode are the period length D and the reserved capacity level μ. We aim at building a modeling framework and a decision support system for the MSP operating under the periodic capacity sell-back mode in the commoditized system environment. In Subsection 3.6.1 we describe the dynamics of the pre-admission and post-admission rental hiring as well as the capacity sell-back option, and introduce the additional decision variables and give the problem formulation. In Subsection 3.6.2, the derivation and the analysis of the total relevant cost per unit time are given. Finally in Subsection 3.6.3, we describe the experimental setting, the search procedure and provide the results of the numerical study with a particular focus on the sensitivity of the cost/policy parameters and the savings under this capacity sell-back mode compared to the best cost performance under the reference, fixed capacity mode in the commoditized environment ${TRC_{Fr}}^*$.

3.6.1 Model, Capacity and the Problem Formulation

The MSP operates in the same environment that is explained in the previous sections. Recall that the repair shop is modeled as a single-server queuing system, where the failures occur following a stationary Poisson demand and each defective system requires an exponentially distributed amount of dedicated repair service time in the shop in order to regain its *good as new* status.

In this final mode, upon a system failure, a substitute system is sent immediately from the rental provider. However, the defective system that has failed is not sent to the repair shop immediately, but its shipment is postponed until the start of the next regular interval. The length of each interval, D, is an important decision parameter. For a given interval length D, periodic admission points to the repair shop are introduced at equidistant times: $D, 2D, 3D, \ldots$

In such a case, all the parts that failed within an interval are sent to the repair shop at the end of that interval simultaneously and the operations at the repair shop level resulting from the periodic admission structure can be modeled as a $D^X/M/1$ queue, where X is a discrete random variable which denotes the number of parts failed within an interval. Due to the Poisson arrival stream for the part failures, X also follows a Poisson distribution with mean λD. Note that as D goes to 0, this model transmutes itself to the reference $M/M/1$ model. The analysis and the characteristics of this gated queue type were already given in the previous chapter, Subsection 2.6.2. Further in this section, we will provide an extended analysis that will yield us the distribution of the time that a failed system spends in the repair shop, which is a key performance measure to be used in the service providers' decision making. Under this queuing formalism, the processing rate of the single server queue, μ, represents the capacity level of the repair shop, which is sustained by the capacity agency indefinitely.

Recall that a periodic, capacity sell-back policy, $\mathbf{C} = [D, \mu]$, consists of a period length D, and a processing rate of μ. Under policy $\mathbf{C}$, the MSP closes a contract with a capacity agency. Under this contract, the capacity is provided to the repair shop indefinitely, however frequent job switching and searching for ad-hoc assignments for the idle capacity create an economic factor that lead to an opportunity cost $o_c(D, \Delta, \alpha)$. This opportunity cost decreases with the period length and is dependent on two other cost parameters: Δ and α. The cost structure of $o_c(D, \Delta, \alpha)$ is similar to the structures described in Table 2-1 and Figure 2-2. Due to the $o_c(D, \Delta, \alpha)$, the unit capacity cost rate for the repair shop c_c, is greater than or equal to the c_p. When all the part repairs in the shop are completed before the end of an interval, the capacity is temporarily sold to the capacity agency until the next interval. The sell-back price is lower than c_p: $c_p R$, where $0 < R < 1$, because we assume that there is a risk of not finding an appropriate ad-hoc task and even in the presence of a temporary task, that temporary task can be less profitable than the core repair shop activity.

Suppose the repair shop operates under a stable (i.e. $\mu > \lambda$) policy, $\mathbf{C} = [D, \mu]$, for an infinite horizon. For a given per time unit permanent and contingent capacity costs, c_p and c_c, and a given capacity sell-back cost reduction rate R (for $0 < R < 1$), the average capacity related costs, $CRC(\mathbf{C})$ can be directly calculated as follows:

$$CRC(\mathbf{C}) = c_c\,\mu - c_p\lambda - (c_p R)(\mu - \lambda) = (c_c - c_p R)\mu - c_p\lambda(1 - R) \tag{3.18}$$

In (3.18), we excluded the costs pertaining to the baseline costs ($c_p\lambda$) from the amount that is paid to the capacity agency ($c_c\,\mu$), for deploying μ level of capacity. The repair shop capacity is sold to the capacity agency during the idle times. Therefore the repair shop gains a revenue of: $(\mu - \lambda)c_p\, R$ per time unit.

We assume that the capacity agency offers a set of possible period lengths, θ, from which the service provider can choose the best period length considering the pros and cons of periodic admission and capacity sell-back options.

The rental provider agrees to provide a substitute system during the pre-admission and post-admission hiring durations. The pre-admission hiring duration is random, uncontrollable but it is dependent on the period length D. On the other hand, the post-admission hiring duration, L, is controllable and is a uniform rental hiring duration that is to be determined by the MSP. The rental provision mechanism and the underlying principles of the uniform rental hiring duration/pricing were already discussed in Section 3.3. Parallel to this provision mechanism, it is assumed that the rental provider charges h_r per unit time for the post-admission hiring duration, whereas, particular to the periodic sell-back capacity mode, for the pre-admission hiring duration, a higher rate of h_{rr} is charged, due to the uncertainty in the duration length. Due to the differentiation between pre- and post-admission hiring durations, the reader is strongly recommended to keep in mind that pre-admission hiring duration should not be overlooked during a comparison of hiring durations among different capacity modes in the commoditized environment.

The down-time cost rate is equal to B per time unit, and we assume that $B > h_{rr} \geq h_r$. The hiring rental duration decision is the post-admission hiring duration L. In this section, we assume that all of the down-time during the random pre-admission time is serviced by a substitute hired from the 3rd party supplier at an inflated cost. The capacity related decisions, $\mathbf{C}$, are twofold:

1) Length of the period: D

2) The size of the repair shop capacity level: μ

The total relevant cost function, TRC, can be represented by $\mathbf{C}$ and L, and is the sum of capacity related costs ($CRC(\mathbf{C})$), down-time costs ($DTC(\mathbf{C}, L)$) and hiring related costs ($HC(L)$). Given these cost components and the decision variables, the problem of the MSP can be formulated as follows:

$$\min_{\mathbf{C},L} TRC(\mathbf{C}, L) = CRC(\mathbf{C}) + DTC(\mathbf{C}, L) + HC(L)$$
$$L > 0$$
$$\mathbf{C} = [D, \mu]$$
$$\mu > \lambda, D \in \theta, \tag{3.19}$$

Motivated by the problem formulation above, in the next subsection, we derive the necessary cost functions used in (3.19) and give some of the analytical properties of the components of $TRC(\mathbf{C}, L)$.

3.6.2 Derivation & Analysis of the Cost Functions

In this subsection, we derive the components of the total relevant costs per unit time. The capacity related costs per unit time, $CRC(\mathbf{C})$, for given cost parameters (c_p , Δ, α and R) and policy parameters (μ and D) are already calculated in (3.18), from the previous subsection. In the previous subsection, we have discussed that a defective system first spends time outside the queue until it is admitted, and afterwards it spends time in the repair shop until the repair of that system is completed. The pre-admission hiring time covers for the first waiting time outside the repair shop and the post-admission hiring time covers for the time when the defective system spends in the repair shop. The pre-admission time is different for each case, whereas the post-admission time, L, is uniform and fixed. For a given L, and capacity policy $\mathbf{C} = [D, \mu]$, hiring related costs can be written as follows:

$HC(L) = \lambda \times \left(h_{rr}\frac{D}{2} + h_r L\right)$. The first part of $HC(L)$ corresponds to the pre-admission hiring duration. Since the system failures occur according to a Poisson process, given a failure occurs, time until the start of the next interval follows a uniform distribution between $(0, D)$ (Ross 1983). Therefore the average pre-admission duration length is equal to $\frac{D}{2}$ units of time. The second part of $HC(L)$ is related to the post-admission hiring duration. Although the time that a defective system spends in the repair shop is unknown, a substitute system is hired from a rental supplier for L time units to alleviate the down-time related costs.

Finally, in order to derive the last remaining cost component, which is the down-time related costs per unit time, $DTC(\mathbf{C}, L)$, we need to delve into the detailed modeling of the operations at the repair shop level such as the sojourn time distribution of a defective system in the repair shop, $P\{S_d \geq L | \mathbf{C}\}$ under a given capacity policy, $\mathbf{C}$.

3.6.2.1 Sojourn Time Distribution in a $D^X/M/1$ Queue

Next, we analyze the sojourn time distribution of the $D^X/M/1$ queuing model. The analysis on the sojourn time distribution builds on the preliminary analysis given in Section 2.6.2, therefore we use the same notation in this part. Note that in this part, "customers" refer to the number of defective systems and the capacity level of the repair shop corresponds to the processing rate of the queue.

Recall that at any time point t, the total number of customers, $N_d(t)$, is the sum of the number of customers in the queue (including the one in the service), $N_q(t)$, and the number of customers outside the queue, $N_o(t)$, that are waiting to be admitted. At the start of each period, all customers outside the queue are admitted into the queue based on their arrival order, therefore we have $N_d(nD) = N_q(nD)$ and $N_o(nD) = 0$ for $n = 1,2,\ldots$

In Section 2.6, from equations (2.25-2.28), we have already derived $P\{{N_q}^t = i\}$ and $P\{{N_o}^t = i\}$, which are the limiting probabilities that there will be "i" customers after "t" time units from the start of an arbitrary interval, in the queue and outside the queue waiting to be admitted, respectively, under policy $\mathbf{C} = [D, \mu]$ for $i = 0,1,2,\ldots$ and $t \leq D$.

In the derivation of the sojourn time distribution of an arbitrary customer, we are primarily interested in the number customers when: $t \to D^-$, i.e. at the end of an interval, just before the admission time point. From the previous analysis in Section 2.6, we can derive $P\{N_q{}^{D^-} = i\}$ under any capacity policy $\mathbf{C} = [D, \mu]$. We also know that $P\{N_o{}^{D^-} = i\}$ is independent from the capacity level μ and is Poisson distributed with a mean equal to λD.

The position of an arbitrary arriving customer in the queue, determines that customer's sojourn time in the queue. The position of an arriving customer in the queue depends on the number of customers left over from the previous period as well as the size of the group of customers, which the arriving customer belongs to. Under a given capacity policy **C** = $[D, \mu]$, if the position of a given arriving customer in the queue is n, then the sojourn time of that arriving customer is the sum of n service times in the queue. Since the service time of each customer is exponentially distributed with mean $(1/\mu)$, the sojourn time of an arriving customer whose position in the queue upon arrival is n, will be $Erlang(n, \mu)$ distributed. Due to its critical importance in the derivation of the sojourn time distribution, next we derive the steady state probability that an arbitrary arriving customer's position in the queue is n.

Lemma 3.1. Suppose G denotes the size of the group of customers that an arbitrarily chosen customer belongs to. The probability that an arbitrary arriving customer belongs to a group of customers that has a size of g, $P\{G = g\}$ can be written as:

$$P\{G = g\} = \frac{g \times P\left(N_o{}^{D^-} = g\right)}{E\left(N_o{}^{D^-}\right)} = \frac{g \times \dfrac{e^{-\lambda D}(\lambda D)^g}{g!}}{\lambda D} \tag{3.20}$$

Proof: Suppose the number of the customer group sizes that are admitted to the queue are recorded during a large number (M) of intervals. By the law of large numbers (Ross 1983), for large enough M, the fraction of the groups of size g would go to $P\{N_o{}^{D^-} = g\}$, the number of groups having a size of g would be equal to $M \times P\{N_o{}^{D^-} = g\}$, and the total number of customers coming from a group having size g would be $g \times M \times P\{N_o{}^{D^-} = g\}$. Similarly, the sum of all of the customers from these M groups would be $\sum_{g=0}^{\infty} g \times M \times P\{N_o{}^{D^-} = g\}$. Without loss of generality, assuming that $\binom{0}{1} = 0$, if M goes to infinity, and a randomly customer is picked, the probability that the picked customer is coming from a group of size g can be written for $g = 0,1,2,\ldots$ as below:

$$P\{G = g\} = \lim_{M\to\infty} \frac{\binom{M \times g \times P\{N_o{}^{D^-} = g\}}{1}}{\binom{\sum_{g=0}^{\infty} g \times M \times P\{N_o{}^{D^-} = g\}}{1}} = \frac{g \times \dfrac{e^{-\lambda D}(\lambda D)^g}{g!}}{\lambda D}$$

From Lemma 3.1, we can proceed to the derivation of the probability of the order of an arbitrary arriving customer.

Theorem 3.1. Suppose O denotes the position of an arbitrary arriving customer in the queue. The probability that an arbitrary arriving customer's position is equal to n for $n > 0$ can be written as follows:

$$P\{O = n\} = \sum_{g=n-i}^{\infty} \sum_{i=0}^{n-1} P\{N_q{}^{D^-} = i\} P\{G = g\} \frac{1}{g} = \sum_{g=n-i}^{\infty} \sum_{i=0}^{n-1} P\{N_q{}^{D^-} = i\} \frac{\frac{e^{-\lambda D}(\lambda D)^g}{g!}}{\lambda D} \tag{3.21}$$

Proof: The position of an arriving customer in the queue consists of the number of customers who are left over from the previous period ($N_q{}^{D^-}$) and the newly arrived group of customers, G, of which the corresponding customer belongs to. In order to have an arriving customer having a position of n, we need to have $N_q{}^{D^-} < n$ and $G > \left(n - N_q{}^{D^-}\right)$ and the order of the arbitrary customer in the arriving group should be equal to $\left(n - N_q{}^{D^-}\right)$. Note that if the size of a group is equal to g ($G = g$), the probability that a given customer's order is $\left(n - N_q{}^{D^-}\right)$ within that group is $1/g$. By summing all possible $N_q{}^{D^-}$ and G combinations, we obtain $P\{O = n\}$ in 3.21.

From Theorem 3.1, we can derive $P\{S_d \geq L |\, \mathbf{C}\}$ for any given $\mathbf{C} = [D, \mu]$ as follows:

$$\begin{aligned} P\{S_d \geq L |\, \mathbf{C}\} &= \sum_{n=1}^{\infty} P\{O = n\} P\{S_d \geq L |\, \mathbf{C} \,\&\, O = n\} \\ &= \sum_{n=1}^{\infty} \sum_{g=n-i}^{\infty} \sum_{i=0}^{n-1} P\{N_q{}^{D^-} = i\} \frac{\frac{e^{-\lambda D}(\lambda D)^g}{g!}}{\lambda D} P\{S_d \geq L |\, \mathbf{C} \,\&\, O = n\} \end{aligned} \tag{3.22}$$

Where $(S_d \geq x |\, \mathbf{C} \,\&\, O = n) = \sum_{j=0}^{n-1} e^{-\mu x} \frac{(\mu x)^j}{j!}$, which is the tail distribution of an $Erlang(n, \mu)$ distributed random variable.

Let $f_{S_d}(x|\mathbf{C}) = \frac{\partial(1 - P\{S_d \geq L |\, \mathbf{C}\})}{\partial x}$. $f_{S_d}(x|\mathbf{C})$ can be obtained from 3.22 easily. In the light of discussions above, for a given $\mathbf{C}$ and L , we can derive $DTC(\mathbf{C}, L)$:

$$DTC(\mathbf{C}, L) = \lambda B\left(\mathrm{E}((S_d - L)^+ |\mathbf{C})\right) = \lambda B \int_{x=L}^{\infty} (x - L)\, f_{S_d}(x|\mathbf{C}) \mathrm{d}x \tag{3.23}$$

which finalizes the derivation of $TRC(\mathbf{C}, L)$. Next, we examine some of the properties of the $TRC(\mathbf{C}, L)$ function and the optimization problem in 3.19. These observations/properties are useful since their implications can be exploited during the numerical search procedure that will be given in the next section.

3.6.2.2 Analytical Properties of Total Relevant Cost Function

We know that the optimal rental hiring duration L^* has to satisfy the following first order condition:

$$\frac{\partial TRC(\mathbf{C}, L)}{\partial L} = \lambda \times \left(h_r + B \times P\{S_d \geq L |\, \mathbf{C}\,\}\right) = 0 \rightarrow P\{S_d \geq L |\, \mathbf{C}\,\} = \frac{h_r}{B} \tag{3.24}$$

Therefore, if $P\{S_d \geq L |\, \mathbf{C}\,\} = \frac{h_r}{B}$ is added to the optimization problem in (3.19) as a constraint, there will not be any implications on the optimal solution, since the optimal hiring duration would already satisfy $P\{S_d \geq L |\, \mathbf{C}\,\} = \frac{h_r}{B}$. On the other hand, it may have important consequences in terms of decomposability.

Note that in all of our numerical results, given L and D, we observe that $P\{S_d \geq L |\, \mathbf{C}\,\}$ is increasing with μ. We can observe this behavior in Figure 3-9, where the $P\{S_d \geq L |\, \mathbf{C}\,\}$ values are given for increasing values of μ levels, for $\lambda = 1,\ L = 2$ and $D = 2.5$.

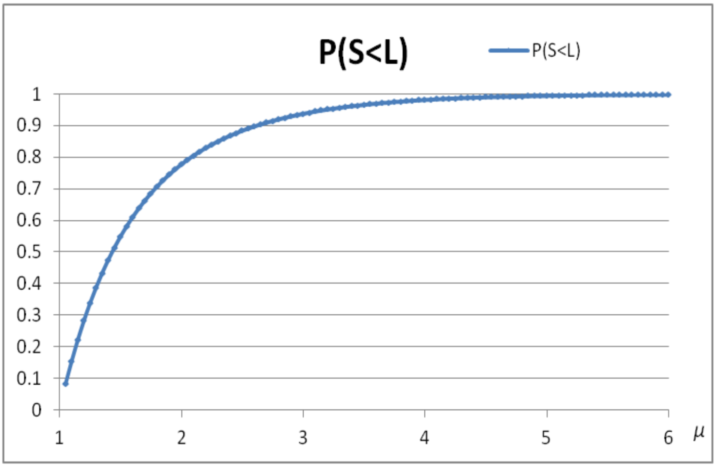

Figure 3-9: The figure depicts how $P\{S_d \geq L |\, \mathbf{C}\,\}$ values are increasing for increasing values of μ levels, when $\lambda = 1, L = 2$ and $D = 2.5$ are given.

Suppose μ^* denotes the capacity, where $P\{S_d \geq L |\, \mathbf{C}\,\} = \frac{h_r}{B}$ when $\mathbf{C} = [D, \mu^*]$, for given $L > 0$ and $D \in \theta$. For a given D, we can always find a μ^* such that $P\{S_d \geq L |\, \mathbf{C}\,\} = \frac{h_r}{B}$, when $\mathbf{C} = [D, \mu^*]$. After μ^* is found for all $L > 0$, $D \in \theta$ candidates, the overall optimal policy parameters can be found from (3.25):

$$\min_{D,L} TRC(\mathbf{C}, L) = CRC(\mathbf{C}) + DTC(\mathbf{C}, L) + HC(L)$$
$$s.t. \quad L \geq 0$$
$$\mathbf{C} = [D, \mu^*]$$
$$D \in \theta$$

(3.25)

3.6.3 Numerical Study

In this section, we use the analysis and the results provided in the previous section in order to assess the performance of the capacity sell-back mode under rental availability. In the end, we compare the cost performance of this mode with the minimum cost rate achieved under the fixed capacity mode with rental availability, ${TRC_{Fr}}^*$. The outline of this section is as follows.

First, we describe the characteristics and the test bed of the computational study. Second, we present the search procedure in order to find the near-optimal policy parameters for the problem given in (3.25). Finally, we discuss the results of the search procedure. First we assess the potential cost benefits that can be gained in capacity sell-back mode in comparison to the total costs under the optimal policy under the fixed capacity mode with rental availability. Afterwards, we examine the sensitivity of and the interactions among the policy and system parameters afterwards, the accuracy of the finite waiting room approximation in the analytical calculations will be assessed.

In our computational study, we normalize the arrival rate for the system failures, $\lambda = 1$ (failures per time unit) as well as the unit cost per time unit for contingent capacity, $c_p = 1$. The other parameters are scaled according to these normalized λ and c_p parameters. Similar to 3.4.3, we analyze a total of 9 scenarios with three different B/h_r values for 3 different h_r values. The test bed are already given in Table 3-1.

For each of these 9 $\left(\frac{B}{h_r}, h_r\right)$ scenarios and different (Δ, α, R) and h_{rr} combinations, we execute the Search Procedure-V to find the capacity policy parameters, (D, μ) and the rental duration L that yield the minimum total costs.

The periodic admission D determines the frequency of defective system admissions as well as the frequency of the capacity sell-back actions, which is chosen from θ, the set of candidate period lengths offered. In our thesis we assume $\theta = \{0.5, 1, 1.5, \dots, 4.5, 5\}$, which are scaled to the normalized inter-arrival time $\frac{1}{\lambda} = 1$.

Note that all the derivations/equations in 3.6.2 pertain to infinite state variables, which can be a limitation of the predictive power of the computational study. In the numerical calculations the infinite state system of equations is truncated to a large enough finite (with a large waiting room of K) ones. Simulation studies that we conduct, which will be explained at the end of this subsection, suggest that the analytical finite system approximation mimics the performance of the infinite state system almost perfectly.

Unlike the fixed capacity mode, we cannot find the global optimal hiring duration and the optimal capacity policy parameters that minimize the total costs in problem (3.25) analytically. Therefore, we develop a solution procedure, which is specifically designed for the optimization problem (3.25). The fixed capacity results $({L_{Fr}}^*, {\mu_{Fr}}^*)$ obtained from Subsection 4.4.3 serve as a reference point to assess the overall cost performance (savings performance of capacity sell-back mode with respect to ${TRC_{Fr}}^*$).

In this capacity mode, we can derive the total relevant costs $TRC(\mathbf{C}, L)$ for a given capacity policy $\mathbf{C}$ and a given hiring duration L. Due to the fact that the optimal capacity policy satisfies $P\{S_d \geq L| \mathbf{C}\} = h_r/B$, we use this constraint to decompose the main optimization problem into the sub-problems with different D and L. Each sub-problem finds the $\mu^*(D, L)$ that satisfies $P\{S_d \geq L| \mathbf{C}\} = h_r/B$ for given D and L. After $\mu^*(D, L)$ is found for all possible D and L, their total cost performance under these candidate policies are obtained from discrete event simulation, because it is computationally more economical to analytical computation of $TRC(\mathbf{C}, L)$, which requires numerical integration until infinity.

In the commoditized environment, rental hiring duration L is a continuous decision variable. The computational requirements for deriving the total cost performance in this section is more demanding than the previous chapter, which enforces us to focus on a limited number of hiring duration possibilities. Therefore, an affordable set for candidate hiring durations has to be generated before performing the search procedure developed to solve (3.25). In this section, we use a uniform candidate hiring duration set ψ for each of the 9 different $\left(\frac{B}{h_r}, h_r\right)$ combinations, which consists of equidistant points (we took a distance of 0.1 in our thesis), in the $(0, L_{max})$ interval. We take $L_{max} = 20$. Considering the similarity and the correspondence between stock level S and the hiring duration L, $L_{max} = 20$ can be considered as a safe maximum cap. As we have determined the candidate hiring duration set, ψ, next we give our search procedure followed by the description of its underlying mechanism.

Search Procedure-V

---

0. For every (L, D) combinations in $L \in \psi$ and $D \in \theta$:
    a. Find $\mu^*(D, L)$ such that $P\{S_d \geq L| \mathbf{C}\} = \frac{h_r}{B}$, when $\mathbf{C} = [D,\ \mu^*(D, L)]$
1. After finding $\mu^*(D, L)$ for all $D \in \theta$, we can find the minimum cost for given L, $TRC(\mathbf{C}^*(L), L)$ as follows:

$$TRC(\mathbf{C}^*(L), L) = \min_{D \in \theta,}(TRC(\mathbf{C}, L): \mathbf{C} = [D,\ \mu^*(D, L)])$$

2. After $TRC(\mathbf{C}^*(L), L)$ for all $L \in \psi$ are found, we can find L^*, and $\mathbf{C}^*(L^*)$ = [D^*, $\mu^*(D^*, L^*)$] values, which would give the global minimum cost rate for problem (4.25): $TRC^* = TRC(\mathbf{C}^*(L^*), L^*)$ for capacity sell-back mode.

   L^* can be found from brute force search as follows:

$$TRC(\mathbf{C}^*(L^*), L^*) = \min_{L \in \psi} TRC(\mathbf{C}^*(L), L)$$

In the search procedure above, at step 0, we find the $\mu^*(D, L)$ that satisfies: $P\{S_d \geq L | \mathbf{C}\}$, when $\mathbf{C} = [D, \mu^*(D, L)]$ for a given $L \in \psi$. After $\mu^*(D, L)$ is found for each candidate hiring duration $L \in \psi$ and for each period length $D \in \theta$, we find the optimal parameters (L^*, D^* and $\mu^*(D^*, L^*)$) by brute force search in steps 1 & 2. This search completes the solution procedure for our problem.

Next, we present and discuss the results of the numerical study that is conducted, where the optimal costs for problem (3.25) are obtained from the Search Procedure-V that is described above. We first discuss the savings in total relevant costs when the best capacity sell-back policy is employed compared to the best fixed capacity system where the rental availability is present (${TRC_{Fr}}^*$ in short), afterwards the accuracy of the finite waiting room will be checked via a simulation study.

3.6.3.1 Savings Compared to the TRC under the Optimal Single Level Capacity ($TRC^*{}_F$)

In this section, we envisage the cost saving prospects of the periodic sell-back capacity mode. Therefore the optimal costs achieved under the third capacity mode will be compared to the $TRC^*{}_{Fr}$ achieved from the single capacity mode. It has been observed that, the capacity sell-back option can reduce the total costs up to 66%. Even under the hypothetical case where the provider agrees to pay the total of c_p per unit time during the repair shop's idle times (i.e. $R = 1$), the reduction can be up to 98% under some instances. On the other hand, for some of the cost parameter instances, especially when the maximum opportunity cost is high and insensitive to the period length and when there is no (or very limited) sell-back opportunity (i.e. $R = 0$), the capacity sell-back policy may lead to losses rather than savings in TRC. The cost increase can be drastic (up to 142%), therefore the implementation of the third capacity mode and the feasibility of the cost parameters have to be scrutinized additionally, taking the cost parameters into account, due to the potential serious consequences.

In Table 3-11, we give the maximum percentage savings that capacity sell-back flexible policies can bring for all 9 different B & h_r scenarios (which are already listed in Table 3-1) with $\Delta = 0, 0.25, 0.5$ and 1 and $\alpha = 0{,}1$ and 2, when $h_r = h_{rr}$, and when c_c is an inversely proportional, exponential and linear function of the period length respectively, for $R = 0, 0.1, 0.5, 0.9$ and 1. Let ${TRC_{3r}}^*$ represent the minimum total costs that can be achieved from Search Procedure-V, for given cost parameters ($B, h_r, R, \Delta, h_{rr}$ and α), and a functional form for c_c. After $TRC_{3r}{}^*$ is found, the % savings in Table 3-11 can be calculated from: $\frac{(TRC_{Fr}{}^* - TRC_{3r}{}^*)}{TRC_{Fr}{}^*}$. The cells are color coded according to their percentage saving values. If $TRC_{3r}{}^*$ is higher than $TRC_{Fr}{}^*$, the corresponding cell is white. On the other hand, if $TRC_{3r}{}^*$ is smaller, the corresponding cell is shaded in gray, where the higher percentage savings have darker gray tones.

From Table 3-11, similar to the specialized system setting, we can observe that the periodic capacity sell-back mode is outperformed to the fixed capacity mode, especially for low sell back rates (e.g. $R = 0, 0.1$), high opportunity cost factors (e.g. $\Delta = 1$) and low time elasticity (e.g. $\alpha = 0$).

The effects of the cost parameters on the cost performance of the optimal sell-back capacity policy in the commoditized system environment seem to resemble the effects of the cost parameters on the optimal cost performance under sell-back mode, in the specialized system environment. For instance, percentage savings/losses seem to decrease with hiring cost rate h_r. This can due to the fact that for higher hiring cost rate h_r, although the absolute change in costs due to the sell-back capacity mode gets bigger, percent wise it gets smaller, since the reference cost parameter, $TRC_{Fr}{}^*$, is greater and the share of the holding costs (HC) in $TRC_{Fr}{}^*$ is bigger for larger h_r. Parallel to our observations before, the percentage savings increase/percentage losses generally decrease with the elasticity factor α. This can be explained as follows: the more elastic the contingent capacity cost gets (with respect to the period length D), the cheaper contingent capacity becomes, which leads to additional savings or alleviation of the losses. On the other hand, the maximum opportunity cost, Δ, has an adversary effect, since higher Δ causes the contingent capacity to be more expensive, which leads to an increase in total costs , $TRC_{3r}{}^*$.

From Table 3-11, we can also see that an increase in B most of the time accompanies an increase in percentage savings / a decrease in the percentage losses.

The functional form of c_c plays an important role in the cost performance of the capacity sell-back mode when $\Delta > 0$ and $\alpha > 0$, because in the absence of the maximum opportunity cost, or absence of the elasticity to the period length, all three functional forms yield the same c_c and therefore the same $TRC^*{}_{3r}$. On the other hand, for $\Delta > 0$ and $\alpha > 0$, under all B, h_r, h_{rr}, Δ & α , the linear functional form appears to be the form that results in the greatest savings (or the least loss) and the inverse proportional form appears to be the form that results in the least savings (or greatest loss).

From Table 3-11, it is remarkable that in all of the 9 B & h_r combinations and under all of the Δ & α parameters, the sell-back rate R is the primary determinant of the cost savings (losses). For each B, h_r, Δ & α quartet, when $R = 0$, the $TRC^*{}_{3r}$ is always surpassed by $TRC^*{}_{Fr}$ (the % difference can be up to -147%), and as R increases, the gap between $TRC^*{}_{3r}$ and $TRC^*{}_{Fr}$ decreases. When $\alpha > 0$, there is a threshold R, after which $TRC^*{}_{3r}$ starts to outperform $TRC^*{}_{Fr}$, and the cost savings of the capacity sell-back policy increases after that threshold R value. This threshold R value is lower, when the capacity price , c_c is lower, i.e. when Δ is small, when α is high and when c_c has a linear cost structure.

Unlike in the two-level capacity mode, in the periodic sell-back mode, we notice that the percentage cost savings are higher in the commoditized system environment than the cost savings in the specialized system environment given in Table 2-12. This can be misleading, because in the commoditized system environment, in Table 3-11, the hiring in the pre-admission period is charged identically as the post-admission period (i.e. $h_r = h_{rr}$). However in real-life environments, we may expect that $h_r < h_{rr}$, since the pre-admission hiring duration is uncertain and ad-hoc. Therefore, the savings in Table 3-11 should be interpreted as upper bounds, instead.

To give a more representative idea on the savings from periodic sell-back capacity policy, we can also describe an upper bound for the $TRC^*{}_{3r}$, when we set $h_{rr} = B$. As it can be seen in Table 3-12, when $h_{rr} = B$, the percentage cost savings have decreased substantially compared to the cost savings in Table 3-11 for the exponential cost structure. A similar decrease in costs is also evident for linear and inverse proportional cost structures, as well. Yet, we present the percentage savings of the sell-back capacity mode only for the exponential cost structure due to the space limitations.

If we compare the percentage savings of the commoditized environment with the specialized environment when $h_{rr} = B$, we observe that, parallel to the two-level capacity mode, the percentage savings in the commoditized environment are lower compared to the savings in the specialized setting, which can be traced back in Table 2-12. This can be explained by the fact that the cost parameters related to the operating systems differ greatly for the commoditized setting from the specialized setting. For instance the unit time holding cost per a critical subsystem is lower than the unit time hiring cost of a substitute system and down-time costs are higher in specialized system environment. Higher B/h and c_p/h ratios (compared to B/h_r and c_p/h_r) make the role of the flexible capacity policies in the specialized system more critical than in the commoditized system.

Finally, we explore further how the optimal policy parameters change under the optimal capacity sell-back mode compared to the single level capacity mode for different cost parameter settings. In Table 3-13 we show how the optimal periodic sell-back capacity mode policy parameters (L^*, D^* and μ^*) differ with various (Δ, α) combinations and 4 different B & h scenarios. The L^* and μ^* values from the periodic sell-back capacity mode, which are higher than the reference optimal L^* and μ^* values from the optimal single-level capacity policies are highlighted.

The data in Table 3-13 illustrate that, under the optimal policies pertaining to the periodic sell-back capacity mode, the rental hiring duration is not necessarily lower than the rental hiring duration under the optimal single-level capacity policy. Contrary, we observe that when the sell-back rate R is low, the sell-back capacity policy most of the time leads to higher rental durations than the fixed capacity policy. For longer R parameters, the rental hiring duration tends to decrease and when $R = 1$, if there is no maximum opportunity costs, the optimal post-admission rental hiring duration goes to

0. The pre-admission rental hiring still exists, because it is independent of the capacity level, but is affected by only the failure arrival rate and the period length.

The declining of L^* with R is actually due to the incline of μ^* with R. The optimal capacity level, μ^* increases with R, because when there are more selling-back possibilities, the MSP is incentivized to buy more capacity in the periodic sell-back capacity mode than it would necessitate under the fixed capacity mode. An interesting hypothetical case is when there is no opportunity cost and when the full sell-back rate is possible, the MSP would choose to install infinite amount of capacity to achieve instantaneous repair, since the idle capacity can be immediately sold back without loss of any revenue.

In the two-level capacity policy, we have observed that both the average deployed capacity ($ACU(\mathbf{C})$) and the optimal hiring duration are lower than the optimal capacity level ${\mu_F}^*$ and the optimal hiring duration ${L_{Fr}}^*$ under the fixed capacity mode. This is different in the periodic sell-back capacity mode, where at least one of the optimal capacity and the hiring duration is higher than the optimal fixed capacity parameters.

The response of the optimal hiring duration to different cost parameters resemble the response of the optimal stock level under sell-back capacity mode in specialized system. L^* tends to decrease with higher h_r and higher α (parallel to other capacity modes) and with higher sell-back rate R values; and tends to increase with higher B and Δ. The optimal capacity level μ^* under the periodic sell-back capacity mode tends to increase with almost all cost parameters (h_r, B, α) except for the maximum opportunity cost Δ. Although μ^* from the periodic sell-back capacity mode can be a lot higher than the ${\mu^*}_{Fr}$ from the single-level capacity mode, one should always keep it in mind that, in the periodic sell-back capacity mode, more capacity does not always lead to higher capacity related costs (CRC), since the excess idle capacity can be sold back to the agency and due to this sell-back opportunity, it may be more profitable to deploy higher capacity compared to the ${\mu^*}_{Fr}$ from the fixed capacity mode. The main explanation of these responses is as follows: when the changed cost parameter is increasing, the capacity unit cost per time directly, MSP installs less capacity, which leads to longer hiring duration to cover the halted operations of the system-owner. Similarly if the changed cost parameter is directly increasing the hiring unit cost per time, MSP would hire a substitute for smaller periods which leads an increase in the capacity level to compensate for shorter hiring lengths. Besides these, there can be cost parameters that do not directly affect neither the capacity nor the hiring duration costs, but increase μ^*, L^* simultaneously.

After this brief comparison between the optimal policy parameters of the fixed capacity policy and the periodic sell-back capacity policy, in the next subsection, we check the accuracy of using the finite waiting room approximation for the repair shop.

Inverse Proportional

h_r = 0.1	B=0.5		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-142%	-138%	-123%	-105%	-100%
		α=1	-58%	-54%	-33%	-5%	4%
		α=2	-42%	-37%	-16%	14%	26%
	Δ=0.5	α=0	-74%	-70%	-52%	-29%	-22%
		α=1	-38%	-33%	-12%	18%	30%
		α=2	-28%	-23%	-1%	32%	45%
	Δ=0.25	α=0	-39%	-34%	-14%	14%	23%
		α=1	-24%	-19%	3%	35%	49%
		α=2	-18%	-13%	9%	44%	59%
	Δ=0	all α	-3%	2%	27%	66%	98%

h_r = 0.1	B=1		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-130%	-127%	-112%	-94%	-89%
		α=1	-53%	-49%	-28%	-1%	9%
		α=2	-38%	-34%	-12%	18%	29%
	Δ=0.5	α=0	-68%	-64%	-46%	-23%	-17%
		α=1	-34%	-30%	-8%	22%	33%
		α=2	-25%	-21%	1%	34%	47%
	Δ=0.25	α=0	-36%	-31%	-11%	17%	26%
		α=1	-22%	-17%	5%	37%	51%
		α=2	-16%	-11%	11%	45%	61%
	Δ=0	all α	-2%	3%	27%	66%	98%

h_r = 0.1	B=2		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-122%	-119%	-104%	-86%	-81%
		α=1	-49%	-45%	-25%	3%	12%
		α=2	-36%	-31%	-10%	20%	32%
	Δ=0.5	α=0	-64%	-60%	-42%	-19%	-12%
		α=1	-32%	-28%	-6%	24%	35%
		α=2	-23%	-19%	3%	35%	49%
	Δ=0.25	α=0	-34%	-29%	-8%	19%	28%
		α=1	-20%	-15%	6%	39%	52%
		α=2	-15%	-10%	12%	46%	62%
	Δ=0	all α	-2%	3%	27%	66%	98%

h_r = 0.2	B=1		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-114%	-111%	-95%	-78%	-73%
		α=1	-58%	-54%	-34%	-8%	1%
		α=2	-43%	-38%	-17%	12%	22%
	Δ=0.5	α=0	-61%	-57%	-38%	-16%	-9%
		α=1	-37%	-33%	-11%	17%	28%
		α=2	-28%	-23%	-2%	30%	42%
	Δ=0.25	α=0	-33%	-28%	-8%	20%	29%
		α=1	-23%	-18%	4%	35%	47%
		α=2	-18%	-14%	9%	42%	57%
	Δ=0	all α	-4%	1%	25%	64%	97%

h_r = 0.2	B=2		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-106%	-103%	-87%	-70%	-65%
		α=1	-53%	-49%	-29%	-3%	5%
		α=2	-39%	-35%	-14%	15%	25%
	Δ=0.5	α=0	-57%	-52%	-34%	-12%	-5%
		α=1	-34%	-30%	-9%	20%	31%
		α=2	-26%	-21%	1%	32%	44%
	Δ=0.25	α=0	-31%	-26%	-5%	22%	31%
		α=1	-22%	-17%	5%	37%	49%
		α=2	-17%	-12%	11%	44%	58%
	Δ=0	all α	-3%	2%	26%	65%	97%

h_r = 0.2	B=4		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-101%	-97%	-81%	-64%	-59%
		α=1	-50%	-46%	-26%	0%	9%
		α=2	-37%	-32%	-11%	17%	28%
	Δ=0.5	α=0	-53%	-49%	-31%	-9%	-2%
		α=1	-32%	-28%	-7%	22%	33%
		α=2	-24%	-20%	2%	33%	46%
	Δ=0.25	α=0	-29%	-24%	-4%	24%	33%
		α=1	-20%	-15%	7%	38%	51%
		α=2	-16%	-11%	12%	45%	60%
	Δ=0	all α	-3%	2%	26%	65%	97%

h_r = 1	B=5		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-81%	-78%	-62%	-45%	-40%
		α=1	-58%	-54%	-35%	-13%	-6%
		α=2	-46%	-41%	-22%	4%	12%
	Δ=0.5	α=0	-47%	-43%	-24%	-2%	5%
		α=1	-35%	-30%	-11%	15%	23%
		α=2	-28%	-24%	-3%	24%	35%
	Δ=0.25	α=0	-28%	-24%	-3%	24%	33%
		α=1	-22%	-17%	4%	34%	45%
		α=2	-19%	-14%	8%	40%	52%
	Δ=0	all α	-9%	-4%	20%	59%	92%

h_r = 1	B=10		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-77%	-74%	-58%	-41%	-36%
		α=1	-54%	-50%	-32%	-9%	-3%
		α=2	-42%	-38%	-19%	7%	15%
	Δ=0.5	α=0	-44%	-40%	-22%	1%	7%
		α=1	-33%	-28%	-9%	17%	25%
		α=2	-27%	-22%	-2%	26%	37%
	Δ=0.25	α=0	-27%	-22%	-2%	26%	35%
		α=1	-21%	-16%	6%	36%	46%
		α=2	-18%	-13%	9%	41%	53%
	Δ=0	all α	-8%	-3%	21%	60%	93%

h_r = 1	B=20		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-74%	-70%	-55%	-37%	-33%
		α=1	-52%	-48%	-29%	-7%	0%
		α=2	-41%	-36%	-17%	9%	17%
	Δ=0.5	α=0	-42%	-38%	-20%	2%	9%
		α=1	-31%	-27%	-7%	18%	27%
		α=2	-25%	-21%	0%	27%	38%
	Δ=0.25	α=0	-25%	-21%	0%	27%	36%
		α=1	-20%	-15%	7%	36%	47%
		α=2	-17%	-12%	10%	42%	54%
	Δ=0	all α	-8%	-2%	22%	61%	94%

Exponential

h_r = 0.1	B=0.5		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-142%	-138%	-123%	-105%	-100%
		α=1	-29%	-24%	-1%	34%	61%
		α=2	-16%	-11%	12%	50%	78%
	Δ=0.5	α=0	-74%	-70%	-52%	-29%	-22%
		α=1	-23%	-19%	4%	40%	65%
		α=2	-13%	-8%	15%	52%	80%
	Δ=0.25	α=0	-39%	-34%	-14%	14%	23%
		α=1	-18%	-13%	10%	46%	69%
		α=2	-11%	-6%	18%	55%	82%
	Δ=0	all α	-3%	2%	27%	66%	98%

h_r = 0.1	B=1		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-130%	-127%	-112%	-94%	-89%
		α=1	-26%	-22%	1%	37%	64%
		α=2	-14%	-10%	14%	51%	80%
	Δ=0.5	α=0	-68%	-64%	-46%	-23%	-17%
		α=1	-21%	-16%	6%	42%	68%
		α=2	-12%	-7%	16%	53%	82%
	Δ=0.25	α=0	-36%	-31%	-11%	17%	26%
		α=1	-16%	-11%	12%	48%	72%
		α=2	-10%	-5%	19%	56%	83%
	Δ=0	all α	-2%	3%	27%	66%	98%

h_r = 0.1	B=2		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-122%	-119%	-104%	-86%	-81%
		α=1	-24%	-20%	3%	39%	66%
		α=2	-13%	-8%	15%	52%	81%
	Δ=0.5	α=0	-64%	-60%	-42%	-19%	-12%
		α=1	-20%	-15%	8%	43%	70%
		α=2	-11%	-6%	17%	54%	83%
	Δ=0.25	α=0	-34%	-29%	-8%	19%	28%
		α=1	-15%	-10%	13%	49%	73%
		α=2	-9%	-4%	19%	57%	84%
	Δ=0	all α	-2%	3%	27%	66%	98%

h_r = 0.2	B=1		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-114%	-111%	-95%	-78%	-73%
		α=1	-35%	-30%	-8%	27%	51%
		α=2	-20%	-15%	8%	45%	72%
	Δ=0.5	α=0	-61%	-57%	-38%	-16%	-9%
		α=1	-27%	-22%	0%	35%	57%
		α=2	-17%	-12%	12%	49%	74%
	Δ=0.25	α=0	-33%	-28%	-8%	20%	29%
		α=1	-19%	-15%	8%	42%	62%
		α=2	-13%	-8%	16%	52%	77%
	Δ=0	all α	-4%	1%	25%	64%	97%

h_r = 0.2	B=2		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-106%	-103%	-87%	-70%	-65%
		α=1	-32%	-27%	-5%	30%	55%
		α=2	-18%	-13%	10%	46%	74%
	Δ=0.5	α=0	-57%	-52%	-34%	-12%	-5%
		α=1	-25%	-20%	3%	37%	60%
		α=2	-15%	-10%	13%	50%	76%
	Δ=0.25	α=0	-31%	-26%	-5%	22%	31%
		α=1	-18%	-13%	10%	44%	65%
		α=2	-12%	-6%	17%	53%	79%
	Δ=0	all α	-3%	2%	26%	65%	97%

h_r = 0.2	B=4		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-101%	-97%	-81%	-64%	-59%
		α=1	-30%	-25%	-3%	32%	58%
		α=2	-17%	-12%	11%	48%	76%
	Δ=0.5	α=0	-53%	-49%	-31%	-9%	-2%
		α=1	-23%	-18%	4%	38%	63%
		α=2	-14%	-9%	14%	51%	78%
	Δ=0.25	α=0	-29%	-24%	-4%	24%	33%
		α=1	-17%	-12%	11%	45%	67%
		α=2	-10%	-5%	18%	54%	80%
	Δ=0	all α	-3%	2%	26%	65%	97%

h_r = 1	B=5		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-81%	-78%	-62%	-45%	-40%
		α=1	-48%	-44%	-24%	5%	20%
		α=2	-31%	-26%	-4%	29%	50%
	Δ=0.5	α=0	-47%	-43%	-24%	-2%	5%
		α=1	-33%	-28%	-8%	21%	33%
		α=2	-23%	-19%	3%	37%	56%
	Δ=0.25	α=0	-28%	-24%	-3%	24%	33%
		α=1	-21%	-16%	5%	36%	48%
		α=2	-16%	-11%	11%	44%	62%
	Δ=0	all α	-9%	-4%	20%	59%	92%

h_r = 1	B=10		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-77%	-74%	-58%	-41%	-36%
		α=1	-45%	-41%	-20%	9%	25%
		α=2	-28%	-24%	-2%	32%	54%
	Δ=0.5	α=0	-44%	-40%	-22%	1%	7%
		α=1	-31%	-26%	-6%	24%	37%
		α=2	-22%	-17%	5%	39%	59%
	Δ=0.25	α=0	-27%	-22%	-2%	26%	35%
		α=1	-20%	-15%	7%	37%	50%
		α=2	-15%	-10%	12%	45%	65%
	Δ=0	all α	-8%	-3%	21%	60%	93%

h_r = 1	B=20		R=0	R=0.1	R=0.5	R=0.9	R=1
	Δ=1	α=0	-74%	-70%	-55%	-37%	-33%
		α=1	-43%	-38%	-18%	12%	30%
		α=2	-27%	-22%	0%	34%	57%
	Δ=0.5	α=0	-42%	-38%	-20%	2%	9%
		α=1	-29%	-25%	-4%	25%	40%
		α=2	-21%	-16%	6%	41%	62%
	Δ=0.25	α=0	-25%	-21%	0%	27%	36%
		α=1	-19%	-14%	8%	38%	51%
		α=2	-14%	-9%	13%	46%	67%
	Δ=0	all α	-8%	-2%	22%	61%	94%

Linear

h_r	B=0.5		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.1	Δ=1	α=0	-142%	-138%	-123%	-105%	-100%
		α=1	-6%	-1%	23%	62%	94%
		α=2	-3%	2%	27%	66%	98%
	Δ=0.5	α=0	-74%	-70%	-52%	-29%	-22%
		α=1	-3%	2%	27%	66%	98%
		α=2	-3%	2%	27%	66%	98%
	Δ=0.25	α=0	-39%	-34%	-14%	14%	23%
		α=1	-3%	2%	27%	66%	98%
		α=2	-3%	2%	27%	66%	98%
	Δ=0	all α	-3%	2%	27%	66%	98%

h_r	B=1		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.1	Δ=1	α=0	-130%	-127%	-112%	-94%	-89%
		α=1	-5%	0%	24%	63%	94%
		α=2	-2%	3%	27%	66%	98%
	Δ=0.5	α=0	-68%	-64%	-46%	-23%	-17%
		α=1	-2%	3%	27%	66%	98%
		α=2	-2%	3%	27%	66%	98%
	Δ=0.25	α=0	-36%	-31%	-11%	17%	26%
		α=1	-2%	3%	27%	66%	98%
		α=2	-2%	3%	27%	66%	98%
	Δ=0	all α	-2%	3%	27%	66%	98%

h_r	B=2		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.1	Δ=1	α=0	-122%	-119%	-104%	-86%	-81%
		α=1	-4%	0%	24%	63%	95%
		α=2	-2%	3%	27%	66%	98%
	Δ=0.5	α=0	-64%	-60%	-42%	-19%	-12%
		α=1	-2%	3%	27%	66%	98%
		α=2	-2%	3%	27%	66%	98%
	Δ=0.25	α=0	-34%	-29%	-8%	19%	28%
		α=1	-2%	3%	27%	66%	98%
		α=2	-2%	3%	27%	66%	98%
	Δ=0	all α	-2%	3%	27%	66%	98%

h_r	B=1		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.2	Δ=1	α=0	-114%	-111%	-95%	-78%	-73%
		α=1	-9%	-4%	21%	60%	91%
		α=2	-4%	1%	25%	64%	97%
	Δ=0.5	α=0	-61%	-57%	-38%	-16%	-9%
		α=1	-4%	1%	25%	64%	97%
		α=2	-4%	1%	25%	64%	97%
	Δ=0.25	α=0	-33%	-28%	-8%	20%	29%
		α=1	-4%	1%	25%	64%	97%
		α=2	-4%	1%	25%	64%	97%
	Δ=0	all α	-4%	1%	25%	64%	97%

h_r	B=2		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.2	Δ=1	α=0	-125%	-122%	-108%	-92%	-87%
		α=1	-41%	-37%	-18%	7%	15%
		α=2	-29%	-25%	-6%	22%	31%
	Δ=0.5	α=0	-65%	-61%	-44%	-24%	-18%
		α=1	-27%	-23%	-3%	24%	34%
		α=2	-20%	-16%	5%	34%	46%
	Δ=0.25	α=0	-34%	-30%	-11%	14%	23%
		α=1	-18%	-13%	7%	37%	48%
		α=2	-13%	-9%	12%	43%	57%
	Δ=0	all α	-2%	3%	25%	61%	90%

h_r	B=4		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.2	Δ=1	α=0	-101%	-97%	-81%	-64%	-59%
		α=1	-7%	-2%	22%	61%	93%
		α=2	-3%	2%	26%	65%	97%
	Δ=0.5	α=0	-53%	-49%	-31%	-9%	-2%
		α=1	-3%	2%	26%	65%	97%
		α=2	-3%	2%	26%	65%	97%
	Δ=0.25	α=0	-29%	-24%	-4%	24%	33%
		α=1	-3%	2%	26%	65%	97%
		α=2	-3%	2%	26%	65%	97%
	Δ=0	all α	-3%	2%	26%	65%	97%

h_r	B=5		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 1	Δ=1	α=0	-81%	-78%	-62%	-45%	-40%
		α=1	-20%	-15%	9%	49%	80%
		α=2	-9%	-4%	20%	59%	92%
	Δ=0.5	α=0	-47%	-43%	-24%	-2%	5%
		α=1	-9%	-4%	20%	59%	92%
		α=2	-9%	-4%	20%	59%	92%
	Δ=0.25	α=0	-28%	-24%	-3%	24%	33%
		α=1	-9%	-4%	20%	59%	92%
		α=2	-9%	-4%	20%	59%	92%
	Δ=0	all α	-9%	-4%	20%	59%	92%

h_r	B=10		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 1	Δ=1	α=0	-77%	-74%	-58%	-41%	-36%
		α=1	-18%	-13%	11%	51%	82%
		α=2	-8%	-3%	21%	60%	93%
	Δ=0.5	α=0	-44%	-40%	-22%	1%	7%
		α=1	-8%	-3%	21%	60%	93%
		α=2	-8%	-3%	21%	60%	93%
	Δ=0.25	α=0	-27%	-22%	-2%	26%	35%
		α=1	-8%	-3%	21%	60%	93%
		α=2	-8%	-3%	21%	60%	93%
	Δ=0	all α	-8%	-3%	21%	60%	93%

h_r	B=20		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 1	Δ=1	α=0	-74%	-70%	-55%	-37%	-33%
		α=1	-17%	-12%	12%	52%	84%
		α=2	-8%	-2%	22%	61%	94%
	Δ=0.5	α=0	-42%	-38%	-20%	2%	9%
		α=1	-8%	-2%	22%	61%	94%
		α=2	-8%	-2%	22%	61%	94%
	Δ=0.25	α=0	-25%	-21%	0%	27%	36%
		α=1	-8%	-2%	22%	61%	94%
		α=2	-8%	-2%	22%	61%	94%
	Δ=0	all α	-8%	-2%	22%	61%	94%

Table 3-11 The % cost savings of the capacity sell-back mode compared to the fixed capacity, when $h_r = 0.1, 0.2,\ 1$ and $B/h_r = 5, 10$ and 20, $h_r = h_{rr}$ for Δ= 0,0.25,0.5 and 1, $\alpha = 0, 1$ and 2 when $R = 0, 0.1, 0.5, 0.9, 1$ when c_c :inverse proportional, exponential and linear structure.

Exponential

h_r	B=0.5		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.1	Δ=1	$\alpha=0$	-151%	-148%	-133%	-115%	-110%
		$\alpha=1$	-71%	-67%	-46%	-15%	-3%
		$\alpha=2$	-45%	-40%	-18%	17%	33%
	Δ=0.5	$\alpha=0$	-84%	-80%	-61%	-39%	-32%
		$\alpha=1$	-52%	-47%	-26%	4%	16%
		$\alpha=2$	-35%	-30%	-8%	27%	43%
	Δ=0.25	$\alpha=0$	-49%	-44%	-24%	4%	13%
		$\alpha=1$	-34%	-30%	-8%	22%	35%
		$\alpha=2$	-26%	-21%	2%	35%	52%
	Δ=0	all α	-12%	-7%	17%	56%	88%

h_r	B=1		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.1	Δ=1	$\alpha=0$	-150%	-146%	-131%	-114%	-109%
		$\alpha=1$	-93%	-89%	-69%	-44%	-35%
		$\alpha=2$	-63%	-58%	-36%	-5%	7%
	Δ=0.5	$\alpha=0$	-88%	-84%	-65%	-43%	-36%
		$\alpha=1$	-62%	-58%	-38%	-12%	-3%
		$\alpha=2$	-47%	-42%	-21%	9%	23%
	Δ=0.25	$\alpha=0$	-56%	-51%	-30%	-3%	6%
		$\alpha=1$	-42%	-38%	-16%	15%	26%
		$\alpha=2$	-34%	-30%	-7%	26%	39%
	Δ=0	all α	-22%	-17%	7%	46%	78%

h_r	B=2		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.1	Δ=1	$\alpha=0$	-160%	-157%	-141%	-124%	-119%
		$\alpha=1$	-114%	-110%	-93%	-72%	-66%
		$\alpha=2$	-85%	-81%	-62%	-37%	-28%
	Δ=0.5	$\alpha=0$	-102%	-97%	-79%	-57%	-50%
		$\alpha=1$	-78%	-73%	-53%	-27%	-19%
		$\alpha=2$	-63%	-58%	-37%	-8%	3%
	Δ=0.25	$\alpha=0$	-71%	-67%	-46%	-19%	-10%
		$\alpha=1$	-59%	-54%	-32%	-2%	9%
		$\alpha=2$	-52%	-47%	-24%	9%	22%
	Δ=0	all α	-40%	-35%	-10%	28%	60%

h_r	B=1		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.2	Δ=1	$\alpha=0$	-128%	-125%	-109%	-92%	-87%
		$\alpha=1$	-79%	-74%	-55%	-28%	-18%
		$\alpha=2$	-52%	-47%	-25%	6%	19%
	Δ=0.5	$\alpha=0$	-75%	-71%	-52%	-30%	-23%
		$\alpha=1$	-53%	-48%	-28%	-2%	8%
		$\alpha=2$	-39%	-34%	-13%	18%	33%
	Δ=0.25	$\alpha=0$	-47%	-42%	-22%	6%	15%
		$\alpha=1$	-35%	-31%	-9%	22%	33%
		$\alpha=2$	-28%	-24%	-1%	32%	45%
	Δ=0	all α	-18%	-13%	11%	50%	83%

h_r	B=2		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.2	Δ=1	$\alpha=0$	-134%	-131%	-115%	-97%	-93%
		$\alpha=1$	-95%	-91%	-74%	-53%	-46%
		$\alpha=2$	-71%	-66%	-47%	-20%	-9%
	Δ=0.5	$\alpha=0$	-84%	-80%	-62%	-39%	-33%
		$\alpha=1$	-64%	-59%	-40%	-14%	-5%
		$\alpha=2$	-51%	-46%	-25%	4%	14%
	Δ=0.25	$\alpha=0$	-58%	-54%	-33%	-6%	3%
		$\alpha=1$	-48%	-43%	-21%	9%	20%
		$\alpha=2$	-41%	-36%	-14%	19%	33%
	Δ=0	all α	-31%	-26%	-2%	37%	69%

h_r	B=4		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 0.2	Δ=1	$\alpha=0$	-154%	-150%	-134%	-117%	-112%
		$\alpha=1$	-117%	-113%	-95%	-74%	-68%
		$\alpha=2$	-94%	-89%	-70%	-45%	-38%
	Δ=0.5	$\alpha=0$	-106%	-102%	-84%	-62%	-55%
		$\alpha=1$	-87%	-83%	-63%	-37%	-28%
		$\alpha=2$	-75%	-71%	-49%	-20%	-10%
	Δ=0.25	$\alpha=0$	-82%	-77%	-57%	-30%	-20%
		$\alpha=1$	-72%	-67%	-46%	-15%	-4%
		$\alpha=2$	-66%	-61%	-38%	-5%	8%
	Δ=0	all α	-56%	-51%	-27%	12%	44%

h_r	B=5		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 1	Δ=1	$\alpha=0$	-112%	-109%	-93%	-76%	-71%
		$\alpha=1$	-85%	-81%	-64%	-43%	-37%
		$\alpha=2$	-68%	-64%	-45%	-20%	-12%
	Δ=0.5	$\alpha=0$	-78%	-74%	-55%	-33%	-26%
		$\alpha=1$	-64%	-59%	-39%	-13%	-4%
		$\alpha=2$	-54%	-50%	-28%	1%	11%
	Δ=0.25	$\alpha=0$	-59%	-55%	-34%	-7%	2%
		$\alpha=1$	-52%	-47%	-25%	5%	16%
		$\alpha=2$	-47%	-42%	-20%	13%	27%
	Δ=0	all α	-40%	-35%	-11%	28%	61%

h_r	B=10		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 1	Δ=1	$\alpha=0$	-139%	-135%	-120%	-102%	-98%
		$\alpha=1$	-114%	-109%	-92%	-71%	-65%
		$\alpha=2$	-97%	-93%	-73%	-49%	-41%
	Δ=0.5	$\alpha=0$	-106%	-102%	-84%	-61%	-55%
		$\alpha=1$	-92%	-88%	-68%	-42%	-34%
		$\alpha=2$	-84%	-79%	-58%	-29%	-18%
	Δ=0.25	$\alpha=0$	-89%	-84%	-64%	-36%	-27%
		$\alpha=1$	-81%	-77%	-55%	-25%	-13%
		$\alpha=2$	-77%	-72%	-50%	-17%	-3%
	Δ=0	all α	-70%	-65%	-41%	-2%	31%

h_r	B=20		R=0	R=0.1	R=0.5	R=0.9	R=1
h_r = 1	Δ=1	$\alpha=0$	-193%	-189%	-173%	-156%	-151%
		$\alpha=1$	-168%	-164%	-147%	-126%	-119%
		$\alpha=2$	-152%	-148%	-129%	-104%	-96%
	Δ=0.5	$\alpha=0$	-161%	-157%	-139%	-116%	-110%
		$\alpha=1$	-148%	-143%	-124%	-98%	-89%
		$\alpha=2$	-140%	-135%	-114%	-85%	-74%
	Δ=0.25	$\alpha=0$	-144%	-140%	-119%	-92%	-83%
		$\alpha=1$	-137%	-133%	-111%	-81%	-69%
		$\alpha=2$	-133%	-128%	-106%	-73%	-59%
	Δ=0	all α	-126%	-121%	-97%	-58%	-25%

Table 3-12: The % cost savings of the capacity sell-back mode compared to the fixed capacity, when $h_r = 0.1, 0.2$ and 1 and $B/h_r = 5, 10$ and 20, $h_{rr} = B$ for Δ= 0,0.25,0.5 and 1, $\alpha = 0, 1$ and 2 when $R = 0, 0.1, 0.5, 0.9$ and 1 when c_c has exponential structure.

B =2, h_r=0.2	**Single Level**		L_F*			2.83	μ_F*			1.81	
			R=0			**R=0.5**			**R=1**		
	Two-Level		L*	D*	μ*	L*	D*	μ*	L*	D*	μ*
inv. Proportional	Δ=0	all α	2.80	0.5	1.83	2.00	0.5	2.17	0.00	0.5	inf.
	Δ=0.25	α=0	3.30	0.5	1.71	2.50	0.5	1.94	1.50	0.5	2.58
		α=1	3.00	0.5	1.78	2.30	0.5	2.02	1.20	0.5	2.99
		α=2	3.00	0.5	1.78	2.30	0.5	2.02	1.10	0.5	3.17
	Δ=0.5	α=0	3.40	0.5	1.68	2.80	0.5	1.83	2.00	0.5	2.17
		α=1	3.30	0.5	1.71	2.70	0.5	1.86	1.70	0.5	2.39
		α=2	3.30	0.5	1.71	2.50	0.5	1.94	1.50	0.5	2.58
	Δ=1	α=0	4.20	0.5	1.55	3.40	0.5	1.68	2.80	0.5	1.83
		α=1	3.50	0.5	1.67	3.00	0.5	1.78	2.30	0.5	2.02
		α=2	3.40	0.5	1.68	2.80	0.5	1.83	2.00	0.5	2.17

B =1, h_r=0.1	**Single Level**		L_F*			4.01	μ_F*			1.57	
			R=0			**R=0.5**			**R=1**		
	Two-Level		L*	D*	μ*	L*	D*	μ*	L*	D*	μ*
inv. Proportional	Δ=0	all α	4.20	0.5	1.55	2.80	0.5	1.83	0.00	0.5	inf.
	Δ=0.25	α=0	4.60	0.5	1.50	3.40	0.5	1.68	2.00	0.5	2.17
		α=1	4.20	0.5	1.55	3.30	0.5	1.71	1.70	0.5	2.39
		α=2	4.20	0.5	1.55	3.30	0.5	1.71	1.50	0.5	2.58
	Δ=0.5	α=0	5.00	0.5	1.46	4.20	0.5	1.55	2.80	0.5	1.83
		α=1	4.60	0.5	1.50	3.50	0.5	1.67	2.30	0.5	2.02
		α=2	4.60	0.5	1.50	3.40	0.5	1.68	2.00	0.5	2.17
	Δ=1	α=0	5.70	0.5	1.41	5.00	0.5	1.46	4.20	0.5	1.55
		α=1	5.20	0.5	1.45	4.10	1.0	1.58	2.90	1.0	1.83
		α=2	5.00	0.5	1.46	4.10	1.0	1.58	2.40	1.0	2.02

B =2, h_r=0.1	**Single Level**		L_F*			4.74	μ_F*			1.63	
			R=0			**R=0.5**			**R=1**		
	Two-Level		L*	D*	μ*	L*	D*	μ*	L*	D*	μ*
inv. Proportional	Δ=0	all α	4.80	0.5	1.63	3.50	0.5	1.87	0.00	0.5	inf.
	Δ=0.25	α=0	5.60	0.5	1.54	4.30	0.5	1.70	2.40	0.5	2.27
		α=1	5.30	0.5	1.57	3.90	0.5	1.78	1.90	0.5	2.61
		α=2	5.10	0.5	1.59	3.80	0.5	1.80	1.70	0.5	2.81
	Δ=0.5	α=0	5.60	0.5	1.54	4.80	0.5	1.63	3.50	0.5	1.87
		α=1	5.60	0.5	1.54	4.30	0.5	1.70	2.70	0.5	2.13
		α=2	5.60	0.5	1.54	4.30	0.5	1.70	2.40	0.5	2.27
	Δ=1	α=0	6.70	0.5	1.45	5.60	0.5	1.54	4.80	0.5	1.63
		α=1	6.20	0.5	1.49	5.30	0.5	1.57	3.90	0.5	1.78
		α=2	5.60	0.5	1.54	4.80	0.5	1.63	3.50	0.5	1.87

B =4, h_r=0.2	**Single Level**		L_F*				3.35	μ_F*			1.58
			R=0			**R=0.5**			**R=1**		
	Periodic Sell-back		L*	D*	μ*	L*	D*	μ*	L*	D*	μ*
inv. Proportional	Δ=0	all α	3.50	0.5	1.87	2.40	0.5	2.27	0.00	0.5	inf
	Δ=0.25	α=0	3.80	0.5	1.80	3.10	0.5	1.98	1.70	0.5	2.81
		α=1	3.80	0.5	1.80	2.70	0.5	2.13	1.50	0.5	3.06
		α=2	3.60	0.5	1.84	2.70	0.5	2.13	1.20	0.5	3.60
	Δ=0.5	α=0	4.30	0.5	1.70	3.50	0.5	1.87	2.40	0.5	2.27
		α=1	3.90	0.5	1.78	3.10	0.5	1.98	1.90	0.5	2.61
		α=2	3.80	0.5	1.80	3.10	0.5	1.98	1.70	0.5	2.81
	Δ=1	α=0	4.80	0.5	1.63	4.30	0.5	1.70	3.50	0.5	1.87
		α=1	4.30	0.5	1.70	3.80	0.5	1.80	2.70	0.5	2.13
		α=2	4.30	0.5	1.70	3.50	0.5	1.87	2.40	0.5	2.27

Table 3-13: The optimal periodic sell-back capacity mode policy parameters (L^*, D^* and μ^*) under different 4 B & h_r scenarios (**1:** $B = 2, h_r = 0.2$ **2:** $B = 1, h_r = 0.1$ **3:** $B = 2, h_r = 0.1$ and **4:** $B = 4, h_r = 0.2$) and various (Δ, α) combinations when $c_p = 1$ and $R = 0, 0.5$ and 1.

3.6.3.2 Accuracy of Finite Waiting Room Approximation

In this subsection, we examine the accuracy of our finite waiting room approximation by comparing the $P\{S_d \leq L|\,\mathbf{C}\}$ values from the analytical model (having a finite waiting room of 40) with the $P\{S_d \leq L|\,\mathbf{C}\}$ values obtained by simulating the real environment having a repair shop that has an infinite waiting room.

In our simulations, we used a run length of 5,000,000,000 defective part arrivals (when $\lambda = 1$) in a single replication, where the average total cost rate TRC is calculated under a policy: $\mathbf{C} = [D, \mu]$.

We investigated a total of 5955 different scenarios with different B, c_p, R, Δ and α resulting policy parameters. The percentage error, $\%err$, of using the analytical approximation for $P\{S_d \leq L|\,\mathbf{C}\}$ in a scenario can be found as:

$$\%err(P\{S_d \leq L|\,\mathbf{C}\}) = 100 \times \frac{\left(P\{S_d \leq L|\,\mathbf{C}\}_{sim} - P\{S_d \leq L|\,\mathbf{C}\}_{app}\right)}{P\{S_d \leq L|\,\mathbf{C}\}_{sim}} \tag{3.26}$$

Table 3-14 summarizes the accuracy of the finite waiting room approximations using the $P\{S_d \leq L|\,\mathbf{C}\}$ generated from simulation.

	average %|*err*|	**min *%err***	**median *%err***	**max *%err***
$P\{S_d \leq L\|\,\mathbf{C}\}$	0.18%	-1.29%	-0.02%	1.76%

Table 3-14 Accuracy of the approximation for the $P\{S_d \leq L|\,\mathbf{C}\}$ values

In Table 3-14, the absolute value, minimum, median and the maximum for the percentage errors are listed, respectively. From the table, we can see that the approximation can mimic the performance of the original, infinite waiting room environment almost perfectly, which demonstrates the accuracy of our method.

This simulation study finalizes this section for the sell-back capacity mode. In the next section, Section 3.7, we discuss and summarize our findings in a nutshell and provide our conclusive remarks on the applicability of these capacity modes in the commoditized environment and link our findings to the previous chapter, retrospectively. In the next chapter we will briefly summarize the future research topics and extensions that are strongly related to the thesis.

3.7 Concluding Remarks

3.7.1 Summary

In this chapter, we studied the integrated down-time service and capacity management problem of a MSP who is running a repair shop and is responsible for the availability of different commoditized systems in an environment. In order to decrease the down-time costs, the repair shop closes a contract with a 3rd party supplier, and upon each system failure the 3rd party supplier agrees to lend a substitute to the MSP straightaway, for a uniform hiring duration length, at a fixed hiring price that is given. If the repair of a substitute system takes longer than this hiring duration, the down-time during this residual time is not covered (or covered with the same substitute system with an inflated price). In this commoditized system operating environment, the MSP is also offered the possibility to close a contract with a capacity agency that allows the MSP to make use of different sorts of capacity modes for the repair shop. The fixed, two-level and sell-back capacity modes have already been explained in the specialized system environment.

In this section, we developed a decision support system for both two-level and sell-back capacity modes which integrates the down-time and capacity decisions of the MSP in order to minimize its total relevant costs. We compared the savings and the optimal policy parameters of both of these capacity modes with the optimal fixed capacity mode results first, and with each other afterwards. Different from the specialized system environment, the down-time service decision of the MSP in this environment is the hiring duration of the substitute system.

For both two-level and periodic sell-back capacity modes, we analyze the performance of the MSP and develop computational approaches based on the decomposition of the overall problem in a number of sub problems that can be potentially solved as convex optimization sub-problems. Moreover, we performed a computational study to investigate the possible benefits of closing a contract with a capacity agency given certain values for the cost parameters (down-time costs, substitute hiring costs and permanent/ contingent capacity costs) for both of the flexible capacity modes.

3.7.2 Results

3.7.2.1 Cost Savings of the Flexible (2 Level and Periodic Sell-back) Capacity Modes and Intra-Environment, Inter-Mode Comparisons

In this section, we wrap up our findings on the cost performance of the capacity modes. First, we observe that, under the fixed capacity mode, in the commoditized system environment, contracting with a 3rd party supplier to provide a substitute for a hiring duration, during the system down-time, immediately after a failure, is the preferred strategy of a small-medium sized MSP. Integrating the hiring duration decision with the capacity level decision reduces the total relevant costs of the MSP substantially, compared to the situation in which the MSP does not hire a substitute but optimizes its capacity level only. The intervention of the hiring duration can decrease the total relevant costs up to 55%. In the commoditized system environment, under the optimal capacity policy, for all down-time and hiring cost parameter combinations, the capacity relevant costs always compose half of the total relevant costs. Therefore the cost saving prospects of the flexible capacity modes are explored.

In the two-level capacity mode, the results show that the maximum savings (with respect to the costs under the optimal fixed capacity policy) can range from 18% to 55%. The biggest savings occur when there is no opportunity cost for the contingent capacity and when the hiring costs are the lowest. In case of positive maximum opportunity costs, the relative savings start to decrease, but even when it reaches its greatest value (i.e. when opportunity costs get as high as the permanent capacity costs), a substantial part of the savings (10% to 20%) are still maintained.

In the periodic sell-back capacity mode, the cost-performance of the optimal periodic sell-back capacity policy seem to be much more unstable than the two-level capacity policy. In addition, the performance of the periodic sell-back capacity mode is highly dependent on how much mark-up the 3rd party supplier puts on the post-admission hiring rate during the provision in the pre-admission duration. Our numerical study reveals that the maximum savings can vary from negative values (up to twice higher costs than fixed capacity mode) to saving values of 97%. The maximum savings occur when there is no opportunity cost for the contingent capacity. However, for sell-back rates lower than 0.1, even if there is no contingent capacity cost, the periodic sell-back capacity mode can be still outperformed by the fixed capacity mode, due to the burstiness of arrivals caused by the periodic admission. Similar to the two-level capacity mode, the data from the numerical results further illustrate that the cost performance of the sell-back mode improves with lower pre-admission & post-admission hiring rates and with lower maximum opportunity costs for the contingent capacity.

For both two-level and periodic sell-back capacity modes, linear contingent capacity cost structure leads to the highest savings and the inverse proportional structure leads to the smallest savings.

Among the investigation of the three capacity modes, it is witnessed that the optimal performance of the fixed capacity mode is surpassed by the optimal two-level capacity policy in all of the cost parameter realizations in the studied commoditized system environment test-bed. We observe that the cost savings of the two- level capacity mode (with respect to the fixed capacity mode) in the total costs derive from both shorter hiring durations and lower capacity deployment.

In contrast, there are a lot of cost parameter instances, which lead to an underperformance of the periodic capacity sell-back mode. A typical instance is where the sell-back rate is zero or very small, periodic sell-back capacity mode is always the least economical mode. As the sell-back rate increases, the gap between the costs of two capacity modes decrease and after a threshold value, periodic sell-back capacity mode starts to outdo the fixed capacity mode. Contrary to the two-level capacity mode, in the periodic capacity sell-back mode, at least one of the capacity policy parameters (capacity or hiring duration) is higher than those in the optimal fixed capacity mode. However, since the capacity can be sold back in the sell-back mode, a higher capacity does not necessarily mean underperformance.

When making a pair-wise comparison between the two-level and periodic sell-back capacity modes, we observe that pre-admission hiring rate is an important driver. If the pre-admission hiring rate is close to the down-time cost, then periodic sell-back mode hardly never surpasses the two-level capacity mode. On the other hand, if the pre-admission hiring rate is similar to the post-admission hiring rate, we observe that under high sell back rates (higher than 90%) the sell-back capacity mode becomes the most economical capacity mode among the three.

As it was discussed in the specialized system environment, the comparability of the maximum opportunity costs between two capacity modes should be checked, since there can be different explaining factors behind. Similarly, if there are shipment costs, the milk-run/shipment pooling possibilities arising from the periodic admission should be reflected in the calculations. In the next section, we set out to compare the results in between two environments, speculate over a scenario where hiring a substitute and holding a substitute per unit time have identical unit costs and introduce the hybrid down-time strategy in which both hiring and stock keeping options can be utilized simultaneously.

3.8 Inter-Environment Comparisons

In this section, we first briefly mention about the inter-environment comparison of the capacity modes. First we evaluate the saving performance results of the capacity modes in the specialized system environment and in the commoditized system environment. Afterwards, in order to understand the differences between the operating dynamics of two strategies, we speculate over a scenario where hiring a substitute and holding a substitute per unit time have identical unit costs. Then, we compare the results of choosing one of the two strategies and explain the differences between them. Finally we introduce the hybrid down-time strategy, in which both hiring and stock keeping options can be taken simultaneously.

When we explore the cost performances in the specialized and in the commoditized environments, the key findings that we come across can be listed as follows:

1. In the fixed capacity mode, the percentage savings of the integration of the capacity and the down-time service decisions is higher in specialized system environment (with respect to adjusting only the capacity level)

2. For the two level capacity mode, the relative performance of the total costs with respect to the fixed capacity mode is better in the specialized system environment

3. When the pre-admission hiring rate is expensive, the periodic sell-back capacity also performs better (in terms of total costs relative to the fixed capacity mode) in the specialized environment, however if the pre-admission hiring is the same as the post-admission hiring rate, one can observe that the savings in the commoditized setting becomes slightly higher.

These differences in percentage savings are explainable by the differences in the test beds. The holding cost in the specialized system setting is quite lower than the hiring cost in the commoditized system setting, which makes the base reference costs in the specialized system setting lower. Therefore, even though the differences in the nominal savings between two environments are not that wide, the percentage savings' gap increases due to the differences at the baseline. Only in periodic sell-back capacity mode, if the pre-admission hiring rate is the same as the post-admission rate, periodic sell-back capacity mode may generate slightly better percentage savings, however we create an unfair situation for the specialized system, which suffers from such a safety net in case of stock-outs.

The sensitivity and the interactions of the hiring duration and the stock level are very similar in all of the three capacity modes. Given a capacity mode and policy, bot hiring duration and stock level decisions can be modeled as newsvendor problems. However the level of controllability and how these two strategies work are different from each other. That's why, given the identical arrival rates, the comparison of the optimal total relevant costs between two strategies ($TRC_F{}^*$ is the cost resulting from keeping a spare unit stock strategy and $TRC_{Fr}{}^*$ resulting from using the rental hiring strategy), under the same B and c_p cost parameters and identical holding cost and hiring rental rates ($h = h_r$), can provide interesting managerial insights.

In particular it is necessary to distinguish the differences between the guaranteed availability due to the agreed rental hiring option and the expected availability due to keeping additional stock is very interesting. Therefore, in Table 3-15, we first give the $S_F{}^*$, $\mu^*(S_F{}^*)$ and $TRC_F{}^*$ which are calculated from the Search Procedure-I in Section 2.4, for $c_p = 1$, and for B and h values follow Table 3-1 when ($h = h_r$).

	$h = 0.1$			$h = 0.2$			$h = 1$		
	$S_F{}^*$	$\mu^*(S_F{}^*)$	$TRC_F{}^*$	$S_F{}^*$	$\mu^*(S_F{}^*)$	$TRC_F{}^*$	$S_F{}^*$	$\mu^*(S_F{}^*)$	$TRC_F{}^*$
$B/h = 5$	3	1.532	1.09	2	1.772	1.58	1	2.730	3.789
$B/h = 10$	5	1.554	1.25	3	1.861	1.82	2	2.740	4.506
$B/h = 20$	6	1.611	1.40	4	1.924	2.04	2	3.171	5.087

Table 3-15 : $S_F{}^*$, $\mu^*(S_F{}^*)$ and $TRC_F{}^*$ which are calculated from the Search Procedure-I in Section 2.4, for $c_p = 1$, and for B/h & h values from Table 3-1, when $h = h_r$.

Afterwards, in Table 3-16, we give the nominal ($TRC_F{}^* - TRC_{Fr}{}^*$) and percentage cost differences $100 \times \left(\frac{TRC_F{}^* - TRC_{Fr}{}^*}{TRC_F{}^*}\right)$ % between the total relevant costs due to the rental hiring option and keeping stock option for 9 different B/h_r & h_r scenarios when ($h = h_r$) and $c_p = \lambda = 1$.

From Table 3-16, for the identical failure arrival ($\lambda = 1$), $c_p = 1$ and B/h_r & h_r scenarios when ($h = h_r$), we can see that the optimal total relevant costs in the commoditized setting, with the rental availability, ($TRC_{Fr}{}^*$) is always less (up to 21.41%) than the optimal total relevant costs in the specialized setting, with stock availability, ($TRC_F{}^*$). These differences can be explained as follows. Firstly, the rental hiring of a substitute system brings about a guaranteed and certain availability during the rental duration, whereas keeping a stock may bring about the same availability on the average, but without any certainty on the availability times.

	$h_r = 0.1 \| h = 0.1$		$h_r = 0.2 \| h = 0.2$		$h_r = 1 \| h = 1$	
	Nominal	Percentage	Nominal	Percentage	Nominal	Percentage
$B/h_r = 5$	0.07173	6.56%	0.139684	8.82%	0.557919	14.72%
$B/h_r = 10$	0.10382	8.28%	0.195956	10.76%	0.870907	19.32%
$B/h_r = 20$	0.13401	9.59%	0.25202	12.36%	1.089304	21.41%

Table 3-16: The nominal ($TRC_F{}^* - TRC_{Fr}{}^*$) and percentage differences $100 \times \left(\frac{TRC_F{}^* - TRC_{Fr}{}^*}{TRC_F{}^*}\right)$ % between the optimal total relevant costs due to the rental hiring option and keeping stock option for 9 different B/h_r & h_r scenarios when $h = h_r$.

Secondly, the rental hiring duration L cancels out the sclerotic response of the stock level decision since S^* can take only positive integer values. The ability of L to take any positive values contributes to the controllability of the environment. We observe that both the nominal and percentage differences between the costs in two environments grow larger for higher B and higher h_r. Because of the controllability advantages of the rental hiring of a substitute system, even at a more expensive unit rental hiring rate compared to the unit holding cost rate ($h_r > h$), the optimal total relevant costs under the rental availability, $TRC_{Fr}{}^*$ can still outperform the optimal total relevant costs under the stock availability $TRC_F{}^*$. In , we give the maximum affordable rental cost rate for each of the 9 (B& c_p) scenarios, above which keeping stock becomes a more economical alternative when $h = 1$.

	$h = 0.1$	$h = 0.2$	$h = 1$
$B/h = 5$	0.126	0.27	1.751
$B/h = 10$	0.130	0.28	1.912
$B/h = 20$	0.132	0.29	1.942

Table 3-17: The upper limits for the rental hiring cost rate h_r, above which the total relevant costs with stock availability, $TRC_F{}^*$, becomes more economical compared to renting a substitute.

It is noteworthy that the MSP would afford more expensive rental availability (can be up to 95% more expensive compared to the holding cost rate $h = 1$), especially when B is higher and c_p is lower. Next we set out to describe the hybrid down-time strategy.

3.8.1 The Hybrid Down-time Strategy

In the previous sections, we have analyzed two different system environments which resulted in two different availability strategies. Namely, in Chapter 2, we focused on the MSP's operations in a specialized system environment and in the previous sections of this chapter, we focused on the same problem in the (partly) commoditized system environment.

In the specialized system environment, we assumed that the repair shop keeps a spare part inventory whereas in the commoditized environment, we assumed that the service provider purely makes use of the rental providers and tries to cover the down-time of the defective systems by uniform rental hiring of another substitute system for a fixed duration of time.

In this section we introduce a hybrid strategy, where the repair shop holds a spare part stock but in the case of a stock-out, if a failure occurs, a new substitute system/part is sent to the customer for a fixed duration length. The rented substitute system stays until the end of the hiring duration no matter when the repair of the corresponding defective system is completed. The down-time costs are incurred if there is no inventory on hand and the repair time took longer than the hiring rental duration. In this section, we analyze only the fixed capacity mode, as we believe that it will give the main message behind the hybrid mechanism. Also we believe that the analysis of the hybrid mechanism under other flexible capacity policies will be rather more involved.

The capacity cost per unit time is c_p in this fixed capacity mode since all the repair shop capacity is permanent (or supplied indefinitely). We pay h per unit time for each spare part in the stock/in the repair shop. The down-time costs due to the backorder of the spare parts is equal to B per time unit, and we assume that $B > h$. We also assume a fixed cost of h_r per unit time in order to hire a substitute from the rental shop.

From now on, we use the notation of $\mathbf{C}$ to denote the capacity policy. In the fixed capacity mode, $\mathbf{C}$ is a single variable, since the only capacity related decision is the processing rate μ. The inventory related decision is the base-stock level of the spare part inventory (S), and the hiring related decision is how long to hire a rental substitute (L) when there is no spare in the inventory. The total relevant cost function, TRC, can be represented as a function of $\mathbf{C}$, S and L and it is the sum of capacity related costs $\big(CRC(\mathbf{C})\big)$, down-time costs $\big(DTC(\mathbf{C},S,L)\big)$, holding costs $\big(HC(S)\big)$ and hiring related costs $(HRC(\mathbf{C},S,L))$. Given these cost components and the decision variables, the problem of the MSP can be formulated as follows:

$$\min_{\mathbf{C},S,L} TRC(\mathbf{C},S,L) = CRC(\mathbf{C}) + DTC(\mathbf{C},S,L) + HC(S) + HRC(\mathbf{C},S,L)$$

$$s.t.\;\; S \in N = \{0,1,2,\ldots\}$$

$$L \geq 0$$

$$\mathbf{C} = \mu > \lambda$$

(3.27)

Given the problem formulation above, we first derive the necessary cost functions used in (3.27), give the analytical properties of $TRC(\mathbf{C},S,L)$, and present the optimization procedure for the problem.

The capacity related cost per unit time is a linear function of the excess capacity $\mu - \lambda$, since the baseline capacity level, (λ), is unaffected by the capacity policy. Per time unit cost of the capacity is constant and equal to c_p. Therefore, we have $CRC(\mathbf{C}) = c_p(\mu - \lambda)$. The holding cost per unit time is also a linear function of the base stock level S, since we have an additional S number of spare parts tied up in the stock/ repair shop and the holding cost rate per unit part is h per time. Hence, we have $HC(S) = h \times S$.

We hire a substitute part/system when there is no stock in the inventory. If there are more than S defective systems/parts, then we know that there is no part/system in the stock. From (3.2), we know that the number of defective parts/systems in the repair shop, when $\mathbf{C} = \mu$, can be calculated from: $\mathrm{P}\{N_d = n|\mu\} = (1-\rho)\rho^n$. Due to the PASTA property, the probability that there is no spare part/system in the stock upon a new failure can be found from:

$$P\{No\ Stock\} = \sum_{n=S}^{\infty}(1-\rho)\rho^n = \rho^S$$

(3.28)

Hiring from rental takes place upon a "No Stock" event and each hiring endures L units of time. Therefore we have the following Hiring Related Costs ($HRC(\mathbf{C}, \mathrm{S}, L) = \lambda\rho^S\, h_r L$).

The down-time also occurs upon a "No Stock" event, as well. However due to our hybrid strategy, we can take a remedy action upon being a backorder instance and hire a rental substitute for a fixed duration L. Down-time happens, if the repair of a new part is delayed more than L time units after the stock-out. Let S_{Dr} denotes the time that a customer waits for a new part to arrive after a stock-out.

The sojourn time in the repair shop is exponentially distributed. Due to reversibility (Kelly 1979) and the memoryless property of the exponential distribution, the wait time of a customer given there is a stock-out, S_{Dr} is also exponentially distributed with rate $(\mu - \lambda)$.

An alternative explanation of the derivation of the distribution is as follows:

Let $P\{S_{dr} > x|No\ Stock\}$ denote the probability that a customer is going to wait more than x time units. Similarly let $P\{S_{dr} > x|N_d = n\}$ denote the probability that a customer is going to wait more than x time units given he sees n other customers in the queue upon arrival.

Then we have:

$$P\{S_{dr} > x|No\ Stock\} = \frac{\sum_{n=S}^{\infty} P\{S_{dr} > x|N_d = n\}\, P(N_d = n)}{P\{No\ Stock\}}$$
$$= \frac{\sum_{n=S}^{\infty} P\{B_{n-S+1} > x\}\,(1-\rho)\rho^n}{\rho^S} = \sum_{n=1}^{\infty} P\{B_n > x\}\,(1-\rho)\rho^{n-1}$$
$$= e^{-(\mu-\lambda)x}$$

(3.29)

Where $B_n \sim Erlang\ n$ distributed random variable. Then we have the following for per time down-time related cost, $DTC(\mathbf{C}, S, L)$:

$$DTC(\mathbf{C}, S, L) = \lambda B(P\{No\ Stock\})\mathrm{E}((S_{Dr} - L)^+|\mu) = \lambda\rho^S B \frac{\mathrm{e}^{-(\mu-\lambda)\mathrm{L}}}{(\mu-\lambda)}$$

Hence, we have the following:

$$TRC(\mathbf{C}, S, L) = hS + c_p(\mu - \lambda) + \lambda\rho^S h_r L + \lambda\rho^S B \frac{e^{-(\mu-\lambda)L}}{(\mu - \lambda)}$$

(3.30)

Note that for a given $\mu > \lambda$, and S, $L^*(\mu, S) = \frac{-ln\left(\frac{h_r}{B}\right)}{(\mu-\lambda)}$ minimizes the $TRC(\mathbf{C}, S, L)$. If we plug this $L^*(\mu, S)$ into (3.30) then we would obtain:

$$TRC\big(\mathbf{C}, S, L^*(\mu, S)\big) = hS + c_p(\mu - \lambda) + \lambda\rho^S h_r \left(\frac{1 - \ln\left(\frac{h_r}{B}\right)}{(\mu - \lambda)}\right)$$

(3.31)

If we use $TRC\big(\mathbf{C}, S, L^*(\mu, S)\big)$ rather than $TRC(\mathbf{C}, S)$ defined in (3.4) in the Search Procedure I, we can obtain the optimal μ^*, S^* and $L^*(\mu^*, S^*)$ parameters.

Table 3-18, we give the optimal μ^*, S^* and $L^*(\mu^*, S^*)$ and the minimum total costs *(TRC^*_{Fh})* for $B = 5{,}10$ and 20 and $c_p = 1{,}5$ and 10 for $h_r/h = 1$ (top) and $h_r/h = 1.5$ (bottom).

	$h = 1$				$h = 0.2$				$h = 0.1$			
	*S**	*L**	*μ**	*TRC**	*S**	*L**	*μ**	*TRC**	*S**	*L**	*μ**	*TRC**
$B/h = 5$	*1*	1.21	2.33	3.17	1	2.43	1.66	1.34	2	3.59	1.45	0.93
$B/h = 10$	*1*	1.57	2.46	3.38	2	3.51	1.66	1.42	3	5.05	1.46	0.99
$B/h = 20$	*1*	1.90	2.58	3.56	2	4.23	1.71	1.49	3	6.12	1.49	1.04
	$h = 1$				$h = 0.2$				$h = 0.1$			
	*S**	*L**	*μ**	*TRC**	S^*	*L**	*μ**	*TRC**	*S**	*L**	*μ**	*TRC**
$B/h = 5$	*1*	0.82	2.46	3.38	2	1.83	1.66	1.42	3	2.64	1.46	0.99
$B/h = 10$	*1*	1.16	2.63	3.64	2	2.59	1.73	1.53	3	3.75	1.51	1.06
$B/h = 20$	*1*	1.45	2.78	3.87	3	3.70	1.7	1.60	4	5.19	1.50	1.11

Table 3-18 : The optimal μ^*, S^* and $L^*(\mu^*, S^*)$ and the minimum total costs for $B = 5{,}10$ and 20 and $c_p = 1{,}5$ and 10 for $h_r/h = 1$ (top) and $h_r/h = 1.5$ (bottom).

When we compare the data in Table 3-18 to the data in Table 3-15 and Table 3-2, we can immediately see that the hybrid policy results in remarkable savings in total costs TRC^*_{Fh} compared to the TRC^*_{F} in the specialized environment and TRC^*_{Fr} in the commoditized environment.

Compared to the specialized system environment, hybrid policy results in lower base-stock values and lower capacity levels. Compared to the commoditized environments, hybrid policy results in higher rental durations and lower capacity levels Compared to the specialized system environment, hybrid policy results in lower base-stock values and lower capacity levels. Compared to the commoditized environments, hybrid policy results in higher rental durations and lower capacity levels.

Keep in mind that hiring rental is less frequent in the hybrid policy, since it occurs after a stock-out, whereas in the commoditized environment hiring is the only availability action.The savings of the hybrid policy $TRC^{*}{}_{Fh}$ compared to the $TRC^{*}{}_{F}$ in the specialized environment and $TRC^{*}{}_{Fr}$ in the commoditized environment are tabulated in Table 3-19 with $h_r = 1$ and $h_r = 1.5$.

	commoditized ($h_r = 1$)			specialized ($h_r = 1$)		
	$h = 1$	$h = 0.2$	$h = 0.1$	$h = 1$	$h = 0.2$	$h = 0.1$
$B/h = 5$	1.82%	7.51%	9.38%	16.28%	15.66%	15.33%
$B/h = 10$	7.02%	12.44%	13.81%	24.99%	21.86%	20.95%
$B/h = 20$	10.95%	16.39%	18.01%	30.02%	26.72%	25.87%
	commoditized ($h_r = 1.5$)			specialized ($h_r = 1.5$)		
	$h = 1$	$h = 0.2$	$h = 0.1$	$h = 1$	$h = 0.2$	$h = 0.1$
$B/h = 5$	7.04%	12.47%	13.83%	10.77%	10.16%	9.37%
$B/h = 10$	12.61%	18.06%	19.80%	19.13%	16.12%	15.62%
$B/h = 20$	16.65%	22.28%	24.18%	23.96%	20.92%	20.42%

Table 3-19 Percentage savings of the hybrid policy in total costs in specialized and commoditized environments

From Table 3-19, we can see that the percentage savings increase by B both compared to commoditized $(TRC^{*}{}_{Fr} - TRC^{*}{}_{Fh})/\, TRC^{*}{}_{Fr}$ and specialized $(TRC^{*}{}_{F} - TRC^{*}{}_{Fh})/\, TRC^{*}{}_{F}$ environments. We observe that percentage savings compared to commoditized environment costs: $(TRC^{*}{}_{Fr} - TRC^{*}{}_{Fh})/\, TRC^{*}{}_{Fr}$, increase by c_p and compared to specialized environment costs: $(TRC^{*}{}_{F} - TRC^{*}{}_{Fh})/\, TRC^{*}{}_{F}$, decrease by c_p. In general, the hybrid policy seem to display significant cost savings and can be thought in other contexts, as well.

4 Conclusions and Future Research

4.1 Overview of Results

In this thesis, we studied the integrated capacity and the availability management of a MSP, which is running a repair shop and is responsible to keep the systems' operations going. The systems are prone to failure, and operated by different customers. We primarily focused on two environments.

In the first environment, the systems are partly specialized and they include a common critical part. We assume the service provider is responsible for the repair and the availability of this critical part and keeps a stock of spare parts near the repair shop. Upon a system failure, a new (non-defective) part is sent immediately to the failed system to replace the defective part, where the defective part is sent to the repair shop to be repaired. After the repair, the repaired part is restored in the spare part stock as an "as good as new" part, to be sent to another failed system in the future.

In the second environment, we assume that the systems are partly commoditized, therefore there are rental suppliers available where a substitute of the failed system with almost the same functionality can be hired. We assume that a long-term agreement is achieved with a rental supplier, which provides a rental substitute system for a fixed duration at a fixed hiring rate. This uniform hiring upon system failure replaces the spare part holding strategy in the commoditized environment.

The repair shop capacity determines how fast the defective parts/systems are repaired. In most of the cases, capacity related costs constitute the biggest component of the total costs, therefore we analyzed the saving prospects of the capacity costs due to capacity volume flexibility arrangements. We assume that actions which lead to capacity flexibility can be taken at periodic instances, and an external/internal capacity agency is needed for the coordination/transaction of these actions. We primarily focused on three capacity modes, which are fixed capacity mode, two-level capacity mode and capacity sell-back mode. In the fixed capacity mode, the capacity is deployed once and its level is fixed. We used the optimal cost performance of the fixed capacity mode as a reference to assess the cost performance of the other two capacity modes.

In the two-level flexible capacity mode, the repair shop capacity is classified into permanent and contingent capacity, and the contingent capacity is supplied from the capacity agency at equidistant intervals for the whole duration of the interval. The length of the interval determines the per time price of the unit contingent capacity. A shorter length implies more frequent decision making and less job security for the contingent capacity, therefore it increases the per time price of the unit contingent capacity. The deployment decision of the contingent capacity is taken by the repair shop at the start of each period, and is based on the number of defective parts/systems waiting to be repaired.

In the capacity sell-back mode, the defective systems/parts are admitted to the repair shop at regular intervals. Due to this periodic admission structure, once the repair shop is finished with the repairs on hand, it is known that the shop capacity will remain idle at least until the start of the next period. This idle time certainty enables the realization of the capacity sell-back mode, in which the capacity agency acts as the trader of the repair shop capacity at its idle times. The capacity agency buys the idle capacity at a reduced rate and assigns the idle capacity for external tasks, if available. There is a fixed opportunity cost per period due to the emergent task search (since the start of the idle time is not known in the beginning) and the risk of not being able to find an external task.

From the analysis we have observed that the integrated management of the availability (spare part stock in the specialized environment and the rental hiring duration in the commoditized environment) and the repair shop capacity in the presence of capacity flexibility can lead to remarkable savings in costs. Mostly, in the two level flexible capacity mode, savings compared to the best fixed capacity are more significant with lower contingent opportunity costs and higher time elasticity. In the capacity sell-back mode, the savings compared to the best fixed capacity are more visible for higher sell-back rates and lower opportunity costs. However in both of the capacity modes, we can observe cost parameter combinations, in which the best fixed capacity policy outperforms other policies in the other flexible modes.

The period length plays a central role and determines the frequency of the capacity actions. In our thesis, we observe that the capacity actions should be taken rather frequently with respect to the mean inter-arrival times. We observe that when the period length is more than 10 times higher than the mean inter-arrival time, in many instances, the cost savings due to the flexibility start to disappear. In our thesis, we focused on the repair-maintenance environments, which are most of the time lower-demand environments (for instance compared to the consumer goods/products environment). Therefore the findings of our thesis are adaptable to other make-to-order/make-to-stock settings with some reservations, however the implications of the profitable period length/mean inter- arrival time ratio should be assessed realistically. Another important aspect is that our analyses are conducted in the stationary demand environment. This can be realistic to some extent for the maintenance environments, however in many real-life situations, we see non-stationary demand patterns. We expect that the savings in total costs would increase in the non-stationary demand environment since the workload will be more erratic in the non-stationary environment, which would require a higher baseline capacity in the fixed capacity mode, and the flexibility options enable us to adapt the capacity to the workload better.

Next we briefly discuss over the future research topics.

4.2 Discussion and Future Research

In the remainder, we summarize the most important findings in the bullet points and afterwards, we will elaborate on the future topics.

First we wrap up the key findings

- Specialized system environment and the commoditized system environment are different from each other.
- Due to the commoditization and the following rentalization, the systems in the commoditized market can be replaced by a substitute covering for the agreed hiring duration. This can be attractive for a small-medium sized MSP, who doesn't want to
- The hiring costs (for a specialized system) are quite high compared to keeping a unit of spare critical part.
- Hiring duration and stock level decision parameters responses to cost parameters are very similar.
- In the single level capacity mode, capacity related costs constitute the major part of the costs.
- Two-level capacity policy outperforms the fixed capacity level costs in both of the environments. However the percentage saving figures are stronger in the specialized system environment. (up to 75%)
- The savings of the two level capacity mode are due to lower stock/hiring duration and lower average deployed capacity.
- The performance of the periodic sell-back capacity mode improves with the sell-back rate. When there is no sell-back rate possibility , it is the least economical option of the three. After a threshold value of the sell-back rate, it starts to surpass the fixed capacity mode, and if the same opportunity and time elasticity is applied for both of the capacity modes, there exists a second threshold, after which the periodic sell-back capacity becomes the most economical option.
- Under the optimal periodic sell-back rate policy, at least one of the decision parameters (i.e. stock level, hiring duration or capacity level) is higher than the optimal decision parameters under the fixed capacity mode.
- In the commoditized system environment, pre-admission hiring rate determines whether the periodic capacity sell-back is cost-effective. If it is close to down-time costs, it is surpassed by the two-level two-level capacity mode.
- The opportunity costs in the two-level and in the sell-back capacity mode can be caused by different factors. One has to be sure about the opportunity costs' compatibility before comparing the capacity modes.
- If choosing either one of the down-time service strategies is mandated, and the costs for keeping a stock or hiring for a system are the same, the hiring strategy results in better costs due to increased/more refined controllability

- In the hybrid policy, both options can be used simultaneously and the resulting cost performance is better than both of the previous experiences
- Period length plays a central role in the cost performance of a flexible capacity policy. If there is no opportunity cost or no elasticity of the capacity costs, the shortest possible period length is most of the time optimal, otherwise, intermediate period lengths can be optimal, as well.

Besides the points mentioned under the bullet points above, now the possible future research questions will be elaborated

- Extension to multi-item/multi-echelon structure for the specialized system environment
- Robustness tests of the results for each capacity mode (i.e what are the implications if the demand rate is miscalculated & how much worse the resulting costs would be than the costs under the optimal policy)
- Dynamic hiring durations based on the number of defective systems, rather than 1 uniform hiring duration. Using the information on the number of
- Analytical approaches to calculate the performance of the existing policies approximately. Some more efficient approximations can be explored
- Cost saving opportunities due to the shipment of defective and repaired systems in the periodic sell-back capacity mode
- Performance of the Flexible capacity mode under the hybrid strategy

Further research can be directed towards one of the topics that is listed above.

Appendix

Proof for Theorem 2.3

Note that for a given S, total holding related cost, hS, is constant and $\left(c_c - c_p R\right)\mu - c_p \lambda(1-R)$ part is linearly increasing with μ. Therefore, for proving the convexity of $TRC(\mathbf{C}, S)$ with μ, it is enough to prove that $\mathrm{E}((N_d - S)^+|\mathbf{C})$ is convex in μ.

For the convexity proof, we use the stochastic convexity and sample path stochastic convexity concepts introduced by (Shaked & Shanthikumar 1994). The following definitions are taken from (Shaked & Shanthikumar 1994):

Definitions

In the following definitions $SI, SCX, SCV, SICX, SIL, SDL, SDCV$, and so forth, stand, respectively, for stochastically increasing, stochastically convex, stochastically concave, stochastically increasing and convex, stochastically increasing and linear, stochastically decreasing, stochastically decreasing and concave, and so forth.

Let $\{X(\theta), \theta \in \Theta\}$ be a set of random variables.

(1) $\{X(\theta), \theta \in \Theta\} \in SI$ [or SD] if $\mathrm{E}(\varphi(X(\theta)))$ is increasing [or decreasing] for all increasing functions φ,

(2) $\{X(\theta), \theta \in \Theta\} \in SCX$ [or SCV] if $\mathrm{E}(\varphi(X(\theta)))$ is convex [or concave] for all convex [or concave] functions φ,

(3) $\{X(\theta), \theta \in \Theta\} \in SICX$ [or $SICV$] if $\{X(\theta), \theta \in \Theta\} \in SI$ and $\mathrm{E}(\varphi(X(\theta)))$ is increasing convex [or concave] in θ for all increasing convex [or concave] functions φ,

(4) $\{X(\theta), \theta \in \Theta\} \in SDCX$ [or $SDCV$] if $\{X(\theta), \theta \in \Theta\} \in SD$ and $\mathrm{E}(\varphi(X(\theta)))$ is decreasing convex [or concave] in θ for all increasing convex [or concave] functions φ,

(5) $\{X(\theta), \theta \in \Theta\} \in SIL$ if $\{X(\theta), \theta \in \Theta\} \in SI$ and $\mathrm{E}(\varphi(X(\theta)))$ is increasing convex in θ for all increasing convex functions φ, and is increasing concave in θ for all increasing concave functions φ,

(6) $\{X(\theta), \theta \in \Theta\} \in SDL$ if $\{X(\theta), \theta \in \Theta\} \in SD$ and $\mathrm{E}(\varphi(X(\theta)))$ is decreasing convex in θ for all increasing convex functions φ, and is decreasing concave in θ for all increasing concave functions φ.

Next we discuss the sample path convexity definitions. In the following definitions $SI(sp), SCX(sp), SCV(sp), SICX(sp), SIL(sp), SDL(sp), SDCV(sp)$, and so forth, stand, respectively, for stochastically increasing, stochastically convex, stochastically concave, stochastically increasing and convex, stochastically increasing and linear, stochastically decreasing, stochastically decreasing and concave, and so forth.

Consider a family $\{X(\theta), \theta \in \Theta\}$ of random variables. Let $\theta_i \in \Theta, i = 1,2,3,4$, be any four values such that $\theta_1 \leq \theta_2 \leq \theta_3 \leq \theta_4$ and $\theta_1 + \theta_4 = \theta_2 + \theta_3$.

If there exist four random variables $\hat{X}_i,\ i = 1,2,3,4$, defined on a common probability space, such that $\hat{X}_i =_{st} X(\theta_i), i = 1,2,3,4$, and

(7) $(i)\ max[\hat{X}_2, \hat{X}_3] \leq \hat{X}_4$ a.s. (almost surely) and $(ii)\ \hat{X}_2 + \hat{X}_3 \leq \hat{X}_1 + \hat{X}_4$ a.s., then $\{X(\theta), \theta \in \Theta\}$ is said to be stochastically increasing and convex in the sample path sense (denoted by $\{X(\theta), \theta \in \Theta\} \in SICX(sp)$);

(8) $(i)\ \hat{X}_1 \leq min[\hat{X}_2, \hat{X}_3]$ a.s. and $(ii)\ \hat{X}_1 + \hat{X}_4 \leq \hat{X}_2 + \hat{X}_3$ a.s. , then $\{X(\theta), \theta \in \Theta\}$ is said to be stochastically increasing and concave in the sample path sense (denoted by $\{X(\theta), \theta \in \Theta\} \in SICV(sp)$);

(9) $(i)\ \hat{X}_1 \geq max[\hat{X}_2, \hat{X}_3]$ a.s. and $(ii)\ \hat{X}_1 + \hat{X}_4 \geq \hat{X}_2 + \hat{X}_3$ a.s., then $\{X(\theta), \theta \in \Theta\}$ is said to be stochastically decreasing and convex in the sample path sense (denoted by $\{X(\theta), \theta \in \Theta\} \in SDCX(sp)$);

(10) $(i)\ \hat{X}_4 \leq min[\hat{X}_2, \hat{X}_3]$ a.s. and $(ii)\ \hat{X}_1 + \hat{X}_4 \leq \hat{X}_2 + \hat{X}_3$ a.s., then $\{X(\theta), \theta \in \Theta\}$ is said to be stochastically decreasing and concave in the sample path sense (denoted by $\{X(\theta), \theta \in \Theta\} \in SDCV(sp)$);

(11) $(i)\ max[\hat{X}_2, \hat{X}_3] \leq \hat{X}_4$ a.s. and $(ii)\ \hat{X}_1 + \hat{X}_4 = \hat{X}_2 + \hat{X}_3$ a.s., then $\{X(\theta), \theta \in \Theta\}$ is said to be stochastically increasing and linear in the sample path sense (denoted by $\{X(\theta), \theta \in \Theta\} \in SIL(sp)$);

(12) $(i)\ \hat{X}_1 \geq max[\hat{X}_2, \hat{X}_3]$ a.s. and $(ii)\ \hat{X}_1 + \hat{X}_4 = \hat{X}_2 + \hat{X}_3$ a.s., then $\{X(\theta), \theta \in \Theta\}$ is said to be stochastically decreasing and linear in the sample path sense (denoted by $\{X(\theta), \theta \in \Theta\} \in SDL(sp)$).

After these definitions, we give the following reformulation:

$$\mathrm{E}((N_d - S)^+|\mathbf{C}) = \lim_{n\to\infty} \int_{t=0}^{D} \frac{1}{D} \mathrm{E}((N_d(nD+t) - S)^+|\mathbf{C})\, dt$$

$$= \int_{t=0}^{D} \frac{1}{D} \mathrm{E}\left(\lim_{n\to\infty} (N_d(nD+t) - S)^+|\mathbf{C}\right) dt = \int_{t=0}^{D} \frac{1}{D} \mathrm{E}\left(\left({N_d}^t - S\right)^+|\mathbf{C}\right) dt$$

We start with proving the convexity of $\mathrm{E}\left(\left(N_d{}^t - S\right)^+ |\mathbf{C}\right)$ for any $0 < t \leq D$. This can be accomplished by showing that $N_d(nD) \in SDCX(sp)$, and then $N_d(nD + t) \in SDCX(sp)$ for any $n = 0,1,2 \ldots$

Recall the recursive relation of $N_d(nD)$ in (2.22) for $n = 0,1,2, \ldots$

$$N_d(nD) = \left(N_d\big((n-1)D\big) - C^n(D)\right)^+ + N_o(nD^-) \tag{A.1}$$

It is known that the $C^n(D)$ and $N_o(nD^-)$ in this recursive relation are independent Poisson distributed random variables with means of (μD) and (λD) respectively. From 8.A.1 on p.358, (Shaked & Shanthikumar 1994), we know that $C^n(D) \in SIL(sp)$, as a Poisson random variable with a mean linearly increasing with μ. Therefore we can also deduce that:

$-C^n(D) \in SDCX(sp)$. Also, for any initial number of defective parts, $N_d(0)$, is unaffected by μ and therefore $N_d(0) \in SDCX(sp)$. This also holds true for $N_o(nD^-)$, since λ is fixed and unaffected.

Now suppose $N_d\big((n-1)D\big) \in SDCX(sp)$. Since $(.)^+$ is an increasing convex function, Theorem 8.B.10 and 8.B.8 on p.370, (Shaked & Shanthikumar 1994) jointly imply the following:

$$N_d(nD) = \left(N_d\big((n-1)D\big) - C^n(D)\right)^+ + N_o(nD^-) \in SDCX(sp).$$

Therefore, for any $n = 1, 2, \ldots$ we have proved that $N_d(nD) \in SDCX(sp)$ by induction.

From $N_d(nD)$, we can proceed to $N_d(nD + t)$ for $0 < t \leq D$ as follows:

$$N_d(nD + t) = \left(N_d(nD) - C^n(t)\right)^+ + N_o(nD + t) \tag{A.2}$$

Similar to $C^n(D)$, $C^n(t)$ is also a Poisson distributed random variable, having a mean of $\mu \times t$, which is linearly increasing with μ. Therefore, we have $-C^n(t) \in SDCX(sp)$, as well. Likewise, $N_o(nD + t) \in SDCX(sp)$ due to the fact that λ is fixed and not dependent to μ. Since we have shown that $N_d(nD) \in SDCX(sp)$, we can see from (6.2) that $N_d(nD + t) \in SDCX(sp)$, following the same reasoning used in proving the stochastic convexity of $N_d(nD)$.

Since the base stock level S is a given positive integer, $(N_d{}^n(t) - S)^+ \in SDCX(sp)$ holds true, too. From Proposition 2.11 in (Shaked & Shanthikumar 1988), we can conclude that the limit of $(N_d(nD + t) - S)^+$, which is:

$$\lim_{n\to\infty}((N_d(nD + t) - S)^+ |\mathbf{C}) = \left(N_d{}^t - S\right)^+ \in SDCX.$$

Finally, the last result leads us to the convexity of: $\mathrm{E}\left(\left(N_d{}^t - S\right)^+ |\mathbf{C}\right)$ with μ, which is deduced from the stochastic convexity definition. The convexity property is preserved after the integral operation, therefore we can state that

$\int_{t=0}^{D} \frac{1}{D} \mathrm{E}\left(\left(N_d{}^t - S\right)^+ |\mathbf{C}\right) dt = \mathrm{E}((N_d - S)^+|\mathbf{C})$ is convex in μ for a given period length D and a stock level S. *QED*

References

Aberdeen Group, 2003. *Service Parts Management - Unlocking Value and Profits in the Service Chain*,

Adan, I.J.B.F., Sleptchenko, A. & Van Houtum, G.J., 2009. REDUCING COSTS OF SPARE PARTS SUPPLY SYSTEMS VIA STATIC PRIORITIES. *Asia-Pacific Journal of Operational Research (APJOR)*, 26(04), pp.559–585. Available at: http://www.worldscientific.com/doi/abs/10.1142/S0217595909002377 [Accessed August 5, 2012].

Agnihothri, S.R., 1989. Interrelationships between performance measures for the machine-repairman problem. *Naval Research Logistics*, 36(3), pp.265–271. Available at: http://doi.wiley.com/10.1002/1520-6750(198906)36:3<265::AID-NAV3220360304>3.0.CO;2-F [Accessed November 13, 2011].

Albright, S.C., 1980. Optimal maintenance-repair policies for the machine repair problem. *Naval Research Logistics Quarterly*, 27(1), pp.17–27. Available at: http://doi.wiley.com/10.1002/nav.3800270103 [Accessed November 13, 2011].

Albright, S.C. & Gupta, A., 1993a. Steady-state approximation of a multiechelon multi-indentured repairable-item inventory system with a single repair facility. *Naval Research Logistics*, 40(4), pp.479–493.

Albright, S.C. & Gupta, A., 1993b. Steady-state approximation of a multiechelon multi-indentured repairable-item inventory system with a single repair facility. *Naval Research Logistics*, 40(4), pp.479–493.

Albright, S.C. & Soni, A., 1988. Markovian multiechelon repairable inventory system. *Naval Research Logistics*, 35(1), pp.49–61. Available at: http://doi.wiley.com/10.1002/1520-6750(198802)35:1<49::AID-NAV3220350105>3.0.CO;2-U [Accessed November 24, 2011].

Alexander, W.L. et al., 2002. The secret life of factory service centers. *The McKinsey Quartetly*, 3, pp.106–115.

Alp, O. & Tan, T., 2008. Tactical capacity management under capacity flexibility. *IIE Transactions*, 40(3), pp.221–237. Available at:

http://www.tandfonline.com/doi/abs/10.1080/07408170701488052 [Accessed August 26, 2011].

Altiok, T., 1997. *Performance Analysis of Manufacturing Systems*, New York, NY: Springer.

Bekker, R. et al., 2004. Queues with Workload-Dependent Arrival and Service Rates. *Queueing Systems*, 46(3/4), pp.537–556. Available at: http://www.springerlink.com/openurl.asp?id=doi:10.1023/B:QUES.0000027998.95375.ee [Accessed November 15, 2011].

Bekker, R. & Boxma, O., 2007. An M/G/1 Queue with Adaptable Service Speed. *Stochastic Models*, 23(3), pp.373–396. Available at: http://www.tandfonline.com/doi/abs/10.1080/15326340701470945 [Accessed November 15, 2011].

Bertrand, J.W.M., 2003. Supply Chain Design: Flexibility Considerations. In A. G. De Kok & S. C. Graves, eds. *Handbooks in Operations Research and Management Science*. Amsterdam, North-Holland: Supply ChainManagement,, pp. 133–198.

Bitran, G.R., Haas, E.A. & Hax, A.C., 2011. Hierarchical Production Planning: A Two-Stage System. *Operations Research*, 30(2), pp.232–251.

Bradley, J.R. & Glynn, P.W., 2002. Managing Capacity and Inventory Jointly in Manufacturing Systems. *Management Science*, 48(2), pp.273–288.

Buyukkaramikli, N.C., Bertrand, J. Will M. & Ooijen, H.P.G., 2011. Flexible hiring in a make to order system with parallel processing units. *Annals of Operations Research*. Available at: http://www.springerlink.com/index/10.1007/s10479-011-0958-4 [Accessed November 15, 2011].

Buyukkaramikli, N.C.;, Bertrand, J.W.M.; & Ooijen, H.P.G., 2011. Periodic capacity management under a lead-time performance constraint. *OR Spectrum*. Available at: http://www.springerlink.com/index/10.1007/s00291-011-0261-4 [Accessed November 15, 2011].

Buzacott, J.A. & Shanthikumar, G.J., 1993. *Stochastic Models of Manufacturing Systems*, Englewoog Cliffs, New Jersey: Prentice-Hall.

Carr, N.G., 2003. IT Doesn ’ t Matter. *Harvard Business Review*, 81(5), pp.41–49.

Chandra, M.J., 1986. A Study of Multiple Finite-Source Queueing Models. *The Journal of Operational Research Society*, 37(3), pp.275–283.

Chandra, M.J. & Shanthikumar, J.G., 1983. On a machine interference problem with several types of machines attended by a single repairman. *International Journal of Production Research*, 21(4), pp.529–541. Available at: http://www.tandfonline.com/doi/abs/10.1080/00207548308942388 [Accessed November 13, 2011].

Chenery, H.B., 1952. Overcapacity and the acceleration principle. *Econometrica*, 20, pp.1–28. Available at: http://www.jstor.org/stable/1907804.

Cho, D.I. & Parlar, M., 1991. A survey of maintenance models for multi-unit systems. *European Journal of Operational Research*, 51(1), pp.1–23. Available at: http://doi.wiley.com/10.1002/nav.3800280104 [Accessed November 13, 2011].

Coetzee, J., 2004. *Maintenance*, Trafford on Demand Publishers.

Cohen, M.A.., N., A. & Agrawal, V., 2006. Winning in the Aftermarket. *Harvard Business Review*, pp.129–138.

Crabill, T.B., 1974. Optimal Control of a Maintenance System with Variable Service Rates. *Operations Research*, 22(4), pp.736–745.

Dana Jr, J.D. & Spier, K.E., 2001. Revenue sharing and vertical control in the video rental industry. *The Journal of Industrial Economics*, 49(3), pp.223–245. Available at: http://onlinelibrary.wiley.com/doi/10.1111/1467-6451.00147/abstract [Accessed November 28, 2011].

Davenport, T.H., 2005. The Coming Commoditization of Processes The Coming Commoditization of Processes. *Harvard Business Review*, June, pp.100–108.

Deloitte, 2006. *The Service Revolution in Global Manufacturing Industries*,

Van Der Duyn Schouten, F.A. & Wartenhorst, P., 1993. A two-machine repair model with variable repair rate. *Naval Research Logistics*, 40(4), pp.495–523. Available at: http://doi.wiley.com/10.1002/1520-6750(199306)40:4<495::AID-NAV3220400406>3.0.CO;2-V [Accessed November 13, 2011].

Diaz, A. & Fu, Michael C., 1997. Models for multi-echelon repairable item inventory systems with limited repair capacity. *European Journal of Operational Research*, 97(3), pp.480–492. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0377221796002792 [Accessed November 13, 2011].

Du, Y. & Hall, R., 1997. Fleet sizing and empty equipment redistribution for center-terminal transportation networks. *Management Science*, 43(2), pp.145–157. Available at: http://www.jstor.org/stable/10.2307/2661679 [Accessed November 28, 2011].

Ealey, L. A.; Troyano-Bermudez, L., 1996. Are automobiles the next commodity? *The McKinsey Quaterly*, 4, pp.62–75.

Eberly, J. & van Mieghem, J., 1997. Multi-factor Dynamic Investment under Uncertainty,. *Journal of Economic Theory*, 75(2), pp.345–387. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0022053196922814 [Accessed November 14, 2011].

Evans, J.S., 1991. STRATEGIC FLEXIBILITY FOR HIGH TECHNOLOGY MANOEUVRES: A CONCEPTUAL FRAMEWORK. *Journal of Management Studies*, 28(1), pp.69–89. Available at: http://doi.wiley.com/10.1111/j.1467-6486.1991.tb00271.x [Accessed October 18, 2011].

Feeney, G. & Sherbrooke, C.C., 1966. The (s-1, s) inventory policy under compound Poisson demand. *Management Science*, 12(5), pp.391–411. Available at: http://www.jstor.org/stable/10.2307/2627824 [Accessed December 6, 2011].

Fleischmann, M., 1997. Quantitative models for reverse logistics: A review. *European Journal of Operational Research*, 103(1), pp.1–17. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0377221797002300 [Accessed August 10, 2011].

Fu, M.C., Marcus, S.I. & Wang, I., 2000. Monotone optimal policies for a transient queueing staffing problem. *Operations Research*, 48(2), pp.327–331. Available at: http://www.jstor.org/stable/10.2307/223149 [Accessed November 15, 2011].

Gans, N., Koole, G. & Mandelbaum, A., 2003. Commissioned Paper: Telephone Call Centers: Tutorial, Review, and Research Prospects. *Manufacturing and Service Operations Management*, 5(2), pp.79–141. Available at: http://msom.journal.informs.org/cgi/doi/10.1287/msom.5.2.79.16071.

Geraerds, W.M.J., 1992. The EUT maintenance model. *International Journal of Production Economics*, 24(3), pp.209–216. Available at: http://linkinghub.elsevier.com/retrieve/pii/092552739290132Q [Accessed November 13, 2011].

Geraghty, M.K. & Johnson, E., 1997. Revenue management saves national car rental. *Interfaces*, pp.107–127. Available at: http://www.jstor.org/stable/10.2307/25062214 [Accessed November 28, 2011].

Gerchak, Y., Cho, R.K. & Ray, S., 2006. Coordination of quantity and shelf-retention timing in the video movie rental industry. *IIE Transactions*, 38(7), pp.525–536. Available at: http://www.tandfonline.com/doi/abs/10.1080/07408170590961166 [Accessed September 19, 2011].

Gits, C.W., 1992. Design of maintenance. *International Journal of Production Economics*, 24, pp.217–226.

Goheen, L.C., 1977. On the Optimal Operating Policy for the Machine Repair Problem When Failure and Repair Times Have Erlang Distribution. *Operations Research*, 25(3), pp.484–492.

De Graaf-Zijl, M., 2012. COMPENSATION OF ON-CALL AND FIXED-TERM EMPLOYMENT: THE ROLE OF UNCERTAINTY. *The Manchester School*, 80(1), pp.6–27. Available at: http://doi.wiley.com/10.1111/j.1467-9957.2011.02248.x [Accessed August 5, 2012].

Graves, S.C., 1982. Using Lagrangean techniques to solve hierarchical production planning problems. *Management Science*, 28(3), pp.260–275. Available at: http://www.jstor.org/stable/10.2307/2630880 [Accessed November 15, 2011].

Gross, D., Miller, D.R. & Soland, R.M., 1983. A Closed Queueing Network Model for Multi-Echelon Repairable Item Provisioning. *IIE Transactions*, 15(4), pp.344–352. Available at: http://www.tandfonline.com/doi/abs/10.1080/05695558308974658 [Accessed November 13, 2011].

Gupta, A. & Albright, S.C., 1992. Steady-state approximations for a multi-echelon multi-indentured repairable-item inventory system. *European Journal of Operational Research*, 62(3), pp.340–353. Available at: http://linkinghub.elsevier.com/retrieve/pii/037722179290123Q [Accessed November 13, 2011].

de Haas, H., 1995. *The coordination of initial stock and flexible manpower in repairable item systems* PhD Thesis., Eindhoven: TU/e.

Hagen, T., 2002. Do Temporary Workers Receive Risk Premiums? Assessing the Wage Effects of Fixed-term Contracts in West Germany by a Matching Estimator Compared with Parametric Approaches. *Labour*, 16(4), pp.667–705. Available at: http://doi.wiley.com/10.1111/1467-9914.00212 [Accessed August 5, 2012].

Holland, J., 2000. *Emergence: From chaos to order*, Oxford University Press.

Holmes, A., 2008. *Commoditization and the Strategic Response*, Gower Publishing, Ltd.

Holt, C.C.. et al., 1960. *Planning production, inventories and work force.*, Prentice-Hall (Englewood Cliffs, NJ).

Horn, R.. & Johnson, C., 1985. *Matrix Analysis*, Cambridge University Press.

Houseman, S.N., 2001. Why Employers Use Flexible Staffing Arrangements: Evidence from an Establishing Survey. *Industrial and Labor Relations Review*, 55(1), pp.149–170.

Iravani, S.M.R. & Krishnamurthy, V., 2007. Workforce Agility in Repair and Maintenance Environments. *Manufacturing & Service Operations Management*, 9(2), pp.168–184. Available at: http://msom.journal.informs.org/cgi/doi/10.1287/msom.1060.0132 [Accessed November 13, 2011].

Jack, E.P. & Raturi, A., 2002. Sources of volume flexibility and their impact on performance. *Journal of Operations Management*, 20(5), pp.519–548. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0272696301000791 [Accessed September 23, 2011].

Kalleberg, A.L., Reynolds, J. & Marsden, P.V., 2003. Externalizing employment: flexible staffing arrangements in US organizations. *Social Science Research*, 32(4), pp.525–552. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0049089X03000139 [Accessed October 4, 2011].

Kameda, H., 1982. A Finite-Source Queue with Different Customers. *Journal of the ACM*, 29(2), pp.478–491. Available at: http://portal.acm.org/citation.cfm?doid=322307.322320 [Accessed November 13, 2011].

Keizers, J., 2000. *Subcontracting as a Capacity Management Tool in Multi-Project Repair Shops* PhD Thesis., TU/e.

Keizers, J.M., Bertrand, J. Will M. & Wessels, J., 2009. Diagnosing Order Planning Performance at a Navy Maintenance and Repair Organization, using Logistic Regression. *Production and Operations Management*, 12(4), pp.445–463. Available at: http://doi.wiley.com/10.1111/j.1937-5956.2003.tb00214.x [Accessed November 24, 2011].

Kelly, F., 1979. *Reversibility and Stochastic Networks*, John Wiley.

Kouvelis, P. & Milner, J.M., 2002. Supply chain capacity and outsourcing decisions: the dynamic interplay of demand and supply uncertainty. *IIE Transactions*, 34(8), pp.717–728.

Kulkarni, V.G., 1995. *Modeling and Analysis with Stochastic Systems*, London, UK: Chapman& Hall.

Ledermann, W. & Reuter, G.E.H., 1954. Spectral theory for the differential equations of simple birth and death processes. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 246(914), pp.321–369. Available at: http://www.jstor.org/stable/10.2307/91569 [Accessed November 28, 2011].

Lindley, D.V., 2008. The theory of queues with a single server. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(02), pp.277–289.

Lippman, S., 1975. Applying a new device in the optimization of exponential queuing systems. *Operations Research*, 23(4), pp.687–710. Available at: http://www.jstor.org/stable/10.2307/169851 [Accessed November 15, 2011].

Love, C.E., 1985. Source/sink inventory control for vehicle rental systems. *Computers & Operations Research*, 12(4), pp.349–364. Available at: http://linkinghub.elsevier.com/retrieve/pii/0305054885900334 [Accessed November 28, 2011].

Manatayev, Y.Y., 2004. Commoditization of the Third Party Logistics Industry.

McLean, T., 2007. Commoditization of the international teleradiology market. *Journal of health services research & policy*, 12(2), pp.120–2. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17407664 [Accessed November 15, 2011].

Milner, J.M. & Pinker, Edieal J, 2001. Contingent Demand Labor and Contracting Under Supply Uncertainty. *Management Science*, 47(8), pp.1046–1062.

Muckstadt, J.A., 1973. A Model for a Multi-Item, Multi-Echelon, Multi-Indenture Inventory System. *Management Science*, 20(4), pp.472–481.

Muckstadt, J.A., 2005. *Analysis and Algorithms for Service Parts Supply Chains*, Springer Series in Operations Research and Financial Engineering.

Neuts, M.F., 1981. *Matrix-Geometric Solutions in Stochastic Models- An Algorithmic Approach*, Baltimore: The John Hopkins University Press.

ORACLE, 2000. *ORACLE Capacity User Guide*, Available at: http://docs.oracle.com/cd/B11454_01/11.5.9/acrobat/115crpug.pdf.

Oliva, R. & Kallenberg, R., 2003. Managing the transition from products to services. *International Journal of Service Industry Management*, 14(2), pp.160–172.

Olson, E.G. & Sharma, D., 2008. Beating the commoditization trend: a framework from the electronics industry. *Journal of Business Strategy*, 29(4), pp.22–28. Available at: http://www.emeraldinsight.com/10.1108/02756660810886971 [Accessed November 15, 2011].

Pachon, J.E. et al., 2003. A Synthesis of Tactical Fleet Planning Models for the Car Rental Industry. *IIE Transactions*, 35(9), pp.907–916. Available at: http://www.tandfonline.com/doi/abs/10.1080/07408170304409 [Accessed November 28, 2011].

Papier, F. & Thonemann, U.W., 2009. Capacity Rationing in Stochastic Rental Systems with Advance Demand Information. *Operations Research*, 58(2), pp.274–288. Available at: http://or.journal.informs.org/cgi/doi/10.1287/opre.1090.0745 [Accessed October 14, 2011].

Papier, F. & Thonemann, U.W., 2008. Queuing Models for Sizing and Structuring Rental Fleets. *Transportation Science*, 42(3), pp.302–317. Available at: http://transci.journal.informs.org/cgi/doi/10.1287/trsc.1070.0225 [Accessed September 9, 2011].

Parikh, S.C., 1977. On a Fleet Sizing and Allocation Problem. *Management Science*, 23(9), pp.972–977. Available at: http://mansci.journal.informs.org/cgi/doi/10.1287/mnsc.23.9.972 [Accessed November 28, 2011].

Perlman, Y., Mehrez, A.. & Kaspi, M., 2001. Setting expediting repair policy in a multi-echelon repairable-item inventory system with limited repair capacity. *Journal of the Operational Research Society*, 52(2), pp.198–209. Available at: http://www.jstor.org/stable/10.2307/254148 [Accessed November 24, 2011].

Pierskalla, W.P. & Voelker, J.A., 1976. A survey of maintenance models: The control and surveillance of deteriorating systems. *Naval Research Logistics Quarterly*, 23(3), pp.353–388. Available at: http://doi.wiley.com/10.1002/nav.3800230302 [Accessed November 13, 2011].

Pinker, E.J., 1996. *Models of flexible workforce management in uncertain environments* PhD Thesis. MIT, ed., Cambridge, MA.

Pinker, E.J. & Larson, R.C., 2003. Optimizing the use of contingent labor when demand is uncertain. *European Journal of Operational Research*, 144(1), pp.39–55. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0377221701003782 [Accessed November 14, 2011].

Puterman, M.L., 1994. *Markov Decision Processes—Discrete Stochastic Dynamic Programming*, New York, NY: Wiley.

Reimann, M., Schilke, O. & Thomas, J.S., 2010. Toward an understanding of industry commoditization: Its nature and role in evolving marketing competition. *International Journal of Research in Marketing*, 27(2), pp.188–197. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167811609000858 [Accessed November 15, 2011].

Rogers, R.D. & Monsell, S., 1995. Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2), pp.207–231. Available at: http://psycnet.apa.org/psycinfo/1995-31890-001 [Accessed August 5, 2012].

Rosen, S., 1986. Chapter 12 The theory of equalizing differences. In *Handbook of Labor Economics*. pp. 641–692.

Ross, S., 1983. *Stochastic Processes*, New York, NY: Wiley.

Safaei, N, Banjevic, D & Jardine, a K.S., 2010. Bi-objective workforce-constrained maintenance scheduling: a case study. *Journal of the Operational Research Society*, 62(6), pp.1005–1018. Available at: http://www.palgrave-journals.com/doifinder/10.1057/jors.2010.51 [Accessed August 5, 2012].

Safaei, Nima, Banjevic, Dragan & Jardine, A.K.S., 2011. Workforce-constrained maintenance scheduling for military aircraft fleet: a case study. *Annals of Operations Research*, 186(1), pp.295–316. Available at: http://www.springerlink.com/index/10.1007/s10479-011-0885-4 [Accessed August 5, 2012].

Savin, S., Cohen, M. & Gans, N., 2005. Capacity management in rental businesses with two customer bases. *Operations Research*, 53(4), pp.617–631. Available at: http://www.jstor.org/stable/10.2307/25146897 [Accessed November 28, 2011].

Schmenner, R.W., 1995. *Service Operations Management*, Prentice-Hall (Englewood Cliffs, NJ).

Scudder, G. & Chua, R., 1987. Determining overtime policies for a repair shop. *Omega*, 15(3), pp.197–206. Available at: http://linkinghub.elsevier.com/retrieve/pii/0305048387900703 [Accessed November 13, 2011].

Scudder, G.D., 1985. An evaluation of overtime policies for a repair shop. *Journal of Operations Management*, 6(1), pp.87–98. Available at: http://linkinghub.elsevier.com/retrieve/pii/0272696385900373 [Accessed August 5, 2012].

Scudder, Gary D. & Hausman, W.H., 1982. Spares stocking policies for repairable items with dependent repair times. *Naval Research Logistics Quarterly*, 29(2), pp.303–322. Available at: http://doi.wiley.com/10.1002/nav.3800290211 [Accessed November 13, 2011].

Sennott, L.I., 1998. *Stochastic Dynamic Programming and the Control of Queueing Systems* 1st Editio., Wiley-Interscience.

Shaked, M.. & Shanthikumar, J.G., 1994. *Stochastic Orders and Their Applications*, San Diego, CA: Academic Press Inc.

Shaked, M.. & Shanthikumar, J.G.., 1988. Stochastic Convexity and its Applications. *Advances in Applied Probability*, 20, pp.427–446.

Shawky, A., 1997. The single-server machine interference model with balking, reneging and an additional server for longer queues. *Microelectronics and Reliability*, 37(2), pp.355–357. Available at: http://linkinghub.elsevier.com/retrieve/pii/0026271494000794 [Accessed November 13, 2011].

Sherbrooke, C., 1992. *Optimal Inventory Modeling of Systems*, New York, NY: John Wiley & Sons.

Sherbrooke, Craig C, 1967. METRIC : A Multi-Echelon Technique for Recoverable Item Control. *Operations Research*, 16(1), pp.122–141.

Sherbrooke, Craig C, 1986. Vari-Metric: Improved Approximations for Multi-Indenture, Multi-Echelon Availability Models. *Operations Research*, 34(2), pp.311–319.

Sherif, Y.S. & Smith, M.L., 1981. Optimal maintenance models for systems subject to failure–A Review. *Naval Research Logistics Quarterly*, 28(1), pp.47–74. Available at: http://doi.wiley.com/10.1002/nav.3800280104 [Accessed November 13, 2011].

Sleptchenko, A., van der Heijden, M.C. & van Harten, A., 2003. Trade-off between inventory and Repair Capacity in Spare Part Networks. *Journal of the Operational Research Society*, 54, pp.263–272.

Smith, A., 1776. *The Wealth of the Nations* 1979th ed. A. Skinner, ed., Baltimore: Baltimore: Baltimore: Penguin.

Stecke, K. & Aronson, J.E., 1985. Review of operator/machine interference models. *International Journal of Production Research*, 23(1), pp.129–151.

Stern, T., 1979. Approximations of Queue Dynamics and Their Application to Adaptive Routing in Computer Communication Networks. *IEEE Transactions on Communications*, 27(9), pp.1331–1335. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1094546 [Accessed November 26, 2011].

StidhamJr, S., 2002. Analysis, Design, and Control of Queueing Systems. *Operations Research*, 50(1), pp.197–216.

Suarez, F.F., Cusumano, M.A.. & Fine, C.H., 1995. An Empirical Study of Flexibility in Manufacturing. *Sloan Management Review*, 37(1), pp.25–32.

Tainiter, M., 1964. Some stochastic inventory models for rental situations. *Management Science*, 11(2), pp.316–326. Available at: http://www.jstor.org/stable/10.2307/2627604 [Accessed November 28, 2011].

Tang, C. & Deo, S., 2008. Rental price and rental duration under retail competition. *European Journal of Operational Research*, 187(3), pp.806–828. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0377221706007843 [Accessed June 28, 2011].

Tijms, H., 1994. *Stochastic Models, An Algorithmic Approach*, Chichester, UK: Wiley.

Turnquist, M.A. & Jordan, W.C., 1986. Fleet Sizing under Production Cycles and Uncertain Travel Times. *Transportation Science*, 20(4), pp.227–236. Available at: http://transci.journal.informs.org/cgi/doi/10.1287/trsc.20.4.227 [Accessed November 28, 2011].

van der Veen, J.A.A. & Venugopal, V., 2005. Using revenue sharing to create win–win in the video rental supply chain. *Journal of the Operational Research Society*, 56(7), pp.757–762.

Vliegen, I., 2009. *Integrated planning for service tools and spare parts for capital goods*. Technische Universiteit Eindhoven. Available at: http://www.narcis.nl/publication/RecordID/oai:library.tue.nl:653236 [Accessed August 5, 2012].

Wang, H., 2002. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139(3), pp.469–489. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0377221701001977 [Accessed November 13, 2011].

Weil, H., 1996. Commoditization of Technology Based Products and Services: a generic Model of Market Dynamics.

Whisler, W.D., 1967. A stochastic inventory model for rented equipment. *Management Science*, 13(9), pp.640–647. Available at: http://www.jstor.org/stable/10.2307/2628083 [Accessed November 28, 2011].

Winston, W., 1977. Optimal Control of Discrete and Continuous Time Maintenance Systems with Variable Service Rates. *Operations Research*, 25(2), pp.259–268.

Winston, W., 1978. Optimality of Monotonic Policies for Multiple-Server Exponential Queuing Systems with State-Dependent Arrival Rates. *Operations Research*, 26(6), pp.1089–1094.

Yan, S., Yang, T.-H. & Chen, H.-H., 2004. Airline short-term maintenance manpower supply planning. *Transportation Research Part A: Policy and Practice*, 38(9-10), pp.615–642. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0965856404000679 [Accessed August 5, 2012].

Yang, T.-H., Yan, S. & Chen, H.-H., 2003. An airline maintenance manpower planning model with flexible strategies. *Journal of Air Transport Management*, 9(4), pp.233–239. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0969699703000139 [Accessed August 5, 2012].

Yoo, J., 1996. *Queueing Models for Staffing Service Operations* PhD Thesis., College Park: University of Maryland.

Zijm, W.H. & Avsar, Z.M., 2003. Capacitated two-indenture models for repairable item systems. *International Journal of Production Economics*, 81-82, pp.573–588. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0925527302003602 [Accessed November 13, 2011].

Zijm, W.H.M. & Buitenhek, R., 1996. Capacity planning and lead time management. *International Journal of Production Economics*, 46-47, pp.165–179. Available at: http://linkinghub.elsevier.com/retrieve/pii/0925527395001611 [Accessed December 14, 2011].

Summary

In the last decades, after-sales services have become increasingly important since service is a source of differentiation as well as a lucrative business opportunity due to the substantial amount of revenue that can be generated from the products in use throughout their life cycle. Following this trend, many after-sales service providers have emerged in the market or evolved as semi-autonomous units within the OEM (Original Equipment Manufacturer) companies.

In this thesis, we focus on the maintenance aspect of after-sales services. We assume that a maintenance service provider (MSP) is running a repair shop in an environment with numerous operating systems that are prone to failure. The MSP is responsible for keeping all systems in an environment up and working. We mainly focus on two types of environments: 1) Specialized System Environment 2) Commoditized System Environment.

The systems in the first environment are highly customized. They are designed and built specifically following the owners' precise requirements. Defense systems, specific lithography systems, mission aircrafts or other advanced/complex, engineer-to-order capital goods are examples of such specialized systems. Due to the diversity of owners' requirements, each system develops many unique characteristics, which make it hard, if not impossible, to find a substitute for the system, in the market as a whole.

In the second environment, the systems are more generic in terms of their functionality. Trucks, cranes, printers, copy machines, forklifts, computer systems, cooling towers, some common medical devices (i.e. anesthesia, x-ray and ultrasound machines, etc…), power systems are examples of such more commoditized systems. Due to the more generic features of the owners' requirements, it is easier to find a substitute for a system in the market, with more or less the same functionality, for short-term hiring purposes.

Upon a system breakdown, the defective unit (system/subsystem) is sent to the repair shop. MSP is responsible for the repair and also liable for the costs related to the down time. In order to alleviate the down-time costs, there are chiefly two different downtime service strategies that the MSP can follow, depending on the environment the repair shop is operating in. In the specialized system environment, the MSP holds a spare unit inventory for the critical subsystem that causes most of the failures. The downtime service related decision in such a case would be the inventory level of the critical spare subsystems. On the other hand, in the commoditized system environment, rather than keeping a spare unit inventory, the MSP hires a substitute system from an agreed rental store/3rd party supplier. The downtime service related decision in this case is the hiring duration.

Next to the above downtime service related decisions, repair shop's capacity level is the other primary determinant of the systems' uptime/availability. Since maintenance is a labor-intensive industry, the capacity costs constitute a large portion of the total costs. Increasing pressure on profitability and the growing role of External Labor Supplier Agencies motivate service provider firms to scrutinize the prospects and possibilities of capacity flexibility by using contingent workforce. For various reasons, flexible capacity practices in real life are often periodic, and the period length is both a decision parameter and a metric for flexibility.

A shorter period length implies more frequent adapting possibilities and a better tailoring of the capacity. On the other hand, the flexible capacity cost per unit time is higher for shorter period lengths due to the compensating wage differentials, which models the relation between the wage rate and the unpleasantness, risk or other undesirable attributes of the job. Certainly, short period length in this context is an undesirable attribute for the flexible capacity resource, as it mandates the resource to switch tasks and to be ready/available more frequently, without the guarantee that s/he will be actually employed. Therefore, we propose several empirically testable functional forms for the cost rate of a flexible capacity unit, which are decreasing with the period length and, in the limit, approaches to the cost rate of a permanent capacity unit from above.

In the light of discussions above, we investigate three different capacity modes in this dissertation:

- *Fixed Capacity Mode:* In this mode, all of the capacity is permanent and ready for use in the repair shop. This mode serves as a reference point in order to assess the benefits of other flexible capacity modes. The relevant capacity decision in this mode is the single capacity level of the repair shop.
- *Periodic Two-Level Capacity Mode:* In this mode, we assume two levels of repair shop capacity: permanent and permanent plus contingent capacity levels. The permanent capacity is always available in the system, whereas the deployment of the contingent capacity is decided at the start of each period based on the number of units waiting to be repaired in the shop. The relevant capacity decisions in this mode are the permanent and contingent capacity levels, the period length and the states (in terms of number of defective units waiting) where the contingent capacity is deployed.
- *Periodic Capacity Sell-Back Mode:* In this mode, the failed units are sent to the repair shop at regular intervals in time. Due to this admission structure, when the repair of all the defective units in the repair shop are completed in a period, it is known that no new defective parts will arrive to the shop at least until the start of the next period. This certainty in idle times allows for a contract, where the repair shop capacity is sold at a reduced price to the capacity agency where it is assigned to other tasks until the start of the next period. The original cost of the multi-skilled repair shop capacity per time unit is higher than the permanent capacity cost that is mentioned in previous modes due to the compensation

factors such as additional skills, frequent task switching and transportation/transaction costs. Similar to the previous capacity mode, the compensation decreases with the length of the period length. The relevant capacity decisions in this mode are the capacity level and the period length.

The primary goal of this thesis is to develop quantitative models and methods for taking optimal capacity decisions for the repair shop in the presence of the capacity modes described above and to integrate these decisions with the other downtime service decisions of the MSP for two different types of system environments (specialized vs. commoditized). After the introduction of the problem, concepts and literature review are given in Chapters 1.

In Chapter 2, we focus on the use of capacity flexibility in the repair operations of the MSP in specialized system environment. The capacity related decisions are integrated with the decision on the stock level of the spare unit inventory for all three capacity modes. In Chapter 3 we investigate the same three capacity modes in a (partially) commoditized system environment, where hiring a substitute system for a pre-determined, uniform duration becomes the conventional method upon a failure. In this chapter the decision on the hiring duration is integrated with the other capacity related decisions. Then we provide some preliminary analysis and give the early results on the hybrid strategy where both “keeping stock” and “hire substitute” strategies are followed. Finally in Chapter 4, we summarize our results, give the conclusion and discuss the topics covered in this thesis with a brief exploration on the future research.

The numerical results reveal that, in both specialized and commoditized system environments, substantial cost savings (up to 70%) can be achieved under periodic two-level capacity and periodic capacity sell-back modes compared to the fixed capacity mode. However, both period length and the compensation scheme of the capacity resources greatly influence the savings, even in some cost instances, flexible modes (periodic two-level and capacity sell-back) become less economical compared to the fixed capacity mode. Cost parameter instances in which each of the 3 capacity modes becomes cost-optimal, the characteristics of the cost savings and the sensitivity analysis of cost/policy parameters are investigated in both of the system environments in Chapter 2 and Chapter 3, respectively.

In the commoditized system environment, under the same cost parameter settings, the hiring substitute from an external supplier for a fixed duration causes a better, more refined and certain control compared to keeping an inventory. Hybrid strategy, in which a substitute is hired after a stock-out instance, is applicable in commoditized as well as commoditizing (previously specialized systems that are in the ongoing commoditization process) system environments. Hybrid strategy outperforms both "only keeping stock" and "only hiring substitute" alternatives; however, in the commoditized system environment, a MSP may still have a proclivity to employ the "hiring substitute" strategy only, because it does not require any initial investment, which is convenient for SMEs. These issues will be explicated further in Chapter 5.

We believe that the framework, the design and analysis of the problems addressed as well as the results and the insights obtained in this dissertation can help and motivate other researchers/practitioners to further investigate the cost saving prospects from capacity flexibility in maintenance service operations. We also anticipate that the commoditization framework described in this thesis will be increasingly useful in the future, since the commoditization of the parts/machines will be much more widespread, pushing all the after-sales service providers to compete on the efficiency of their operation

Acknowledgements

The realization of this thesis would not have been possible without the help of others. Firstly, I would like to express my gratitude to my first promoter Will Bertrand. I have learned a lot from you and enjoyed your wisdom. Your enthusiasm, intellectuality, open mindedness and flexibility made me one of the luckiest candidates. I am greatly indebted to you and more than proud to be your last PhD student.

I would also like to express my gratefulness to my second promoter Ton de Kok and daily supervisor Henny van Ooijen. Both of you gave me great ideas and critical feedback on the previous versions of the thesis.

I would like to thank Nesim Erkip, Philippe Chevalier, Geert Jan van Houtum and Ivo Adan for kindly accepting the invitation to take place in my defense committee. Again, they provided very useful comments on the earlier version of this thesis.

I enjoyed being a member of the OPAC and EURANDOM teams. Thanks for all the cozy times in the last four years. I am grateful to my new colleagues at PHARMERIT as they supported me under stressful times many times. I am very thankful to my other friends and my family, especially my mother and my father, as I wouldn’t be here without them. Last but not the least, I thank to my better half, for completing my life, supporting me when I am down, bearing with me when I am stressed. I appreciate all the things you ‘ve done.

Curriculum Vitae

Nasuh Çagdas Büyükkaramikli was born on May 18th , 1983 in Ankara, Turkey. He has a Bachelor (2004) and a Master (2006) degree in Industrial Engineering from Bilkent University, Turkey.

His master thesis is about the coordination of inventory replenishment and transportation operations in a two echelon supply chain with multiple retailers under transportation capacity. In 2007, he started his PhD project at Eindhoven University of Technology, of which the results are presented in this dissertation. During his PhD study, he was also a part-time member of EURANDOM and collaborated with EURANDOM members on several projects. Since 2011 he is employed at Pharmerit as a research consultant.