Combinatorial optimization and Markov decision process for planning MRI examinations

Na Geng

N° d'ordre : 2010 EMSE 0572

THÈSE

PRESENTEE PAR

Na GENG

POUR OBTENIR LE GRADE DE
DOCTEUR DE L'ÉCOLE NATIONALE SUPERIEURE DES MINES DE SAINT-ÉTIENNE

SPECIALITE : GENIE INDUSTRIEL

Combinatorial optimization and Markov decision process for planning MRI examinations

soutenue à Shanghai, Chine, le 29 Avril 2010

Membres du jury

Président	Alexandre DOLGUI	Professeur, École des Mines de Saint-Étienne
Rapporteur:	Maria DI MASCOLO	Chargée de recherche CNRS-HDR, Grenoble
Rapporteur:	Liming LIU	Professeur, Hong Kong Polytechnic University
Examinateur:	Alexandre DOLGUI	Professeur, École des Mines de Saint-Étienne
Examinateur:	Xiaobo ZHAO	Professeur,Tsinghua University, Beijing
Directeur:	Xiaolan XIE	Professeur, École des Mines de Saint-Étienne
Co-directeur :	Zhibin JIANG	Professeur, Shanghai Jiao Tong University

Spécialités doctorales :	**Responsables :**
SCIENCES ET GENIE DES MATERIAUX	J. DRIVER Directeur de recherche – Centre SMS
MECANIQUE ET INGENIERIE	A. VAUTRIN Professeur – Centre SMS
GENIE DES PROCEDES	G. THOMAS Professeur – Centre SPIN
SCIENCES DE LA TERRE	B. GUY Maître de recherche – Centre SPIN
SCIENCES ET GENIE DE L'ENVIRONNEMENT	J. BOURGOIS Professeur – Centre SITE
MATHEMATIQUES APPLIQUEES	E. TOUBOUL Ingénieur – Centre G2I
INFORMATIQUE	O. BOISSIER Professeur – Centre G2I
IMAGE, VISION, SIGNAL	JC. PINOLI Professeur – Centre CIS
GENIE INDUSTRIEL	P. BURLAT Professeur – Centre G2I
MICROELECTRONIQUE	Ph. COLLOT Professeur – Centre CMP

Enseignants-chercheurs et chercheurs autorisés à diriger des thèses de doctorat (titulaires d'un doctorat d'État ou d'une HDR)

AVRIL	Stéphane	MA	Mécanique & Ingénierie	CIS
BATTON-HUBERT	Mireille	MA	Sciences & Génie de l'Environnement	SITE
BENABEN	Patrick	PR 2	Sciences & Génie des Matériaux	CMP
BERNACHE-ASSOLANT	Didier	PR 0	Génie des Procédés	CIS
BIGOT	Jean-Pierre	MR	Génie des Procédés	SPIN
BILAL	Essaïd	DR	Sciences de la Terre	SPIN
BOISSIER	Olivier	PR 2	Informatique	G2I
BOUCHER	Xavier	MA	Génie Industriel	G2I
BOUDAREL	Marie-Reine	MA	Génie Industriel	DF
BOURGOIS	Jacques	PR 0	Sciences & Génie de l'Environnement	SITE
BRODHAG	Christian	DR	Sciences & Génie de l'Environnement	SITE
BURLAT	Patrick	PR 2	Génie industriel	G2I
COLLOT	Philippe	PR 1	Microélectronique	CMP
COURNIL	Michel	PR 0	Génie des Procédés	DF
DAUZERE-PERES	Stéphane	PR 1	Génie industriel	CMP
DARRIEULAT	Michel	IGM	Sciences & Génie des Matériaux	SMS
DECHOMETS	Roland	PR 1	Sciences & Génie de l'Environnement	SITE
DESRAYAUD	Christophe	MA	Mécanique & Ingénierie	SMS
DELAFOSSE	David	PR 1	Sciences & Génie des Matériaux	SMS
DOLGUI	Alexandre	PR 1	Génie Industriel	G2I
DRAPIER	Sylvain	PR 2	Mécanique & Ingénierie	SMS
DRIVER	Julian	DR	Sciences & Génie des Matériaux	SMS
FEILLET	Dominique	PR 2	Génie Industriel	CMP
FOREST	Bernard	PR 1	Sciences & Génie des Matériaux	CIS
FORMISYN	Pascal	PR 1	Sciences & Génie de l'Environnement	SITE
FORTUNIER	Roland	PR 1	Sciences & Génie des Matériaux	SMS
FRACZKIEWICZ	Anna	DR	Sciences & Génie des Matériaux	SMS
GARCIA	Daniel	CR	Génie des Procédés	SPIN
GIRARDOT	Jean-Jacques	MR	Informatique	G2I
GOEURIOT	Dominique	MR	Sciences & Génie des Matériaux	SMS
GRAILLOT	Didier	DR	Sciences & Génie de l'Environnement	SITE
GROSSEAU	Philippe	MR	Génie des Procédés	SPIN
GRUY	Frédéric	MR	Génie des Procédés	SPIN
GUILHOT	Bernard	DR	Génie des Procédés	CIS
GUY	Bernard	MR	Sciences de la Terre	SPIN
GUYONNET	René	DR	Génie des Procédés	SPIN
HERRI	Jean-Michel	PR 2	Génie des Procédés	SPIN
INAL	Karim	MR	Microélectronique	CMP
KLÖCKER	Helmut	MR	Sciences & Génie des Matériaux	SMS
LAFOREST	Valérie	CR	Sciences & Génie de l'Environnement	SITE
LERICHE	Rodolphe	CR	Mécanique et Ingénierie	SMS
LI	Jean-Michel	EC (CCI MP)	Microélectronique	CMP
LONDICHE	Henry	MR	Sciences & Génie de l'Environnement	SITE
MOLIMARD	Jérôme	MA	Mécanique et Ingénierie	SMS
MONTHEILLET	Frank	DR 1 CNRS	Sciences & Génie des Matériaux	SMS
PERIER-CAMBY	Laurent	PR1	Génie des Procédés	SPIN
PIJOLAT	Christophe	PR 1	Génie des Procédés	SPIN
PIJOLAT	Michèle	PR 1	Génie des Procédés	SPIN
PINOLI	Jean-Charles	PR 0	Image, Vision, Signal	CIS
STOLARZ	Jacques	CR	Sciences & Génie des Matériaux	SMS
SZAFNICKI	Konrad	CR	Sciences & Génie de l'Environnement	DF
THOMAS	Gérard	PR 0	Génie des Procédés	SPIN
TRIA	Assia	(Alès)		CMP
VALDIVIESO	François	MA	Sciences & Génie des Matériaux	SMS
VAUTRIN	Alain	PR 0	Mécanique & Ingénierie	SMS
VIRICELLE	Jean-Paul	MR	Génie des procédés	SPIN
WOLSKI	Krzysztof	MR	Sciences & Génie des Matériaux	SMS
XIE	Xiaolan	PR 1	Génie industriel	CIS

Glossaire :

PR 0	Professeur classe exceptionnelle
PR 1	Professeur 1ère catégorie
PR 2	Professeur 2ème catégorie
MA(MDC)	Maître assistant
DR (DR1)	Directeur de recherche
Ing.	Ingénieur
MR(DR2)	Maître de recherche
CR	Chargé de recherche
EC	Enseignant-chercheur
IGM	Ingénieur général des mines

Centres :

SMS	Sciences des Matériaux et des Structures
SPIN	Sciences des Processus Industriels et Naturels
SITE	Sciences Information et Technologies pour l'Environnement
G2I	Génie Industriel et Informatique
CMP	Centre de Microélectronique de Provence
CIS	Centre Ingénierie et Santé

Dernière mise à jour le : 12 janvier 2010

Abstract

This research is motivated by our collaborations with a large French university teaching hospital in order to reduce the Length of Stay (LoS) of stroke patients treated in the neurovascular department. Quick diagnosis is critical for stroke patients but relies on expensive and heavily used imaging facilities such as MRI (Magnetic Resonance Imaging) scanners. Therefore, it is very important for the neurovascular department to reduce the patient LoS by reducing their waiting time of imaging examinations.

From the neurovascular department perspective, this thesis proposes a new MRI examinations reservation process in order to reduce patient waiting times without degrading the utilization of MRI. The service provider, i.e., the imaging department, reserves each week a certain number of appropriately distributed contracted time slots (CTS) for the neurovascular department to ensure quick MRI examination of stroke patients. In addition to CTS, it is still possible for stroke patients to get MRI time slots through regular reservation (RTS).

This thesis first proposes a stochastic programming model to simultaneously determine the contract decision, i.e., the number of CTS and its distribution, and the patient assignment policy to assign patients to either CTS or RTS. To solve this problem, structure properties of the optimal patient assignment policy for a given contract are proved by an average cost Markov decision process (MDP) approach. The contract is determined by a Monte Carlo approximation approach and then improved by local search. Computational experiments show that the proposed algorithms can efficiently solve the model. The new reservation process greatly reduces the average waiting time of stroke patients. At the same time, some CTS cannot be used for the lack of patients.

To reduce the unused CTS, we further explore the possibility of the advance cancellation of CTS. Structure properties of optimal control policies for one-day and two-day advance cancellation are established separately via an average-cost MDP approach with appropriate modeling and advanced convexity concepts used in control of queueing systems. Computational experiments show that appropriate advance cancellations of CTS greatly reduce the unused CTS with nearly the same waiting times.

Acknowledgments

I would like to express my gratitude to Professor Xiaolan XIE and Professor Zhibin JIANG, my directors, for their joint guidance, support and patience. This work would not have been possible without their direction and financial support.

I also would like to thank the faculties and colleagues in Ecole des Mine de Saint Etienne, France and Shanghai Jiao Tong University, China. They give me a lot of helps during my stay.

Finally, I am very grateful to my family for their support and encouragement throughout my long study. In particular, Dr. Yuying SHAO, my husband, gives me a lot of support and help.

Content

List of tables

List of figures

Chapter 1

Planning of MRI examinations: Introduction

1.1 Introduction

This thesis is motivated by our collaborations with a large French university teaching hospital in order to reduce the Length of Stay (LoS) of stroke patients treated in the neurovascular department.

A stroke is a sudden loss of the brain function caused by lack of blood supply to the brain (ischemic stroke) or rupture of blood vessels in the brain (hemorrhagic stroke). Stroke patients may suffer from the inability to speak or speak clearly, walk, or move a limb because of the lack of blood supply to the brain. The brain cannot tolerate long periods without blood flow and stroke patients need the appropriate treatment as soon as possible.

Before starting the treatment, a number of examinations are needed for diagnosis purpose. Significant delays are observed as many key examinations rely on expensive and heavily used imaging facilities such as MRI (Magnetic Resonance Imaging) scanners facing demand from all medical units of the hospital. Therefore, it is very important to reduce the LoS of stroke patients by reducing their waiting time for imaging examinations.

In this thesis, we restrict ourselves to MRI examinations for two reasons. First, delays for MRI examinations are observed as the longest ones, from 30 to 40 days. Second, joint optimization of all medical examinations is fairly complex and will be subject of our further research. Insights gained from this thesis will be exploited in joint optimization of all medical examinations.

1.2 MRI examination reservation for stroke patients

1.2.1 Stroke and MRI scan

A stroke (sometimes called an acute cerebrovascular attack) is a sudden loss of the brain function due to disturbance in the blood supply to the brain. Strokes can be grouped into two major classes: ischemic and hemorrhagic (Kidwell and Warach (2003)), as shown in Fig.

1.1. Ischemic strokes are due to block of the blood supply to the brain, whereas hemorrhagic strokes are due to rupture of a blood vessel or an abnormal vascular structure in the brain. The affected area of the brain is unable to function, which leads to the inability to move one or more limbs in one side of the body, inability to understand or speak clearly, or inability to see one side of the visual field.

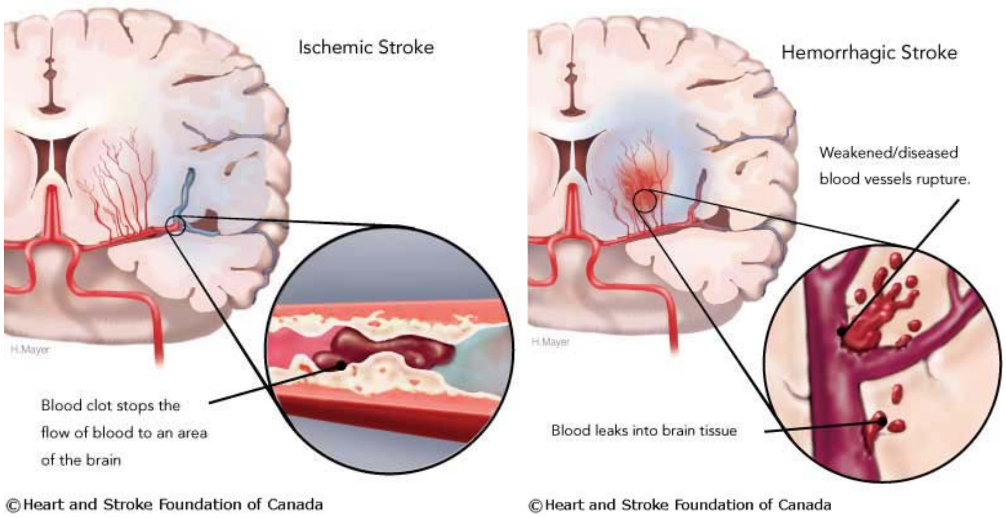


FIG. 1.1 Photos for ischemic and hemorrhagic stroke

Photos from (http://www.strokegenomics.org/index.php?page=about-stroke-genetics).

A stroke is a medical emergency which can cause permanent neurological damage, complications, and death. It is the leading cause of adult disability in the United States and Europe. It is the number two cause of death worldwide and may soon become the leading cause of death worldwide (Donnan et al. (2008)).

Stroke diagnosis needs rapid access to medical personnel and diagnosis facilities. So all tests can be done timely, and the right diagnosis can be made, and appropriate treatment can be provided. The diagnosis of stroke itself is clinical, with assistance from the imaging techniques in finding the causes of stroke. There are two major imaging techniques: Computed Tomography (CT) scanner and MRI scanner.

When stroke patients are diagnosed, many other examinations have to be performed in order to determine the underlying etiology. Commonly used techniques include:

- an ultrasound/doppler study of the carotid arteries (to detect carotid stenosis) or dissection of the precerebral arteries
- an electrocardiogram (ECG) and echocardiogram (to identify arrhythmias and resultant clots in the heart which may spread to the brain vessels through the bloodstream)

- a Holter monitor study to identify intermittent arrhythmias
- an angiogram of the cerebral vasculature (if a bleed is thought to have originated from an aneurysm or arteriovenous malformation)
- blood tests to determine hypercholesterolemia, bleeding diathesis and some rarer causes such as homocysteinuria

Among all the examinations, MRI scan is one of the most helpful tests in the diagnosis of stroke because it can detect strokes within minutes of their onset and is superior to CT. As shown in Fig. 1.2, MRI scanner, or nuclear magnetic resonance imaging (NMRI) scanner, is primarily a medical imaging technique most commonly used in radiology to visualize detailed internal structure and limited function of the body. MRI provides much greater contrast between different soft tissues of the body than CT does, which makes it especially useful in neurological, musculoskeletal, cardiovascular, and oncological imaging. Unlike CT, it uses no ionizing radiation, but uses a powerful magnetic field to align the nuclear magnetization of hydrogen atoms in water in the body.

A new MRI scanner is very expensive, cost about $2 million, with a commensurate cost for building and preparing the space it needed. Therefore, hospital managers are under great pressure to keep high utilization ratio of such facilities, which makes patients to wait long time for imaging examinations.

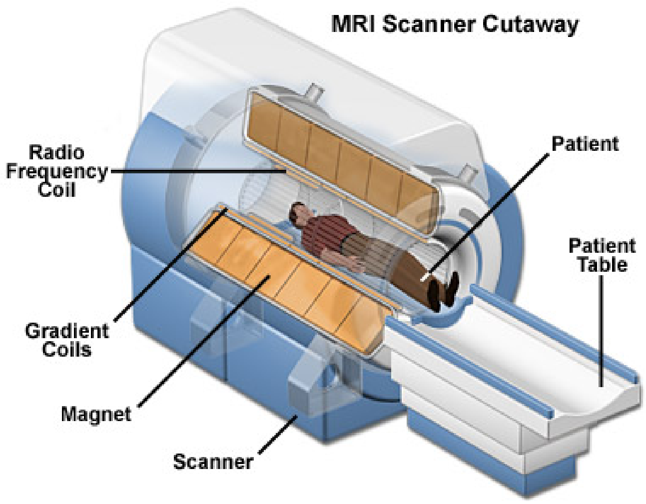


FIG. 1.2 The photo for MRI scanner

Photo from http://www.magnet.fsu.edu/education/tutorials/magnetacademy/mri/.

1.2.2 MRI examination reservation process

The pathway of stroke patients is shown in Fig. 1.3. Nearly all stroke patients arrive at the hospital through the emergency department. When a potential stroke patient arrives at the hospital, CT scan and/or Echocardiography is scheduled to identify the type of the stroke. After this examination, the patient is transferred to the neurovascular department of the hospital. The first checkup is performed by an intern. Then a senior physician examines the patient. Patient LoS is initially determined by neurologists. A certain number of medical examinations are needed and requested by neurologists during the visits.

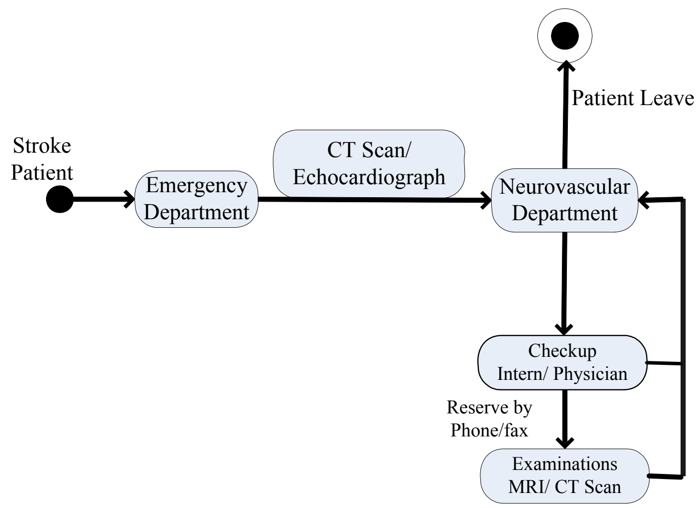


FIG. 1.3 Pathway of stroke patients

If the stroke is not so heavy, the patients may be allowed to leave the hospital before getting all examinations. In this case, the LoS of patients depends on the waiting time for imaging examinations. At the end of his/her stay, the patients go back home or is transferred to a special medical unit for rehabilitation. The transfer time also depends on the time of medical examinations.

In the related French hospital, the same-day examinations are requested by phone for emergency patients. For the patients with stable status, appointments are made by the secretaries by fax. If patients are available, then examinations are scheduled. The delay for examinations depends heavily on the experiences of the secretary and the personal relationship with the secretary of the imaging department. Often, appointments for patients that no longer need the examination for various reasons are not cancelled but kept by the neurovascular department for other patients.

This examination reservation process seems simple but time-consuming as it needs the informal collaboration of neurovascular department, medical imaging department and patients. In France, the actual delay for obtaining a time slot for examination is very long, more than 30 days for a regular MRI request. This negatively impacts the LoS but most importantly threatens the life of patients.

1.3 Problems and contributions

Imaging examination delays can be improved either from the service provider side (the imaging department), i.e. better schedule of examination demands, or from the client side, i.e. the neurovascular department side.

The management of a diagnostic facility from imaging department side consists of two interrelated tasks (Green et al. (2006)): establishing an appointment schedule for outpatients, and designing a system of dynamic priority rules for admitting patients into service in real time. Appointment scheduling refers to the determination of the duration, number, and timing of time slots for a particular day. Dynamic priority rules provide the real-time control of access to the facility by priorities of different patients.

The improvement of the operations of the imaging department involves the whole hospital and concerns too many medical services. For this reason, we start from the neurovascular department perspective and develop solutions.

We performed six-month field observation of the neurovascular department and collected data concerning patient arrival, medical examinations requested for each patient, delays of these examinations, and LoS of the patient. A detailed analysis of the historical data, as shown in Fig. 1.4, reveals that the neurovascular department has a large but rather stable weekly demand for medical examinations. The neurovascular department is actually the largest customer of the imaging department. Further, MRI examination of stroke patients takes nearly the same time, i.e. one time slot of about 30 minutes.

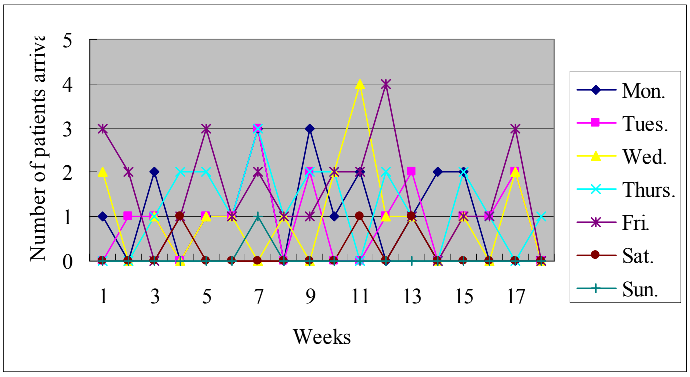


FIG. 1.4 Historical data collected from the neurovascular department

Based on the observation in the neurovascular department in the hospital under study, we propose a contract-based MRI examination reservation process. The imaging department reserves each week a certain number of appropriately distributed time slots (contracted time slots, CTS) for the neurovascular department. When needed, patients can still get extra MRI time slots through regular reservation (regular time slots, RTS) which takes much longer time with a delay of over 30 days.

The contract-based examination reservation process is characterized by the following decisions and control policies:

Contract decisions, i.e., the number of CTS and its distribution over time;

Patient assignment control policy, which assigns patients to either CTS or RTS. If the patient is assigned to CTS, then he/she will wait for CTS in the following days. Else, his/her examination will be reserved through regular MRI examination reservation process.

Advance CTS cancellation policy, which cancels the CTS in advance when there are no enough stroke patients to fill contracted time slots.

This thesis provides a mathematical analysis of the contract-based approach under the following assumptions:

Assumption 1: Only MRI examinations are considered and each patient requires one MRI time slot. Each patient can be assigned to either one CTS or one RTS.

Assumption 2: Emergency patients are not considered in this thesis. All patients have equal priority.

Assumption 3: Patient arrivals vary during a week but are stationary from one week to another. Further, the number of arrivals in one day is independent of the arrivals of other days.

Assumption 4: The same contract is used for different weeks, i.e. $n_t = n_{t+7}$ for all t where n_t denotes the number of CTS of day t. As a result, the contract can be represented by a 7-entry integer-valued vector $\mathbf{n}= \{n_1,\ldots, n7\}$.

Another major assumption is the focus on the neurovascular department which is the most important consumer of MRI examinations. The impact of the MRI examination reservations from this department on other departments sharing the MRI facilities is neglected.

The contributions are described as follows:

The first chapter introduces the stroke, MRI scanner, and regular MRI examination reservation process for stroke patients. We also define the problems solved in this thesis.

The second chapter reviews the state-of-the-art of methods and approaches used for the related problems.

The third chapter proposes a stochastic programming model to simultaneously determine the contract decision and patient assignment policy. In order to solve this model, patient assignment policy is first established via a Markov Decision Process (MDP) method. Then, Monte Carlo optimization and local search are used to determine the contract decisions. Computational experiments show that the proposed algorithms can efficiently solve the model. The new reservation process can greatly reduce the average waiting time of stroke patients at the expense of some unused CTS.

The fourth chapter exploits the possibility of one-day advance CTS cancellation to improve the contract-based MRI examination reservation process. An average-cost MDP formulation is proposed to determine patient assignment and one-day advance CTS cancellation control policies at the same time. Local search is used to improve the contract decisions with the known policies. Computational results show that one-day advance CTS cancellation can greatly reduce the unused CTS ratio with a little increase in average delay.

In the fifth chapter, we extend the one-day advance cancellation to two-day advance cancellation. The patient assignment, one-day advance cancellation, and two-day advance cancellation control policies for average-cost MDP are established via discounted-cost MDP and advanced convexity concepts. Contract decisions are further improved by local optimization. Computational results show that the criterion values can be further reduced.

Chapter 6 concludes the thesis.

Chapter 2

Literature review

This chapter presents a literature review for existing methods and approaches related to the MRI examination reservation problem. There are several streams relevant for this thesis, including the management of diagnostic services, appointment scheduling, capacity allocation of hospital resources; and similar problems in other fields.

2.1 Managing diagnostic services

The management of diagnostic devices, such as computer tomography (CT) and MRI scanners, has received limited coverage. The two earliest contributions are Walter (1973) and Lev et al. (1976). Simulation studies were performed in Walter (1973) to investigate the effect of scheduling rules on patient waiting time for X-rays in a radiology department. Lev et al. (1976) pointed out that the design of the management systems and scheduling techniques were the emphasis for improving radiology services.

Vasanawala and Desser (2005) used queuing theory to predict the optimal number of schedule slots to reserve for urgent CT and ultrasonography. Emergency studies were modeled as a Poisson process; and slots were reserved such that the rate of rescheduling of routine studies to accommodate emergencies is below a certain level.

Effective allocation of expensive imaging diagnosis capacity among several classes of patients within a day was addressed in Green et al. (2006). Three classes of patients were considered: inpatients, outpatients, and emergency patients. They considered two interrelated problems: the outpatient appointment schedule and the dynamic priority rules for admitting patients into service. The problem was formulated as a finite-horizon dynamic program and properties of the optimal policies have been identified, in order to design the outpatient appointment schedule, and establish dynamic priority rules for admitting patients into services.

A simple approach for dividing the available diagnosis capacity between emergency and inpatients on the one hand and outpatients on the other was proposed in Patrick and Puterman (2007). The authors looked at the benefit of reserving space for carrying over a percentage of non-emergency inpatient demand to the next day. Patrick et al. (2008) addressed the admission of multi-priority patients on a waiting queue to a diagnostic resource. They used an MDP framework to model the dynamical scheduling problem of multi-priority patients to a diagnosis facility in a public health care setting and proposed an approximate dynamic programming approach to overcome the state space explosion problem.

A dynamic capacity allocation problem for several priority classes patients was considered in Erdelyi and Topaloglu (2009) with protection level policies. Protection levels were used to “protect” a part of the capacity from the lower priority jobs so as to make it available for the future higher priority jobs. A simulation-based optimization approach was proposed to find a good set of protection levels. It combined a perturbation analysis technique to evaluate the gradient with respect to the protection levels and a stochastic approximation approach to determine the optimal protection levels.

The allocation of two CT-scanners was considered in Kolish (2008) by providing medical services to three patient groups with different arrival patterns and cost-structures. The problem was formulated as an MDP with the aim of allocating the available resources dynamically to patients of the three groups such that the expected total reward was maximized. Sickinger and Kolisch (2008) pursued the previous work to determine the optimal number of outpatients to be scheduled and assign the outpatients to a variable-block/fixed-interval appointment schedule. An MDP approach was proposed in Schutz and Kolisch (2009) to decide whether to accept requests for MRI examinations from patients with different priorities such as inpatients and outpatients. Different examination types, cancellations, no-shows and over-booking, and same-day demand were considered.

2.2 Appointment scheduling

Appointment scheduling is the problem of assigning a specific time when the patient is scheduled to start receiving care (Gupta and Denton (2008)). The appointment scheduling in general was reviewed in Mondschein and Weintraub (2003), while Magerlein and Martin (1978) and Blake and Carter (1997) summarized articles on surgery scheduling, and Cayirli and Veral (2003) provided excellent reviews on outpatient appointment systems, and Gupta and Denton (2008) surveyed appointment scheduling in health care system. Here we focus on the two latest reviews and some latest articles which are not included in the review.

Cayirli and Veral (2003) classified the outpatient appointment scheduling literature as follows: 1) static vs. dynamic appointment scheduling; 2) performance measures; 3) system design; and 4) methodology.

The most common appointment system in health care is the static appointment scheduling, where all decisions must be made before the start of a clinic session. The dynamic case can adjust the appointment time based on the current state of the system, which is most applicable in situations where patients are already admitted to a hospital or clinic. Most of the literature focuses on the static case, because the outpatient schedule for a session must be finished before the session begins. The simplest case is when all scheduled patients arrive on time and a single doctor serves them with stochastic processing times. If more doctors and more services are considered, the problem becomes more difficult. The patient lateness, no-shows, walk-ins, and emergencies may make it more complex. A representative set of papers on static appointment scheduling includes Vanden Bosch and Dietz (2000), Denton and Gupta (2003), and Robinson and Chen (2003) et al.

There is numerous performance criteria used to evaluate a given schedule. They are classified as time, congestion, and fairness. Time based measures include patients' waiting time, physician idle time, and staff overtime. Congestion based measures mainly refer to the mean number of patients in the queue. Fairness based measures focus on the even distribution of patients' waiting time over the day. Mondschein and Weintraub (2003) provide a detailed review of performance measures.

The design of the appointment system can be decomposed into three decisions: a) the appointment rule, b) the use of patient classification, and c) the adjustments made to reduce the disruptive impact of walk-ins, no-shows, and/or emergency patients. The appointment rule is typically specified by three parameters, the "block-size", i.e., the number of patients scheduled to this block, the "initial-block", i.e., the number of patients given the same appointment time at the start of a session, and "appointment interval", i.e., the interval between two appointments. Any combination of the above parameters is a possible appointment rule. Patient classification in outpatient scheduling can be used for two purposes: to sequence patients at the time of booking; and/or make the adjustment for the intervals based on the different service time of different patient class. The design of an appointment system must consider the possible adjustments for no-shows, walk-ins, urgent patients, and /or emergencies. This problem is addressed in Ho and Lau (1992), Cayirli (2006), Harper and Gamlin (2003), Wijewickrama and Takakuwa (2005) and Wijewickrama (2006).

Methods can be classified as analytical and simulation-based. Analytical methods include queuing theory and mathematical programming methods. The simulation studies include Babes and Sarma (1991),Cayirli (2006), Harper and Gamlin (2003), and Rohleder and Klassen (2002), etc., while the analytical include Robinson and Chen (2003), Vanden Bosch et al. (1999), and Vanden Bosch and Dietz (2000), etc.

Gupta and Denton (2008) described the appointment scheduling in three different environment, including primary care appointment scheduling, specialty clinic appointment scheduling, and scheduling elective surgery appointments. Complicating factors were discussed, including arrival and service time variability, patient and care provider preferences, and available information technology, etc. The articles were characterized into three themes according to the complicating factors. The readers are referred to Gupta and Denton (2008) for details.

Cayirli et al. (2008) investigated two approaches to patient classification by simulation. The first one used patient classification only for sequencing patient appointments at the time of booking. The second used patient classification for both sequencing and appointment interval adjustment, in which appointment intervals were adjusted to match the consultation time characteristics of different patient classes. Simulation results showed that new appointment systems that used interval adjustment for patient class were efficient in improving doctors' idle time, doctors' overtime and patients' waiting times without any trade-offs. Practical guidelines have been developed for managers responsible for designing appointment systems.

To deal with the problem of last-minute cancellation or "no-shows", Green and Savin (2008) proposed a new conception of appointment system, which was similar to a single-server queuing system. In this system, customers to enter service had a state-dependent probability of not being served and might rejoin the queue. Experimental results showed that the queuing models could provide efficient guidance in identifying patient panel sizes for medical practices that were trying to implement a policy of "advanced access".

Muthuraman and Lawley (2008) developed an appointment scheduling policy for outpatient clinics with overbooking used to compensate for the possible patient no-show. The objectives were to minimize patient wait times, maximize resource utilization, and minimize the number of patients waiting at the end of the day. Patients should be served during over time if there were patients waiting at the end of day. Conditions under which the objective evolution is unimodal have been derived and the behavior of the scheduling policy has been investigated under a variety of conditions.

A Markov decision process model was proposed in Gupta and Wang (2008) to solve the capacity management problem of a the clinic, i.e. the determination of which appointment requests to accept in order to maximize revenue. In this model, the patients' choice behavior was modeled explicitly. When the clinic is served by a single physician, the optimal policy has been identified as a threshold-type policy as long as the choice probabilities satisfy a weak condition. For a multiple-doctor clinic, the structure of the optimal policy has been partially characterized. Several heuristics and an upper bound were proposed. Numerical experiments demonstrated that the two heuristics based on the partial characterization of the optimal policy were quite accurate.

A semiclosed migration network was used in Lee and Zenios (2009) to capture patient flow into a clinic and between the clinic and hospital. Temporary patient absences were considered. A simple class of stationary control policies for patient admissions was proposed and algorithms were provided for selecting the one that maximizes long-run average earnings.

2.3 Planning and allocation of hospital resources

2.3.1 Planning and allocation of operating rooms

Decisions pertaining to the planning and allocation of operating rooms are among the most critical day-to-day problems faced by hospitals. These decisions have influence not only on the quality of care of patients but also on the crucial relationship between the hospital and physicians who work there.

Cardoen et al. (2010) provided a comprehensive review of recent operational research literature on operating room planning and scheduling. The related articles have been classified into 7 classes based on the criteria, including patient characteristics, performance measures, decision levels, type of analysis, solution techniques, uncertainty, and applicability of research. Some interesting articles are reviewed according to the above classifications.

1) In the literature of operating room planning and scheduling, two major patient classes are considered: elective and non-elective patients. The former refers to patients whose surgery can be planned in advance, and the latter usually refers to emergency patients, needing surgery as soon as possible, and urgent patients, needing surgery within a short period. Elective patients can be divided into inpatients and outpatients.

The surgery operation scheduling problem with elective patients considered was addressed in Perdomo et al. (2006) and Augusto et al. (2008), with two types of resources considered, including operating rooms and recovery beds. The problem was formulated as the assignment of patients to operating rooms and recovery beds with the objective of minimizing the sum over all patients of one defined function of their completion times. A Lagrangian relaxation approach was proposed to determine a near optimal schedule and a tight lower bound.

The allocation of medical service capacity between distinct demand streams was analyzed in Gerchak et al. (1996) in the setting of an operating room where the capacity was shared between elective and emergency surgeries. This reference focused on the reservation-planning policy for elective patients by determining at the start of each day how many additional elective surgeries to assign for that day. A stochastic dynamic programming model was proposed for this problem. The nature of the optimal policy was analyzed and characterized, which was not necessarily of a control-limit type.

2) There are many performance measures used to evaluate the performance of the planning and scheduling methods. The waiting time of patients or surgeons is one common evaluation measures. The other criteria include throughput, utilization, leveling, makespan, patient deferral/refusal, finance, and preferences, etc.

Denton et al. (2007) examined how case sequencing influenced patient waiting time, operating room idling and overtime. A stochastic optimization model was formulated and some practical heuristics were proposed for computing operating room schedules that hedged against the uncertainty in surgery durations. Sequencing surgeries and scheduling start times were also considered by a simple sequencing rule based on surgery duration variance. The rule was used to generate substantial reductions in total surgeon and operating room team waiting, operating room idling, and overtime costs.

3) There are many solution techniques, including mathematical programming, simulation, constructive heuristics, improvement heuristics, dedicated branch-and-bound, and analytical procedures.

Mathematical methods are well applied in planning and scheduling of operating rooms.

Guinet and Chaabane (2003) addressed the operating theatre planning over a medium term horizon (one or two weeks). The operating theatre under consideration was composed of several operating rooms and one recovery room where several beds were available. An extension of the Hungarian method has been developed to calculate the operating theatre planning.

Belien and Demeulemeester (2007) proposed and evaluated a number of models for building surgery schedules with leveled bed occupancy. A number of mixed integer programming based heuristics and a metaheuristic have been developed to minimize the expected total bed shortage.

A stochastic programming model for operating room planning with two types of surgery demands: elective and emergency was proposed in Lamiri et al. (2008a). A Monte Carlo optimization method was used to solve this model. Lamiri et al. (2008b) addressed the problem of scheduling patients in a hospital operating theatre, where three types of resources were considered: porters, operating rooms and recovery beds. The problem was formulated as the assignment of patients to the different resources in order to minimize a criterion function of patients' completion times. Column generation was used as a decomposition approach to solve the scheduling problem. Numerical results have illustrated that column generation is a promising decomposition approach for the scheduling problem.

A methodology was developed in Zhang et al. (2009) for allocating operating room capacity to specialties. A finite-horizon mixed integer programming (MIP) model was built to determine a weekly operating room allocation template that minimizes inpatients' cost measured as their length of stay.

Wang and Xu (2008) developed a fuzzy multi-objective programming model to optimize the operating room scheduling. A multi-objective combinatorial optimization problem was addressed in Cardoen et al. (2009), which determined the sequence of patients within the operating rooms of a freestanding ambulatory surgical center. Mixed integer linear programming solution approaches have been developed.

Discrete-event simulation is an effective tool for planning and allocation of resources to improve patient flow, while minimizing health care delivery costs and increasing patient satisfaction.

Intensive care unit (ICU) is a limited and critical resource. Ridge et al. (1998) developed a simulation model for bed capacity planning in intensive care. By this model, they have found non-linear relationship exists between numbers of beds, average occupancy level and the numbers of patients that have to be transferred due to the lack of beds. The compromise between bed occupancy and the number of transfers was also considered.

Kim et al. (1999) utilized queuing and simulation models to analyze the admission-and-discharge processes of one particular ICU. The beds of an ICU are scarce resources. The stochastic demands and random service times make it difficult to manage that resource. The admission of elective-surgery patients can be delayed. In order to minimize the number of cancelled surgeries, Kim et al. (2000) proposed a simulation model to evaluate the bed-reservation schemes by reserving the ICU beds for the exclusive use of the elective-surgery patients.

Shmueli et al. (2003) presented a model for optimizing admissions to an ICU in order to maximize the expected incremental number of lives saved from operating the ICU. Queuing theory was used to model the probability distribution of the number of occupied ICU beds. Three different admissions policies have been considered: first come first served, first come first served for all referrals whose expected incremental survival benefits gained from ICU admission exceed some threshold, and first come first served for all referrals whose expected incremental survival benefits exceed a bed specific threshold that depends upon the number of occupied beds. Experimental results showed that the last two methods could save more lives.

Persson and Persson (2009) analyzed the operating room planning at a department of orthopedic surgery in Sweden. A discrete-event model was used to solve the problem of handling uncertainty in patient arrival and surgery duration and at the same time maximizing the utilization of operating room. The experiments have demonstrated that the operating room department can perform much better by applying a different policy in reserving operating room capacity for emergency cases together with a policy to increase staff in stand-by.

4) In terms of decision levels, there are different combinations between type, i.e., date, time, room, capacity, and level, i.e., discipline, surgeon, and patient level. For example, Blake et al. (2002) and Blake and donald (2002) used integer programming model for a decision concerning date and room, i.e., the determination of each specialty what operating room types were assigned to what days of the week. At the surgeon level, Belien et al. (2006) introduced a software tool in order to decide when and where the surgeries had to be performed.

5) The types of analysis mainly include optimization, scenario analysis, and complexity analysis. The combinatorial optimization methods are either exact, for instance, Belien and Demeulemeester (2008) and Calichman (2005), or heuristic, for example, Blake and donald (2002) and Marcon et al. (2003). Scenario analysis refers to those focusing on the impact of different operating room setting. For example, Niu et al. (2007) used a simulation model with different scenarios of resource capacities. The computational complexity of the combinatorial problem or the corresponding solution approach was also analyzed. For example, Lamiri et al. (2008a) used the 3-partition problem in order to prove that their problem is strongly NP-hard and very difficult to solve.

6) With respect to uncertainty, there are two classes: deterministic and stochastic. Deterministic planning and schedule approaches ignore the uncertainties inherent to surgical services, for example, Adan and Vissers (2002) and Arenas et al. (2002); whereas uncertain approaches try to include such uncertainties. Two types of uncertainty are mainly considered, i.e., arrival uncertainty and duration uncertainty. The former refers to the uncertain arrival of emergency patients and the lateness of surgeons at the start of the surgery session, and the latter represents deviations between the actual and the planned durations of activities associated with the surgical process. For example, Harper (2002) included the patient arrival uncertainty in a detailed hospital capacity simulation model and Persson and Persson (2007) considered both patient arrival uncertainty and surgery duration variability to study how resource allocation policies affected the waiting time and utilization of emergency resources.

2.3.2 Planning and allocation of other hospital resources

Mathematical method is one of the most popular methods in this field.

Hsu et al. (2003) presented a deterministic approach to schedule patients in an ambulatory surgical center with the objective of minimizing the number of post anesthesia care unit nurses at the center. The patient scheduling problem was formulated as new variants of the no-wait, two-stage process shop scheduling problem. A tabu search-based heuristic algorithm has been proposed to solve the patient scheduling problem.

With respect to the problem of nurse scheduling, Cheng et al. (2008) modeled the daily nursing care scheduling problems and proposed an efficient scheduling method based on simulated annealing algorithm.

Olivares et al. (2008) applied the newsvendor model to a hospital that tried to balance the costs of reserving too much or too little operating room capacity to cardiac surgery cases. Results have shown that the hospital placed more emphasis on the tangible costs of having idle capacity than on the costs of schedule overrun and long working hours for the staff.

The efficient radiotherapy patient scheduling, within oncology departments, was addressed in Conforti et al. (2008). Novel optimization models have been proposed with the objective of minimizing patients' waiting time and maximizing the equipment utilization. Experimental results have shown that the proposed methods perform better than the human experts (i.e., the number of patients that begin the radiotherapy treatment is maximized).

Conforti et al. (2010) proposed an approach based on a well tailored integer linear optimization program, modeling a non-block scheduling strategy in order to minimize the mean waiting time or maximize the number of new scheduled patients.
An integer programming formulation was proposed in Gunes and Yaman (2010) for the hospital re-planning problem which arose after hospital network mergers. The model has found the best re-allocation of resources among hospitals, the assignment of patients to hospitals and the service portfolio to minimize the system costs subject to quality and capacity constraints.

An MDP method was proposed in Thompson et al. (2009) to allocate and reallocate patients to different floors of a hospital during demand surges. Decisions such as patient assignment and reactive or proactive patient transfers have been considered.
Min and Yih (2010) addressed a scheduling problem where patients with different priorities were scheduled for elective surgery in a surgical facility with a limited capacity. A stochastic dynamic programming model was formulated to schedule the patients. A structural analysis of the proposed model has been conducted to understand the properties of an optimal schedule policy. Based on the structural analysis, bounds on feasible actions have been incorporated into a value iteration algorithm, and a brief computation experiment has shown the improvement in computational efficiency.

Except for mathematical method, there are some other methods used in this field.
Ho and Lau (1999) used simulation to evaluate the impact on appointment schedules of the environmental factors, including probability of no-show, the coefficient of variation of service times, and the number of customers per service session. Jun et al. (1999) surveyed the application of discrete-event simulation modeling to health care systems.

Vissers (1994) developed an approach to allocate resources to specialties according to demand. The balanced utilization of the resources was considered. The proposed approach has utilized a set of models to support hospital managerial decision making on resource allocation issues. The allocation of inpatient resources was considered in Vissers (1998) within a hospital setting in the form of case study. An allocation procedure has been described that takes patient flows as its starting point and enables an evaluation of combined influences on the different resources concerned.
Bharadwaj et al. (1999) proposed a knowledge-based approach for solving the scheduling problem in a large cardiac center. System architecture has been derived that integrates

principles of opportunistic planning and reason maintenance. The former allows for incremental schedule construction, while the latter makes sure that the system records the reasons for scheduling decisions and revises the schedule whenever conflicts or new opportunities for schedule improvement arise.

2.4 Related problems in other fields

Allocating service capacity among competing customer classes has been studied in diverse applications including airlines seat management (Belobaba (1989), Barut and Sridharan (2004), Haerian et al. (2006)), hotels management (Liberman and Yechiali (1977), Bitran and Gilbert (1996)), car rental (Carroll and Grimes (1995), Geraghty and Johnson (1997)), telecommunications(Ross and Tsang (1989), Altman et al. (2001), Ormeci and van der Wal (2006)), and call center management (Perros and Elsayed (1996),Gans et al. (2003)). The capacity allocation in the last two research streams is often modeled as a dynamic priority queuing control problem. The service in some of the business environment, for example, hotel management, call center, etc., cannot be delayed. However, the service for patients can be delayed with some penalties.

Capacity protection is one of the most popular methods to deal with the capacity allocation problem with multiple priorities of demands. A dynamic capacity allocation problem was considered in Erdelyi and Topaloglu (2009), where jobs of different priorities arrived randomly over time and a decision was required on which jobs should be scheduled on which days. The authors have identified a class of policies defined by a set of protection levels, which protects a portion of the capacity from the lower priority jobs so as to make it available for the future higher priority jobs. A stochastic approximation method has been developed to find a good set of protection levels. Shumsky and Zhang (2009) examined a multi-period capacity allocation model with upgrading. In this reference, multiple product types and multiple classes of demand were considered. The optimal allocation policy is a simple two-step algorithm: First, use any available capacity to satisfy same-class demand, and then upgrade customers until capacity reaches a protection limit, so that in the second step the higher-level capacity is rationed.

2.5 Conclusion

This thesis differs from the previous studies on diagnostic facility management by investigating the problem from a totally different perspective and explores solutions from the client side, i.e. from the neurovascular department side. The use of contract gives a long term view of diagnostic capacity available and the neurovascular department can better manage the priority of the stroke patients and reduce waiting time for MRI examination. From the perspective of the imaging department, although the use of contract potentially leads to unused time slots, it also gives the imaging department stable and known demands which can be used to improve staff scheduling and MRI facility scheduling.

From a methodological point of view, our approach seems related to capacity allocation such as staffing in call center management (Gans et al. (2003)). Capacity allocation in this context also has to take into account random demands. The major difference with our problem is the acceptable waiting time. In the call center case, the acceptable waiting time is fairly short and, as a result, customers overflowing from one capacity planning time slot to another one can be neglected. In our case, the waiting of several days for MRI examinations is common and the planning of contract decisions has to take into account patients untreated overflowing from one day to the next. This makes the contract decisions closely linked to CTS waiting queue control.

Note that this thesis is an extension of our preliminary work, Augusto and Xie (2009), which analyzed the contract design problem by discrete event simulation and experimental design. This thesis provides an in-depth mathematical analysis of the optimal control and proposes efficient contract optimization approaches.

Chapter 3

Contract planning and patient assignment policy

This chapter proposes a stochastic programming model in order to determine the contract decisions and patient assignment policy at the same time. To solve this problem, structure properties of the optimal control policy for a given contract are proved by using an average cost MDP approach. The contract optimization is solved by a two-step approach. A Monte Carlo approximation approach is proposed to determine an initial contract decision which is then improved by local search. Extensive numerical experiments are performed to show the efficiency of the proposed approach and to investigate the impact of different parameters on the contract decisions and the control policy. To avoid the "unlucky" patients assigned to RTS, this chapter proposes an improved method by replacing patient assignment policy with RTS reservation policy. Numerical experiments show that the distribution of patient waiting times can be improved.

Papers relevant with this chapter: Geng et al. (2009a), Geng et al. (2009b), Geng et al. (2010a)

3.1 Introduction

Based on the observations in the hospital under study, we propose a new contract-based reservation process: the neurovascular department reserves each day some CTS to ensure the quick examination for stroke patients. RTS is still possible in case of arrival surges of stroke patients. The efficiency of the new reservation process greatly depends on two closely related decisions:

1) The contract planning decisions, i.e. the number of CTS and its distribution over time, which is related to capacity planning and allocation. The contract is characterized by an integer vector of 7-entries, each corresponding to the number of CTS for a weekday or weekend and the same contract is applied for different weeks;
2) The patient assignment policy for assigning incoming patients to either CTS or RTS, which is a decision to be made at the real time level, i.e. each day.

This problem is difficult as it involves simultaneously two decisions at different levels, the contract at the tactical level with decisions in discrete integer space and the optimal control policy at the real-time level. A stochastic programming model is proposed to solve this problem. In order to solve this model, the optimal control policy is firstly identified for any given contract. Then Monte Carlo approximation and local search are used to determine the contract decisions. Finally, an improved implementation method is proposed in order to avoid the “unlucky” patients who are assigned to RTS with long waiting times.

The rest of this chapter is organized as follows. The problem formulation is described in Section 3.2. Section 3.3 proposes an average cost MDP model for exploring the structure properties of the optimal control policy for any given contract. Contract optimization is addressed in Section 3.4. Section 3.5 presents computational results to show the efficiency of the proposed approach and the impact of different problem parameters on the contract decisions and the control policy. Implementation issues are discussed in Section 3.6. Conclusions and perspectives are given in Section 3.7.

3.2 Problem formulation

A contract-based MRI examination reservation process is proposed to reduce the stroke patient waiting times. This approach is characterized with the contract decision, i.e., the number and the distribution of CTS, and patient assignment control policy which assigns patient to either CTS or RTS. The contract decisions and the control policy are related to each other. The optimal control policy depends on the contract decisions, while the control policy has an impact in the determination of CTS. An integrated decision model is proposed to determine these decisions based on the assumption 1- assumption 4.

The MRI examination reservation problem is defined by the following notation:

Indices:

t: the index of days, $t=1,\ldots,T$;

i: the index of days in one week, $i=1,\ldots,7$, denoting Monday, …, Sunday; note that the day $i\pm j$ is the day in one week of j days after or before day i. For example, 1+1 is 2, i.e., Tuesday, and 1-1 is 7, i.e., Sunday.

$d(t)$: the day in the week corresponding to day t with $d(t)\in\{1, \ldots,7\}$;

Data:

T^R: average number of days for a patient to have his/her MRI examination through regular reservation process with $T^R > 1$;

c: penalty factor of an unused CTS. It serves as a weighting factor in order to balance the waiting time and unused MRI time slots;

a_t: number of patients arrived in day t. By assumption 3, daily arrivals a_t for $t \in IN$ are mutually independent random variables and weekly arrivals (a_{7j+1}, a_{7j+2}, …, a_{7j+7}) are identically distributed for all j = 0, 1, …. As a result, the arrival process is characterized by probability matrix $\boldsymbol{P} = [P_{ij}]$ for i = 1, …, 7 and for all $j \geq 0$ with P_{ij} denoting the probability of j arrivals in day i;

Decision variables:

n_t: number of CTS of day t;

x_t: number of patients waiting for CTS at the end of day t, which is also called CTS queue, x_0 is a given constant. Note that x_t does not include patients that are directed to RTS.

$y_t = f_t(x_{t-1}+a_t)$: IN→IN: number of patients directed to RTS at the end of day t, who will have an average delay of T^R days for the MRI examination. $f_t(x_{t-1}+a_t)$ is the unknown function of the total number of patients after new arrival.

The sequence of events during each day t is as follows. First, the CTS queue length x_{t-1} at the beginning of the day is known. The number a_t of new incoming patients during the day becomes known. MIN$\{n_t, x_{t-1} + a_t\}$ patients are served by the n_t CTS of the day and MAX$\{0, n_t - x_{t-1} - a_t\}$ CTS cannot be filled. $y_t = f_t(x_{t-1} + a_t)$ patients are directed to RTS and will have the MRI examination after an average delay of T^R days. The number x_t =MAX$\{0, x_{t-1} + a_t - n_t - y_t\}$of remaining patients will wait for CTS in the subsequent days.

Remark 3-1: The waiting time of the stroke patients served by RTS is approximated by a contract-independent constant T^R. The use of a constant T^R is reasonable because (i) MRI facility is shared by all medical units of the hospital and (ii) the sensitivity analysis performed in Section 3.5 shows that the optimal contract is quite insensitive to the change of T^R.

Remark 3-2: No constraint on the number of CTS is made in this thesis. The contract determined this way best reflects the demands of the neurovascular department. Nevertheless, all results of this paper still hold under constraints about the maximum number of CTS for each week and each day.

Under Assumption 4, the control policy will be proved stationary over weeks, i.e. $f_{t+7}(.) = f_t(.)$, in the next section.

Under the above assumptions, the MRI examination reservation problem can be formally stated by the following stochastic programming model:

Model 3-1:

$$\underset{\mathbf{n},f}{MIN} \quad \mathrm{E}\left[\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left(T^R y_t + x_t + c\left(n_t - x_{t-1} - a_t\right)^+\right)\right] \tag{3.1}$$

Subject to:

$$y_t = f_t\left(x_{t-1} + a_t\right) \le x_{t-1} + a_t \tag{3.2}$$

$$x_t = \left(x_{t-1} + a_t - y_t - n_t\right)^+ \tag{3.3}$$

$$\left(n_1, n_2, ..., n_7\right) \in IN^7, f_t : IN \to IN. \tag{3.4}$$

where $\left(\xi\right)^+ = \begin{cases} \xi & \text{if } \xi \ge 0 \\ 0 & \text{else} \end{cases}$.

In this formulation, the criteria contain three terms: the first two terms are respectively the average delays of patients using RTS and CTS, and the last term corresponds to the average penalty cost of unused CTS. Constraint (3.2) defines the control policy for use of RTS. Constraint (3.3) updates the number of patients in CTS queue. This model is quite difficult to solve because it contains some unknown $f_t(x_{t-1} + a_t)$. To identify this function, the optimal control policy $\boldsymbol{f}$ must be established. In the following, we first investigate the structure properties of the optimal control policy and then proposed an optimization method for determining the contract decisions.

3.3 Structure properties of the optimal control policy

This section considers the optimal patient assignment policy for the average cost MDP under any given contract **n**. The structure properties of the average cost MDP are established via discounted cost MDP. Let $z_t = x_{t-1} + a_t$ denote the state variable, which is the CTS queue length after patient arrivals. History-dependent policies are considered in this chapter. Let $h_t = (z_i, x_i, \ldots, z_{t-1}, x_{t-1}, z_t)$ be the full history by stating from initial state z_i at the beginning of day i. For ease of notation, we equivalently choose the CTS queue length at the end of each day as the control variable. The patient assignment policy is denoted as $\pi = \{\pi_1, \pi_2, ...\}$ where the CTS queue length at the end of day t is $x_t = \pi_t(h_t)$ with $0 \le x_t \le \left(z_t - n_t\right)^+$. This definition of control policy is equivalent to that of relation (3.2) as a result of relations (3.2)-(3.3).

The objective is to minimize over all history-dependent policies $\pi=\{\pi_1,\pi_2,...\}$ the following long-run average cost

$$J_\pi(i,z)=\lim_{T\to\infty}\frac{1}{T}E\left\{\sum_{t=i}^{T+i}g_{d(t)}(z_t,x_t)\middle|z_i=z\right\} \quad (3.5)$$

for any given initial state $z_i = z$ with $i = 1, \ldots 7$ where $g_{d(t)}(z_t,x_t)$ is the cost incurred in day t with

$$g_{d(t)}(z_t,x_t)=c\left(n_{d(t)}-z_t\right)^+ + x_t + T^R\left[\left(z_t-n_{d(t)}\right)^+ - x_t\right] \quad (3.6)$$

In the following, when no confusion is possible, $g_{d(t)}(.)$ and $n_{d(t)}$ are written as $g_t(.)$ and n_t for convenience.

Theorem 3-1: There exists an optimal average cost policy such that $x_t \le \bar{x}$ for all $t > 0$ with $\bar{x}=\left\lceil\left(T^R+c\right)n^*\right\rceil$ where $n^* = \text{MAX}\ \{n_1, \ldots, n_7\}$ and $\lceil X \rceil$ is the least integer greater or equal to X.

Proof: Without loss of generality, assume that the system starts from state z in day $i = 1$. Let (x_t, $\forall t$=1, 2,…) be an optimal control policy. Construct another history-dependent policy $\left(x_t^{'}\right)$ such that $x_t^{'}=\min\left(\left(x_{t-1}^{'}+a_t-n_t\right)^+,x_t,\bar{x}\right)$. By definition, $x_t^{'}\le x_t$. Hence, the control policy $\left(x_t^{'}\right)$ is feasible.

Note that the criterion value does not depend on patient schedule as long as the number of patients remaining in the CTS queue is equal to x_t. The following history dependent patient scheduling policy is used for (x_t). Patients are either marked or unmarked. When ($x_t - x'_t$) increases by Δ which is clearly smaller than a_t, Δ patients arrived in day t are marked and added to marked patient queue. When ($x_t - x'_t$) decreases by Δ, Δ marked patients are assigned to RTS and will have their MRI examination in T^R days if $x'_t > 0$. Otherwise, δ marked patients will be assigned CTS of day t and $\Delta-\delta$ marked patients assigned RTS with $\delta=\min\left(x_{t-1}+a_t,n_t\right)-\min\left(x_{t-1}^{'}+a_t,n_t\right)$. By definition, the CTS queue length of unmarked patients is exactly x'_t. Since $x_t\le\left(x_{t-1}+a_t-n_t\right)^+$ and $x_t^{'}\le x_t$, marked patients arrive in day t such that $x_t^{'}=\bar{x}$.

By optimality of (x_t), $\left(x_t^{'}\right)$ is also optimal and the Theorem holds if

$$\lim_{T\to\infty}\Phi_T\equiv\frac{1}{T}\sum_{t=1}^{T}\left(g_{d(t)}(z_t,x_t)-g_{d(t)}(z'_t,x'_t)\right)\ge 0 \quad (3.7)$$

By construction,

$$\Phi_T = \frac{1}{T}\sum_{j=1}^{Q(T)}\left(G_j(T) - T^R\right) \tag{3.8}$$

where $Q(T)$ is the total number of marked patients by day T and $G_j(T)$ is the cost incurred by j-th marked patient till day T including waiting time in the marked patient queue plus eventually a cost of T^R if is served by RTS or the cost of $-c$ if it is served by CTS.

From (3.8),

$$\Phi_T \geq \frac{1}{T}\left(\sum_{j=1}^{Q\left(T-T^R\right)}\left(G_j(T) - T^R\right) - \left(Q(T) - Q\left(T - T^R\right)\right)T^R\right)$$

By definition $Q(T) - Q(T\text{-}T^R)$ is upper bounded by the number of patients arrived in the last T^R days and is hence finite. As a result, $(Q(T) - Q(T\text{-}T^R))/T$ tends to 0 as T increases. As a result, relation (3.7) holds and the proof is completed if $G_j(T) - T^R \geq 0$ for all $j \leq Q(T\text{-}T^R)$.

Three cases are possible. Case 1: marked patient j is not yet served till day T and hence has been waiting for at least T^R days, which implies $G_j(T) - T^R \geq 0$. Case 2: marked patient j has been assigned to RTS and by definition, $G_j(T) - T^R \geq 0$. Case 3: marked patient j has been served by a CTS. Note that marked patient j arrives in a day t such that $x_t^{'} = \overline{x}$ and will be served by a CTS in day τ such that $x_\tau^{'} = 0$. As a result, marked patient j waits at least $\overline{x}/n^*$ days. Hence, $G_j(T) - T^R \geq \overline{x}/n^* - c - T^R \geq 0$.

Q.E.D

Due to Theorem 3-1, we can make the following assumption without loss of generality.

Assumption 3-A1: $x_t \leq \overline{x}$ for all $t > 0$.

3.3.1 Discounted cost problem

According to relation (3.5), the corresponding α-discounted cost MDP is defined as follows:

$$J_{\alpha,\pi}(i,z) = \lim_{T\to\infty} E\left[\sum_{t=i}^{T}\alpha^{t-i} g_{d(t)}\left(z_t, x_t\right) \middle| z_i = z\right] \tag{3.9}$$

for any given initial state $z_i = z$ with $i = 1, \ldots 7$ with discount factor α such that $0 < \alpha < 1$.

Consider the following optimal cost function

$$V_\alpha(i,z) = \underset{\pi}{MIN}\, J_{\alpha,\pi}(i,z)$$

In the remaining, for simplicity, the notation α is omitted in this subsection where only discounted cost problem with a given α is considered.

Theorem 6.10.4 in Puterman (1994) is used to establish the optimality equation. It will be shown in the following Remark 3-3 that all conditions needed for application of Theorem 6.10.4 are satisfied. Since the set of states (i, z) is countable and the control constraint set is finite as $x_t \le z_t$ for each z_t, Theorem 6.10.4 in Puterman (1994) implies that the optimal cost function is the unique solution of the following optimality equation:

$$V(i,z) = \min_{0\le x\le\left((z-n_i)^+\wedge\bar{x}\right)} \left\{ \begin{array}{l} c(n_i - z)^+ + x + T^R(z-n_i-x)^+ \\ +\alpha\sum_a P_{i+1,a} V(i+1, x+a) \end{array} \right\}, \forall i = 1,...,7 \tag{3.10}$$

where $x \wedge y = \min(x, y)$. The optimal control policy is given by the argument x that reaches the minimum in (3.10) and the optimal cost function is the limiting function of the following value iteration:

$$V^t(z_t) = \min_{0\le x_t\le\left((z_t-n_t)^+\wedge\bar{x}\right)} \left\{ c(n_t - z_t)^+ + x_t + T^R(z_t - n_t - x_t)^+ + \alpha\sum_a P_{t+1,a} V^{t+1}(x_t + a) \right\} \tag{3.11}$$

$$V^0(z) = 0 \tag{3.12}$$

for t = 0, −1, − 2, … where $n_t = n_{d(t)}$ and $P_{t+1,a} = P_{d(t+1),a}$ are shorthand notation with $d(t)$ denoting the corresponding weekday or weekend with $d(0) = 7$, $d(-1) = 6$, … As a result,

$$V(i,z) = \lim_{n\to\infty} V^{-7n+i}(z) \tag{3.13}$$

Relation (3.11) can be rewritten as

$$V^t(z_t) = c(n_t - z_t)^+ + T^R(z_t - n_t)^+ + \min_{0\le x_t\le\left((z_t-n_t)^+\wedge\bar{x}\right)} \left\{ U^{t+1}(x_t) - (T^R - 1)x_t \right\} \tag{3.14}$$

where

$$U^{t+1}(x_t) = \alpha\sum_a P_{t+1,a} V^{t+1}(x_t + a) \tag{3.15}$$

Similarly, optimality equation (3.10) can be put in similar form.

$$V(i,z) = c(n_i - z)^+ + T^R(z - n_i)^+ + \min_{0\le x\le\left((z-n_i)^+\wedge\bar{x}\right)} \left\{ U(i+1, x) - (T^R - 1)x \right\}, \forall i = 1,...,7 \tag{3.16}$$

where

$$U(i+1, x) = \alpha\sum_a P_{i+1,a} V(i+1, x+a) \tag{3.17}$$

Remark 3-3: From Theorem 6.10.4 of Puterman (1994), the optimality equations (3.10) actually has a unique solution as Assumption 6.10.1 and condition (6.10.11) of Puterman (1994) hold with $g_t(z_t, x_t) \leq w(z_t) \equiv cn^* + T^R z_t$ and

$$\sum_a P_{i+1,a} w\left(x_t + a\right) \leq cn^* + T^R x_t + T^R E\left[a_{t+1}\right] \leq cn^* + T^R z_t + T^R a^* = w\left(z_t\right) + T^R a^*$$

where $n^* = \mathrm{MAX}\{n_1, \ldots, n_7\}$, $a^* = \mathrm{MAX}\{\mathrm{E}[a_1], \ldots, \mathrm{E}[a_7]\}$.

Property 3-1: In the value iteration by (3.11) or equivalently by (3.14), the optimal x_t is nondecreasing in z_t.

Proof: Denote x_t^1 and x_t^2 as the optimal x_t for $V^t\left(z_t+1\right)$ and $V^t\left(z_t\right)$. From relation (3.14),

$$0 \leq x_t^1 \leq \left((z_t+1-n_t)^+ \wedge \bar{x}\right) \leq \left((z_t-n_t)^+ + 1 \wedge \bar{x}\right)$$

$$0 \leq x_t^2 \leq \left((z_t-n_t)^+ \wedge \bar{x}\right)$$

Therefore, x_t^1 is equal to either x_t^2 or $(z_t-n_t)^+ + 1$ and hence $x_t^1 \geq x_t^2$. This completes the proof.

Property 3-2: In the value iteration by (3.11) or equivalently by (3.14), $-c \leq V^t(z_t+1) - V^t(z_t) \leq T^R$, for any z_t and t.

Proof: The proof is made by induction. First the property is clearly true for $t = 0$. Assume that the property holds for t+1 and we prove that it also holds for t. Let x_t^1 and x_t^2 as the optimal x_t for $V^t\left(z_t+1\right)$ and $V^t\left(z_t\right)$.

From Property 3-1, $x_t^1 \geq x_t^2$. As a result, by (3.14) for z_t+1 with $x_t = x_t^2$,

$$V^t\left(z_t+1\right) \leq c\left(n_t - z_t - 1\right)^+ + x_t^2 + T^R\left(\left(z_t+1-n_t\right)^+ - x_t^2\right) + U^{t+1}\left(x_t^2\right)$$

By definition,

$$V^t\left(z_t\right) = c\left(n_t - z_t\right)^+ + x_t^2 + T^R\left(\left(z_t-n_t\right)^+ - x_t^2\right) + U^{t+1}\left(x_t^2\right)$$

Combining the two relations,

$$V^t\left(z_t+1\right) - V^t\left(z_t\right) \leq c\left(n_t - z_t - 1\right)^+ - c\left(n_t - z_t\right)^+ + T^R\left(z_t+1-n_t\right)^+ - T^R\left(z_t-n_t\right)^+ \leq T^R$$

Three cases are considered for the proof of the first inequality of this Property.

Case: $z_t + 1 - n_t \leq 0$. In this case, $x_t^1 = x_t^2 = 0$. As a result,

$$V^t\left(z_t+1\right) = c\left(n_t - z_t - 1\right) + U^{t+1}\left(0\right)$$

$$V^t\left(z_t\right)=c\left(n_t-z_t\right)+U^{t+1}\left(0\right)$$

$$V^t\left(z_t+1\right)-V^t\left(z_t\right)=-c$$

Case: $z_t+1-n_t>0$ and $0<x_t^1<\left(z_t+1-n_t\wedge\overline{x}\right)$. From (3.14), there is no unused CTS for both $V^t\left(z_t+1\right)$ and $V^t\left(z_t\right)$ and $x_t^2=x_t^1$. Hence,

$$V^t\left(z_t+1\right)=x_t^1+T^R\left(\left(z_t+1-n_t\right)-x_t^1\right)+U^{t+1}\left(x_t^1\right)$$

$$V^t\left(z_t\right)=x_t^1+T^R\left(\left(z_t-n_t\right)-x_t^1\right)+U^{t+1}\left(x_t^1\right)$$

$$V^t\left(z_t+1\right)-V^t\left(z_t\right)=T^R\geq -c$$

Case: $z_t+1-n_t>0$ and $x_t^1=z_t+1-n_t$. From (3.14), there is no unused CTS for both $V^t\left(z_t+1\right)$ and $V^t\left(z_t\right)$. Hence,

$$V^t\left(z_t+1\right)=x_t^1+U^{t+1}\left(x_t^1\right)$$

Take x_t^1-1 as the feasible control policy for $V^t\left(z_t\right)$,

$$V^t\left(z_t\right)\leq x_t^1-1+U^{t+1}\left(x_t^1-1\right)$$

Combining the two relations with (3.15),

$$V^t\left(z_t+1\right)-V^t\left(z_t\right)\geq 1+\alpha\sum_a P_{t+1,a}\left(V^{t+1}\left(x_t^1+a\right)-V^{t+1}\left(x_t^1+a-1\right)\right)$$

By induction assumption, $V^{t+1}\left(x_t^1+a\right)-V^{t+1}\left(x_t^1+a-1\right)\geq -c$ and hence,

$$V^t\left(z_t+1\right)-V^t\left(z_t\right)\geq 1-\alpha c\geq -c\,.$$

Definition: A function $\phi(x):\mathbb{Z}\rightarrow\mathbb{R}$ is said convex if $\phi(x+1)-\phi(x)\geq\phi(x)-\phi(x-1)$, for all x.

Property 3-3: In the value iteration by (3.11) or equivalently by (3.14), $V^t\left(z_t\right)$ is convex in z_t. As a result, $U^{t+1}\left(x_t\right)$ is convex in x_t.

Proof: First, the property holds for t = 0 as $V^t\left(z_t\right)$= 0. Assume that $V^{t+1}\left(z_{t+1}\right)$ is convex and we prove the property for *t*. From (3.15), the convexity of $V^{t+1}\left(z_{t+1}\right)$ implies the convexity of $U^{t+1}\left(x_t\right)$. Hence, $U^{t+1}\left(x_t\right)-\left(T^R-1\right)x_t$ is also convex. Let

$$L_t = \underset{0\le x_t \le \left((z_t-n_t)^+ \wedge \bar{x}\right)}{\arg\min} \left(U^{t+1}(x_t) - \left(T^R-1\right)x_t\right)$$

Equation (3.14) can be written as

$$V^t(z_t) = \begin{cases} c(n_t - z_t) + U^{t+1}(0) & \text{if } z_t \le n_t \\ z_t - n_t + U^{t+1}(z_t - n_t) & \text{if } n_t \le z_t \le L_t + n_t \\ L_t + T^R(z_t - n_t - L_t) + U^{t+1}(L_t) & \text{if } z_t \ge L_t + n_t \end{cases} \tag{3.18}$$

where

$$L_t = \underset{x_t \ge 0}{\arg\min}\left(U^{t+1}(x_t) - \left(T^R-1\right)x_t\right)$$

By convexity of $U^{t+1}(x_t)$, $V^t(z_t)$ is convex in z_t in the following internal [0, n_t), (n_t, L_t+n_t), and (L_t+n_t, +∞). We still need to prove the convexity of $V^t(z_t)$ for $z_t = n_t$ and $z_t = L_t+n_t$.

The convexity of $V^t(z_t)$ at $z_t = n_t$ holds as, by Property 3-2,

$$V^t(z_t+1) - V^t(z_t) \ge -c = V^t(z_t) - V^t(z_t-1).$$

The convexity of $V^t(z_t)$ at $z_t = L_t+n_t$ holds as, by Property 3-2,

$$V^t(z_t) - V^t(z_t-1) \le T^R = V^t(z_t+1) - V^t(z_t).$$

By induction, this completes the proof.

Theorem 3-2: The value functions V(i, z) and U(i, x) are convex functions respectively in z and x for all i = 1, …7. Further in the optimal patient assignment control policy, 1) the optimal number of patients assigned to CTS queue is

$$x^* = \begin{cases} 0 & \text{if } z - n_i \le 0 \\ z - n_i & \text{if } 0 \le z - n_i \le L_i \\ L_i & \text{if } z - n_i \ge L_i \end{cases} \tag{3.19}$$

where $L_i = \underset{x_i \ge 0}{\arg\min}\left(U(i+1, x_i) - \left(T^R-1\right)x_i\right)$.

2) The optimal number of patients assigned to RTS is $y^* = \left((z-n_i)^+ - L_i\right)^+$.

Proof: The theorem is a direct consequence of relations (3.13), (3.16)-(3.17) and Property 3-3. Q.E.D

3.3.2 Average cost problem

In this subsection the optimality equation and the form of the optimal control policy will be established via the α-discounted problem.

Even through x_t is bounded from Assumption 3-A1, $z_t = x_{t-1} + a_t$ can be unbounded. For the sake of readability, we first consider the case of bounded z_t and then establish properties of optimal control for the unbounded case.

3.3.2.1 Bounded patient arrival

Assumption 3-A2: The number of patients arriving in any day is bounded from above by M, i.e. $a_t \leq M$ for some given positive integer M.

Assumptions 3-A1 and 3-A2 imply that z_t is bounded from above by $z_{\max} = M + T^R n^*$. Therefore, $g_t(z_t, x_t)$ is bounded from above and $0 \leq g_t(z_t, x_t) \leq cn_t + T^R(z_{\max} - n_t) \leq cn^* + T^R z_{\max}$.

Property 3-4: Under Assumption 4, Assumption 3-A1, and 3-A2, there exists $\Gamma > 0$ such that $\left|V_\alpha(i,z) - V_\alpha(7,0)\right| \leq \Gamma$, for all i = 1, …, 7 and for all z.

Proof**:** From Property 3-2,

$$-c \leq V_\alpha^t(z+1) - V_\alpha^t(z) \leq T^R$$

which, together with $0 \leq z \leq z_{max}$ implies, for all z and z',

$$-cz_{\max} \leq V_\alpha^t(z) - V_\alpha^t(z') \leq T^R z_{\max}$$

Combining with (3.13),

$$-Cz_{\max} \leq V_\alpha(i,z) - V_\alpha(i,z') \leq Cz_{\max} \tag{3.20}$$

Where $C = \max(T^R, c)$ This establish the property for i = 7. Consider now the case i = 1, …6. From the optimality equations (3.10) and let π be the optimal control policy,

$$V_\alpha(i,z) = g_i(z, \pi(z)) + \alpha \sum_a P_{i+1,a} V_\alpha(i+1, \pi(z)+a), \forall i = 1,...,7$$

where $g_i(z,x) = c(n_i - z)^+ + x + T^R(z - n_i - x)^+$. From the words following Assumption 3-A2,, $g_i(z,x) \leq B$ with $B = cn^* + T^R z_{\max}$. As a result,

$$V_\alpha(i,z) \leq B + \sum_a P_{i+1,a} V_\alpha(i+1, \pi(z)+a), \forall i = 1,...,7$$

Repeat the above relations for t subsequent days leads to:

$$V_\alpha(i,z) \le tB + \sum_{z'} Q^{\pi}_{(i,z),(t+i,z')} V_\alpha(t+i,z') \quad (3.21)$$

where $Q^{\pi}_{(i,z),(t+i,z')}$ is the probability of reaching state z' at the beginning of day $t+i$ by starting from state z at day i under policy π. Combining (3.20) and (3.21) with $t+i=7$,

$$V_\alpha(i,z) \le 6B + \sum_{z'} Q^{\pi}_{(i,z),(7,z')} V_\alpha(7,z') \le 6B + V_\alpha(7,0) + Cz_{\max}$$

Similarly,

$$V_\alpha(7,0) \le 6B + \sum_{z'} Q^{\pi}_{(7,0),(7+i,z')} V_\alpha(i,z') \le 6B + V_\alpha(i,z) + Cz_{\max}$$

The above two properties concludes the proof.

Theorem 3-3. Under Assumption 4, Assumptions 3-A1, and 3-A2, there exists an optimal stationary control policy, the same with that in Theorem 3-2, for the average cost model (3.5). Further the optimal average cost is independent of the initial state (i, z).

Proof. From Proposition 4.2.6 in Bertsekas (1996) and Property 3-4, the optimal average cost per day exists and has the same value λ for all initial states, and λ satisfies

$$\lambda = \lim_{\alpha \to 1} (1-\alpha) V_\alpha(i,z) \quad (3.22)$$

The differential cost functions

$$\psi(i,z) = \lim_{\alpha \to 1} \left(V_\alpha(i,z) - V_\alpha(0,0) \right) \quad (3.23)$$

satisfy the following optimality equations:

$$\lambda + \psi(i,z) = c(n_i - z)^+ + T^R(z-n_i)^+ + \min_{0 \le x \le ((z-n_i)^+ \wedge \bar{x})} \left\{ H(i+1,x) - (T^R - 1)x \right\}, \forall i = 1,...,7 \quad (3.24)$$

$$H(i,x) = \alpha \sum_a P_{i,a} \psi(i, x+a) \quad (3.25)$$

Further, the optimal control policy is defined by the argument x that reaches the minimum in (3.24)-(3.25). From Property 3-3, equations (3.13) and (3.23), $\psi(i, z)$ is convex in z and $H(i, x)$ is convex in x for all i = 1, …, 7. This implies that the optimal control policy for the average cost problem is of the form (3.19). Q.E.D.

3.3.2.2 Unbounded patient arrival

In this subsection, assumption 3-A2 is relaxed.

Theorem 3-4. Under Assumptions 4 and 3-A1, (a) there exists a constant λ satisfying (3.22) for all (i, z), a matrix ψ (i, z) satisfying (3.24)-(3.25), (b) the optimal control policy is defined by the argument x that reaches the minimum in (3.24)-(3.25), (c) there exists an optimal stationary control policy of the form of equation (3.19) for the average cost model (3.5).

Proof: The proof is based on Theorem 8.10.7 of Puterman (1994) and the conditions that need to be checked are the following ones:

C1: For each state (i, z), the stage cost is such that $-\infty < R \le g_i(z, x_i) < \infty$.

C2: For each (i, z) and $\alpha < 1$, $V_\alpha(i,z) < \infty$.

C3: There exists $\varphi > -\infty$ such that, for each (i, z), $\psi_\alpha(i,z) \equiv V_\alpha(i,z) - V_\alpha(7,0) \ge \varphi, \forall \alpha < 1$.

C4: There exists a non-negative function W(i, z) such that

a) $W(i, z) < \infty$;

b) for each (i, z), $\psi_\alpha(i,z) \le W(i,z), \forall \alpha < 1$; and

c) for each (i, z) and x_i,

$$\sum_a P_{i+1,a} W(i+1, x_i + a) < \infty.$$

According to Theorem 8.10.7 of Puterman (1994), as the control constraint set for each state (i, z) is finite as a result of Assumption 3-A1, (a) and (b) of the Theorem hold. Further $\psi(i, z)$ is the limit of a sequence $\psi_{\alpha_m}(i,z)$ such that α_m converges to 1 and $\psi_{\alpha_m}(i,z)$ converges for all (i, z). From Property 3-3, equations (3.13) and (3.23), $\psi(i, z)$ is convex in z and (c) of the Theorem can be proved as for Theorem 3-3.

Let us now prove conditions C1-C4.

Condition C1 clearly holds as $g_i(z, x_i) \ge 0$.

Condition C2 holds as well as

(i) by Assumption 3-A1, $x_{t-1} \le \bar{x}$ and $z_t \le x_{t-1} + a_t \le \bar{x} + a_t$;

(ii) $g_i(z_i, x_i) \le cn^* + T^R z$, $E\left[g_t(z_t, x_t)\right] \le E\left[cn^* + T^R z_t\right] \le cn^* + T^R(\bar{x} + a^*)$;

(iii) $V_\alpha(i,z) \le cn^* + T^R z + \dfrac{\alpha}{1-\alpha}\left(cn^* + T^R(\bar{x} + a^*)\right)$.

Condition C3 holds as

(i) by Property 3-2 and relation (3.13), $-c \le V_\alpha(i, z+1) - V_\alpha(i, z) \le T^R$;

(ii) by Assumption 3-A1 and Theorem 3-2, the control threshold $L_i \le \bar{x}$.

(iii) by relations (3.18) and (3.13), $V_\alpha(i, z)$ is increasing for $z \ge L_i + n_i$;

(iv) By (i)-(iii), $-C(T^R n^* + n^*) \le V_\alpha(i, z) - V_\alpha(i, z')$ with $C = max(c, T^R)$;

(v) $V_\alpha(7,0) \le icn^* + T^R \sum_{s=1}^{i-1} \sum_{\tau=1}^{s} E[a_\tau] + \sum_{z'} Q^\pi_{(7,0),(7+i,z')} V_\alpha(i, z')$ where $Q^\pi_{(i,z),(t+i,z')}$ is the probability of reaching state z' at the beginning of day t+i by starting from state z at day i under policy π.

(vi) Combining with (iv), $V_\alpha(7,0) \le icn^* + T^R ia^* + i^2 a^* + \left(V_\alpha(i, z) + C(T^R n^* + n^*)\right)$ which proves C3.

Condition C4 holds as

(i) By starting from state (i, z), the average stage cost of any period t is bounded from above by $cn^* + T^R\left(z + \sum_{\tau=1}^{t} E[a_{\tau+i}]\right) \le cn^* + T^R(z + ta^*)$.

(ii) From (i), $V_\alpha(i, z) \le \sum_{\tau=0}^{t-1}\left(cn^* + T^R(z + \tau a^*)\right) + \sum_{z'} Q^\pi_{(i,z),(i+t,z')} V_\alpha(i+t, z')$; .

(iii) Combing (ii) with i+t = 7 and Property 3-2,

$$V_\alpha(i, z) \le 6cn^* + 6T^R z + 36T^R a^* + V_\alpha(7,0) + T^R E[z_7];$$

(iv) Combining (iii) with $E[z_7] \le z + \sum_{\tau=i}^{7} E[a_\tau] \le z + 6a^*$ leads to $\psi_\alpha(i, z) \le W(i, z)$ with $W(i, z) \le 7T^R z + 6cn^* + 42T^R a^*$. Condition C4.a-C4.c clearly holds. This proves C4.

Q.E.D.

In the following, we will restrict us to threshold policies and denote each policy by its threshold vector **L**.

Assumption 3-A3: For each day i = 1, …7, the probability of no patient arrival is non null, i.e. $P_{i0} \ge \delta$, for some $\delta > 0$.

Remark 3-4: Assumption 3-A3 is not restrictive in this chapter. If number of patient arrivals of a day i is at least one, i.e. $a_i \ge 1$, there will be at least one CTS for day i in the optimal contract **n**, i.e. $n_i \geq 1$. Further, the contract **n**' with $n'_i = n_i - 1$ and $n'_j = n_j$, for all $j \neq i$ is the optimal contract for patient arrivals **a**' with $a'_i = a_i - 1$ and $a'_j = a_j$. The reverse is also true.

Property 3-5: Under Assumption 4, Assumptions 3-A1, and 3-A3, for any stationary control policy **L**, i.e. $L(t) = L(t+7)$, the underlying stochastic process $(d(t), z_t)$ and $(d(t), x_t)$ are Markov chains with a unique positive recurrent class including all states $(i, 0)$ for i = 1, …, 7.

Proof: Under the assumptions of the Property, it is clear that $(d(t), x_t)$ is a finite state Markov chain and $(d(t), z_t)$ is a Markov chain. This property obviously holds for the case **n** = 0 and we assume **n** ≠ 0, i.e. $n^* > 0$, in the following. Starting from any initial state (i, x), for any control policy **L**, any state $(d(t), x_t) = (i, 0)$ can be reached in at most $7T^R + 6$ days with probability δ^{7T^R+6} as $x \leq \overline{x}$. As a result, $(d(t), x_t)$ is a finite state Markov chain with a unique positive recurrent class including all states $(i, 0)$. Since $z_t = x_{t-1} + a_t$, Assumption 3-A1 and the property of a_t imply that $(d(t), z_t)$ is a Markov chain with a unique positive recurrent class including all states $(i, 0)$.

3.3.3 Computation and implementation of the optimal control policy

From Theorems 3-3 and 3-4, the average-cost MDP problem has the same optimal control policy with the discounted-cost MDP. Further, the optimal control policy for a given contract **n** can be determined by solving optimality equations (3.24)-(3.25). This can be either done either by value iteration or the linear program (Puterman (1994)). The linear program (LP) model is as follows:

Model 3-2:

$$F_{\boldsymbol{n}} = \text{Maximize } \lambda$$

subject to

$$\lambda + \psi(i,z) \leq c(n_i - z)^+ + x + T^R(z - n_i - x)^+ + \alpha \sum_a P_{i,a} \psi(i+1, x+a),$$

$$\forall z, \forall x \leq \left((z-n_i)^+ \wedge \overline{x}\right), \forall i = 1, \dots, 7$$

F_n is the optimal average cost and, for each state (i, z), the optimal control is given by x reaching equality in the above relations. Further, the optimal control policy is characterized by a control threshold vector ***L*** with one threshold for each day**.** From relation $x = \min\left(L_i, (z - n_i)^+\right)$, **L** can be easily determined.

The existence of optimal threshold control makes the implementation easy. According to the relation (3.19), the implementation of the **L** control policy can be divided into three cases, as shown in Fig. 3.1-3.3:

Case 1: If z_t, CTS queue after new patients' arrival, is smaller than n_t, the number of CTS in the same day, then there exists the number n_t–z_t of unused CTS and there are no patients waiting for the future examinations.

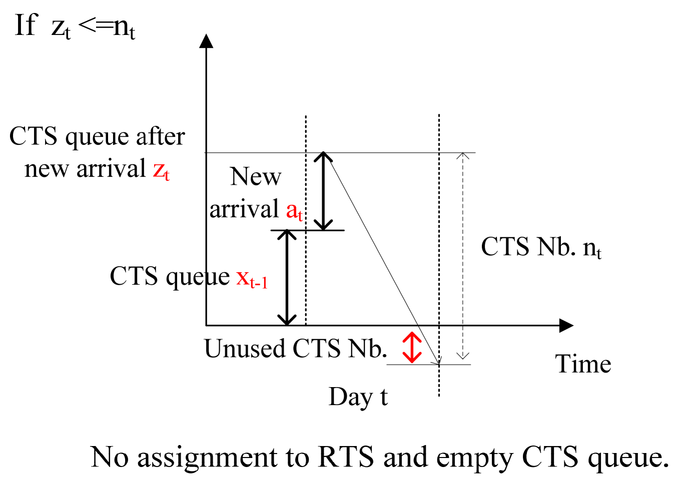


FIG. 3.1 The optimal control if $z_t <= n_t$

Case 2: If z_t is between the values of n_t and $n_t + L_t$, then all the number z_t–n_t of remaining patients are kept in the CTS queue and no patients are assigned to RTS.

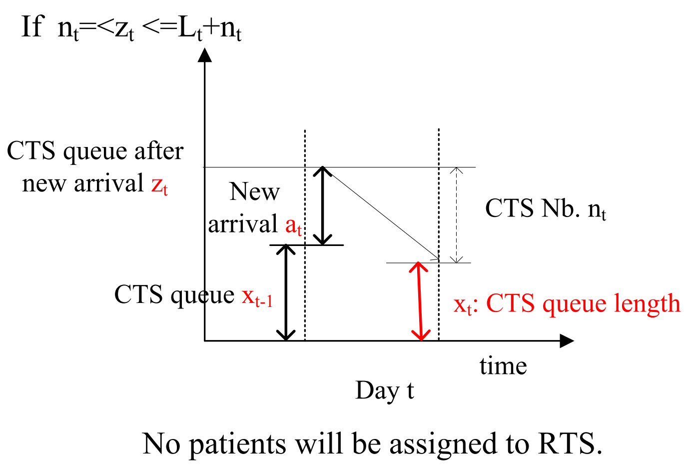


FIG. 3.2 The optimal control if $n_t =< z_t <= L_t + n_t$

Case 3: If z_t is greater than the values of $n_t + L_t$, then the number L_t of patients are kept in the CTS queue and the remaining patients are assigned to RTS.

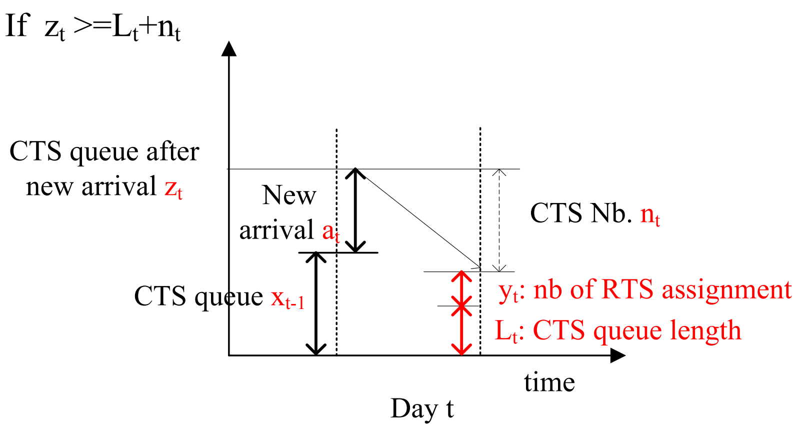


FIG. 3.3 The optimal control if $z_t >= L_t + n_t$

Remark 3-5: The implementation of contract-based MRI examination reservation process needs both the control policy and patients scheduling methods. The control policy is used to reduce the average criterion value, while patients scheduling is used to reduce the variance of patients waiting time. This part of research does not consider patients scheduling. More work is needed to explore the scheduling method which can better the distribution of patients' waiting times.

3.4 Contract optimization

In this Section, we propose a two-step approach for optimization of the contract **n**. First a Monte Carlo approximation approach is used to identify an initial contract. This contract is further improved with a local optimization.

3.4.1 Monte Carlo approximation

The contract optimization problem stated in (3.1)-(3.4) is still difficult to solve as it involves integer variables and random demands. The structure properties of the optimal control policies of Theorems 3-2 and 3-3 lead to the following equivalent reformulation of Model 3-1, i.e., relation (3.1)-(3.4):

Model 3-3:

$$F^* \equiv \underset{\mathbf{n},\mathbf{L}}{MIN} \quad F_{\mathbf{n},\mathbf{L}} \equiv \mathrm{E}\left[\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left(T^R y_t + x_t + c\left(n_{d(t)} - x_{t-1} - a_t\right)^+\right)\right] \tag{3.26}$$

subject to:

$$y_t = \left(x_{t-1} + a_t - n_{d(t)} - L_{d(t)}\right)^+ \tag{3.27}$$

$$x_t = \left(x_{t-1} + a_t - y_t - n_{d(t)}\right)^+ \tag{3.28}$$

$$\mathbf{n}, \mathbf{L} \in IN^7. \tag{3.29}$$

Theorem 3-5: Any contract **n** such that $\sum_{i=1}^{7} n_i > \Pi$ with $\Pi = \frac{T^R + c}{c}\sum_{i=1}^{7} E\left[a_i\right]$ cannot be optimal.

Proof: First, for the contract **n** with $n_i = 0$, all patients are sent to RTS and hence:

$$F_{0,0} = T^R \frac{\sum_{i=1}^{7} E\left[a_i\right]}{7}.$$

For any other contract **n**,

$$F_{\mathbf{n},\mathbf{L}} \geq c\frac{\sum_{i=1}^{7} n_i - \sum_{i=1}^{7} E\left[a_i\right]}{7}$$

The combination of the above two relations completes the proof. Q.E.D.

To simplify the problem, we convert the Model 3-3, i.e., equation (3.26)-(3.29) into a deterministic optimization problem by using a single given but long enough patient arrival sample path $\mathbf{a} = (a_1, a_2, \ldots, a_T)$. This together with Theorem 3-1 and 3-5 leads to the following Monte Carlo approximation:

$$F_T\left(\mathbf{a}\right) \equiv \underset{\mathbf{n},\mathbf{L}}{MIN} \quad F_{T,\mathbf{n},\mathbf{L}}\left(\mathbf{a}\right) \equiv \left(\sum_{t=1}^{T}\left(T^R y_t + x_t + c\left(n_{d(t)} - x_{t-1} - a_t\right)^+\right) + K\left(x_T;\mathbf{n}\right)\right)/T \tag{3.30}$$

subject to: (3.27)-(3.29), $n_i \leq \Pi$, $L_i \leq (T^R+c)\Pi$ where $K(.)$ is the total waiting time of patients in the CTS queue remaining at the end of the horizon T.

Theorem 3-6: With probability 1, $F_T(\mathbf{a})$ converges to F^* when T goes to infinity.

Proof: From Theorem 3-1 and Theorem 3-5, conditions $n_i \leq \Pi$ and $L_i \leq (T^R+c)\Pi$ do not exclude any optimal contract and optimal control policy. For the contract $\mathbf{n} = \mathbf{0}$, $F_{T,0,\mathbf{L}}\left(\mathbf{a}\right) \geq F_{T,0,\mathbf{0}}\left(\mathbf{a}\right)$, $F_{0\mathbf{L}} \geq F_{00}$ and

$$\lim_{T\to\infty} F_{T,0,0}(\mathbf{a}) = \lim_{T\to\infty}\left(\sum_{t=1}^{T}\left(T^R a_t\right)\right)/T = \sum_{i=1}^{7} T^R E\left[a_i\right]/7 = F_{0,0}$$

As a result, we only need to show the convergence of $F_{T,\mathbf{n},\mathbf{L}}(\mathbf{a})$ for each ($\mathbf{n}$, $\mathbf{L}$) with $\mathbf{n} \neq \mathbf{0}$. From Property 3-5, ($d(t)$, x_t) forms a finite state Markov chain with unique positive recurrent class including (7, 0) which implies that ($d(t)$, x_t) is a regenerative process with (7, 0) as a regeneration point. Hence,

$$\lim_{T\to\infty}\sum_{t=1}^{T}\left(T^R y_t + x_t + c\left(n_{d(t)} - x_{t-1} - a_t\right)^+\right)/T = F_{\mathbf{n},\mathbf{L}}\text{, with probability 1.}$$

Since $L_i \leq (T^R+c)\Pi$, $x_T \leq (T^R+c)\Pi$. Since $\mathbf{n} \neq \mathbf{0}$,

$$K\left(x_T;\mathbf{n}\right) \leq \sum_{j=x_T}^{1} 7j = 7x_T\left(x_T+1\right)/2 \leq 7\left(T^R+c\right)\Pi\left(\left(T^R+c\right)\Pi+1\right)/2$$

which implies $F_{T,\mathbf{n},\mathbf{L}}(\mathbf{a})$ converges to $F_{\mathbf{n},\mathbf{L}}$ as T goes to infinity. Q.E.D.

The above Monte Carlo optimization problem is still difficult to solve due to the non linear constraint (3.27) related to the control policy. We further omit this constraint and consider the following relaxed Monte Carlo optimization problem:

Model 3-4:

$$LB(\mathbf{a}) = \min\left(\sum_{t=1}^{T} T^R y_t + \sum_{t=1}^{T+D} x_t + c\sum_{t=1}^{T} u_t\right)/T \tag{3.31}$$

subject to

$$u_t = \left(n_t - (a_t + x_{t-1})\right)^+ \qquad \forall t = 1,\ldots,T \tag{3.32}$$

$$x_t = x_{t-1} + a_t - y_t - (n_t - u_t) \qquad \forall t = 1,\ldots,T \tag{3.33}$$

$$u_t = \left(n_t - x_{t-1}\right)^+ \qquad \forall t = T+1,\ldots,T+D \tag{3.34}$$

$$x_t = x_{t-1} - (n_t - u_t) \qquad \forall t = T+1,\ldots,T+D \tag{3.35}$$

$$x_t, y_t, u_t, n_t \in \mathrm{IN}, \qquad \forall t = 1, \ldots, T+D \tag{3.36}$$

where u_t denotes the number of unused CTS used in day t, D is the extra days introduced to determine the waiting times of patients remaining at the end of time horizon T. As $x_T \leq (T^R+c)\Pi$ for any optimal control policy, we can set $D = 7(T^R+c)\Pi$.

This formulation provides a lower bound of the Monte Carlo optimization problem (3.30) as any feasible solution of problem (3.30) corresponds to a feasible solution of the above problem. Further the decision variable y_t is determined with the full knowledge of the demand, both past demand and future demand. This contradicts the requirement of the so-called non-anticipativeness of any feasible control policy. However, we expect that the contract determined by this relaxed Monte Carlo approximation is a good contract. This statement will be confirmed by our numerous numerical experiments. From Theorem 3-6, LB(**a**) becomes a lower bound of the optimal average cost when the sample path **a** is longer enough.

The two nonlinear constraints (3.32) and (3.34) can be made linear. Note that reducing variables u_t leads to the reduction of both x_t and y_t, which results in the reduction of the criterion value. Thus, constraints (3.32) and (3.34) can be replaced by the following equivalent constraints:

$$u_t \geq n_t - (a_t + x_{t-1}) \text{ and } u_t \geq 0 \quad \forall t = 1,...,T \tag{3.37}$$

$$u_t \geq n_t - x_{t-1} \text{ and } u_t \geq 0 \quad \forall t = T+1,...,T+D \tag{3.38}$$

As a result, Model 3-4 can be equivalently defined as follows.

Model 3-5:

$$LB(\mathbf{a}) = \min\left(\sum_{t=1}^{T} T^R y_t + \sum_{t=1}^{T+D} x_t + c\sum_{t=1}^{T} u_t\right) / T$$

Subject to

$$\begin{aligned}
x_{t-1} \quad\quad + u_t &\geq n_t - a_t, && \forall t = 1,...,T \\
x_t - x_{t-1} + y_t - u_t &= a_t - n_t, && \forall t = 1,...,T \\
x_{t-1} \quad\quad + u_t &\geq n_t, && \forall t = T+1,...,T+D \\
x_t - x_{t-1} \quad\quad - u_t &= -n_t, && \forall t = T+1,...,T+D
\end{aligned}$$

$$x_t, y_t, u_t, n_t \in \mathrm{IN}, \quad\quad \forall t = 1, \ldots, T{+}D$$

We further show that integrity constraints of variables x_t, y_t, u_t can be relaxed. This greatly reduces the computation effort for solving LB(**a**) which contains only seven integer variables for contract **n**.

Property 3-6 (Ghouila-Houri (1962)). An $m*n$ matrix A is total unimodular if and only if for every $C \subseteq \{1,...,n\}$ there exists a partition (C^1, C^2) of C such that

$$\left|\sum_{j\in C^1} a_{ij} - \sum_{j\in C^2} a_{ij}\right| \leq 1 \text{ for } i=1,\ldots.m.$$

Theorem 3-7. The constraint matrix of the left hand side terms of the constraints of Model 3-5 is total unimodular. As a result, the integrity constraints of variables x_t, y_t, u_t can be relaxed.

Proof: From Property 3-6, it is enough to prove that, for any subset C of variables x_t, y_t, u_t, the condition of Property 3-6 holds. This is established by the following partition. All variables $x_t \in C$ belong to set C^1. For every t, variables $y_t, u_t \in C$ are partitioned as follows:

- $y_t \in C^1$, if $x_{t-1} \in C$, $u_t \notin C$
- $y_t, u_t \in C^2$, if $x_{t-1} \in C$, $u_t \in C$
- $y_t \in C^2$, if $x_{t-1} \notin C$, $u_t \notin C$
- $y_t, u_t \in C^1$, if $x_{t-1} \notin C$, $u_t \in C$. Q.E.D.

3.4.2 Improvement of the contract by local search

Starting from the contract obtained from the solution of Model 3-5, this subsection presents a local search approach to further improve the contract.

The following notation is needed:

- $F_{\mathbf{n}}$: optimal average cost under contract **n**, i.e. $F_{\mathbf{n}} \equiv \underset{\mathbf{L}}{MIN} \; F_{\mathbf{n},\mathbf{L}}$. $F_{\mathbf{n}}$ and the related optimal control policy **L**(**n**) can be determined by solving Model 3-2;
- $F_{\mathbf{n},\mathbf{L}}(\mathbf{a})$: average cost of policy **L** under contract **n** and sample path **a** estimated by relation (3.30). Note that compared with relation (3.30), index T is omitted for simplicity;
- $\mathbf{e}_i$: a seven dimension vector with i-th entry equal to 1 and all other entries equal to 0.

By definition, LB(**a**) $\leq F_{\mathbf{n},\mathbf{L}}(\mathbf{a})$.

The local search starts from the contract **n** determined by the Monte Carlo approximation. It then iteratively improves this contract. At each iteration, it determines the best neighbor solution among the set of contracts: $\mathbf{n} + \mathbf{e}_k$ (add one time slot in day k), $\mathbf{n} - \mathbf{e}_k$ (remove one time slot in day k), $\mathbf{n} - \mathbf{e}_k + \mathbf{e}_j$ (move one time slot from day k to day j). This process repeats until no improvement can be found.

The overall algorithm for the contract optimization is summarized as follows:

Algorithm 1 (Contract optimization)

1. Generate a long enough sample path **a** of patient arrivals;

2. Solve Model 3-5, the relaxed Monte Carlo approximation problem to determine LB(**a**) and an initial contract **n0**;

3. Determine the optimal control policy **L**(**n0**) and the optimal average cost $F_{\mathbf{n}0}$ under contract **n0** by solving Model 3-2;

4. Let **n*** = **n0**; $F_{\mathbf{n}^*} = F_{\mathbf{n}0}$;

5. Determine the neighbor solution **n'** with the smallest average cost as follows:

$$\mathbf{n}' = \underset{\mathbf{n}\in\{\mathbf{n}+e_k;\mathbf{n}-e_k;\mathbf{n}-e_k+e_j:1\le k,j\le 7,i\ne j\}\cap IN^7}{\arg\min} F_{\mathbf{n}}$$

6. Determine the optimal control policy **L**(**n'**) and the optimal average cost $F_{\mathbf{n}'}$ as in Step 3 if necessary;

7. If $F_{\mathbf{n}'} < F_{\mathbf{n}^*}$, set **n*** = **n'** and go to step 5.

8. The final contract is **n*** and the final control policy is **L**($\mathbf{n}^*$).

For high demand case with high patient arrival rate, the state space is large and solving the optimality relations (3.24)-(3.25) for determining $F_{\mathbf{n}}$ is time consuming. In order to reduce the computation burden, $F_{\mathbf{n},\mathbf{L}(\mathbf{n}^*)}(\mathbf{a})$ where **a** is the sample path of Step 1 can be used to replace $F_{\mathbf{n}}$ in step 5. This leads to *Algorithm 2*. Note that the contract **n'** selected in each iteration is still evaluated exactly in *Algorithm 2*.

To summarize, the solution strategies for contract optimization are summarized as follows:

1) Generating a long enough sample path **a** of patient arrivals, the relaxed Monte Carlo approximation model (Model 3-5) is solved to determine LB(**a**) and an initial contract **n0**;

2) Based on this **n0**, local optimization 1 or 2 is used to improve the contracts ($\mathbf{n}^*$ is the final contract), determine the final control policy **L**(**n***), and obtain the exact average cost $F_{\mathbf{n}^*}$.

To evaluate the performance of this solution strategy, the global optimal contract and control policies can also be determined by exhaustive search by comparing $F_{\mathbf{n}}$ for all **n** in a certain range containing the optimal solution. The global optimal solution is recorded as $F^* = \text{MIN}\{F_{\mathbf{n}}: \forall\mathbf{n}\}$. This is possible when the demand is low but too time consuming to obtain when the demand is high.

Remark 3-6: Although the Monte Carlo approximation and relaxations are used for determination of the contract decisions, all our algorithms use MDP to find the exact criterion value of the resulting contracts. Further, both Algorithms 1 and 2 start with the same contract provided by the solution of the same model, i.e., Model 3-5.

3.5 Computational results

This section presents numerical experiments performed to evaluate the performance of the solution strategies and to investigate how the contract and control policies depend on problem parameters. The performance of the solution strategies is given by comparing with the exact optimum F^* obtained by exhaustive search for small-size problems and with F^{best} for large-size problems, where F^{best} is the best solution of all independent runs of Algorithms 1 and 2 for the same problem instance. We then perform sensitivity analysis to show how the optimal contract depends on different factors such as average RTS delay T^R and unused CTS penalty c, patient arrival patterns and patient arrival rates. All numerical results are performed on a Pentium® 4 PC running at 3.21 GHz with 1.0 GB RAM. LP models are solved by the CPLEX 11 solver.

3.5.1 Numerical experiments

We first describe the base case corresponding to our real case study. From the data collected from the neurovascular department under study, the average numbers of patient arrivals during the week are as follows: {1, 0.89, 0.95, 1.16, 1.53, 0.16, 0.05}. The number of patients arrived each day is assumed to follow a Poisson distribution. The average waiting time for RTS is in the range of 30~40 days with an average of T^R = 35 days. The weight, c, is set to 15. These data define the base case shown in Table 3.1.

Average number of patients arrived							T^R	C
Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.		
1	0.89	0.95	1.16	1.53	0.16	0.05	35	15

TAB 3.1 Base case data

The above base case is then modified to investigate the impact of parameters average RTS delay T^R and unused CTS penalty c, patient arrival pattern and patient arrival rate. More precisely, the following numerical experiments are considered:

- Case 1: base case but with different weighing factor $c \in \{1, 5, 10, 15, 20\}$ (impact of unused CTS penalty, c);
- Case 2: base case but with different delay for RTS $T^R \in \{25, 30, 35, 40, 45\}$ (impact of average RTS delay, T^R);
- Case 3: base case but with different patient arrival patterns (impact of patient arrival pattern). The peak demand of the base case occurs on Friday. To see the impact of

the patient arrival patterns on the contract and control policy, we interchange the demand in Friday and the demand of another weekday. We also consider the case of stationary demand for all weekdays.

- Case 4: base case but with different patient arrival rate (impact of patient arrival rate). The base case is considered as low demand case. Two other case termed medium and high demand instances are derived from the base case by multiplying the patient arrival rate by 5 and 10 respectively.

The sample path **a** needed for the Monte Carlo optimization is generated on a time horizon of 1000 weeks, i.e. T=7000 days, nearly 20 years. The number of patients arrival a_t for t=1, …, 7000 is generated according to a Poisson distribution truncated at 20, 40, and 60 for respectively low, medium, and high demand instances. D = 100 extra days are used in the Monte Carlo approximation to determine waiting time of patients in the queue at the end of the planning horizon. Model 3-2 is solved by restricting the queue length x_t to a given parameter LL set by trial-and-error. LL = 30, 40, and 40 for low, medium, and high demand instances. We also use very long-time simulation for the evaluation of different performance measures of final contracts and control policies determined by exhaustive search for low demand instances and by optimization algorithms for medium and high demand instances.

For every problem instance, 10 independent patient arrival sample paths **a** are generated, Algorithms 1-2 are run using each sample path to investigate the performance of these two algorithms.

3.5.2 Performances of the proposed algorithms and impact of average RTS delay T^R and unused CTS penalty c

This subsection considers Cases 1-2 to investigate the performances of Algorithms 1-2 and to show the impact of average RTS delay T^R and unused CTS penalty c. The patient arrival rate is very low and exhaustive search of optimal contract can be done in acceptable time.

The exhaustive search is performed for all contract **n** with $n_i \in \{0, 1, 2, 3\}$. The range was selected based on observation of different contracts provided by the proposed two-step approaches Algorithms 1-2. The optimal control policy and the average cost $F_{\mathbf{n}}$ of each contract are determined by solving the Model 3-2. The cost of the resulting optimal contract is denoted as F^*.

Table 3.2 summarizes optimal solutions given by exhaustive search. First note that the exhaustive search takes more the 2.5 hour for each instance. The total number of CTS decreases and the criterion value increases when the weight factor c of unused CTS increases. The contracts and control policy are sensitive to the change of *c*. With the increase of T^R, the objective criterion *F** increases and the control threshold L_i increases, but contracts do not change. This means that the optimal contract is insensitive to change of T^R.

T^R	*c*	*F**	CPU Time (s)	n_1	n_2	n_3	n_4	n_5	n_6	n_7	L_1	L_2	L_3	L_4	L_5	L_6	L_7
35	1	0.945	9257	2	1	2	2	2	1	0	22	22	22	21	21	21	22
35	5	2.484	9249	1	1	1	2	2	1	0	13	14	14	13	13	12	13
35	10	3.589	9225	1	1	1	1	3	0	0	10	10	10	10	8	9	10
35	15	4.501	9189	1	1	1	1	3	0	0	11	11	11	11	9	10	10
35	20	5.410	9187	1	1	1	1	3	0	0	11	12	12	12	10	11	11
25	15	4.471	9203	1	1	1	1	3	0	0	9	9	9	9	7	8	9
30	15	4.489	9224	1	1	1	1	3	0	0	10	10	10	10	8	9	10
35	15	4.501	9189	1	1	1	1	3	0	0	11	11	11	11	9	10	10
40	15	4.510	9279	1	1	1	1	3	0	0	11	12	12	12	10	11	11
45	15	4.516	9433	1	1	1	1	3	0	0	12	12	13	13	11	12	12

TAB 3.2 Optimal solutions of Cases 1-2 by exhaustive search

Table 3.3 shows the impact of T^R and *c* on average delay (*Delay*), percentages of unused CTS (*Unused CTS Ratio*) and percentages of patients using RTS (*RTS Perc.*) with the optimal contracts and control policies. When *c* decreases, the corresponding waiting time decreases while the unused CTS ratio increases. Therefore, the choice of *c* allows balancing between the average waiting time and the unused CTS ratio. There is no obvious trend in the change of RTS percentage. However, as the contract is insensitive to the change of T^R, these performance measures are also insensitive to the change of T^R.

T^R	*c*	*Delay(days)*	*Unused CTS Ratio* (%)	*RTS Perc.*(%)
35	1	0.41	42.60	0.00
35	5	1.06	28.26	0.00
35	10	2.14	18.32	0.38
35	15	2.16	18.22	0.26
35	20	2.17	18.16	0.18
25	15	2.08	18.46	0.56
30	15	2.13	18.32	0.38
35	15	2.16	18.22	0.26
40	15	2.17	18.22	0.26
45	15	2.20	18.11	0.13

TAB 3.3 Performances of contract-based reservation process for different average RTS delay and unused CTS penalty

Table 3.4 summarizes results of Algorithms 1 and 2 for the same cases. For each problem instance, each algorithm is applied 10 times with 10 different sample paths. Let **n1** and **n2** be the contract provided by respectively Algorithms 1 and 2. Columns "$F_{\mathbf{n}k}$" give the minimum, average, and maximum of the exact criterion values of contract **n***k* of 10 different runs of Algorithm *k,* where *k*=1, 2. Columns "*Gapk*" show the minimum, average, and maximum of the average deviations of $F_{\mathbf{n}k}$ from F^*, i.e. $(F_{\mathbf{n}k} - F^*)/F_{\mathbf{n}k}$. Columns "*Movek*" are the minimum, average, and maximum of local search moves in Algorithms *k*. *RT*1 and *RT*2 are the average CPU time. The optimal contract is always found by Algorithm 1 except for the instance $c = 1$ for which the 10 criterion values of Algorithm 1 are all within 1% of the optimum. Results in column "*Move*1" show that the contract of the Monte Carlo approximation is close to the optimum as the optimal contract can be obtained with less than 2 local moves. Recall that, from Remark 3-6, Algorithms 1 and 2 actually start from the same Monte Carlo solution for each of the 10 sample paths. Results of Algorithm 2 are good with average deviation from the optimum of less than 2.2% and with a deviation of about 11% for one run of $c = 15$. However the CPU time of Algorithm 2 is much smaller than that of Algorithm 1. From Table 3.4, Algorithm 2 is at least six times faster than Algorithm 1. The quality of the best solution of six independent runs of Algorithm 2 is fairly close to that of Algorithm 1 of one run.

T^R	c	$F_{\mathbf{n}1}$	$F_{\mathbf{n}2}$	*Gap*1(%)	*Gap*2(%)	*Move*1	*Move*2	*RT*1(s)	*RT*2(s)
35	1	[0.945,0.947,0.948]	[0.945,0.951,0.982]	[0,0.20,0.32]	[0,0.53,3.70]	[0,0.1,1]	[0,0,0]	126	7
35	5	[2.484,2.484,2.484]	[2.484,2.504,2.671]	[0,0,0]	[0,0.77,7.02]	[0,0.7,2]	[0,0.3,2]	190	13
35	10	[3.589,3.589,3.589]	[3.589,3.592,3.604]	[0,0,0]	[0,0.08,0.43]	[0,0.3,1]	[0,0.1,1]	140	11
35	15	[4.501,4.501,4.501]	[4.501,4.557,5.056]	[0,0,0]	[0,1.11,10.98]	[0,0.7,1]	[0,1.3,3]	191	27
35	20	[5.410,5.410,5.410]	[5.410,5.430,5.450]	[0,0,0]	[0,0.37,0.73]	[1,1,1]	[0,1,2]	230	15
25	15	[4.471,4.471,4.471]	[4.471,4.540,4.608]	[0,0,0]	[0,1.48, 2.96]	[1,1,1]	[0,1,2]	205	13
30	15	[4.489,4.489,4.489]	[4.489,4.594,4.840]	[0,0,0]	[0,2.18, 7.25]	[1,1,1]	[0,1.5,3]	217	25
35	15	[4.501,4.501,4.501]	[4.501,4.557,5.056]	[0,0,0]	[0,1.11,10.98]	[0,0.7,1]	[0,1.3,3]	191	27
40	15	[4.510,4.510,4.510]	[4.510,4.510,4.516]	[0,0,0]	[0,0.02,0.14]	[0,0.6,1]	[0,1.1,3]	186	31
45	15	[4.516,4.516,4.516]	[4.516,4.518,4.522]	[0,0,0]	[0,0.04,0.14]	[0,0.5,1]	[0,0.6,2]	176	28

TAB 3.4 Performances of Algorithms 1 and 2 for different average RTS delay and unused CTS penalty

We now analyze the quality of the sample path lower bound LB(**a**). Fig. 3.4 and Fig. 3.5 shows the deviation gap between LB(**a**) and $F_{\mathbf{n}^*,\mathbf{L}(\mathbf{n}^*)}(\mathbf{a})$ where n* is the optimal contract obtained by exhaustive search, i.e. $(F_{\mathbf{n}^*,\mathbf{L}(\mathbf{n}^*)}(\mathbf{a})$ - LB(**a**))/$F_{\mathbf{n}^*,\mathbf{L}(\mathbf{n}^*)}(\mathbf{a})$. As it can be seen, the quality of the lower bound becomes worse when the weighting factor c increases but it becomes tighter when T^R increases.

Even though the sample path lower bound LB(**a**) is loose, the contract of the Monte Carlo approximation is actually close to the optimal contract and is at most two local moves away from the optimal contract in all our numerical experiments.

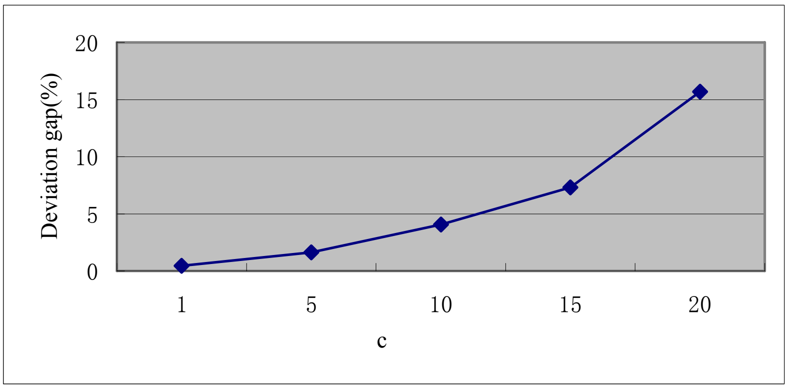


FIG. 3.4 Gap of LB(**a**) for different c

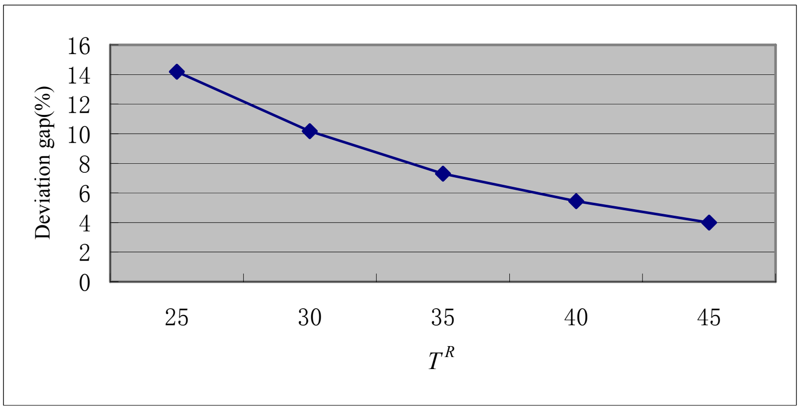


FIG. 3.5 Gap of LB(**a**) for different T^R

Remark 3-7: In order to check whether Assumption 4 requiring stable weekly contract is strong, we apply Algorithm 1 to optimize bi-week contract and the related optimal control policy, i.e. for contract **n** and control **L** defined over 14 days. Numerical experiments are performed for Cases 1 and 2. The resulting bi-week contracts and the optimal criterion values remain the same as weekly contracts of Table 3.2. This implies that Assumption 4 is not really strong in these cases.

3.5.3 Impact of patient arrival pattern

This subsection considers Case 3 to show the impact of patient arrival pattern. Table 3.5 summarizes the optimal solutions obtained by exhaustive search for different patient arrival patterns. The row “Mon.” corresponds to patient arrival pattern derived from the base case by exchanging the patient arrival rate of Monday with that of Friday (actual peak arrival). The next four rows are defined similarly. The row “Ave.” corresponds to the case of stationary arrival with the same arrival rate for all workday with the same weekly patient arrival rate. From this table, the total number of CTS of the optimal contract decisions does not change with respect to the patient arrival patterns. However, one CTS moves from Friday to the day of peak arrival and there are still 2 CTS for Friday in order to serve patient arrival during the week. In the case of stationary weekday arrival, one CTS moves from Friday to Wednesday. The control policy seems to be insensitive to the patient arrival patterns.

Table 3.6 summarizes average delay time, unused CTS ratios, and RTS percentages for different arrival patterns. These performance measures seem to be insensitive to the change of patient arrival pattern. Table 3.7 presents results of Algorithms 1 and 2 for the same cases. The same observations can be made as in Section 3.5.2.

Peak arrival	F^*	CPU Time (s)	n_1	n_2	n_3	n_4	n_5	n_6	n_7	L_1	L_2	L_3	L_4	L_5	L_6	L_7
Mon.	4.506	9274	2	1	1	1	2	0	0	10	10	11	11	10	10	11
Tues.	4.496	9277	1	2	1	1	2	0	0	11	10	10	11	10	10	11
Wed.	4.487	9332	1	1	2	1	2	0	0	11	11	10	11	9	10	11
Thurs.	4.476	9283	1	1	1	2	2	0	0	11	11	11	10	9	10	10
Fri.	4.501	9189	1	1	1	1	3	0	0	11	11	11	11	9	10	10
Ave.	4.517	9246	1	1	2	1	2	0	0	11	11	10	11	10	10	11

TAB 3.5 Optimal solutions of different patient arrival patterns by exhaustive search

Peak arrival	*Delay(days)*	*Unused CTS Ratio* (%)	*RTS Perc.*(%)
Mon.	2.16	18.19	0.24
Tues.	2.16	18.20	0.26
Wed.	2.16	18.18	0.28
Thurs.	2.13	18.20	0.27
Fri.	2.16	18.22	0.26
Ave.	2.13	18.59	0.22

TAB 3.6 Performances of contract-based reservation process for different patient arrival patterns

Peak arrival	$F_{\mathbf{n}1}$	$F_{\mathbf{n}2}$	*Gap*1(%)	*Gap*2(%)	*Move*1	*Move*2	*RT*1(s)	*RT*2(s)
Mon.	[4.506,4.506,4.506]	[4.506,4.624,5.080]	[0,0,0]	[0,2.33,11.30]	[0,0.9,1]	[0,1.3,2]	211	25
Tues.	[4.496,4.496,4.496]	[4.496,4.504,4.555]	[0,0,0]	[0,0.17,1.30]	[0,1,2]	[0,1.6,2]	222	28
Wed.	[4.487,4.487,4.487]	[4.487,4.491,4.522]	[0,0,0]	[0,0.08,0.77]	[0,1.2,2]	[0,1.7,3]	246	34
Thurs.	[4.476,4.476,4.476]	[4.476,4.536,5.075]	[0,0,0]	[0,1.18,11.79]	[0,0.7,1]	[0,1.2,2]	192	27
Fri.	[4.501,4.501,4.501]	[4.501,4.557,5.056]	[0,0,0]	[0,1.11,10.98]	[0,0.7,1]	[0,1.3,3]	191	27
Ave.	[4.517,4.517,4.517]	[4.517,4.567,4.999]	[0,0,0]	[0,1.01,9.64]	[1,1,1]	[0,2.1,3]	220	28

TAB 3.7 Performance of solution strategies for different patient arrival patterns

3.5.4 Impact of patient arrival rate

As the patient arrival rate increases, the optimal solution is hard to obtain by exhaustive search because it needs too long time to search within a large solution space. We limit ourselves to Algorithms 1 and 2 in this experiment.

Tables 3.8 and 3.9 summarize best contracts and corresponding control policies and the performances of the two algorithms for different patient arrival rates where three scenarios are considered "Low" (base case), "Medium" (patient arrival rates 5 times larger), "High" (patient arrival rates 10 times larger). "*Gap*1" and "*Gap*2" in this case are the deviation gap of $F_{\mathbf{n}k}$ from F^{best}, i.e. $Gapk = (F_{\mathbf{n}k} - F^{\text{best}})/F_{\mathbf{n}k}$ where F^{best} is the best solution of 20 solutions of the ten runs of Algorithms 1 and 2.

From these tables, Algorithm 1 is always able to find the best solutions given in Table 3.8 whatever the sample path used while the quality of the contract decisions obtained by Algorithm 2 is more sensitive to the sample path used. *Move*1 shows that the contract decisions obtained by Monte Carlo approximation is very close to the best contract and is at most two local moves from the best contract. This highlights the quality of contract decisions given by the Monte Carlo Approximation as 61 CTS (resp. 31 and 7 CTS) are needed for high (resp. medium and low) demand instance. Finally, Algorithm 1 becomes too slow for high demand instance while Algorithm 2 is much faster and is able to provide a good solution with a reasonable CPU time.

Arrival rate	n_1	n_2	n_3	n_4	n_5	n_6	n_7	L_1	L_2	L_3	L_4	L_5	L_6	L_7
Low	1	1	1	1	3	0	0	11	11	11	11	9	10	10
Medium	5	5	5	6	9	1	0	21	21	21	21	19	19	21
High	10	9	10	12	17	2	1	31	31	31	31	28	29	30

TAB 3.8 Best contracts and control policies for different patient arrival rates

Arrival rate	F_{n1}	F_{n2}	*Gap*1(%)	*Gap*2(%)	*Move*1	*Move*2	*RT*1(s)	*RT*2(s)
Low	[4.50,4.50,4.50]	[4.50,4.55,5.06]	[0,0,0]	[0,1.11,10.98]	[0,0.7,1]	[0,1.3,3]	191	27
Medium	[9.83,9.83,9.83]	[9.83,9.90,10.40]	[0,0,0]	[0,0.64,5.50]	[1,1.1,2]	[0,1.7,3]	2006	99
High	[13.94,13.94,13.94]	[13.94,14.00,14.10]	[0,0,0]	[0,0.43,1.14]	[1,1.9,2]	[0,2.3,4]	9913	771

TAB 3.9 The performance of solution strategies for different patient arrival rates

Table 3.10 summarizes average delay, unused CTS ratios, RTS percentages for different patient arrival rates with the best contract decisions obtained. All the performance measures decrease when the arrival rate increases.

Arrival rates	*Delay(days)*	*Unused CTS Ratio* (%)	*RTS Perc.* (%)
Low	2.16	18.22	0.26
Medium	1.16	7.68	0.24
High	0.74	6.01	0.11

TAB 3.10 Performances of contract-based reservation process for different patient arrival rates

3.6 Implementation issues

As noticed in Section 3.5, patients directed to RTS will face significantly longer regular reservation time than patients waiting for CTS. Although the new MRI examination reservation process leads to shorter average waiting time, these "unlucky" patients directed to RTS still face long waiting time.

In order to have better waiting time distribution, this subsection proposes an improved method. The new method still makes use of contracts and patient assignment control policy. However, the patient assignment policy is replaced by an **RTS reservation policy**. When the CTS queue length exceeds the threshold L_i, additional time slots are reserved according to the regular reservation process. However patients are not directly assigned to RTS at the same day. CTS and RTS time slots are grouped according to their day of availability and filled by patients. All patients are scheduled to both CTS and RTS in the First in First out (FIFO) order. This method is expected to reduce the longest waiting time of stroke patients and to avoid unlucky patients.

In the following of this subsection, we call the reservation process with patient assignment policy the **old reservation process** and the one with RTS reservation policy the **new reservation process**.

3.6.1 Comparison of control policies

For the old reservation process, the optimal patient assignment policy for any given contract can be determined by solving Model 3-2. Further it was shown that, for a given contract ***n***, the optimal patient assignment policy is a **threshold policy** and there exists a threshold L_i for day i such that new incoming patients are directed to RTS when the ending CTS queue length reaches L_i. The old MRI examination reservation process is as follows. During each day t is as follows, first, the queue length x_{t-1} of patients waiting for CTS at the end of day $t-1$ is known. The number of CTS for day t is n_t. The number of new arrival patients in day t is a_t. The state variable $z_t = x_{t-1} + a_t$ is known. MIN$\{n_t, x_{t-1} + a_t\}$ patients are served by the CTS of the day and $u_t = (n_t - z_t)^+$ CTS of day t cannot be filled and are unused. $y_t = (z_t - n_t - L_t)^+$ new incoming patients with $L_t = L_{d(t)}$ are directed to RTS. $x_t = (z_t - n_t - y_t)^+$ remaining patients will wait for CTS in the subsequent days. Patients in CTS queue are served in FIFO order. The average cost of the old reservation process is as follows:

$$J^{old} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left[cu_t + x_t + T^R y_t \right]$$

For the new reservation process, the structure of the optimal RTS reservation policy is unclear. Instead, we use the optimal patient assignment policy to make RTS reservation. Both CTS and RTS time slots are grouped together and assigned to patients in a FIFO order. More specifically, during each day t, we keep track of an artificial CTS queue length x_{t-1} under patient assignment policy and use it to determine the number y_t of RTS time slots to reserve in day t, i.e. $y_t = (x_{t-1} + a_t - n_t - L_t)^+$ and $x_t = (x_{t-1} + a_t - n_t - y_t)^+$. The real patient queue length, the number of patients waiting for a time slot, at the end of day t-1 is denoted as x_{t-1}^{new} with $x_0^{new} = x_0$. The number of time slots available in day t including both CTS and RTS is $n_t^{new} = n_t + y_{t-T^R}$. The number new incoming patients during day t is a_t. MIN$\{n_t^{new}, x_{t-1}^{new} + a_t\}$ patients are served by the CTS and RTS of the day and $u_t^{new} = \left(n_t^{new} - \left(x_{t-1}^{new} + a_t\right)\right)^+$ time slots of day t cannot be filled. The remaining $x_t^{new} = \left(x_{t-1}^{new} + a_t - n_t^{new}\right)^+$ patients will wait in the patient queue. Patients are served in FIFO order. The average cost of the new reservation process is as follows:

$$J^{new} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left[cu_t^{new} + x_t^{new} \right]$$

Both new and old reservation processes are characterized by a contract ***n*** and a set of control thresholds ***L***.

In order to compare the two control policies, the following notation is needed:

$u_t^{old} = u_t$: number of unused CTS in old reservation process

$x_t^{old} = x_t + \sum_{s=t-T^R}^{t} y_s$: number of patients waiting for a time slot in the old reservation process including both those in the CTS queue and those sending to the RTS but not yet served.

y_t^{old} : number of patients sending to RTS

y_t^{new} : number of RTS reserved

Property 3-7: $y_t^{old} = y_t^{new} = y_t, \forall t$.

Proof: Trivial and follows from the definition. Q.E.D.

For simplicity, let us assume that there is no pending RTS reservation initially, i.e. $y_t^{old} = y_t^{new} = 0, \forall t \le 0$. By patient assignment policy, $x_t \le L_t, \forall t > 0$ and hence $y_t = \left(x_{t-1} + a_t - n_t - L_t\right)^+ \le L_{t-1} + a_t, \forall t > 1$ is a finite random variable. As a result,

$$J^{old} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left[cu_t^{old} + x_t^{old} \right]$$

Property 3-8: $x_t^{old} \ge x_t^{new}, \forall t \ge 0$ **.**

Proof*:* This property is proved by induction. The property clearly holds for *t* = 0. Assume that it holds for *t*–1. We now prove it for *t*. First, for the old reservation process and for day *t*, a_t new patients arrive. $MIN\left\{n_t, x_{t-1} + a_t\right\}$ patients have their MRI exam by CTS time slots and $y_{t-T^R}^{old}$ patients by RTS time slots. Hence,

$$\begin{aligned} x_t^{old} &= x_{t-1}^{old} + a_t - MIN\left\{n_t, x_{t-1} + a_t\right\} - y_{t-T^R}^{old} \\ &= MAX\left\{x_{t-1}^{old} + a_t - n_t - y_{t-T^R}^{old}, x_{t-1}^{old} - x_{t-1} - y_{t-T^R}^{old}\right\} \end{aligned}$$

For the new reservation process,

$$x_t^{new} = MAX\left\{x_{t-1}^{new} + a_t - n_t - y_{t-T^R}^{new}, 0\right\}$$

Combining with Property 3-7, the induction assumption and the fact of $x_{t-1}^{old} \ge x_{t-1} + y_{t-T^R}^{old}$, the above two relations lead to $x_t^{old} \ge x_t^{new}$ which completes the proof. Q.E.D.

Property 3-9: $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}u_t^{old} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}u_t^{new}$.

Proof: By definition,

$$x_t^{old} = x_0 + \sum_{s=1}^{t} a_s - \sum_{s=1}^{t-T^R} y_s^{old} - \sum_{s=1}^{t}\left(n_s - u_s^{old}\right)$$

$$x_t^{new} = x_0 + \sum_{s=1}^{t} a_s - \sum_{s=1}^{t}\left(n_s + y_{s-T^R}^{new} - u_s^{new}\right)$$

As $y_t = 0$ for all $t \le 0$, subtracting the two relations leads to:

$$x_t^{old} - x_t^{new} = \sum_{s=1}^{t} u_s^{old} - \sum_{s=1}^{t} u_s^{new}$$

which proves the property if both x_t^{old} and x_t^{new} are finite numbers. The finiteness of x_t^{old} is true as $x_t \le L_t$, $y_t \le L_{t-1} + a_t$ and $x_t^{old} = x_t + \sum_{s=t-T^R}^{t} y_s \le \left(T^R + 1\right)L* + \sum_{s=t-T^R}^{t} a_s$. The finiteness of x_t^{new} follows from Property 3-8. Q.E.D.

Property 3-10: $J^{old} \ge J^{new}$.

Proof: Direct consequence of Properties 3-9 and 3-10. Q.E.D.

Remark 3-8: From the proof of Property 3-8, more patients can be served by CTS time slots in the new reservation process and hence $x_t^{old} > x_t^{new}$ when $n_t > x_{t-1} + a_t$. In this case, in the old reservation process, CTS queue does not contain enough patients to fill all CTS time slots and patients directed to RTS cannot be redirected. In the new reservation process, as no patients are directed RTS, these extra CTS time slots can be filled by patients that were directed to RTS in the old reservation process.

Remark 3-9: From the proof of Property 3-8 and the proof of Property 3-9, $\sum_{s=1}^{t} u_s^{old} - \sum_{s=1}^{t} u_s^{new} \ge 0$. The new reservation process has less time slot cancellation even though the average cancellation rate is the same.

Property 3-11: The maximum waiting time of the new reservation process is smaller or equal to T^R if the thresholds L_t are such that $L_t \le \sum_{s=t+1}^{t+T^R} n_s$ and $L_t - n_{t+1} \le L_{t+1}$.

Remark 3-10: Condition of Property 3-11 basically assumes that patients in CTS queue of the old reservation process wait no longer than T^R and a patient in the CTS queue will not be directed to RTS. The last condition holds for any optimal control policy. By relaxing the first condition, it is still possible to show that the maximal waiting time of the new reservation process does not exceed that of the old reservation process.

Proof: Consider the last patient arriving in day t in the new reservation process, i.e. the x_t^{new}-th patient in the patient queue. Assume by contradiction that its waiting time exceeds T^R. As patients are served in FIFO order in the new reservation process,

$$x_{t+T^R}^{new} > \sum_{s=t+1}^{t+T^R} a_s$$

Further, for the old reservation process, as $x_t^{old} = x_t + \sum_{s=t-T^R}^{t} y_s$ and $x_t \le L_t$, the conditions of the property ensure that all patients in x_t^{old} have been served by time $t + T^R$. As a result,

$$x_{t+T^R}^{old} \le \sum_{s=t+1}^{t+T^R} a_s$$

which contradicts Property 3-8 and concludes the proof. Q.E.D.

Property 3-12: $D^{old} \ge D^{new}$ where D^{old} and D^{new} correspond to average patient waiting times in old and new contract-based reservation process.

Proof: Since both policies face the same patient arrival rate and both x_t^{old} and x_t^{new} are finite, by Little's law,

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t^{old} = D^{old} \cdot \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} a_t$$

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t^{new} = D^{new} \cdot \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} a_t$$

which, together with Property 3-8, concludes the proof. Q.E.D.

This property shows that, the new reservation process not only reduces the maximum waiting time by avoiding "unlucky" patients directed to longer regular reservation but also reduces the average waiting time.

3.6.2 Experimental results

The data are the same with Section 3.5. For each case, two scheduling methods are considered: (i) old MRI examination reservation process, denoted “Old”; (ii) new MRI examination reservation process, denoted “New”. The contract is provided in the following and the optimal patient assignment policies of the old reservation process are determined by solving the LP model in Remark 3-4. The new reservation is defined with the same contract **n** and the same set of control thresholds **L**. Each reservation process is then evaluated by Monte Carlo simulation over a time horizon of 1 million weeks.

3.6.2.1 Waiting time distribution

We consider first the base case. The average weekly demand is 5.74 MRI time slots. The contract for 6 CTS is: **n** = (1, 1, 1, 1, 2, 0, 0). For this contract, the optimal patient assignment policy is as follows **L** = (6, 6, 6, 6, 5, 6, 6).

Fig.3.6 compares the probability distribution of patient waiting times of two scheduling methods for the base case. From this figure, the waiting time in old method is distributed over the range of 0~7 days plus a high probability spike of 5% at 35 days corresponding to "unlucky" patients directed to RTS. In the new method, the distribution of patients' waiting time is much more smoothed on a range from 0 to about 20 days without "unlucky" patients having to wait for much longer time than patients arriving around the same time. The new method greatly reduces the maximal waiting time.

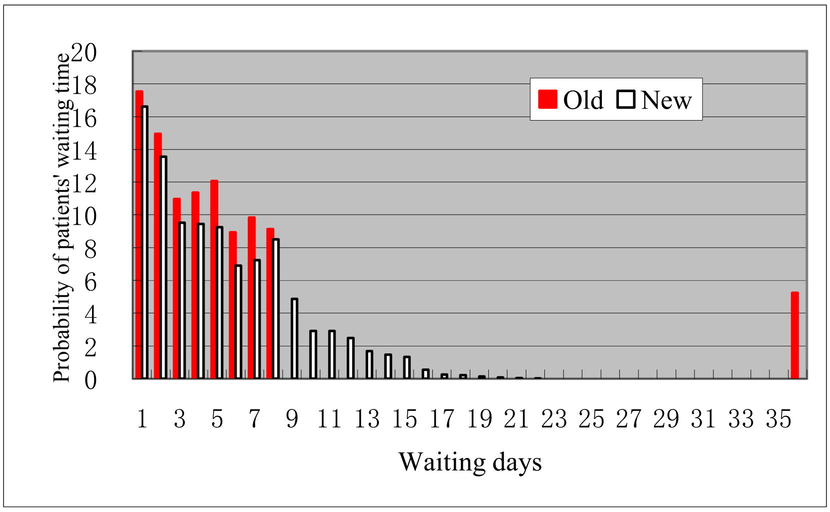


FIG. 3.6 Waiting time distribution of the base case

Probability distributions of waiting time in all other cases are similar. In the following sensitivity analysis are performed with respect to different performance criteria including the average criterion value, the average delays, the variance of waiting times, and the maximal waiting times.

3.6.2.2 Impact of parameters

We now perform the sensitivity analysis with respect to cost *c* of unused time slots by varying *c* in {1, 5, 10, 15, 20}. The contract is **n** = (1, 1, 1, 1, 2, 0, 0). The optimal patient assignment policies are: **L** = (5,5,5,5,4,5,5), (5,5,5,5,4,5,5), (6,6,6,6,5,5,6), (6,6,6,6,5,6,6), (7,6,6,7,6,6,7).

Table 3.11 compares the performance of both scheduling methods. "*OBJ*" is the long run average criterion value, i.e. J^{new} or J^{old}. "*Ave*", "*Dev*" and "*Max*" are the average waiting time, standard deviation and maximum of the waiting times, respectively. "*Gap*" = $(J^{old} - J^{new}) / J^{old}$ is the relative deviation of the criterion value of new method with respect to that of the old method. Compared with those in the old method, the performance indices in new method greatly decreases, which means that the new method improves the performances of contract-based reservation process. With the increase of *c*, *OBJ*, *Ave*, *Dev*, and *Max* increase. This means the value of unused CTS cost greatly affects patients' waiting time. *Gap* decreases from 12.3% to 3.74%, and the average delay increases from 4.04 to 4.54. Maximal waiting time keeps nearly the same, about 26~27 days.

It seems that the new method improves more when more patients are directed to RTS. Indeed, for small *c*, the CTS queue length *L* is smaller and the improvement of the new reservation process is greater.

c	Old				New				Gap (%)
	OBJ	*Ave*	*Dev*	*Max*	*OBJ*	*Ave*	*Dev*	*Max*	
1	3.88	4.62	8.18	35	3.40	4.04	3.65	26	12.30
5	4.24	4.62	8.18	35	3.76	4.04	3.65	26	11.26
10	4.65	4.69	7.53	35	4.36	4.33	3.83	27	6.23
15	5.06	4.70	7.47	35	4.78	4.37	3.84	27	5.41
20	5.45	4.79	7.16	35	5.25	4.54	3.95	27	3.74

TAB 3.11 Performance vs unused CTS penalty costs *c*

We now perform the sensitivity analysis with respect to the average delay of regular reservation by varying T^R in {25, 30, 35, 40, 45}. The contract is **n** = (1, 1, 1, 1, 2, 0, 0). The optimal patient assignment policies are: **L** = (5,5,5,5,4,5,5), (6,6,6,6,5,5,6), (6,6,6,6,5,6,6), (7,6,6,7,6,6,7), (7,7,7,7,6,7,7).

Table 3.12 compares the performance of both reservation processes. The impact of T^R on the performance is nearly the same with the impact of *c*, except the Max and *Gap*. As shown in Table 3.12, *Max* increases from 21 to 30 days, and *Gap* increases from 4.5% to 5.4% with the increase of T^R. This means the new method is more important for longer T^R.

We now perform the sensitivity analysis with respect to the patient arrival pattern. The arrival pattern has no obvious impact on the performances of the reservation processes.

T^R	Old				New				Gap(%)
	OBJ	Ave	Dev	Max	OBJ	Ave	Dev	Max	
25	4.61	3.98	5.83	25	4.40	3.73	3.26	21	4.50
30	4.84	4.42	6.47	30	4.65	4.19	3.64	24	3.92
35	5.06	4.70	7.47	35	4.78	4.37	3.84	27	5.41
40	5.26	5.03	8.17	40	4.98	4.68	4.11	28	5.35
45	5.45	5.31	8.86	45	5.15	4.95	4.34	30	5.40

TAB 3.12 Performance vs average RTS delays T^R

Peak demand	**n**	**L**
Mon.	{2,1,1,1,1,0,0}	{6,6,6,6,6,6,7}
Tues.	{1,2,1,1,1,0,0}	{7,6,6,6,6,6,7}
Wed.	{1,1,2,1,1,0,0}	{6,6,6,6,6,6,6}
Thurs.	{1,1,1,2,1,0,0}	{6,6,6,6,5,6,6}
Fri.	{1,1,1,1,2,0,0}	{6,6,6,6,5,6,6}
Ave	{1,1,2,1,1,0,0}	{7,7,6,6,6,6,7}

TAB 3.13 Contracts and patient assignment policies vs patient arrival patterns

Peak demand	Old				New				Gap (%)
	OBJ	Ave	Dev	Max	OBJ	Ave	Dev	Max	
Mon.	5.13	4.84	7.28	35	4.90	4.55	3.90	26	4.59
Tues.	5.12	4.84	7.24	35	4.90	4.57	3.96	27	4.28
Wed.	5.13	4.83	7.34	35	4.89	4.54	3.95	26	4.67
Thurs.	5.12	4.76	7.54	35	4.83	4.41	3.86	26	5.69
Fri.	5.06	4.70	7.47	35	4.78	4.37	3.84	27	5.41
Ave	5.11	4.82	7.05	35	4.91	4.57	3.94	27	3.89

TAB 3.14 Performance vs patient arrival patterns

We now perform the sensitivity analysis with respect to the patient arrival rate. Three scenarios are considered "Low" (base case), "Medium" (patient arrival rates 5 times larger), "High" (patient arrival rates 10 times larger). The contract is **n** = (1, 1, 1, 1, 2, 0, 0) for the base case, five times larger for medium demand case, and ten times larger for high demand case. The optimal patient assignment policy is as follows: **L** = (6, 6, 6, 6, 5, 6, 6), (17, 16, 16, 17, 14, 15, 17), (27, 27, 26, 28, 22, 25, 27).

From Table 3.15, compared with old method, nearly all performance criteria decrease in the new method. With the increase of arrival rate, *OBJ* greatly increases, whereas *Ave*, *Dev*, and

Max greatly decrease. Further, it seems that the new reservation process improves more for low demand case.

Arrival Rate	Old				New				Gap (%)
	OBJ	AVE	Dev	Max	OBJ	AVE	Dev	Max	
Low	5.06	4.70	7.47	35	4.78	4.37	3.84	27	5.41
Medium	10.55	1.72	3.78	35	10.30	1.66	1.80	14	2.41
High	14.55	1.04	2.37	35	14.40	1.02	1.28	12	1.03

TAB 3.15 Performance vs patient arrival rates

3.7 Conclusion

This chapter proposes a contract-based reservation process of MRI examinations for stroke patients in order to reduce the average waiting time of patients without degrading the utilization ratio of MRI facilities. The new method requires the determination of the number of contracted time slots and the optimal patient assignment control policy for assigning patients to either CTS or RTS. This results a stochastic combinatorial problem that combines combinatorial planning decision variables and dynamic control policies. This chapter first explores the structure properties of the optimal control policies under a given contract and then proposed a two-step approach to obtain efficient contracts. A single sample-path Monte Carlo approximation is used to determine an initial contract which is further improved through local search. Numerical results show that the deviation gap from the optimal solution is rather small which means that the contract and the corresponding control policy are very close to the optimal ones.

This chapter also proposes new strategy to avoid the “unlucky patients” who are assigned to RTS. The new strategy differs from the existing one by reserving regular time slots for the neurovascular department instead for particular patients. This allows us to avoid "unlucky" patients having to wait for much longer time than other patients arriving at the same time. Numerical results show that the new method can greatly reduce the criterion values and better the waiting time distribution.

Future research can be pursued in several directions. To reduce the unused CTS and improve the utilization ratio of MRI, one immediate extension is the development of real time control strategies for advanced cancellation of CTS in case of short CTS queue. There are several other research directions, for example, the determination of the optimal contract with non stationary patient arrival and the combination of multiple classes of patients and several examinations.

Chapter 4

Contract planning and one-day advance cancellation of contracted time slots

This chapter addresses the improvement of contract-based MRI examinations reservation process by allowing one-day advance cancellation of contracted time slots. The contract is now composed of three parts: contract decision, patient assignment policy, and one-day advance cancellation policies. The problem of CTS cancellation and patient assignment is formally formulated as an average cost Markov Decision Process in order to minimize the criterion values including the average patient waiting times, average unused CTS penalty and CTS cancellation penalty. Structure properties of the optimal control policies are established via the discounted cost problem. A local optimization algorithm is proposed to improve a given initial contract. Numerical results show that advance CTS cancellation significantly reduces the unused CTS with slight increase of patients' waiting time.

Papers relevant with this chapter: Geng et al. (2010b), Geng (2009c)

4.1 Introduction

Chapter 3 has proposed a contract-based MRI examination reservation process for stroke patients. The reservation process reduces stroke patients' waiting time but it also leads to unused time slots. To improve the utilization of MRI scanner, we explore the possibility of avoiding unused CTS by canceling CTS in advance when the CTS queue is short. In this way, the possible unused CTS can be released from the contract in advance and the imaging department can arrange other patients to have the examinations on these release time slots. Of course, the earlier a contracted time slot is released, the better the imaging department can make use of it. In this chapter, we limit ourselves to cancellation one day before. More specifically, the contract includes three decisions:

Contract decision, i.e., the number and distribution of CTS;

Patient assignment control policy, i.e., the control policy which assigns patient to either CTS or RTS;

One-day advance cancellation control policy. The neurovascular department can cancel part of CTS in the next day in order to reduce unused CTS when the CTS queue length at the end of some day is too short.

This chapter proposes an MDP approach to simultaneously identify the forms of the optimal patient assignment and one-day advance cancellation policies for each given contract. Then, local search is used to improve the contract decision.

The rest of this chapter is organized as follows: Section 4.2 provides a formal problem setting. Structure properties of the optimal control policies are established via discounted cost MDP in Section 4.3. Section 4.4 proposes local optimization algorithm. Section 4.5 analyzes results of computational experiments. Conclusions and perspectives are given in Section 4.6.

4.2 Problem setting

This section presents an MDP formulation to determine the optimal patient assignment policy and one-day advance CTS cancellation policy. Assumptions 1 to 4 presented in Chapter 1 are made. Further, the following assumption is made throughout the chapter:

Assumption 4-A1: CTS of day t can be cancelled in advance of one day, i.e., at the end of day $t-1$.

Based on the above assumptions, the problem of patient assignment and advance CTS cancellation can be characterized by the following notation:

Indices:

t: index of days, $t=1,\ldots,T$;

i: index of days in one week, $i=1,\ldots,7$, i.e., Monday, …, Sunday; note that the day $i\pm j$ is the weekday of j days after or before day i in one week;

$d(t)$: the day in the week corresponding to day t with $d(t)=t(\text{mod})7+1\in\{1,\ldots,7\}$;

Data:

T^R: average number of days for a patient to have his/her MRI examination through regular reservation with $T^R > 1$;

c: penalty factor of an unused CTS. It serves as a weighting factor in order to balance average waiting time and unused MRI time slots;

b: penalty factor of canceling one CTS with $b < c$;

a_t: number of patients arrived in day t. By assumption 3, daily arrivals a_t for $t \in IN$ are mutually independent random variables and weekly arrivals $(a_{7j+1}, a_{7j+2}, \ldots, a_{7j+7})$ are identically distributed for all j = 0, 1, …. As a result, the arrival process is characterized by probability matrix $\boldsymbol{P} = [P_{ij}]$ for $i = 1, \ldots, 7$ and for all $j \geq 0$ with P_{ij} denoting the probability of j arrivals in day i;

n_t: number of CTS in day t;

Decision variables are:

x_t: number of patients waiting for CTS at the end of day t, which is also called CTS queue, x_0 is a given constant. Note that x_t does not include patients that are directed to RTS.

w_t: number of CTS for day t+1 cancelled at the end of day t; w_0 is known;

State variable:

$z_t = w_{t-1} + x_{t-1} + a_t$: number of CTS to be consumed in day t.

The sequence of events during each day t is as follows. First, the queue length x_{t-1} for CTS is known at the end of day t–1, and the number w_{t-1} of CTS canceled for day t is determined. During day t, the number a_t of new incoming patients during the day becomes known. The state variable z_t is known. MIN{ $n_t - w_{t-1}$, $x_{t-1} + a_t$} patients are served by the remaining CTS of the day and MAX{0, $n_t - z_t$ } CTS of day t cannot be filled. x_t patients will wait for CTS of the subsequent days. MAX{0, $z_t - n_t - x_t$ } remaining patients are directed to RTS and will have the MRI examination after an average of T^R days.

There are two control policies in this problem: CTS cancellation and patient assignment policies. History-dependent policies are considered in this chapter. Let $h_t = (z_i, x_i, w_i, \ldots, z_{t-1}, x_{t-1}, w_{t-1}, z_t)$ be the full history stating from initial state z_i at the beginning of day i. We denote the patient assignment policy by $\boldsymbol{\pi}=\{\pi_1,\pi_2,...\}$, where the CTS queue length at the end of day t is $x_t = \pi_t(h_t)$ with $0 \leq x_t \leq (z_t - n_t)^+$, and the CTS cancellation policy by $\boldsymbol{\mu}=\{\mu_1, \mu_2,\ldots\}$, where the number of CTS cancelled for day t+1 at the end of day t is $w_t = \mu_t(h_t)$ with $0 \leq w_t \leq (n_{t+1} - x_t)^+$.

The objective is to minimize over all history-dependent policies $\boldsymbol{\mu}=\{\mu_1, \mu_2,\ldots\}$ and $\boldsymbol{\pi}=\{\pi_1,\pi_2,...\}$ the long-run average cost

$$J_{\mu\pi}(i,z) = \lim_{T\to\infty} \frac{1}{T} E\left[\sum_{t=i}^{T+i-1} g_{d(t)}(z_t, x_t, w_t) \middle| z_i = z \right] \tag{4.1}$$

for any given initial state $z_i = z$ with $i = 1, \ldots 7$ where

$$g_{d(t)}(z_t, x_t, w_t) = bw_t + c\left(n_{d(t)} - z_t\right)^+ + x_t + T^R\left(z_t - n_{d(t)} - x_t\right)^+$$

is the stage cost, i.e., CTS cancellation cost plus unused CTS penalty plus waiting time for CTS and RTS. In the following, $g_{d(t)}(.)$ are written as $g_t(.)$ for convenience.

Theorem 4-1: There exists an optimal average cost policy such that $x_t \le \bar{x}$ for all $t > 0$ with $\bar{x} = \lceil (T^R + c) n^* \rceil$ where $n^* = \text{MAX}\{n_1, \ldots, n_7\}$.

Proof: This proof is similar with Theorem 3-1. Q.E.D.

Thanks to Theorem 4-1, we can make without loss of generality the following assumption:

Assumption 4-A2: $x_t \le \bar{x}$ for all $t > 0$.

4.3 Properties of optimal control policies

4.3.1 Discounted cost problem

According to relation (4.1), the corresponding discounted cost MDP is as follows:

$$J_{\alpha,\mu\pi}(i,z) = \lim_{T\to\infty} E\left[\sum_{t=i}^{T} \alpha^{t-i} g_t(z_t, x_t, w_t) \middle| z_i = z\right] \tag{4.2}$$

for any given initial state $z_i = z$ with $i = 1, \ldots 7$ with discount factor α such that $0 < \alpha < 1$. Consider the following optimal cost function

$$U_\alpha(i,z) = \underset{\mu\pi}{MIN}\, J_{\alpha,\mu\pi}(i,z) \tag{4.3}$$

In the remaining, for simplicity, the notation α is omitted in this subsection where only discounted cost problem with a given α is considered.

Theorem 6.10.4 in Puterman (1994) is used to established the optimality equation. It will be shown in Remark 4-1 that all conditions needed for application of Theorem 6.10.4 are satisfied. Since the set of states (i, z) is countable and the control constraint set is finite as $x_t \le z_t$ and $w_t \le n_{t+1}$ for each z_t, Theorem 6.10.4 in Puterman (1994) implies that the optimal cost function is the unique solution of the following optimality equation:

$$U(i,z) = \min_{x_i, w_i}\left\{ g_i(z_i, x_i, w_i) + \alpha \sum_a P_{d(i+1),a} U(i+1, x_i + w_i + a) \right\}, \forall i = 1, \ldots, 7 \tag{4.4}$$

where the optimal control policy is stationary deterministic and is given by the argument ***w*** and ***x*** that reach the minimum in (4.4) and the optimal cost function is the limiting function of the following value iteration:

$$U^t\left(z_t\right)=\min_{x_t,w_t}\left\{g_t\left(z_t,x_t,w_t\right)+\alpha\sum_a P_{t+1,a}U^{t+1}\left(x_t+w_t+a\right)\right\} \tag{4.5}$$

$$U^0\left(z\right)=0 \tag{4.6}$$

for t = 0, −1, − 2, … where $0\le x_t\le(z_t-n_t)^+$, $0\le w_t\le(n_{t+1}-x_t)^+$, $n_t=n_{d(t)}$ and $P_{t+1,a}=P_{d(t+1),a}$ are shorthand notation with $d(t)$ denoting the corresponding day with $d(0)=7$, $d(-1)=6$, … As a result,

$$U\left(i,z\right)=\lim_{n\to\infty}U^{-7n+i}\left(z\right) \tag{4.7}$$

Since

$$g_t\left(z_t,x_t,w_t\right)=c\left(n_t-z_t\right)^++T^R\left(z_t-n_t\right)^++bw_t-\left(T^R-1\right)x_t,\forall 0\le x_t\le\left(z_t-n_t\right)^+,$$

relation (4.5) can be rewritten as

$$U^t\left(z_t\right)=c\left(n_t-z_t\right)^++T^R\left(z_t-n_t\right)^++\min_{x_t\in\left[0,\left(z_t-n_t\right)^+\wedge\bar{x}\right]}\left\{V^t\left(x_t\right)-\left(T^R-1\right)x_t\right\} \tag{4.8}$$

$$V^t\left(x_t\right)=\min_{w_t\in\left[0,\left(n_{t+1}-x_t\right)^+\right]}\left\{bw_t+\alpha\sum_a P_{t+1,a}U^{t+1}\left(x_t+w_t+a\right)\right\} \tag{4.9}$$

$U^t(z_t)$ is used to explore the optimal patient assignment policy depending on state variable z_t, whereas $V^t(x_t)$ is used to establish the optimal one-day advance CTS cancellation control policy depending on state variable x_t.

Similarly, relation (4.4) can be rewritten as

$$U\left(i,z\right)=c\left(n_i-z\right)^++T^R\left(z-n_i\right)^++\min_{x\in\left[0,\left(z-n_i\right)^+\wedge\bar{x}\right]}\left\{V\left(i,x\right)-\left(T^R-1\right)x\right\} \tag{4.10}$$

$$V\left(i,x\right)=\min_{w\in\left[0,\left(n_{i+1}-x\right)^+\right]}\left\{bw+\alpha\sum_a P_{i+1,a}U\left(i+1,x+w+a\right)\right\} \tag{4.11}$$

By relation (4.7) and the uniqueness of the optimal value function,

$$V\left(i,x\right)=\lim_{n\to\infty}V^{-7n+i}\left(x\right) \tag{4.12}$$

Remark 4-1: The optimality equation (4.4) requires Assumption 6.10.1 and condition (6.10.11) of Puterman (1994). Assumption 6.10.1 holds as $g_t(z_t, x_t, w_t)\le W(z_t)\equiv(b+c)n^*+T^Rz_t$. Condition (6.10.11) holds as:

$$\sum_a P_{i+1,a}W\left(x_t+w_t+a\right)\le\left(b+c\right)n^*+T^RE\left[x_t+w_t+a_{t+1}\right]$$
$$\le W\left(z_t\right)+\left(b+c\right)n^*+T^R\bar{x}+T^Ra^*$$

where a^* = MAX{E[a_1], …, E[a_7]}. The above relation holds as $x_t + w_t \le \max(x_t, n_{t+1})$ and, by assumption 4-A2, $x_t + w_t \le \bar{x}$.

Property 4-1: In the value iteration by relation (4.5) or equivalently (4.8)-(4.9), $-c \le U^t(z_t) - U^t(z_t - 1) \le T^R$.

Proof: The proof is done by induction on t. The property trivially holds for t=0. Assume that it holds for some t+1≤0 and consider day t. Let (x_t^1, w_t^1) and (x_t^2, w_t^2) be arguments reaching minimum in relation (4.5) or equivalently (4.8)-(4.9) for $U^t(z_t)$ and $U^t(z_t-1)$. From the optimality condition, $x_t^1 \in \left[0, (z_t - n_t)^+ \wedge \bar{x}\right]$, $x_t^2 \in \left[0, (z_t - 1 - n_t)^+ \wedge \bar{x}\right]$.

Taking $(x_t, w_t) = (x_t^2, w_t^2)$ as the feasible control for $U^t(z_t)$ leads to $U^t(z_t) - U^t(z_t - 1) \le T^R$ and proves the right relation of the property.

To prove the left side, the relation trivially holds if $(z_t-n_t)^+=0$ as $U^t(z_t) - U^t(z_t-1) = -c$. Otherwise, two cases are possible.

Case 1: $0 \le x_t^1 < (z_t - n_t)^+$, taking $(x_t, w_t) = (x_t^1, w_t^1)$ as the feasible control for $U^t(z_t - 1)$ gives

$$U^t(z_t) - U^t(z_t - 1) \ge T^R$$

Case 2: $x_t^1 = (z_t - n_t)^+ > 0$, taking $(x_t, w_t) = (x_t^1 - 1, w_t^1)$ as the feasible control for $U^t(z_t - 1)$, then

$$U^t(z_t) - U^t(z_t - 1) \ge 1 + \alpha \sum_a P_{t+1,a} \left(U^{t+1}(x_t^1 + w_t^1 + a) - U^{t+1}(x_t^1 + w_t^1 + a - 1) \right).$$

By induction assumption, $U^t(z_t) - U^t(z_t - 1) \ge 1 - \alpha c \ge -c$.

Q.E.D.

Property 4-2: In the value iteration by (4.8)-(4.9), $-b \le V^t(x_t) - V^t(x_t - 1) \le T^R$ for any x_t.

Proof: The proof is done by induction on t. Since $U^0(z) = 0$, $V^{-1}(x_t)$=0 and the property holds for t = −1. Assume that the property holds for t+1 < 0 and consider t. Let w_t^1 and w_t^2 be the minimizing argument in relation (4.9) for $V^t(x_t)$ and $V^t(x_t - 1)$. By definition, $w_t^1 \in \left(0, (n_{t+1} - x_t)^+\right)$, $w_t^2 \in \left(0, (n_{t+1} - x_t + 1)^+\right)$.

Two cases are considered to prove the left hand relation of the property.

Case 1: $0 \le x_t - 1 < x_t \le n_{t+1}$. Take $w_t = w_t^1 + 1$ as the feasible control for $V^t(x_t - 1)$, $V^t(x_t) - V^t(x_t - 1) \ge -b$.

Case 2: $x_t > x_t - 1 \ge n_{t+1}$. Hence, $w_t^1 = w_t^2 = 0$ and

$$V^t(x_t) - V^t(x_t - 1) = \alpha \sum_a P_{t+1,a}\left(U^{t+1}(x_t + a) - U^{t+1}(x_t - 1 + a)\right)$$

It is enough to prove $U^{t+1}(x_t + a) - U^{t+1}(x_t - 1 + a) \ge -b$ for all *a*. Let x_{t+1}^1 and x_{t+1}^2 be minimizing argument in relation (4.8) for $U^{t+1}(x_t + a)$ and $U^{t+1}(x_t - 1 + a)$. If $0 \le x_{t+1}^1 < x_t + a - n_{t+1}$, taking $x_{t+1} = x_{t+1}^1$ as the feasible control for $U^{t+1}(x_t - 1 + a)$ in relation (4.8) gives

$$U^{t+1}(x_t + a) - U^{t+1}(x_t - 1 + a) \ge T^R \ge -b$$

If $x_{t+1}^1 = x_t + a - n_{t+1}$, taking $x_{t+1} = x_{t+1}^1 - 1$ as the feasible control for $U^{t+1}(x_t - 1 + a)$ in (4.8) gives

$$U^{t+1}(x_t + a) - U^{t+1}(x_t - 1 + a) \ge 1 + \left(V^{t+1}(x_{t+1}^1) - V^{t+1}(x_{t+1}^1 - 1)\right) \ge -b .$$

We now prove the right hand relation. Two cases are considered:

Case 1: $0 \le x_t - 1 < x_t \le n_{t+1}$ and $w_t^2 > 0$, taking $w_t = w_t^2 - 1$ for $V^t(x_t)$ leads to $V^t(x_t) - V^t(x_t - 1) \le -b$.

Case 2: $w_t^2 = 0$. Taking $w_t = 0$ for $V^t(x_t)$ and combining with Property 4-1 lead to

$$V^t(x_t) - V^t(x_t - 1) = \alpha \sum_a P_{t+1,a}\left(U^{t+1}(x_t + a) - U^{t+1}(x_t + a - 1)\right) \le \alpha T^R$$

This completes the proof. Q.E.D.

Property 4-3: In the value iteration by (4.5) or equivalently (4.8)-(4.9), $U^t(z_t)$ is convex in z_t and $V^t(x_t)$ is convex in x_t.

Proof: This proof is made by induction on *t*. First $U^0(z) = 0$ and is hence convex in *z*. Assume that $U^{t+1}(z)$ is convex in *z*.

We first prove that $V^t(x_t)$ is convex in x_t and show the form of the optimal CTS cancellation policy. Let us rewrite relation (4.9) as follows:

$$V^t(x_t) = -bx_t + \min_{x_t + w_t \in [x_t, \max(n_{t+1}, x_t)]} R^t(x_t + w_t)$$

where

$$R^t(y) = by + \alpha \sum_a P_{t+1,a} U^{t+1}(y+a)$$

By induction assumption, $R^t(y)$ is convex in y. Let S_{t+1} be the smallest argument minimizing $R^t(y)$, i.e.

$$S_{t+1} = \arg\min_{y \geq 0} by + \alpha \sum_a P_{t+1,a} U^{t+1}(y+a) \tag{4.13}$$

We show by contradiction that $S_{t+1} \leq n_{t+1}$. If $S_{t+1} > n_{t+1}$, $V^t(S_{t+1}) = -bS_{t+1} + R^t(S_{t+1})$ and $V^t(S_{t+1}-1) = -b(S_{t+1}-1) + R^t(S_{t+1}-1)$. Hence

$$V^t(S_{t+1}) - V^t(S_{t+1}-1) = -b + \left(R^t(S_{t+1}) - R^t(S_{t+1}-1)\right) < -b$$

which contradicts Property 4-2 and prove $S_{t+1} \leq n_{t+1}$.

Since $S_{t+1} \leq n_{t+1}$,

$$V^t(x_t) = \begin{cases} -bx_t + R^t(S_{t+1}), & \forall x_t \leq S_{t+1} \\ -bx_t + R^t(x_t), & \forall x_t > S_{t+1} \end{cases}$$

which can be easily shown to be convex in x_t. The optimal CTS cancellation is characterized by a single threshold S_{t+1}. If the ending CTS queue length in day t is below this threshold, then cancel enough CTS in order for x_t+w_t to reach this threshold. Otherwise, no CTS is cancelled.

We then prove that $U^t(z_t)$ is convex in z_t and show the form of the optimal patient assignment policy. First $V^t(x_t) - (T^R - 1)x_t$ is also convex. Let

$$L_t = \arg\min_{0 \leq x_t \leq \left((z_t - n_t)^+ \wedge \bar{x}\right)} \left(V^t(x_t) - (T^R - 1)x_t\right) \tag{4.14}$$

Equation (4.8) can be written as

$$U^t(z_t) = \begin{cases} c(n_t - z_t) + V^t(0) & \text{if } z_t \leq n_t \\ z_t - n_t + V^t(z_t - n_t) & \text{if } n_t < z_t \leq L_t + n_t \\ L_t + T^R(z_t - n_t - L_t) + V^t(L_t) & \text{if } z_t > L_t + n_t \end{cases}$$

The optimal patient assignment policy is as follows. No patient is sent to RTS if the resulting CTS queue length at the end of day is below the threshold L_t. Otherwise, some patients are sent to RTS to keep the CTS queue length at L_t.

$U^t(z_t)$ is convex in z_t in the following internal $[0, n_t)$, (n_t, L_t+n_t), and $(L_t+n_t, +\infty)$. We still need to prove the convexity of $U^t(z_t)$ for $z_t = n_t$ and $z_t=L_t +n_t$. The convexity of $U^t(z_t)$ in z_t holds when $z_t= n_t$, because $U^t(z_t) - U^t(z_t-1) = -c$ and from Property 4-1, $U^t(z_t) - U^t(z_t-1)\geq -c$. The convexity of U^t (z_t) in z_t holds when $z_t= L_t+n_t$, because $U^t(z_t)- U^t(z_t-1)=T^R$ and from Property 4-1, $U^t(z_t) - U^t(z_t-1)\leq T^R$. Therefore, $U^t(z_t)$ is convex in z_t. Q.E.D.

Theorem 4-2: The optimal value functions $U(i, z)$ and $V(i, x)$ in relation (4.10) and (4.11) are convex in z and x respectively. Further, the optimal control policy for problem (4.3) is of the following form:

1) The optimal CTS queue length at the end of day i is:

$$x_i^* = \begin{cases} 0 & \text{if } z_i - n_i \leq 0 \\ z_i - n_i & \text{if } 0 \leq z_i - n_i \leq L_i \\ L_i & \text{if } z_i - n_i \geq L_i \end{cases} \tag{4.15}$$

where $L_i = \arg\min_{0\leq x\leq \bar{x}} \left(V\left(i,x\right)-\left(T^R-1\right)x\right)$.

The optimal number of patients assigned to RTS at the end of day i is:

$$y_i^* = \left(z_i - n_i - x_i^*\right)^+$$

2) The optimal number of CTS cancelled at the end of day i is:

$$w_i^* = \begin{cases} S_{i+1} - x_i & \text{if } x_i \leq S_{i+1} \\ 0 & \text{if } x_i \geq S_{i+1} \end{cases} \tag{4.16}$$

where $S_{i+1} = \arg\min_{y\geq 0} by + \alpha \sum_a P_{i+1,a} U\left(i+1, y+a\right)$.

Proof: The convexity of $U(i, z)$ and $V(i, x)$ is a direct consequence of relations (4.7), (4.12), and Property 4-3. Note that, as a result of relations (4.7) and (4.12), Properties 4-1, 4-2 also hold for $U(i, z)$ and $V(i, x)$. The form of the optimal control policy can be proved as in the proof of Property 4-3. Q.E.D.

4.3.2 Average cost problem

4.3.2.1 Bounded demand case

The following assumption is also made for the average cost problem case.

Assumption 4-A3. There exists a finite number A such that $a_t\leq A$, for all t.

Remark 4-2: The assumption is not restrictive in practice as A can be chosen large enough.

The combination of Assumptions 4-A2 and 4-A3 implies that the state variable z_t is upper bounded and:

$$z_t \le \overline{z} \equiv \overline{x} + A \tag{4.17}$$

as $x_t + w_t \le \max(x_t, n_{t+1}) \le \overline{x}$. As a result, under Assumptions 4-A2 and 4-A3, the stage cost function is also bounded with

$$g_t\left(z_t, x_t, w_t\right) \le B \equiv \left(b+c\right)n^* + T^R \overline{z} \tag{4.18}$$

Property 4-4: There exists M>0 such that $\left|U_\alpha\left(i,z\right) - U_\alpha\left(7,0\right)\right| \le M$, for all i = 1, …,7 and for all x and z.

Proof: From Property 4-1,

$$-c \le U_\alpha^t\left(z_t\right) - U_\alpha^t\left(z_t - 1\right) \le T^R$$

which, together with the finiteness of the state space,

$$-C\overline{z} \le U_\alpha^t\left(z_t\right) - U_\alpha^t\left(z_t^{'}\right) \le C\overline{z}$$

for all z_t, z'_t with C = Max(T^R, c). Combining with relation (4.7),

$$-C\overline{z} \le U_\alpha\left(i,z\right) - U_\alpha\left(i,z'\right) \le C\overline{z} \tag{4.19}$$

This establish the property for i = 7. Consider now the case i = 1, …, 6. From (4.18),

$$U_\alpha\left(i,z\right) \le B + E_{a_i}\left[U_\alpha\left(i+1,z\right)\right] \tag{4.20}$$

Repeat the relations (4.20) for t subsequent days leads to:

$$U_\alpha\left(i,z\right) \le tB + \sum_{z'} p_{(i,z),(t+i,z')}^{\mu\pi} U_\alpha\left(t+i,z'\right), \forall i = 1,...,7 \tag{4.21}$$

where $p_{(i,z),(t+i,z')}^{\pi}$ is the probability of reaching state z' at the beginning of day t+i by starting from state z at day i under policies μ and π. Combining (4.19) and (4.21) with t+i=7,

$$U_\alpha\left(i,z\right) \le 6B + \sum_{z'} p_{(i,z),(7,z')}^{\pi} U_\alpha\left(7,z'\right) \le 6B + U_\alpha\left(7,0\right) + C\overline{z} \tag{4.22}$$

Similarly,

$$U_\alpha\left(7,0\right) \le 6B + \sum_{z'} p_{(7,0),(7+i,z')}^{\pi} U_\alpha\left(i,z'\right) \le 6B + U_\alpha\left(i,z\right) + C\overline{z} \tag{4.23}$$

Relations (4.22)-(4.23) conclude the proof with $M = 6B + C\overline{z}$. Q.E.D.

Theorem 4-3. There exists an optimal stationary control policy, the same with those in Theorem 4-2, for the average cost model (4.1). Further the optimal average cost is independent of the initial state (i, z).

Proof. From Proposition 4.2.6 in Bertsekas (1996), the optimal average cost per day exists and has the same value λ for all initial states, and λ satisfies

$$\lambda = \lim_{\alpha\to 1}(1-\alpha)U_\alpha(i,z) \tag{4.24}$$

The differential cost functions

$$H(i,z) = \lim_{\alpha\to 1}\left(U_\alpha(i,z) - U_\alpha(7,0)\right) \tag{4.25}$$

satisfy the following optimality equations:

$$\lambda + H(i,z) = \min_{x,w}\left\{g_i(z,x,w) + \sum_a P_{i+1,a}H(i+1,x+w+a)\right\}, \forall i = 1,...,7 \tag{4.26}$$

Relation (4.26) can be rewritten as

$$H(i,z) = \min_{x\le\left((z-n_i)^+\wedge\bar{x}\right)}\left\{c(n_i-z)^+ + x + T^R(z-n_i-x)^+ + F(i,x)\right\}, \forall i = 1,...,7 \tag{4.27}$$

$$F(i,x) \equiv \min_{w\le(n_{i+1}-x)^+}\left\{bw + \sum_a P_{i+1,a}H(i+1,x+w+a)\right\} - \lambda, \forall i = 1,...,7 \tag{4.28}$$

Relations(4.25), (4.28) and (4.11) implies that

$$F(i,x) = \lim_{\alpha\to 1}\left(V_\alpha(i,x) - U_\alpha(7,0)\right) \tag{4.29}$$

Further, the optimal control policy is stationary deterministic and is defined by the argument that reaches the minimum in (4.26) or equivalently (4.27)-(4.28). From Theorem 4-2 and (4.25) and (4.29), $H(i, z)$ is convex in z and $F(i, x)$ is convex in x for all i = 1, …, 7. The optimal control policy is the same with that of Theorem 4-2. Q.E.D.

4.3.2.2 Unbounded demand case

This subsection relaxes assumption 4-A3.

Property 4-5: For any z ≥ z'≥0, $-m \le U_\alpha^t(z) - U_\alpha^t(z') \le m + T^R z$ with $m = (c+b+T^R)\bar{x}$.

Proof: The property trivially holds for t = 0. Consider the case t < 0. Since the "min" term in equation (4.8) is decreasing in z_t, subtracting equation (4.8) with $z_t = z$ by equation (4.8) $z_t = z'$ leads to:

$$U_\alpha^t(z)-U_\alpha^t(z')\le c(n_t-z)^+ +T^R(z-n_t)^+ -c(n_t-z')^+ -T^R(z'-n_t)^+$$
$$\le cn^* + T^R z \le m + T^R z$$

as $n^* \le \bar{x}$. Let *x'* be the argument reaching minimum in (4.8) with $z_t = z'$. As a result,

$$U_\alpha^t(z') = c(n_t-z')^+ + T^R(z'-n_t)^+ + V_\alpha^t(x') - (T^R-1)x' \tag{4.30}$$

Subtracting equation (4.8) with $z_t = z$ by equation (4.30) leads to:

$$\begin{aligned} U_\alpha^t(z)-U_\alpha^t(z') &= c(n_t-z)^+ +T^R(z-n_t)^+ -c(n_t-z')^+ -T^R(z'-n_t)^+ \\ &+ \min_{x\in\left[0,(z-n_t)^+\wedge\bar{x}\right]}\left\{V_\alpha^t(x)-V_\alpha^t(x')-(T^R-1)(x-x')\right\} \\ &\ge -cn^* + \min_{x\in\left[0,(z-n_t)^+\wedge\bar{x}\right]}\left\{V_\alpha^t(x)-V_\alpha^t(x')-(T^R-1)(x-x')\right\} \end{aligned} \tag{4.31}$$

For any $x > x'$, from Property 4-2,

$$V_\alpha^t(x)-V_\alpha^t(x')-(T^R-1)(x-x') \ge -bx - T^R x \tag{4.32}$$

For any $x \le x'$, from Property 4-2,

$$V_\alpha^t(x')-V_\alpha^t(x) \le (T^R-1)(x'-x) \tag{4.33}$$

Combining relations (4.31)-(4.32) leads to:

$$U_\alpha^t(z)-U_\alpha^t(z') \ge -cn^* - b\bar{x} - T^R\bar{x} \ge -(c+b+T^R)\bar{x}\,.$$

Q.E.D.

Property 4-6: There exist *M*>0 and r > 0 such that $-M \le U_\alpha(i,z)-U_\alpha(7,0) \le M + rz$, for all $i = 1, \ldots, 7$ and for all z.

Proof: From Property 4-5,

$$-m \le U_\alpha^t(z)-U_\alpha^t(z') \le m + T^R z, \forall z \ge z'.$$

Combining with relation (4.7),

$$-m \le U_\alpha^t(i,z)-U_\alpha^t(i,z') \le m + T^R z, \forall z \ge z'$$

This establish the property for $i = 7$. Further

$$U_\alpha^t(i,z) \le m + T^R z + U_\alpha^t(i,z'), \forall z, z' \tag{4.34}$$

Consider now the case $i = 1, \ldots, 6$. First,

$$0 \le g_t(z_t,x_t,w_t) = c(n_t-z_t)^+ + x_t + T^R(z_t-n_t-x_t)^+ + bw_t \le (b+c)n^* + T^R z_t \le m + T^R z_t\,.$$

From (4.4),

$$U_\alpha(i,z) \le m + T^R z + E\left[U_\alpha(i+1, z_{i+1})\right] \tag{4.35}$$

Repeat the relations (4.35) for t subsequent days leads to:

$$U_\alpha(i,z) \le \sum_{\tau=i}^{t+i-1}\left(m + T^R E[z_\tau]\right) + E\left[U_\alpha(t+i, z_{t+i})\right], \forall i = 1,...,7 \tag{4.36}$$

Combining (4.34) and (4.36) with $t+i=7$,

$$U_\alpha(i,z) \le \sum_{\tau=i}^{7}\left(m + T^R E[z_\tau]\right) + \left(U_\alpha(7,0) + m + T^R(z+6a^*)\right), \forall i = 1,...,7 \tag{4.37}$$

Since $z_t \le z + a_{i+1} + w_i + \ldots + a_t + w_{t-1} \le z + a_{i+1} + n_{i+1} + \ldots + a_t + n_t$, for $t > i$,

$$\begin{aligned} U_\alpha(i,z) &\le 7m + 49T^R a^* + 7T^R z + m + T^R(z+6a^*) + U_\alpha(7,0) \\ &\le 8m + 55T^R a^* + 8T^R z + U_\alpha(7,0), \forall i = 1,...,7 \end{aligned} \tag{4.38}$$

Similarly,

$$U_\alpha(7,0) \le \sum_{\tau=7}^{7+i}\left(m + T^R E[z_\tau]\right) + \left(U_\alpha(7+i,z) + m + 6T^R a^*\right) \le 8m + 55T^R a^* + U_\alpha(i,z) \tag{4.39}$$

Relations (4.38)-(4.38) conclude the proof with $M = 8m + 55T^R a^*$ and $r = 8T^R$. Q.E.D.

Theorem 4-4. Under Assumptions 4, and 4-A2, (a) there exists a constant λ satisfying (4.24) for all (i, z), a matrix $H(i, z)$ satisfying (4.25)-(4.26), (b) the optimal control policy is defined by the argument that reaches the minimum in (4.26), (c) there exists an optimal stationary control policy of the form of equations (4.15)-(4.16)for the average cost model.

Proof: The proof is based on Theorem 8.10.7 of Puterman (1994) and the conditions that need to be checked are the following ones:

C1: For each state (i, z), the stage cost is such that $-\infty < R \le g_i(z, x_i) < \infty$.

C2: For each (i, z) and $\alpha < 1$, $U_\alpha(i,z) < \infty$.

C3: There exists $K > -\infty$ such that, for each (i, z),

$$H_\alpha(i,z) \equiv U_\alpha(i,z) - U_\alpha(7,0) \ge K, \forall \alpha < 1.$$

C4: There exists a non-negative function W(i, z) such that

a) $W(i, z) < \infty$;

b) for each (i, z), $H_\alpha(i,z) \le W(i,z), \forall \alpha < 1$; and

c) for each (i, z) and x_i,

$$\sum_a P_{i+1,a} W\left(i+1, x_i + a\right) < \infty.$$

According to Theorem 8.10.7 of Puterman (1994), as the control constraint set for each state (i, z) is finite, (a) and (b) of the Theorem hold. Further $H(i, z)$ is the limit of a sequence $H_{\alpha_m}(i,z)$ such that α_m converges to 1 and $H_{\alpha_m}(i,z)$ converges for all (i, z). From Property 4-3, equations (9), $H(i, z)$ is convex in z and (c) of the Theorem can be proved as for Theorem 4-2.

Let us now prove conditions C1-C4. Condition C1 clearly holds as $g_i(z, x_i) \geq 0$. Condition C2 holds as well as

(iv)as in Property 4-6, $0 \leq g_t\left(z_t, x_t, w_t\right) \leq m + T^R z_t$

(v) $E\left[g_t\left(z_t, x_t, w_t\right)\right] \leq E\left[m + T^R z_t\right] \leq m + T^R\left(\bar{x} + a^*\right)$;

(vi) $V_\alpha\left(i,z\right) \leq \left(c+b\right)n^* + T^R z + \dfrac{\alpha}{1-\alpha}\left(m + T^R\left(\bar{x} + a^*\right)\right)$.

Condition C3 is guaranteed by Property 4-6 with K = -M. Condition C4 is a consequence of Property 4-6 with W(i, z) = M + rz. Q.E.D.

4.3.3 Computation and implementation of the optimal control policies

For any given contract n, as proved in Theorem 4-3 and 4-4, there exit the same optimal control policies for average-cost MDP and discounted-cost MDP. The related optimal control policies $\pi(\mathbf{n})$ and $\mu(\mathbf{n})$ can be determined by solving the following LP model:

$$J\left(\mathbf{n}\right) \equiv \text{maximize } \lambda$$

Subject to

$$H\left(i, z_i\right) \leq c\left(n_i - z_i\right)^+ + x_i + T^R\left(z_i - n_i - x_i\right)^+ + F\left(i, x_i\right), \forall i = 1, \ldots, 7$$

$$\lambda + F\left(i, x_i\right) \leq b w_i + \sum_a P_{i+1,a} H\left(i+1, x_i + w_i + a\right), \forall i = 1, \ldots, 7$$

$$\forall x_i \leq \left(z_i - n_i\right)^+ \wedge \bar{x}, \forall w_i \leq \left(n_{i+1} - x_i\right)^+, \forall i = 1, \ldots, 7$$

Where J(**n**) is the optimal average cost for problem (4.1) under contract **n**. The optimal controls are respectively given by x and w reaching equality in the above relations. Further, the optimal control is characterized by two control threshold vectors **L** and **S** for patient assignment and CTS cancellation. From relations x_i=min(L_i,$(z_i$-$n_i)^+$), and w_i=max ((S_{i+1}-$x_i)^+$,0), the optimal control thresholds **L** and **S** can be easily determined.

The existence of optimal control policies makes the implementation easy. For day t, the implementation of the optimal patient assignment policy depends on state variable z_t, while that of the optimal one-day advance cancellation policy depends on x_t, the CTS queue length at the end of day t. Therefore, patient assignment is first made, then CTS is cancelled for next day.

The implementation of the optimal patient assignment control policy can be divided into three cases:

Case 1: As shown in Fig. 4.1, if state variable z_t is smaller than n_t, then there exists the number n_t–z_t of unused CTS, and no patients waiting for the incoming time slots.

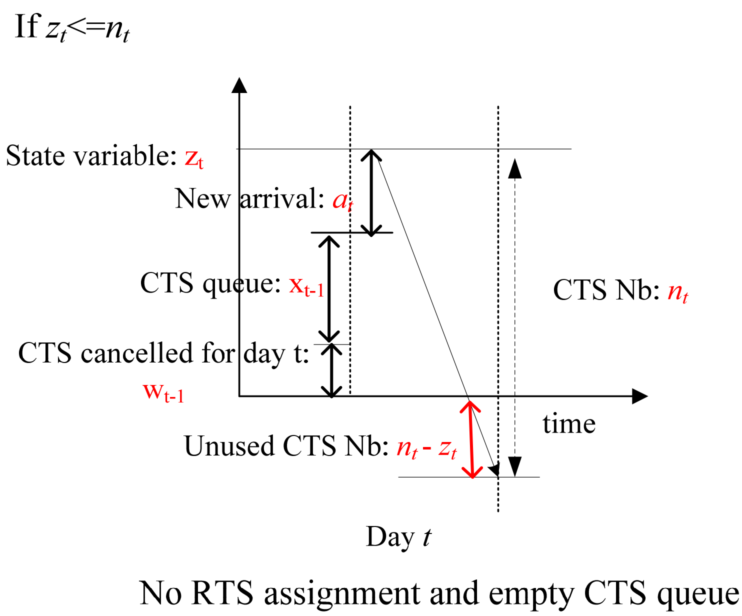
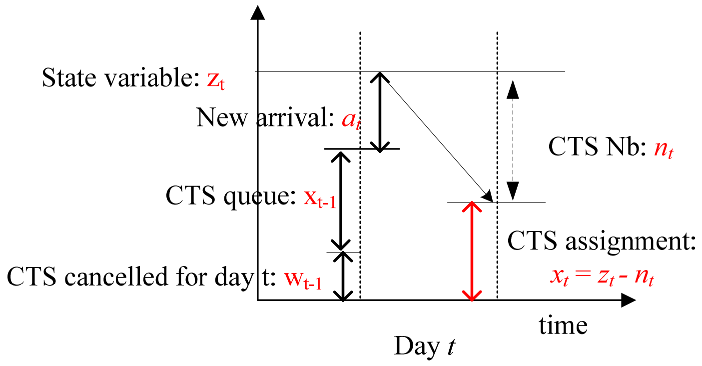


FIG. 4.1 The optimal patient assignment control if z_t <=n_t

Case 2: As shown in Fig. 4.2, if state variable z_t is greater than n_t but smaller than L_t + n_t, then all the remaining patients are kept in the CTS queue and no patients are assigned to RTS.

If n_t=<z_t<=L_t+n_t



No patients are assigned to RTS.

FIG. 4.2 The optimal patient assignment control if n_t =<z_t <= L_t +n_t

Case 3: As shown in Fig. 4.3, if state variable z_t is greater than $L_t + n_t$, then the CTS queue length is L_t, and all the other remaining patients are assigned to RTS.

If z_t>=L_t+n_t

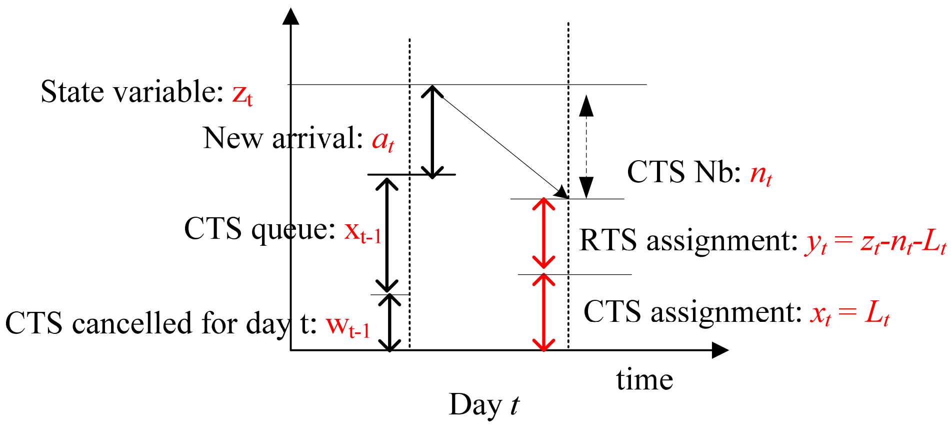


FIG. 4.3 The optimal patient assignment control if z_t >= L_t +n_t

The implementation of the optimal one-day advance cancellation control policy depends on the ending CTS queue at the end of the same day, which can be divided into two cases:

Case 1: As shown in Fig. 4.4, if the ending CTS queue at day t is smaller than S_{t+1}, then the number of CTS cancelled for day t+1 is $w_t = S_{t+1} - x_t$.

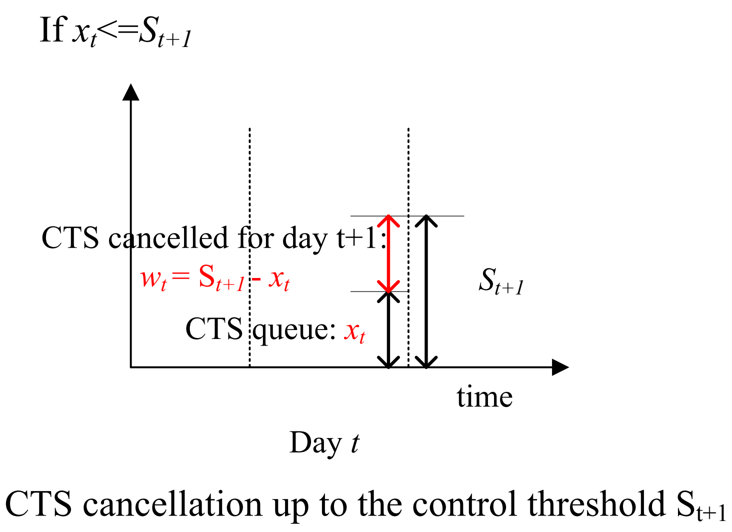


FIG. 4.4 The optimal one-day advance CTS cancellation control if $x_t <= S_{t+1}$

Case 2: As shown in Fig. 4.5, if the ending CTS queue at day t is greater than S_{t+1}, then there is no cancellation.

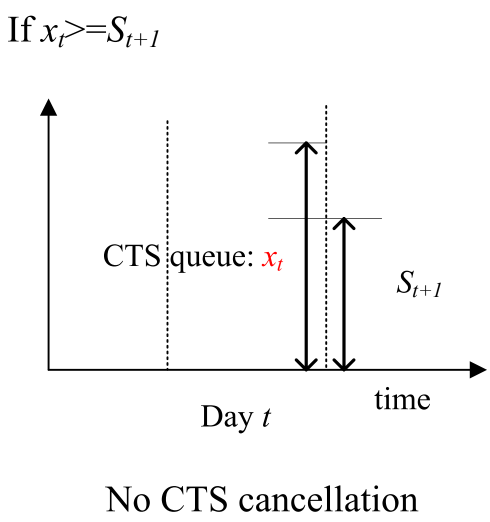


FIG. 4.5 The optimal one-day advance CTS cancellation control if $x_t <= S_{t+1}$

Remark 3-5 can be applied to here. The implementation of this contract-based MRI reservation process can be directly applied to the hospital combined with some patient scheduling policy. However, to improve its performance, more work should be done about patients scheduling in order to reduce the variance of patients waiting times.

4.4 Local Optimization

Starting from a given initial contract, this section presents a local search for improving the contract decisions by taking into account both patient assignment and one-day advance CTS cancellation policies. This local search relies on the structure properties of the previous section especially the LP model in Section 4.3.3 for contract evaluation.

The local search starts from an initial contract $\mathbf{n}^0$. It then iteratively improves this contract. At each iteration, it determines the best neighbor solution among the set of contracts: $\mathbf{n} + \mathbf{e}_k$ (increasing one time slot in day k), $\mathbf{n} - \mathbf{e}_k$ (reducing one time slot in day k), $\mathbf{n} - \mathbf{e}_k + \mathbf{e}_j$ (move one time slot from day k to day j). This process repeats until no improvement can be found.

The overall algorithm for the contract optimization is summarized as follows:

Algorithm (Local optimization)

1. Select an initial contract $\mathbf{n}^0$, determine the optimal control policy $\pi(\mathbf{n}^0)$, $\mu(\mathbf{n}^0)$ and the optimal average cost $J(\mathbf{n}^0)$ under contract $\mathbf{n}^0$ by solving LP model;

2. Let $\mathbf{n}^* = \mathbf{n}^0$; $J(\mathbf{n}^*) = J(\mathbf{n}^0)$;

3. Determine the neighbor solution $\mathbf{n}'$ with the smallest average cost as follows:

$$\mathbf{n}' = \underset{\mathbf{n}\in\{\mathbf{n}+e_k;\mathbf{n}-e_k;\mathbf{n}-e_k+e_j:1\le k,j\le 7,k\ne j\}\cap IN^7}{\arg\min} J(\mathbf{n})$$

4. If $J(\mathbf{n}') < J(\mathbf{n}^*)$, set $\mathbf{n}^* = \mathbf{n}'$ and go to step 3;

5. The final contract is $\mathbf{n}^*$ and the final control policy is $\pi(\mathbf{n}^*)$ and $\mu(\mathbf{n}^*)$.

In the numerical experiments of the next section, the initial contract decisions are obtained by the method proposed in our previous chapter for optimizing the contract decisions without CTS cancellation.

4.5 Computational Results

This section presents numerical results to show the benefit of CTS cancellation control policy and local improvement. All numerical experiments are performed on a Intel(R) Core (TM)2 Duo CPU T7250 based PC running at 2.00 GHz with 3.0 GB of Memory. The optimal control policies for the MDP formulation (4.1) are obtained by solving LP model with CPLEX 11 solver.

The numerical experiments are all derived from the base case corresponding to our real case study. From the data collected from the neurovascular department of our study, the average

numbers of patient arrivals during the week are as follows: {1, 0.89, 0.95, 1.16, 1.53, 0.16, 0.05}. The number of patients arrived each day is assumed to follow a Poisson distribution truncated at A = 20 which is large enough such that the probability of a_i > A can be neglected. The waiting time for RTS varies from 30 to 40 days with an average of T^R = 35 days. The weighting factor of unused CTS is set to c = 15. CTS cancellation cost is taken as half of c, i.e. b = 7.5.
In the following, the impact of CTS cancellation is analyzed with respect to the cancellation cost b, unused CTS cost c, average delay T^R of regular reservation, patient arrival pattern, and the patient arrival rate.
For each case, three solutions are considered: (i) the solution with only patient assignment policy considered for the initial contract $\mathbf{n^0}$. This solution will be denoted **"NoCancel"**; (ii) the solution considering both CTS cancellation and patient assignment policies for the contract $\mathbf{n^0}$. This solution will be denoted **"Cancel"**; (iii) the solution after local optimization starting from $\mathbf{n^0}$, where both policies are considered. This solution will be denoted **"LocalOpt"**.

The three solutions are further compared with respect to different performance criteria including the average delay, the unused CTS ratio, the percentage of patients using RTS, and the percentage of CTS cancelled.

The CPU time for local optimization is less than 6 minutes except for medium and high demand cases considered in Section 4.5.3.

4.5.1 Impact of CTS cancellation and CTS cancellation cost

This subsection considers the impact of CTS cancellation by varying the cancellation cost b from 0.1c to 0.9c.

Fig.4.6 compares the criterion values of the three solutions "NoCancel", "Cancel", "LocalOpt". From this figure, the cancellation cost b greatly impacts on the benefit of CTS cancellation. The gain of CTS cancellation with respect to the contract "NoCancel" is 31.76% for b=0.1c, 14.86% when b=0.4c, and 0% when b=0.9c. Similar, the local optimization further improves both the contract and the control policies. The gain of Local optimization with respect to the solution "Cancel" is 33.77% for b= 0.1c, 5.38% for b=0.4c and 0 for b=0.9c. The total improvement of cancellation and local optimization with respect to the solution "NoCancel" is 54.8% for b = 0.1c, 19.4% for b = 0.4c and 0 for b = 0.9c.

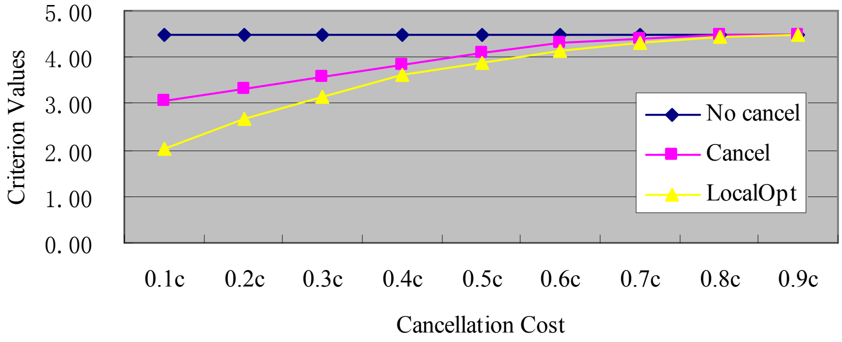


FIG. 4.6 Criterion values vs cancellation costs *b*

Table 4.1 summarizes the performance measures for the three different solution strategies. With respect to the strategy "NoCancel", the CTS cancellation policy greatly reduces the unused CTS ratio and slightly increases the average delay by CTS cancellation, and at the same time more patients assigned to RTS. The reduction of unused CTS is drastic from 18.22% to about 2% when the cancellation cost *b* is small. As *b* increases, less CTS are cancelled and the reduction of unused CTS ratio decreases. The contract "LocalOpt" further takes advantage of CTS cancellation. When *b* is small, all CTS that cannot be used by patients in CTS queue are cancelled and hence there is no unused CTS. Further, enough CTS are planned and no RTS is used.

b	"NoCancel"			"Cancel"				"LocalOpt"			
	Delay (days)	*Unused* (%)	*RTS* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)
0.1*c*	2.16	18.22	0.26	3.15	1.59	0.67	16.96	1.63	0.00	0.00	36.23
0.2*c*	2.16	18.22	0.26	3.15	1.59	0.62	16.93	2.06	0.00	0.01	28.26
0.3*c*	2.16	18.22	0.26	3.15	1.59	0.58	16.89	2.06	0.00	0.01	28.26
0.4*c*	2.16	18.22	0.26	3.15	1.58	0.48	16.82	2.06	0.00	0.00	28.26
0.5*c*	2.16	18.22	0.26	3.16	1.58	0.45	16.80	2.97	0.98	0.44	17.38
0.6*c*	2.16	18.22	0.26	2.79	6.20	0.40	12.13	2.82	2.66	0.39	15.67
0.7*c*	2.16	18.22	0.26	2.28	13.92	0.29	4.32	2.40	9.58	0.30	8.67
0.8*c*	2.16	18.22	0.26	2.29	13.91	0.27	4.32	2.26	12.64	0.28	5.60
0.9*c*	2.16	18.22	0.26	2.16	18.22	0.26	0.00	2.16	18.22	0.26	0.00

TAB. 4.1 Performance measures vs cancellation costs *b*

Table 4.2 summarizes the contracts, the optimal patient assignment policies and the optimal CTS cancellation policies for different cancellation cost *b*. The optimal contract the optimal control policy for the "NoCancel" solution strategy is **n** = (1, 1, 1, 1, 3, 0, 0), **L** = (11, 11, 11, 11, 9, 10, 10) which are identical to the control for "Cancel" strategy with $b = 0.9c$. The optimal patient assignment policy of "Cancel" strategy is somewhat counter-intuitive. One would expect that, with CTS cancellation, a longer waiting queue is needed. However, the opposite happens. The optimal CTS queue threshold **L** is actually smaller than the one for "NoCancel" strategy. More CTS are cancelled when *b* is small. For example, when *b* is smaller than 0.6*c*, all CTS except one for Friday that cannot be used by patients in CTS queue are cancelled. For the optimal contract of "LocalOpt" strategy, more CTS are planned when *b* is small. Further, thanks to the possibility of CTS cancellation, CTS are now planned for Weekend. Further, when *b* is small, large CTS queue are used because of more CTS introduced so as to avoid assignment to RTS.

b	"Cancel"	"LocalOpt"
0.1c	**n**={1, 1, 1, 1, 3, 0, 0}; 7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={9, 9, 9, 10, 8, 8, 9}.	**n**={0, 1, 2, 1, 2, 2, 1}; 9CTS **S**={0, 1, 2, 1, 2, 2, 1}; **L**={18, 18, 18, 19, 18, 17, 17}.
0.2c	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={9, 9, 10, 10, 8, 9, 9}.	**n**={0, 1, 1, 1, 2, 2, 1};8CTS **S**={0, 1, 1, 1, 2, 2, 1}; **L**={14, 14, 14, 14, 14, 13, 12}.
0.3c	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={10, 10, 10, 10, 8, 9, 9}.	**n**={0, 1, 1, 1, 2, 2, 1};8CTS **S**={0, 1, 1, 1, 2, 2, 1}; **L**={14, 14, 15, 15, 14, 13, 13}.
0.4c	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={10, 10, 10, 10, 9, 9, 10}.	**n**={0, 1, 1, 1, 2, 2, 1};8CTS **S**={0, 1, 1, 1, 2, 2, 1}; **L**={14, 15, 15, 15, 15, 13, 13}.
0.5c	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={10, 10, 10, 11, 9, 9, 10}.	**n**={0, 1,1, 1, 2, 2, 0};7CTS **S**={0, 1, 1, 1, 1, 2, 0}; **L**={10, 10, 10, 11, 10, 8, 9}.
0.6c	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={0, 1, 1, 1, 1, 0, 0}; **L**={10, 10, 10, 11, 9, 9, 10}.	**n**={0, 1, 1, 2, 1, 2, 0};7CTS **S**={0, 1, 1, 1, 0, 2, 0}; **L**={10, 11, 11, 10, 10, 9, 9}.
0.7c	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={0, 0, 0, 0, 1, 0, 0}; **L**={10, 11, 11, 11, 9, 10, 10}.	**n**={1, 1, 1, 1, 1, 2, 0};7CTS **S**={0, 0, 0, 0, 0, 2, 0}; **L**={10, 10, 10, 11, 11, 9, 10}.
0.8c	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={0, 0, 0, 0, 1, 0, 0}; **L**={11, 11, 11, 11, 9, 10, 10}.	**n**={1, 1, 1, 1, 2, 1, 0};7CTS **S**={0, 0, 0, 0, 0, 1, 0}; **L**={10, 10, 11, 11, 10, 10, 10}.
0.9c	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={0, 0, 0, 0, 0, 0, 0}; **L**={11, 11, 11, 11, 9, 10, 10}.	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={0, 0, 0, 0, 0, 0, 0}; **L**={11, 11, 11, 11, 9, 10, 10}.

TAB. 4.2 Contracts and control policies vs cancellation costs *b*

4.5.2 Impact of CTS cancellation and unused CTS cost

This subsection considers the relation between the impact of CTS cancellation and the unused CTS cost *c* by varying *c* for the base case with $b = 0.5c$.

Fig.4.7 compares the criterion values of different solution strategies for different unused CTS cost *c*. It is clear that the unused CTS cost has great impact on the contract and the control policies in all solution strategies. When *c* is very small and close to 1, CTS cancellation only provides marginal improvement. CTS cancellation brings larger improvement when *c* is large with 9.3% improvement of "Cancel" over "NoCancel" for $c = 15$ and 15.3% improvement for $c = 20$. Local Optimization further improves the contract and the control policies. The combined improvement over the "NoCancel" strategy reaches 19.2% for $c = 20$.

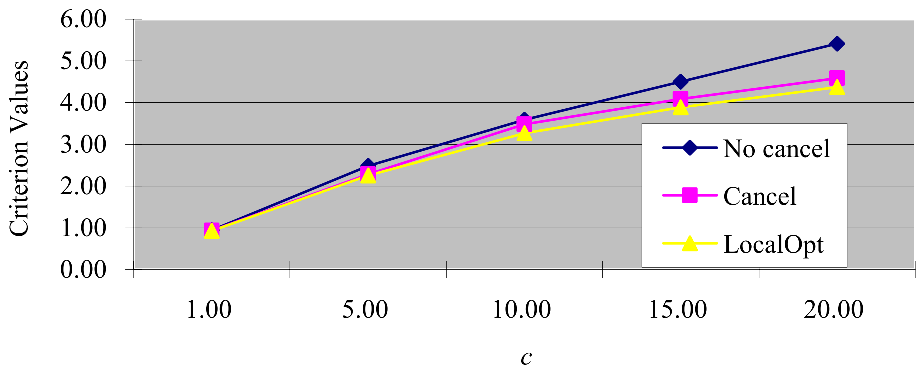


FIG. 4.7 Criterion values vs unused CTS costs *c*

Table 4.3 summarizes the performance measures of different solution strategies. When *c* increases, CTS cancellation significantly reduces the unused CTS ratios with slightly longer delay. There is no obvious trend of CTS cancellation ratio when *c* increases. The "LocalOpt" strategy further improves the "Cancel" strategy by reducing the unused CTS ratio, the average delay with increased CTS cancellation.

Table 4.4 summarizes the contract, patient assignment and CTS cancellation policies. For all solution strategies, the number of CTS decreases when *c* increases. "Cancel" strategy has slightly shorter CTS queue than "NoCancel" strategy. The major difference of the "LocalOpt" strategy with the two other strategies is the CTS planned for the weekend even for the case of large unused CTS cost. "LocalOpt" also allows more CTS cancellation than "Cancel" strategy.

c	"NoCancel"			"Cancel"				"LocalOpt"			
	Delay (days)	*Unused* (%)	*RTS* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)
1	0.41	42.60	0.00	0.47	35.38	0.00	7.22	0.52	28.44	0.00	14.17
5	1.06	28.26	0.00	1.26	15.73	0.00	12.53	1.37	11.45	0.01	16.81
10	2.14	18.32	0.38	2.27	13.99	0.40	4.34	1.76	3.56	0.00	24.70
15	2.16	18.22	0.26	3.16	1.58	0.45	16.80	2.97	0.98	0.44	17.38
20	2.17	18.16	0.18	3.16	1.57	0.39	16.75	2.98	0.98	0.38	17.34

TAB. 4.3 Performance measures vs unused CTS costs *c*

c	"NoCancel"	"Cancel"	"LocalOpt"
1	**n**={2, 1, 2, 2, 2, 1, 0}; 10CTS **L**={22, 22, 22, 21, 21, 21, 22}	**n**={2, 1, 2, 2, 2, 1, 0}; 10CTS **S**={0, 0, 0, 0, 0, 1, 0}; **L**={22, 22, 22, 21, 21, 21, 22}	**n**={1, 2, 1, 2, 2, 1, 1}; 10CTS **S**={0, 0, 0, 0, 0, 1, 1}; **L**={22, 21, 22, 22, 22, 21, 21}
5	**n**={1, 1, 1, 2, 2, 1, 0}; 8CTS **L**={13, 14, 14, 13, 13, 12, 13}	**n**={1, 1, 1, 2, 2, 1, 0}; 8CTS **S**={0, 0, 0, 1, 0, 1, 0}; **L**={13, 14, 14, 13, 13, 12, 13}	**n**={1, 1, 1, 1, 2, 2, 0}; 8CTS **S**={0, 0, 0, 0, 1, 2, 0}; **L**={13, 13, 13, 14, 13, 12, 13}
10	**n**={1, 1, 1, 1, 3, 0, 0}; 7CTS **L**={10, 10, 10, 10, 8, 9, 9}	**n**={1, 1, 1, 1, 3, 0, 0}; 7CTS **S**={0, 0, 0, 0, 1, 0, 0}; **L**={10, 10, 10, 10, 8, 9, 9}	**n**={0, 1, 1, 2, 2, 1, 1}; 8CTS **S**={0, 1, 1, 1, 1, 1, 1}; **L**={14, 15, 15, 14, 14, 13, 13}
15	**n**={1, 1, 1, 1, 3, 0, 0}; 7CTS **L**={11, 11, 11, 11, 9, 10,10}	**n**={1, 1, 1, 1, 3, 0, 0}; 7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={10, 10, 10, 11, 9, 9,10}.	**n**={0, 1, 1, 1, 2, 2, 0}; 7CTS **S**={0, 1, 1, 1, 1, 2, 0}; **L**={10, 10, 10, 11, 10, 8, 9}
20	**n**={1, 1, 1, 1, 3, 0, 0}; 7CTS **L**={11, 12, 12, 12, 10, 11, 11}	**n**={1, 1, 1, 1, 3, 0, 0}; 7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={11, 11, 11, 11, 9, 10, 10}	**n**={ 0, 1, 1, 1, 2, 2, 0}; 7CTS **S**={ 0, 1, 1, 1, 1, 2, 0}; **L**={11, 11, 11, 11, 10, 9, 9}

TAB. 4.4 Contracts and control policies vs unused CTS costs *c*

4.5.3 Impact of CTS cancellation and other data

This subsection investigates the relationship between the impact of CTS cancellation and other problem parameters including (i) the average RTS delay T^R, (ii) the patient arrival pattern, and (iii) the patient arrival rate.

With respect to the average RTS delay T^R, numerical experiments are performed for the base case by varying T^R from 25 to 45. For all three solution strategies, the criterion values, the performance measures, the contract and the control strategies given in the Table 4.5 and 4.6 are fairly insensitive to the change of T^R.

T^R	"NoCancel"			"Cancel"				"LocalOpt"			
	Delay (days)	*Unused* (%)	*RTS* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)
25	2.08	18.46	0.56	3.05	1.61	0.92	17.15	2.87	1.01	0.94	17.77
30	2.13	18.32	0.38	3.11	1.59	0.65	16.95	2.93	0.99	0.65	17.55
35	2.16	18.22	0.26	3.16	1.58	0.45	16.80	2.97	0.98	0.44	17.38
40	2.17	18.22	0.26	3.19	1.57	0.31	16.69	2.85	2.65	0.30	15.60
45	2.20	18.11	0.13	3.21	1.56	0.23	16.63	2.87	2.64	0.21	15.53

TAB. 4.5 Performance measures vs RTS delay T^R

T^R	"NoCancel"	"Cancel"	"LocalOpt"
25	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **L**={9, 9, 9, 9, 7, 8, 9}. J=4.47	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={8, 8, 9, 9, 7, 8, 8}. J=4.03	**n**={ 0, 1, 1, 1, 2, 2, 0};7CTS **S**={ 0, 1, 1, 1, 1, 2, 0}; **L**={8, 8, 8, 9, 8, 7, 7}. J=3.83
30	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **L**={10, 10, 10, 10, 8, 8, 10} J=4.49	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={9, 9, 9, 10, 8, 9, 9}. J=4.06	**n**={ 0, 1, 1, 1, 2, 2, 0};7CTS **S**={ 0, 1, 1, 1, 1, 2, 0}; **L**={9, 9, 9, 10, 9, 7, 8}. J=3.87
35	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **L**={11, 11, 11, 11, 9, 10, 10}. J=4.50	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={10, 10, 10, 11, 9, 9, 10}. J=4.08	**n**={0, 1, 1, 1, 2, 2, 0};7CTS **S**={0, 1, 1, 1, 1, 2, 0}; **L**={10, 10, 10, 11, 10, 8, 9}. J=3.89
40	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **L**={11, 12, 12, 12, 10, 11, 11}. J=4.51	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={11, 11, 11, 12, 10, 10, 11}. J=4.10	**n**={ 0, 1, 1, 2, 1, 2, 0};7CTS **S**={ 0, 1, 1, 1, 0, 2, 0}; **L**={11, 11, 11, 11, 11, 9, 10}. J=3.90
45	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **L**={12, 12, 13, 13, 11, 12, 12}. J=4.52	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={12, 12, 12, 12, 11, 11, 12}. J=4.11	**n**={ 0, 1, 1, 2, 1, 2, 0};7CTS **S**={ 0, 1, 1, 1, 0, 2, 0}; **L**={12, 12, 12, 12, 12, 10, 11}. J=3.91

TAB. 4.6 Contracts and control policies vs RTS delay T^R

To investigate the relation with respect to the patient arrival pattern, the peak demand of the base case occurred on Friday is interchanged with the demand other weekdays. We also consider the case of stationary demand for all weekdays. Again, the criterion values and the performance measures are insensitive to the change of patient arrival patterns. When the peak arrival changes to another weekday, one CTS time slot is move from Friday to the peak arrival day in the "“NoCancel”" strategy. Except for CTS of peak arrival day and Friday, "Cancel" strategy cancels all CTS time slots that cannot be used by patients in CTS queue. For the contract of "LocalOpt", one CTS time slot is moved from Friday to the peak arrival day. Again, one CTS time slot is left by the CTS cancellation policy for the peak arrival day. As for the other case, CTS are planned for the weekend.

Peak arrival	"NoCancel"	"Cancel"	"LocalOpt"
Mon	**n**={2, 1, 1, 1, 2, 0, 0}; 7CTS **L**={10,10,11,11,10,10,11}. J=4.51	**n**={2, 1, 1, 1, 2, 0, 0};7CTS **S**={1, 1, 1, 1, 1, 0, 0}; **L**={10,10,10,10,9,10,10}. J=4.08	**n**={1, 1, 1, 1, 1, 2, 0};7CTS **S**={0, 1, 1, 0, 1, 2, 0}; **L**={10,10,10,10,10,9,9}. J=3.91
Tues	**n**={1, 2, 1, 1, 2, 0, 0};7CTS **L**={11,10,10,11,10,10,11}. J=4.50	**n**={1, 2, 1, 1, 2, 0, 0};7CTS **S**={1, 1, 1, 1, 1, 0, 0}; **L**={10,10,10,10,9,10,10}. J=4.06	**n**={0, 2, 1, 1, 1, 2, 0};7CTS **S**={0, 1, 1, 0, 1, 2, 0}; **L**={10,10,10,10,10,9,9}. J=3.90
Wed	**n**={1, 1, 2, 1, 2, 0, 0};7CTS **L**={11,11,10,11,9,10,11}. J=4.49	**n**={1, 1, 2, 1, 2, 0, 0};7CTS **S**={1, 1, 1, 1, 1, 0, 0}; **L**={10,10,10,10,9, 10,10}. J=4.05	**n**={0, 1, 2, 1, 1, 2, 0};7CTS **S**={0, 1, 1, 0, 1, 2, 0}; **L**={10,10,10,10,10,9,9}. J=3.88
Thurs	**n**={1, 1, 1, 2, 2, 0, 0};7CTS **L**={11,11,11,10,9,10,10}. J=4.48	**n**={1, 1, 1, 2, 2, 0, 0};7CTS **S**={1, 1, 1, 1, 1, 0, 0}; **L**={10,10,10,10,9,9,10}. J=4.03	**n**={0, 1, 1, 2, 1, 2, 0};7CTS **S**={0, 1,1, 1, 1, 2, 0}; **L**={10,10,10,10,10,8,9}. J=3.88
Fri	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **L**={11,11,11,11,9,10,10}. J=4.50	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **S**={1, 1, 1, 1, 2, 0, 0}; **L**={10,10,10,11,9,9,10}. J=4.08	**n**={0, 1, 1, 1, 2, 2, 0};7CTS **S**={0, 1, 1, 1, 1, 2, 0}; **L**={10,10,10,11,10,8,9}. J=3.89
Ave	**n**={1, 1, 1, 1, 3, 0, 0};7CTS **L**={11,11,10,11,10,10,11}. J=4.52	**n**={1, 1, 2, 1, 2, 0, 0};7CTS **S**={1, 1, 1, 1, 1, 0, 0}; **L**={11,11,10,10,9,10,10}. J=4.07	**n**={0, 1, 2, 1, 1, 2, 0};7CTS **S**={0, 1, 1, 0, 1, 2, 0}; **L**={11,11,10,10,10,9,9}. J=3.92

TAB. 4.7 Contracts and control policies vs patient arrival patterns

Peak arrival	"NoCancel"			"Cancel"				"LocalOpt"			
	Delay (days)	*Unused* (%)	*RTS* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)
Mon	2.16	18.19	0.24	3.00	3.23	0.44	15.11	2.90	2.08	0.44	16.27
Tues	2.16	18.20	0.26	2.98	3.32	0.44	15.03	2.90	2.05	0.44	16.30
Wed	2.16	18.18	0.28	2.97	3.27	0.44	15.04	2.85	2.41	0.44	15.90
Thurs	2.13	18.20	0.27	2.94	3.18	0.44	15.16	2.95	1.15	0.46	17.21
Fri	2.16	18.22	0.26	3.16	1.58	0.45	16.80	2.97	0.98	0.44	17.38
AVE	2.13	18.59	0.22	2.95	3.58	0.38	15.13	2.80	3.09	0.37	15.62

TAB. 4.8 Performance measures vs patient arrival patterns

With respect to the patient arrival rate, three instances are considered: (i) the base case that is termed "low demand" instance; (ii) the base case but with patient arrival rate 5 times larger and termed "medium demand" instance; and (iii) the base case with patient arrival rate multiplied by 10 and termed "high demand" instance. Fig.4.8 compared the criterion values of different solution strategies. CTS cancellation brings 9.3% improvement over "NoCancel" strategy for low demand instance, 5% for medium demand and 6.8% for high demand instance. Local optimization further brings 4.8%, 5.8% and 3.7% improvement over "Cancel" strategy and leads to a combined improvement of 13.6%, 10.5% and 10.2% for the three instances. These results show that the CTS cancellation is useful for both low demand and high demand instances.

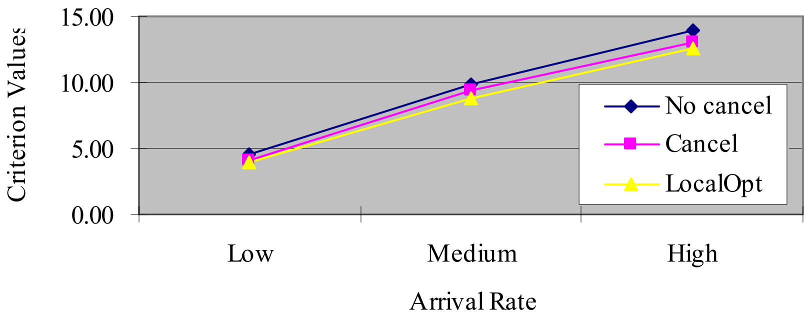


FIG. 4.8 Criterion values vs patient arrival rates

Table 4.9 summarizes the performance measures of different solution strategies. In all solution strategies, increasing the patient arrival rate leads to shorter delay. For "NoCancel", unused CTS ratio becomes smaller. For "Cancel" and "LocalOpt" strategies, CTS cancellation ratio decreases significantly, the percentage of patients assigned to RTS decreases while CTS unused ratio fluctuates but remains small.

CPU time for local optimization is less than 6 minutes for all the low demand instances, 2407s for medium demand instance and 6324s for the high demand instance. The resulting contract and control policies are given in the Table 4.10.

Arrival Rate	"NoCancel"			"Cancel"				"LocalOpt"			
	Delay (days)	*Unused* (%)	*RTS* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)
Low	2.16	18.22	0.26	3.16	1.58	0.45	16.80	2.97	0.98	0.44	17.38
Medium	1.16	7.68	0.24	1.38	3.36	0.34	4.40	1.14	1.61	0.06	8.78
High	0.74	6.01	0.11	0.90	2.53	0.16	3.52	0.77	1.85	0.04	5.61

TAB. 4.9 Performance measures vs patient arrival rates

Arrival Rate	"NoCancel"	"Cancel"	"LocalOpt"
Low	**n**={1,1,1,1,3,0,0};7CTS **L**={11,11,11,11,9,10,10}. J=4.50	**n**={1,1,1,1,3,0,0};7CTS **S**={1,1,1,1,2,0,0}; **L**={10,10,10,11,9,9,10}. J=4.08	**n**={0,1,1,1,2,2,0};7CTS **S**={0,1,1,1,1,2,0}; **L**={10,10,10,11,10,8,9}. J=3.89
Medium	**n**={5,5,5,6,9,1,0};31CTS **L**={21,21,21,21,19,19,21}. J=9.83	**n**={5,5,5,6,9,1,0};31CTS **S**={1,1,1,1,2,1,0}; **L**={20,20,20,20,18,19,20}. J=9.34	**n**={3,5,5,6,8,4,1};32CTS **S**={0,2,2,2,2,4,1}; **L**={25,25,25,25,24,22,23}. J=8.79
High	**n**={10,9,10,12,17,2,1};61CTS **L**={29,30,29,27,22,29,30}. J=13.94	**n**={10,9,10,12,17,2,1};61CTS **S**={1,2,2,2,2,1,1}; **L**={30,30,30,30,27,28,29}. J=13.94	**n**={8,9,10,12,15,7,1};62CTS **S**={0,2,2,2,2,6,1}; **L**={34,35,35,35,35,31,32}. J=12.51

TAB. 4.10 Contracts and control policies vs patient arrival rates

4.6 Conclusions and perspectives

This chapter proposed an MRI examination reservation process between the neurovascular department and the imaging department combining (i) contracted time slots reserved by the imaging department for the neurovascular department, (ii) one-day advance cancellation of contracted time slots, and (iii) patient assignment control policy. An average cost MDP model is proposed to simultaneously determine the optimal patient assignment and CTS cancellation control policies. Structure properties are established via discounted cost problem. A local optimization is used to improve a given initial contract. Numerical results show that the consideration of CTS cancellation can greatly reduce the unused CTS ratio with a little increase in average delay. Local optimization can further decrease the contract and the control policies.

Future research can be pursued in several directions. One immediate extension is to allow CTS cancellation several days before. One research direction is the extension of this work to multiple classes of patients and multiple imaging examinations. Another very challenging issue is the optimization of the operation of the imaging department by considering the quality requirements of medical units.

Chapter 5

Contract planning and two-day advance cancellation of contracted time slots

This chapter exploits the use of two-day advance CTS cancellation to further improve contract for MRI examinations between a neurovascular department treating stroke patients and an imaging department. The contract is composed of four parts: contract decisions, one-day advance CTS cancellation policy, two-day advance CTS cancellation policy, and patient assignment policy. The problem of CTS cancellations and patient assignment are formally formulated as an average cost Markov Decision Process in order to compromise among the patients' waiting time, unused CTS penalty, and CTS cancellation penalty. Structure properties of the optimal control policies are established via the discounted cost problem and some advanced convexity concepts. Local search algorithm is proposed to improve the contract decisions. Numerical results show that advance CTS cancellation and local optimization significantly reduces the criterion value and the ratio of unused CTS.

5.1 Introduction

Chapter 3 and Chapter 4 propose a contract-based MRI examination reservation process in order to reduce the stroke patients' waiting time for MRI examination. The contract decisions and patient assignment control policy are jointly solved in Chapter 3. Chapter 4 proposes an MDP formulation to determine the patient assignment and one-day advance CTS cancellation control policies at the same time in order to further reduce the unused CTS. This new reservation process seems perfect with shorter delay and lower unused CTS ratio. However, for the imaging department side, it would be hard to arrange other patients for the time slots released from contract with short notice of only one day. Therefore, this chapter exploits the possibility of earlier CTS cancellation.

The contract now includes the following four decisions: 1) contract decisions; 2) patient assignment control policy; 3) one-day advance CTS cancellation control policy; and 4) two-day advance CTS cancellation control policy.

This chapter tries to simultaneously explore the structure properties of three control policies and then improve the contract by using local search algorithm. The rest of this chapter is organized as follows: Section 5.2 provides a formal problem setting. Structure properties of the optimal control policies are established via discounted cost MDP in Section 5.3 and 5.4.

Local optimization of the contract is proposed in Section 5.5. Section 5.6 analyzes results of computational experiments. Conclusions and perspectives are given in Section 5.7.

5.2 Problem setting

This section presents an MDP formulation of the problem of patient assignment and one-day and two-day advance CTS cancellation. Assumptions 1 to 4 presented in Chapter 1 are considered. Further, the following assumption is made throughout the chapter:

Assumption 5-A1 CTS of day t can be cancelled in advance of one and two days, i.e., at end of day t–2 or at the end of day t–1.

Based on the above assumptions, the problem of patient assignment and advance CTS cancellation can be characterized by the following notations:

Indices:

t: index of days, t=1,…,T;

i: index of days in one week, i=1,…,7, i.e., Monday, …, Sunday; note that the day $i\pm j$ is the weekday of j days after or before day i in one week;

$d(t)$: day in the week of day t with $d(t)\in\{1, \ldots,7\}$;

Data:

T^R: average number of waiting days for a patient to have his/her MRI examination through regular reservation with $T^R > 1$;

c: penalty factor of an unused CTS. It serves as a weighting factor in order to balance the average waiting time and unused MRI time slots;

b_1: penalty factor of canceling one CTS in advance of one day with $b_1 < c$;

b_2: penalty factor of canceling one CTS in advance of two days with $b_2 < b_1$;

a_t: number of patients arrived in day t. By assumption 3, daily arrivals a_t for $t \in IN$ are mutually independent random variables and weekly arrivals (a_{7j+1}, a_{7j+2}, …, a_{7j+7}) are identically distributed for all j = 0, 1, …. As a result, the arrival process is characterized by probability matrix $\boldsymbol{P} = [P_{ij}]$ for i = 1, …, 7 and for all $j \geq 0$ with P_{ij} denoting the probability of j arrivals in day i;

n_t: number of CTS in day t;

Decision variables:

x_t: number of patients waiting for CTS at the end of day t, which is also called CTS queue, x_0 is a given constant. Note that x_t does not include patients that are directed to RTS.

w_t^1: number of CTS cancelled for day t in advance of one day, i.e., at the end of day t-1;

w_t^2 : number of CTS cancelled for day t in advance of two day, i.e., at the end of day t-2;

State variables:

$u_t = x_t + w_{t+1}^2$: number of CTS to be consumed at day t+1 without the one-day advance CTS cancellation control and new patients' arrival.

$y_t = u_t + w_{t+1}^1 = x_t + w_{t+1}^1 + w_{t+1}^2$: number of CTS to be consumed at day t+1 without new patients' arrival.

$z_t = y_{t-1} + a_t = x_t + w_{t+1}^1 + w_{t+1}^2 + a_t$: number of CTS to be consumed at the day t.

The sequence of events during each day t is as follows. First, the following information is known: the queue length x_{t-1} for CTS at the end of day t–1, the number of two-day advance CTS cancellations for day t w_t^2, the number of one-day advance CTS cancellations for day t w_t^1, and the number of two-day advance CTS cancellations for day t+1 w_{t+1}^2. Based on the above information, state variable $y_{t-1} = x_{t-1} + w_t^1 + w_t^2$ is also known. During day t, the number a_t of new incoming patients during the day becomes known. The state variable z_t is known as $y_{t-1}+ a_t$. MIN$\{n_t - w_t^1 - w_t^2, x_{t-1} + a_t\}$ patients are served by the remaining CTS of the day and MAX$\{0, n_t - z_t\}$ CTS of day t cannot be filled. MAX$\{0, z_t - n_t - x_t\}$ patients are directed to RTS and will have the MRI examination after an average of T^R days. The remaining x_t patients will wait for CTS of the subsequent days. The number of CTS cancelled for day t+1 and day t+2 is determined.

There are three control decisions in this problem: one-day advance CTS cancellation, two-day advance cancellation, and patient assignment policies. The state is represented by the combination of two variables: $\left(z_t, w_{t+1}^2\right)$. History-dependent policies are considered in this chapter. Let $h_t = \left\{h_{t-1}, \left(\left(z_t, w_{t+1}^2\right), \left(x_t, w_{t+1}^1, w_{t+2}^2\right)\right)\right\}$, $h_0 = \left\{\left(z_i, w_{i+1}^2\right), \left(x_i, w_{i+1}^1, w_{i+2}^2\right)\right\}$ be the full history by stating from initial state $\left(z_i, w_{i+1}^2\right)$ at the beginning of day i. We denote the patient assignment policy by $\boldsymbol{\pi}=\{\pi_1, \pi_2, ...\}$ where the CTS queue at the end of day t is $x_t = \pi_t(h_t)$ with $0 \le x_t \le (z_t - n_t)^+$, and one-day advance cancellation control policy by $\boldsymbol{\mu}^1 = \left\{\mu_1^1, \mu_2^1, ...\right\}$ where the number of CTS cancelled for day t+1 is $w_{t+1}^1 = \mu_t^1(h_t)$ with $0 \le w_{t+1}^1 \le (n_{t+1} - x_t - w_{t+1}^2)^+$, and the two-day advance cancellation policy by $\boldsymbol{\mu}^2 = \{\mu_1^2, \mu_2^2, ...\}$ where the number of CTS cancelled for day t+2 is $w_{t+2}^2 = \mu_t^2(h_t)$ with $0 \le w_{t+2}^2 \le n_{t+2}$.

The objective is to minimize over all history-dependent policies $\boldsymbol{\mu}^1 = \{\mu_1^1, \mu_2^1, ...\}$, $\boldsymbol{\mu}^2 = \{\mu_1^2, \mu_2^2, ...\}$ and $\boldsymbol{\pi}$={π_1,π_2,...} the long-run average cost

$$J_{\mu^1\mu^2\pi}(i,z,w) = \lim_{T\to\infty} \frac{1}{T} E\left[\sum_{t=i}^{T+i-1} g_{d(t)}(z_t, w_{t+1}^2, x_t, w_{t+1}^1, w_{t+2}^2) \middle| (z_i = z, w_{i+1}^2 = w)\right] \tag{5.1}$$

for any given initial state $(z_i = z, w_{i+1}^2 = w)$ with i = 1, …7 where

$$g_{d(t)}(z_t, w_{t+1}^2, x_t, w_{t+1}^1, w_{t+2}^2) = b_1 w_{t+1}^1 + b_2 w_{t+2}^2 + c(n_{d(t)} - z_t)^+ + x_t + T^R(z_t - n_{d(t)} - x_t)^+$$

is the stage cost, i.e., CTS cancellation cost plus unused CTS penalty plus waiting time for CTS and RTS. In the following, $g_{d(t)}$(.) are written as g_t(.) for convenience.

Theorem 5-1: There exists an optimal average cost policy such that $x_t \le \bar{x}$ for all t > 0 with $\bar{x} = \lceil (T^R + c) n^* \rceil$ where n^* = MAX{n_1, …, n_7}.

Proof. The proof can use the same method with that for Theorem 3-1. Q.E.D.

Thanks to Theorem 5-1, we can make without loss of generality the following assumption:

Assumption 5-A2: $x_t \le \bar{x}$ for all t > 0.

5.3 Properties of optimal control policies for discounted cost problem

According to relation (5.1), the corresponding α-discounted cost MDP is as follows:

$$J_{\alpha,\mu^1\mu^2\pi}(i,z,w) = \lim_{T\to\infty} E\left[\sum_{t=i}^{T} \alpha^{t-i} g_t(z_t, w_{t+1}^2, x_t, w_{t+1}^1, w_{t+2}^2) \middle| z_i = z, w_{i+1}^2 = w\right] \tag{5.2}$$

for any given initial state $z_i = z, w_{i+1}^2 = w$ with i = 1, …7 with discount factor α such that 0 < α < 1. Consider the following optimal cost function

$$U_\alpha(i,z,w) = \underset{\mu^1\mu^2\pi}{MIN}\, J_{\alpha,\mu^1\mu^2\pi}(i,z,w) \tag{5.3}$$

In the remaining, for simplicity, the notation α is omitted in this subsection where only discounted cost problem with a given α is considered.

Theorem 6.10.4 in Puterman (1994) is used to establish the optimality equation. It will be shown in Remark 5-1 that all conditions needed for application of Theorem 6.10.4 are satisfied. Since the set of states (*i*, *z, w*) is countable and the control constraint set is finite as $x_t \le z_t$ and $w_{t+1}^1 \le n_{t+1}$, and $w_{t+2}^2 \le n_{t+2}$ for each z_t, Theorem 6.10.4 in Puterman (1994) implies that the optimal cost function is the unique solution of the following optimality equation:

$$U(i,z,w) = \min_{x_i, w_{i+1}^1, w_{i+2}^2} \left\{ \begin{array}{l} g_i\left(z_i, w_{i+1}^2, x_i, w_{i+1}^1, w_{i+2}^2\right) \\ +\alpha \sum_a P_{d(i+1),a} U\left(i+1, x_i + w_{i+1}^2 + w_{i+1}^1 + a, w_{i+2}^2\right) \end{array} \right\}, \forall i = 1,...,7 \quad (5.4)$$

The optimal control policy is stationary deterministic and is given by the argument $\boldsymbol{w^1}$, $\boldsymbol{w^2}$ and $\boldsymbol{x}$ that reach the minimum in (5.4) and the optimal cost function is the limiting function of the following value iteration:

$$U^t\left(z_t, w_{t+1}^2\right) = \min_{x_t, w_{t+1}^1, w_{t+2}^2} \left\{ g_t\left(z_t, w_{t+1}^2, x_t, w_{t+1}^1, w_{t+2}^2\right) + \alpha \sum_a P_{t+1,a} U^{t+1}\left(x_t + w_{t+1}^1 + w_{t+1}^2 + a, w_{t+2}^2\right) \right\} \quad (5.5)$$

$$U^0(z) = 0 \quad (5.6)$$

for $t = 0, -1, -2, \ldots$ where $0 \le x_t \le \left(\left(z_t - n_t\right)^+ \wedge \bar{x}\right)$, $w_{t+1}^1 \le \left(n_{t+1} - x_t - w_{t+1}^2\right)^+$, $w_{t+2}^2 \le n_{t+2}$. n_t= $n_{d(t)}$ and $P_{t+1,a}$= $P_{d(t+1),a}$ are shorthand notation with $d(t)$ denoting the corresponding day with $d(0) = 7$, $d(-1) = 6$, … As a result,

$$U(i,z,w) = \lim_{n\to\infty} U^{-7n+i}(z,w) \quad (5.7)$$

Since $g_t\left(z_t, w_{t+1}^2, x_t, w_{t+1}^1, w_{t+2}^2\right) = b_1 w_{t+1}^1 + b_2 w_{t+1}^2 + c\left(n_t - z_t\right)^+ + x_t + T^R\left(z_t - n_t - x_t\right)^+$, relation (5.5) can be rewritten as

$$U^t\left(z_t, w_{t+1}^2\right) = \min_{0 \le x_t \le \left(\left(z_t - n_t\right)^+ \wedge \bar{x}\right)} \left\{ c\left(n_t - z_t\right)^+ + x_t + T^R\left(z_t - n_t - x_t\right)^+ + V^t\left(x_t, w_{t+1}^2\right) \right\} \quad (5.8)$$

$$V^t\left(x_t, w_{t+1}^2\right) = \min_{x_t + w_{t+1}^2 \le y_t \le \max\left\{x_t + w_{t+1}^2, n_{t+1}\right\}} \left\{ W^t\left(y_t\right) + b_1\left(y_t - x_t - w_{t+1}^2\right) \right\} \quad (5.9)$$

$$W^t\left(y_t\right) = \min_{0 \le w_{t+2}^2 \le n_{t+2}} E\left[b_2 w_{t+2}^2 + \alpha U^{t+1}\left(y_t + a_{t+1}, w_{t+2}^2\right) \right] \quad (5.10)$$

$U^t(z_t,w^2_{t+1})$, $V^t(x_t,w^2_{t+1})$ and $W^t(y_t)$ are separately used to identify the optimal patient assignment, one-day advance cancellation, and two-day advance cancellation control policies.

Similarly, relation (5.4) can be rewritten as

$$U(i,z,w)=\min_{0\le x\le\left((z_i-n_i)^+\wedge\bar{x}\right)}\left\{c(n_i-z)^+ + x + T^R(z-n_i-x)^+ + V(i,x,w)\right\} \tag{5.11}$$

$$V(i,x,w)=\min_{x+w\le y\le\max\{x+w,n_{i+1}\}}\left\{W(i,y)+b_1(y-x-w)\right\} \tag{5.12}$$

$$W(i,y)=\min_{0\le w\le n_{i+2}} E\left[b_2 w+\alpha U(i+1,y+a,w)\right] \tag{5.13}$$

By relation (5.7) and the uniqueness of the optimal value function,

$$V(i,x,w)=\lim_{n\to\infty}V^{-7n+i}(x,w) \tag{5.14}$$

$$W(i,y)=\lim_{n\to\infty}W^{-7n+i}(y) \tag{5.15}$$

From relation (5.9), its right hand side is a function of u_t with $u_t=x_t+w^2_{t+1}$. For this reason, in the following, by abuse of notation we replace $V^t\left(x_t,w^2_{t+1}\right)$ by $V^t\left(x_t+w^2_{t+1}\right)$. Relation (5.9) becomes

$$V^t(u_t)=\min_{u_t\le y_t\le\max\{u_t,n_{t+1}\}}\left\{W^t(y_t)+b_1 y_t\right\}-b_1 u_t \tag{5.16}$$

Similar, relation (5.12) becomes

$$V(i,u)=\min_{u\le y\le\max\{u,n_{i+1}\}}\left\{W(i,y)+b_1 y\right\}-b_1 u \tag{5.17}$$

Remark 5-1: The optimality equation (5.4) requires Assumption 6.10.1 and condition (6.10.11) of Puterman (1994). Assumption 6.10.1 holds as $g_t\left(z_t,w^2_{t+1},x_t,w^1_{t+1},w^2_{t+2}\right)\le(b_1+b_2+c)n^*+T^R z_t=\mathrm{K}(z_t)$. Condition (6.10.11) holds as:

$$\begin{aligned}\sum_a P_{i+1,a}K\left(x_t+w^1_{t+1}+w^2_{t+1}+a\right) &\le (b_1+b_2+c)n^*+T^R E\left[x_t+w^1_{t+1}+w^2_{t+1}+a\right]\\ &\le K(z_t)+(b_1+b_2+c)n^*+T^R\bar{x}+T^R a^*+2T^R n^*\end{aligned}$$

Property 5-1. In the value iteration by (5.8) and (5.9) or equivalently by (5.16), $-c\le U^t(z+1,w)-U^t(z,w)\le T^R$, $-b_1\le V^t(u+1)-V^t(u)\le T^R$.

Proof. The proof is done by induction on t. The property trivially holds for t=0. Assume that it holds for some t+1≤0 for $V^{t+1}(u)$ and consider day t.

We first prove $-c \le U^{t+1}(z+1,w)-U^{t+1}(z,w) \le T^R$. Let x^0 and x^1 be arguments reaching minimum in relation (5.8) for $U^{t+1}(z,w)$ and $U^{t+1}(z+1,w)$. From the optimality condition, $x^0 \in \left[0,(z-n_{t+1})^+ \wedge \overline{x}\right]$, and $x^1 \in \left[0,(z+1-n_{t+1})^+ \wedge \overline{x}\right]$.

To prove $U^{t+1}(z+1,w)-U^{t+1}(z,w) \le T^R$ the right side, take the feasible control x^0 for $U^{t+1}(z+1,w)$, then $U^{t+1}(z+1,w)-U^{t+1}(z,w) \le T^R$.

To prove $-c \le U^{t+1}(z+1,w)-U^{t+1}(z,w)$, two cases are considered:

Case 1: $x^1 \le \left((z-n_{t+1})^+ \wedge \overline{x}\right)$ or $x^1 = (z+1-n_{t+1})^+ = 0$, take the feasible control x^1 for $U^{t+1}(z,w)$, then $U^{t+1}(z+1,w)-U^{t}(z,w) \ge -c$.

Case 2: $x^1 = z+1-n_{t+1} > 0$, take the feasible control $x^1-1 = z-n_{t+1}$ for $U^{t+1}(z,w)$, then

$$\begin{aligned} U^{t+1}(z+1,w)-U^{t+1}(z,w) &\ge 1+V^{t+1}(z+1-n_{t+1}+w)-V^{t+1}(z-n_{t+1}+w) \\ &\ge 1-b_1 \ge -c \end{aligned}$$

Now we prove $-b_1 \le V^t(u+1)-V^t(u) \le T^R$. Let y^0 and y^1 be arguments reaching minimum in relation (5.16) for $V^t(u)$ and $V^t(u+1)$. From the optimality condition, $y^0 \in \left[u,\max(u,n_{t+1})\right]$, $y^1 \in \left[u+1,\max(u+1,n_{t+1})\right]$.

To prove $V^t(u+1)-V^t(u) \le T^R$, two cases are considered:

Case 1: $y^0 > u$. Take the feasible control y^0 for $V^t(u+1)$, then $V^t(u+1)-V^t(u) \le -b_1 \le 0$;

Case 2: $y^0 = u$. Take the feasible control $u+1$ for $V^t(u+1)$, then

$$\begin{aligned} V^t(u+1)-V^t(u) &\le W^t(u+1)-W^t(u) \\ &\le E\left[U^{t+1}(z+1,w)-U^{t+1}(z,w)\right] \\ &\le T^R \end{aligned}$$

The first relation is from relation (5.9). The second is from relation (5.10) by applying the optimal two-day advance cancellation of $W^t(u)$ to $W^t(u+1)$. The third is from the induction assumption.

To prove $-b_1 \le V^t(u+1)-V^t(u)$, two cases are considered:

Case 1: $y^1 = u+1 > n_{t+1}$. Take the feasible control u for $V^t(u)$ and then apply the optimal two-day advance cancellation of $W^t(u+1)$ to $W^t(u)$, then

$$\begin{aligned} V^t(u+1) - V^t(u) &\geq W^t(u+1) - W^t(u) \\ &\geq E\left[U^{t+1}(z+1,w) - U^{t+1}(z,w)\right] \end{aligned}$$

In this case, there is no unused CTS for $U^{t+1}(z+1,w)$ and $U^{t+1}(z,w)$. Let $\Delta U = U^{t+1}(z+1,w) - U^{t+1}(z,w)$. We show $\Delta U \geq -b_1$ which yields $-b_1 \leq V^t(u+1) - V^t(u)$. Let x be the optimal control for $U^{t+1}(z+1,w)$. If $x \leq ((z-n_{t+1}) \wedge \bar{x})$, then take the feasible control x for $U^{t+1}(z,w)$ and we have $\Delta U \geq T^R$. If $x = z+1-n_{t+1} < \bar{x}$, use $x-1$ as the feasible control policy for $U^{t+1}(z,w)$, then $\Delta U \geq 1 + V^{t+1}(x+w) - V^{t+1}(x+w-1) \geq 1 - b_1$.

Case 2: $y^1 > u+1$. Take the feasible control y^1 for $V^t(u)$, $V_t(u+1) - V_t(u) \geq -b_1$.

Q.E.D.

Property 5-2. If $W^t(y_t)$ is convex in y_t, then $V^t(u_t)$ is convex in u_t.

Proof. Let $R^t(y) = b_1 y + W^t(y)$. Under the assumption of the Property, $R_t(y)$ is convex.

Let

$$S_t^1 = \arg\min_{y \geq 0} R^t(y).$$

We will show $S_t^1 \leq n_{t+1}$ by contradiction. If $S_t^1 > n_{t+1}$,

$$V^t(S_t^1) - V^t(S_t^1 - 1) = R^t(S_t^1) - R^t(S_t^1 - 1) - b_1 < -b_1$$

which contradicts Property 5-1 and proves $S_t^1 \leq n_{t+1}$.

Since $S_t^1 \leq n_{t+1}$,

$$V^t(u_t) = -b_1 u_t + \begin{cases} R^t(u_t), & \text{if } u_t \geq S_t^1 \\ R^t(S_t), & \text{if } u_t < S_t^1 \end{cases},$$

which is a combination of two convex functions. Hence, $V^t(u_t)$ is convex in u_t.

Q.E.D.

Relation (5.8) can be rewritten as:

$$U^t(z,w) = c(n_t - z)^+ + T^R(z - n_t)^+ + \min_{0 \le x \le \left((z-n_t)^+ \wedge \bar{x}\right)} \left\{V^t(x+w) - (T^R - 1)x\right\}$$

which leads to:

$$\begin{aligned} U^t(z,w) = c(n_t - z)^+ + T^R(z - n_t)^+ + (T^R - 1)w \\ + \min_{w \le u \le w + (z-n_t)^+} \left\{V^t(u) - (T^R - 1)u\right\} \end{aligned} \tag{5.18}$$

Let

$$H^t(u) = V^t(u) - (T^R - 1)u,$$

which is convex since $V_t(u)$ is convex in u. Let

$$L_t = \arg\min_{u \ge 0} \left(V^t(u) - (T^R - 1)u\right).$$

As a result,

$$\overline{H}^t(u) \equiv \begin{cases} H^t(u) & u \le L_t \\ H^t(L) & u > L_t \end{cases} \tag{5.19}$$

is a convex and non-increasing function of u.

Let

$$F^t(z,w) \equiv \min_{w \le u \le w + (z-n_t)^+} \left\{V^t(u) - (T^R - 1)u\right\}$$

which can also be defined as follows:

$$F^t(z,w) = \begin{cases} H^t(w) & \text{if } w \ge L_t \\ \bar{H}^t\left(w + (z - n_t)^+\right) & \text{if } w \le L_t \end{cases} \tag{5.20}$$

As a result, relation (5.18) can be rewritten as

$$U^t(z,w) = c(n_t - z)^+ + T^R(z - n_t)^+ + (T^R - 1)w + F^t(z,w) \tag{5.21}$$

Definition 5-1 (Koole (1998)): A function $f(x)$ is supermodular, denoted Super, if

$$f(x) + f(x + e_i + e_j) \ge f(x + e_i) + f(x + e_j)$$

or equivalently

$$f(x \vee y) + f(x \wedge y) \ge f(x) + f(y)$$

where $(x \vee y) = \max(x, y)$.

Definition 5-2 (Koole (1998)): A function *f*(*x*) is superconvex w.r.t. (*i*,*j*), denoted SuperC(*i*,*j*) if

$$f\left(x+e_i\right)+f\left(x+e_i+e_j\right)\leq f\left(x+e_j\right)+f\left(x+2e_i\right)$$

From the above definitions, it can be proved that a function *f*(*x*) that is Super and SuperC(*i*, *j*) is convex in *i*.

Property 5-3. In the value iteration by (5.8), $U^t\left(z,w\right)$ is supermodular if $W^t\left(y_t\right)$ is convex in y_t,.

Proof.

In order to prove the supermodularity of $U^t\left(z,w\right)$, from relation (5.21), we only need to show the supermodularity of $F^t\left(z,w\right)$, i.e. ,

$$F^t\left(z+1,w+1\right)+F^t\left(z,w\right)\geq F^t\left(z+1,w\right)+F^t\left(z,w+1\right) \tag{5.22}$$

In order to prove relation (5.22), three cases are considered.

Case 1: $w+1\geq L_t$.

$$F^t\left(z+1,w+1\right)=F^t\left(z,w+1\right)=H^t\left(w+1\right)$$

which together with the monotonicity of $F^t(z,w)$ in *z* proves relation (5.22).

Case 2: $w+1<L_t\ \&\ z+1\leq n_t$, then $F^t\left(z+1,w'\right)=F^t\left(z,w'\right)=\overline{H}^t\left(w'\right)$ for both $w'=w+1$ and $w'=w$. Hence (5.22) holds.

Case 3: $w+1<L_t\ \&\ z\geq n_t$, then we have

$$F^t\left(z,w\right)=\overline{H}^t\left(w+z-n_t\right),\ F^t\left(z+1,w\right)=\overline{H}^t\left(w+z+1-n_t\right)$$

$$F^t\left(z,w+1\right)=\overline{H}^t\left(w+1+z-n_t\right),\ F^t\left(z+1,w+1\right)=\overline{H}^t\left(w+1+z+1-n_t\right)$$

Relation (5.22) holds as $\overline{H}^t\left(u_t\right)$ is convex. Q.E.D.

Since $U^t\left(z,w\right)$ is super-modularity, $G^t\left(y,w\right)=E\left[b_2w+U^{t+1}\left(y+a_{t+1},w\right)\right]$ is super-modular.

From Topkis (1979), there exists a series of *w* which can reach the minimum of $G^t\left(y,w\right)$, i.e., $w^*(y)=\arg\min_{0\leq w\leq n_{t+2}} G^t\left(y,w\right)$.

Property 5-4. The maximum and minimal selection of $w^*(y)$, $\overline{w}(y)$ and $\underline{w}(y)$, are non-increasing functions if $W^t(y_t)$ is convex in y_t,.

Proof. By contradiction, assume that $\overline{w}(y+1) \geq \overline{w}(y)$. By supermodularity and optimality, there exists

$$0 \leq G^t(y, \overline{w}(y+1)) - G^t(y, \overline{w}(y)) \leq G^t(y+1, \overline{w}(y+1)) - G^t(y+1, \overline{w}(y)) \leq 0 .$$

As a result, $\overline{w}(y+1)$ is also an optimum of $G^t(y,w)$ which contradicts the definition of $\overline{w}(y)$.

The property that $\underline{w}(y)$ is decreasing function can be proved in the same way. Q.E.D.

Property 5-5. In the value iteration by (5.8), $U^t(z,w)$ is superconvex if $W^t(y_t)$ is convex in y_t,.

Proof. In order to prove the superconvexity of $U^t(z,w)$, we need to prove the following relations:

$$U^t(z+1,w)+U^t(z,w+2) \geq U^t(z+1,w+1)+U^t(z,w+1) \tag{5.23}$$

$$U^t(z,w+1)+U^t(z+2,w) \geq U^t(z+1,w+1)+U^t(z+1,w) \tag{5.24}$$

From relation (5.21), to prove relation (5.23), we only need to prove the following relation:

$$F^t(z+1,w)+F^t(z,w+2) \geq F^t(z+1,w+1)+F^t(z,w+1) \tag{5.25}$$

Three cases are considered to prove the relation (5.25):

Case 1: $w+2 \leq L_t$. Relation (5.25) holds if and only if the following relation exists:

$$\begin{aligned} &\overline{H}^t\left(w+(z+1-n_t)^+\right)+\overline{H}^t\left(w+2+(z-n_t)^+\right) \\ &\geq \overline{H}^t\left(w+1+(z+1-n_t)^+\right)+\overline{H}^t\left(w+1+(z-n_t)^+\right) \end{aligned} \tag{5.26}$$

When $z+1 \leq n_t$, relation (5.26) holds due to convexity of $\overline{H}^t(u)$. When $z+1 > n_t$, relation (5.26) obviously holds.

Case 2: $w+1 = L_t$. Relation (5.25) holds if and only if the following relation exists:

$\overline{H}^t\left(L_t-1+(z+1-n_t)^+\right)+H^t(L_t+1) \geq 2H^t(L_t)$ which clearly holds by definition of L_t.

Case 3: $w \geq L_t$. Relation (5.25) holds if and only if $H^t(w)+H^t(w+2) \geq 2H^t(w+1)$ holds by convexity of $H^t(u)$.

Four cases are considered to prove relation (5.24):

Case 1: $w \geq L_t$. Relation (5.24) can be proved by the following relation:

$$F^t(z,w+1)+F^t(z+2,w) \geq F^t(z+1,w+1)+F^t(z+1,w)$$

or equivalently $H^t(w+1)+H^t(w) \geq H^t(w+1)+H^t(w)$, which obviously holds.

Case 2: $w < L_t \,\&\, z \geq n_t$. Relation (5.24) holds because $\overline{H}^t(w+1+z-n_t)+\overline{H}^t(w+2+z-n_t) \geq \overline{H}^t(w+1+z+1-n_t)+\overline{H}^t(w+z+1-n_t)$ obvious holds.

Case 3: $w < L_t \,\&\, z \leq n_t - 2$. Relation (5.24) holds because of $\overline{H}^t(w+1)+\overline{H}^t(w) \geq \overline{H}^t(w+1)+\overline{H}^t(w)$.

Case 4: $w < L_t \,\&\, z = n_t - 1$. To prove relation (5.24), we need to prove the following relation: $c+\overline{H}^t(w+1)+T^R+\overline{H}^t(w+1) \geq \overline{H}^t(w+1)+\overline{H}^t(w)$, or equivalently, $\overline{H}^t(w+1)-\overline{H}^t(w) \geq -(c+T^R)$.

From Property 5-1,

$$H^t(u_t+1)-H^t(u_t)=V^t(u_t+1)-V^t(u_t)-(T^R-1) \geq -b_1-T^R+1 \geq -(c+T^R).$$

In this case, $\overline{H}^t(w+1)-\overline{H}^t(w)=H^t(w+1)-H^t(w) \geq -(c+T^R)$, which concludes relation (5.24). Q.E.D.

Property 5-6. $U^t(z,w)$ is convex in z and convex in w if $W^t(y_t)$ is convex in y_t,.

Proof. This property is a direct result from properties 5-3 and 5-5. Q.E.D.

The superconvexity of $U^t(z,w)$ implies the superconvexity of $G^t(y,w)$.

$$G^t(y+1,w)+G^t(y,w+2) \geq G^t(y+1,w+1)+G^t(y,w+1) \tag{5.27}$$

$$G^t(y,w+1)+G^t(y+2,w) \geq G^t(y+1,w+1)+G^t(y+1,w) \tag{5.28}$$

The supermodularity and superconvexity of $G^t(y,w)$ implies that $G^t(y,w)$ is convex in y and convex in w.

$$G^t(y,w)+G^t(y,w+2)\geq 2G^t(y,w+1)$$

$$G^t(y+2,w)+G^t(y,w)\geq 2G^t(y+1,w)$$

Property 5-7. $\overline{w}(y_t)\leq\overline{w}(y_t+1)+1$, and $\underline{w}(y_t)\leq\underline{w}(y_t+1)+1$ if $W^t(y_t)$ is convex in y_t.

Proof. Let $w=\overline{w}(y_t+1)$. By definition, $G^t(y+1,w)<G^t(y+1,w+1)$.

From (5.27), $G^t(y,w+1)<G^t(y,w+2)$. This together with the convexity concludes the proof.

The second relation can be proved in the same way. Q.E.D.

Hence,

$$\overline{w}(y+1)\leq\overline{w}(y)\leq\overline{w}(y+1)+1 \tag{5.29}$$

$$\underline{w}(y+1)\leq\underline{w}(y)\leq\underline{w}(y+1)+1\,. \tag{5.30}$$

Property 5-8. $W^{t-1}(y)$ is convex in y if $W^t(y_t)$ is convex in y_t

Proof.

$W^{t-1}(y)$ is convex if $W^{t-1}(y+1)-W^{t-1}(y)$ does not decrease. Note that

$$W^{t-1}(y)=\min_{0\leq w\leq n_{t+1}}G^{t-1}(y,w)=\min_{0\leq w\leq n_{t+1}}E\left[b_2w+U^t(y+a_t,w)\right]$$

From relation (5.21), $U^t(z,w)$ increases in w for $w\geq L_t$. In the following, we only need to consider the case of $w\leq L_t$, i.e.

$$W^{t-1}(y)=\min_{0\leq w\leq n_{t+1}\wedge L_t}G^{t-1}(y,w)=\min_{0\leq w\leq n_{t+1}\wedge L_t}E\left[b_2w+U^t(y+a_t,w)\right]$$

Combining with property 5-7,

$$W^{t-1}(y+1)-W^{t-1}(y)=\min_{w\leq L_t\wedge n_{t+1}}\max_{\substack{w'\leq L_t\wedge n_{t+1}\\ w+1\geq w'\geq w}}\left(G^{t-1}(y+1,w)-G^{t-1}(y,w')\right)$$

To prove the convexity of $W^{t-1}(y)$, we only need to prove

$$\begin{aligned}&\min_{w\leq L_t\wedge n_{t+1}}\max_{\substack{w'\leq L_t\wedge n_{t+1}\\ w+1\geq w'\geq w}}\left(G^{t-1}(y+1,w)-G^{t-1}(y,w')\right)\\ &\qquad\geq\min_{w\leq L_t\wedge n_{t+1}}\max_{\substack{w'\leq L_t\wedge n_{t+1}\\ w+1\geq w'\geq w}}\left(G^{t-1}(y,w)-G^{t-1}(y-1,w')\right)\end{aligned} \tag{5.31}$$

Relation (5.31) holds if

$$G^{t-1}(y+1,w)-G^{t-1}(y,w')\geq G^{t-1}(y,w)-G^{t-1}(y-1,w'),\quad \forall w\leq L_t\wedge n_{t+1}, w'\leq L_t\wedge n_{t+1}, w+1\geq w'\geq w \tag{5.32}$$

By definition of $G^{t-1}(.)$, relation (5.32) holds if

$$U^t(z+1,w)-U^t(z,w')\geq U^t(z,w)-U^t(z-1,w'),\quad \forall w\leq w'\leq w+1, w,w'\leq L_t \tag{5.33}$$

Three cases are considered to prove this relation.

Case 1: $z+1\leq n_t$. Relation (5.33) is equivalent with

$$\overline{H}^t(w)-\overline{H}^t(w')\geq \overline{H}^t(w)-\overline{H}^t(w')$$

which obviously holds.

Case 2: $z-1\geq n_t$. Relation (5.33) is equivalent with

$$\overline{H}^t(w+z+1-n_t)-\overline{H}^t(w'+z-n_t)\geq \overline{H}^t(w+z-n_t)-\overline{H}^t(w'+z-1-n_t)$$

which holds because of the convexity of $\overline{H}^t(u_t)$ and $w\leq w'\leq w+1$.

Case 3: $z=n_t$. Relation (5.33) is equivalent to:

$$T^R+\overline{H}^t(w+1)-\overline{H}^t(w')\geq \overline{H}^t(w)-c-\overline{H}^t(w')$$

which holds if $\overline{H}^t(w)-\overline{H}^t(w+1)\leq T^R+c$. The last relation obviously holds for $w=L_t$ as $\overline{H}^t(w)-\overline{H}^t(w+1)=0$. If $w<L_t$, then

$$\overline{H}^t(w)-\overline{H}^t(w+1)=H^t(w)-H^t(w+1)=V^t(w)-V^t(w+1)+T^R-1\leq b_1+T^R\leq c+T^R$$

where Property 5-1 is applied in the above. Q.E.D.

Remark 5-2: (Property 5-2) – (Property 5-7) hold because of Property 5-8.

Theorem 5-2: The optimal value function *U*(*i*, *z*, *w*) in relation (5.11) is convex in *z* and *w*. *V*(*i*, *u*) in relation (5.12) is convex in *u*. *W*(*i*, *y*) in relation (5.13) is convex in *y*. Further, the optimal control policies for problem (5.4) are as follows:

i) the optimal CTS queue length at the end of day *i* is of the following form:

$$x_i^*=\begin{cases}0 & \text{if } z_i-n_i\leq 0\\ z_i-n_i & \text{if } 0\leq z_i-n_i\leq L_i-w_i^2\\ L_i-w_i^2 & \text{if } z_i-n_i\geq L_i-w_i^2\end{cases} \tag{5.34}$$

And the optimal number of patients assigned to RTS at the end of day i is of the following form:

$$y_i^* = \left(z_i - n_i - x_i^*\right)^+$$

ii) the optimal one-day advance CTS cancellation policy for day i+1 is of the following form:

$$w_{i+1}^{1*} = \begin{cases} S_{i+1}^1 - x_i - w_{i+1}^2 & \text{if } x_i + w_{i+1}^2 \le S_{i+1}^1 \\ 0 & \text{if } x_i + w_{i+1}^2 \ge S_{i+1}^1 \end{cases} \tag{5.35}$$

iii) the optimal two-day advance CTS cancellation policy for day i+2 is of the following form:

$$w_{i+2}^{2*}\left(y_i\right) = S_{i+2}^2\left(y_i\right) \tag{5.36}$$

Where

$$L_i = \underset{w^2 \le u \le w^2 + \left(\left(z_i - n_i\right)^+ \wedge \bar{x}\right)}{\arg\min} \left(V\left(i,u\right) - \left(T^R - 1\right)u\right),$$

$$S_{i+1}^1 = \underset{x_i + w_{i+1}^2 \le y \le \max\left\{x_i + w_{i+1}^2, n_{i+1}\right\}}{\arg\min} \left\{W_i\left(y\right) + b_1 y\right\},$$

$$S_{i+2}^2\left(y\right) = \underset{0 \le w \le n_{i+2}}{\arg\min} \left\{E\left[b_2 w + U_{i+1}\left(y + a_{i+1}, w\right)\right]\right\}.$$

Proof: The convexity of $U(i, z, w)$, $V(i, x+w)$, and $W(i, y)$ is a direct consequence of relations (5.7), (5.14), (5.15) and Property 5-2, 5-6, 5-8. The theorem can be directly derived from the convexity. Q.E.D.

5.4 Properties of optimal control policies for average cost problem

5.4.1 Bounded demand case

The following assumption is also made for the average cost problem case.

Assumption 5-A3. There exists a finite number A such that $a_t \le A$, for all t.

The combination of Assumptions 5-A2 and 5-A3 implies that the state variable z_t is upper bounded and:

$$z_t \le \bar{z} \equiv \bar{x} + A + n^* \tag{5.37}$$

as $x_{t-1}+w_t^1+w_t^2 \le \max(x_{t-1},n_t)+w_t^2 \le \bar{x}+n^*$. The other state variable w is also upper bounded as $w \le n^*$. As a result, under Assumptions 5-A2 and 5-A3, the stage cost function is also bounded with

$$g_t\left(z_t,x_t,w_t\right) \le B \equiv \left(b_1+b_2+c\right)n^* + T^R\bar{z} \tag{5.38}$$

Property 5-9: There exists $M>0$ such that $\left|U_\alpha\left(i,z,w\right)-U_\alpha\left(7,0,0\right)\right| \le M$, for all i = 1, …,7 and for all z and w.

Proof: From Property 5-1 and relation (5.8),

$$-c \le U_\alpha^t\left(z,w\right)-U_\alpha^t\left(z-1,w\right) \le T^R$$

$$-b_1 \le U_\alpha^t\left(z,w\right)-U_\alpha^t\left(z,w-1\right) \le T^R$$

which, together with the finiteness of the state space,

$$-c\bar{z} \le U_\alpha^t\left(z,w\right)-U_\alpha^t\left(z',w\right) \le T^R\bar{z}$$

$$-b_1 n^* \le U_\alpha^t\left(z',w\right)-U_\alpha^t\left(z',w'\right) \le T^R n^*$$

for all z, z', w, and w'. Therefore,

$$-C\left(\bar{z}+n^*\right) \le U_\alpha^t\left(z,w\right)-U_\alpha^t\left(z',w'\right) \le C\left(\bar{z}+n^*\right)$$

with C = Max($2T^R$, b_1+c). Combining with relation (5.7),

$$-C\left(\bar{z}+n^*\right) \le U_\alpha\left(i,z,w\right)-U_\alpha\left(i,z',w'\right) \le C\left(\bar{z}+n^*\right) \tag{5.39}$$

This establish the property for i = 7. Consider now the case i = 1, …, 6. From (5.5) and (5.7),

$$U_\alpha\left(i,z,w\right) \le B + E_a\left[U_\alpha\left(i+1,z,w\right)\right] \tag{5.40}$$

Repeat the relations (5.40) for t subsequent days leads to:

$$U_\alpha\left(i,z,w\right) \le tB + \sum_{z'} p^{\mu^1\mu^2\pi}_{(i,z,w),(t+i,z',w')} U_\alpha\left(t+i,z',w'\right), \forall i=1,...,7 \tag{5.41}$$

where $p^{\mu^1\mu^2\pi}_{(i,z,w),(t+i,z',w')}$ is the probability of reaching state (z',w') at the beginning of day $t+i$ by starting from state (z, w) at day i under policies μ^1, μ^2 and π. Combining (5.39) and (5.41) with $t+i$=7,

$$U_\alpha\left(i,z,w\right) \le 6B + \sum_{z',w'} p^{\pi}_{(i,z,w),(7,z',w')} U_\alpha\left(7,z',w'\right) \le 6B + U_\alpha\left(7,0,0\right) + C\left(\bar{z}+n^*\right) \tag{5.42}$$

Similarly,

$$U_\alpha(7,0,0) \le 6B + \sum_{z',w'} p^{\pi}_{(7,0,0),(7+i,z',w')} U_\alpha(i,z',w') \le 6B + U_\alpha(i,z,w) + C(\overline{z} + n^*) \quad (5.43)$$

Relations (5.42)-(5.43) conclude the proof with $M = 6B + C(\overline{z} + n^*)$. Q.E.D.

Theorem 5-3. There exists an optimal stationary control policy for the average cost model (5.1). Further the optimal average cost is independent of the initial state (*i*, *z, w*).

Proof. From Proposition 4.2.6 in Bertsekas (1996), the optimal average cost per day exists and has the same value λ for all initial states, and λ satisfies

$$\lambda = \lim_{\alpha \to 1} (1-\alpha) U_\alpha(i,z,w) \quad (5.44)$$

The differential cost functions

$$\psi(i,z,w) = U_\alpha(i,z,w) - U_\alpha(7,0,0) \quad (5.45)$$

satisfy the following optimality equations:

$$\lambda + \psi(i,z,w) = \min_{x,y,w'} \begin{pmatrix} c(n_i - z)^+ + x + T^R(z - n_i - x)^+ + b_1(y - x - w) \\ +b_2 w' + \sum_a P_{i+1,a} \psi(i+1, y+a, w') \end{pmatrix} \quad (5.46)$$

Relation (5.46) can be rewritten as

$$\psi(i,z,w) = \min_{0 \le x \le ((z-n_i)^+, \overline{x})} \left\{ c(n_i - z)^+ + x + T^R(z - n_i - x)^+ + \psi^1(i, x+w) \right\} \quad (5.47)$$

$$\psi^1(i,u) = \min_{u \le y \le \max\{u, n_{i+1}\}} \left\{ \psi^2(i,y) + b_1(y-u) \right\} \quad (5.48)$$

$$\psi^2(i,y) = \min_{0 \le w \le n_{i+2}} E\left[b_2 w + \psi(i+1, y+a, w) \right] - \lambda \quad (5.49)$$

Relations (5.45), (5.48)-(5.49) and (5.12)-(5.13) implies that

$$\psi^1(i,u) = \lim_{\alpha \to 1} \left(V_\alpha(i,u) - U_\alpha(7,0,0) \right) \quad (5.50)$$

$$\psi^2(i,y) = \lim_{\alpha \to 1} \left(W_\alpha(i,y) - U_\alpha(7,0,0) \right) \quad (5.51)$$

Further, the optimal control policy is stationary deterministic and is defined by the argument that reaches the minimum in (5.46) or equivalently (5.47)-(5.49). From Theorem 5-2, relation (5.45), (5.50), and (5.51), $\psi(i, z, w)$ is convex in *z* and convex in *w*, $\psi^1(i, u)$ is convex in *u*, $\psi^2(i, y)$ is convex in *y*, for all $i = 1, \ldots, 7$. The optimal control policy can be shown to be of the same form as that of Theorem 5-2. Q.E.D.

5.4.2 Unbounded demand case

Property 5-10: For any $z^1 \geq z^2 \geq 0$, $w^1 \geq w^2 \geq 0$, $-m \leq U_\alpha^t\left(z^1, w^1\right) - U_\alpha^t\left(z^2, w^2\right) \leq m + T^R z^1$ with $m = \left(b_1 + T^R + c\right)\bar{x}$.

Proof: The property trivially holds for $t = 0$. Consider the case $t < 0$. Since the "min" term in equation (5.18) is decreasing in z, subtracting equation (5.18) with $z = z^1$ by equation (5.18) $z_t = z^2$ leads to:

$$U_\alpha^t\left(z^1, w^1\right) - U_\alpha^t\left(z^2, w^1\right) \leq c\left(n_t - z^1\right)^+ + T^R\left(z^1 - n_t\right)^+ - c\left(n_t - z^2\right)^+ - T^R\left(z^2 - n_t\right)^+$$
$$\leq cn^* + T^R z^1$$

Since the "min" term in equation (5.18) is decreasing in w, subtracting equation (5.18) with $w = w^1$ by equation (5.18) $w = w^2$ leads to:

$$U_t\left(z^2, w^1\right) - U_t\left(z^2, w^2\right) \leq \left(T^R - 1\right)w^1 - \left(T^R - 1\right)w^2 \leq \left(T^R - 1\right)n^*$$

Therefore, as $n^* \leq \bar{x}$,

$$U_\alpha^t\left(z^1, w^1\right) - U_\alpha^t\left(z^2, w^2\right) = U_\alpha^t\left(z^1, w^1\right) - U_\alpha^t\left(z^2, w^1\right) + U_\alpha^t\left(z^2, w^1\right) - U_\alpha^t\left(z^2, w^2\right)$$
$$\leq \left(T^R - 1 + c\right)n^* + T^R z^1 \leq m + T^R z^1$$

Let u' be the argument reaching minimum in (5.18) with $z = z^2$, $w = w^1$. As a result,

$$U_\alpha^t\left(z^2, w^1\right) = c\left(n_t - z^2\right)^+ + T^R\left(z^2 - n_t\right)^+ + \left(T^R - 1\right)w^1 + V_\alpha^t\left(u'\right) - \left(T^R - 1\right)u' \quad (5.52)$$

Subtracting equation (5.18) with $z = z^1$, $w = w^1$ by equation (5.52) leads to:

$$U_\alpha^t\left(z^1, w^1\right) - U_\alpha^t\left(z^2, w^1\right) = c\left(n_t - z^1\right)^+ + T^R\left(z^1 - n_t\right)^+ - c\left(n_t - z^2\right)^+ - T^R\left(z^2 - n_t\right)^+$$
$$+ \min_{u \in \left[w^1, w^1 + \left(z^1 - n_t\right)^+ \wedge \bar{x}\right]} \left\{V_\alpha^t(u) - V_\alpha^t\left(u'\right) - \left(T^R - 1\right)\left(u - u'\right)\right\} \quad (5.53)$$
$$\geq cn^* + \min_{u \in \left[w^1, w^1 + \left(z^1 - n_t\right)^+ \wedge \bar{x}\right]} \left\{V_\alpha^t(u) - V_\alpha^t\left(u'\right) - \left(T^R - 1\right)\left(u - u'\right)\right\}$$

Since $u \geq u'$, from Property 5-1,

$$V_\alpha^t(u) - V_\alpha^t\left(u'\right) - \left(T^R - 1\right)\left(u - u'\right) \geq -b_1 u - T^R u$$

Therefore,

$$U_\alpha^t\left(z^1, w^1\right) - U_\alpha^t\left(z^2, w^1\right) \geq cn^* - b_1 u - T^R u \geq cn^* - \left(b_1 + T^R\right)\bar{x} \quad (5.54)$$

Subtracting equation (5.52) by equation (5.18) with $z = z^2$, $w = w^2$ leads to:

$$U_\alpha^t\left(z^2, w^1\right) - U_\alpha^t\left(z^2, w^2\right) = \left(T^R - 1\right)\left(w^1 - w^2\right) - \min_{u \in \left[w^2, w^2 + \left(z^2 - n_t\right)^+ \wedge \overline{x}\right]} \left\{V_\alpha^t(u) - V_\alpha^t(u') - \left(T^R - 1\right)(u - u')\right\}$$

Since $u' \geq u$, from Property 5-1,

$$U_\alpha^t\left(z^2, w^1\right) - U_\alpha^t\left(z^2, w^2\right) \geq \left(T^R - 1\right)\left(w^1 - w^2\right) - b_1\overline{x} \geq -b_1\overline{x} \tag{5.55}$$

Combining relations (5.54)-(5.55) leads to:

$$U_\alpha^t\left(z^1, w^1\right) - U_\alpha^t\left(z^2, w^2\right) \geq -\left(2b_1 + T^R\right)\overline{x} \geq m$$

Q.E.D.

Property 5-11: There exist M>0 and r > 0 such that $-M \leq U_\alpha\left(i, z, w\right) - U_\alpha\left(7, 0, 0\right) \leq M + rz$, for all i = 1, …,7 and for all z.

Proof: From Property 5-10,

$$-m \leq U_\alpha^t\left(z, w\right) - U_\alpha^t\left(z', w'\right) \leq m + T^R z, \forall z \geq z'.$$

Combining with relation (5.7),

$$-m \leq U_\alpha\left(i, z, w\right) - U_\alpha\left(i, z', w'\right) \leq m + T^R z, \forall z \geq z', w \geq w'$$

This establish the property for i = 7. Further

$$U_\alpha\left(i, z, w\right) \leq m + T^R z + U_\alpha\left(i, z', w'\right), \forall z, z', w, w' \tag{5.56}$$

Consider now the case i = 1, …, 6. First,

$$0 \leq g_t\left(z_t, w_{t+1}^2, x_t, w_{t+1}^1, w_{t+2}^2\right) \leq \left(b_2 + c\right)n^* + T^R z_t \leq m + T^R z_t.$$

From (5.7),

$$U_\alpha\left(i, z, w\right) \leq m + T^R z + E\left[U_\alpha\left(i+1, z_{i+1}, w_{i+2}^2\right)\right] \tag{5.57}$$

Repeat the relations (5.57) for t subsequent days leads to:

$$U_\alpha\left(i, z, w\right) \leq \sum_{\tau=i}^{t+i-1}\left(m + T^R E\left[z_\tau\right]\right) + E\left[U_\alpha\left(t+i, z_{t+i}, w_{t+i+1}^2\right)\right], \forall i = 1, ..., 7 \tag{5.58}$$

Combining (5.56) and (5.58) with t+i=7,

$$U_\alpha(i,z,w) \le \sum_{\tau=i}^{7}\left(m+T^R E[z_\tau]\right)+\left(U_\alpha(7,0,0)+m+T^R(z+6a^*)\right), \forall i=1,...,7 \quad (5.59)$$

Since $z_t \le z+a_{i+1}+w_{i+1}^1+w_{i+1}^2+...+a_t+w_t^1+w_t^2 \le z+a_{i+1}+2n^*+...+a_t+2n^*$, for $t > i$,

$$\begin{aligned} U_\alpha(i,z,w) &\le 7m+49T^R\left(a^*+n^*\right)+7T^R z+U_\alpha(7,0,0)+m+T^R(z+6a^*) \\ &\le 8m+T^R(55a^*+49n^*)+8T^R z+U_\alpha(7,0,0), \forall i=1,...,7 \end{aligned} \quad (5.60)$$

Similarly,

$$\begin{aligned} U_\alpha(7,0,0) &\le \sum_{\tau=7}^{7+i}\left(m+T^R E[z_\tau]\right)+\left(U_\alpha(i,z,w)+m+6T^R a^*\right) \\ &\le 7m+49T^R\left(a^*+n^*\right)+m+6T^R a^* U_\alpha(i,z,w) \\ &\le 8m+T^R(55a^*+49n^*) \end{aligned} \quad (5.61)$$

Relations (5.60)-(5.61) conclude the proof with $M=8m+T^R(55a^*+49n^*)$ and $r=8T^R$.

Q.E.D.

Theorem 5-4. Under Assumptions 4, and 5-A2, (a) there exists a constant λ satisfying (5.44) for all (*i*, *z*, *w*), a matrix *ψ*(*i*, *z*, *w*) satisfying (5.45)-(5.46) (b) the optimal control policy is defined by the argument that reaches the minimum in (5.46), (c) there exists an optimal stationary control policy of the form of equations (5.34)-(5.36) for the average cost model.

Proof: The proof is based on Theorem 8.10.7 of Puterman (1994) and the conditions that need to be checked are the following ones:

C1: For each state (*i*, *z*, *w*), the stage cost is such that $-\infty < R \le g_i\left(z,w,x_i,w_{i+1}^1,w_{i+2}^2\right) < \infty$.

C2: For each (*i*, *z*, *w*) and α < 1, $U_\alpha(i,z,w) < \infty$.

C3: There exists K $> -\infty$ such that, for each (*i*, *z*, w^2),

$$\psi_\alpha(i,z,w) \equiv U_\alpha(i,z,w) - U_\alpha(7,0,0) \ge K, \forall \alpha < 1.$$

C4: There exists a non-negative function W(*i*, *z*, *w*) such that

a) W(*i*, *z*, *w*) < ∞;

b) for each (*i*, *z*, *w*), $\psi_\alpha(i,z,w) \le W(i,z,w), \forall \alpha < 1$; and

c) for each (*i*, *z*, *w*) and w_{i+1}^1, w_{i+1}^2 and x_i,

$$\sum_a P_{i+1,a} W\left(i+1, x_i+w_{i+1}^1+w_{i+1}^2+a\right) < \infty.$$

According to Theorem 8.10.7 of Puterman (1994), as the control constraint set for each state (*i*, *z*, *w*) is finite, (a) and (b) of the Theorem hold. Further $\psi(i, z, w)$ is the limit of a sequence $\psi_{\alpha_m}(i,z,w)$ such that α_m converges to 1 and $\psi_{\alpha_m}(i,z,w)$ converges for all (*i*, *z*, *w*). From Property 5-2, 5-6, 5-8, and equations(5.7), (5.14), and (5.15), (c) of the Theorem can be proved as for Theorem 5-2.

Let us now prove conditions C1-C4. Condition C1 clearly holds as $g_i\left(z,w,x_i,w_{i+1}^1,w_{i+2}^2\right)\geq 0$. Condition C2 holds as well as

1) as in Property 5-11, $0\leq g_t\left(z,w,x_t,w_{t+1}^1,w_{t+2}^2\right)\leq\left(b_2+c\right)n^*+T^R z_t\leq m+T^R z_t$

2) $E\left[g_t\left(z,w,x_t,w_{t+1}^1,w_{t+2}^2\right)\right]\leq E\left[m+T^R z_t\right]\leq m+T^R\left(\bar{x}+n^*+a^*\right)$;

3) $U_\alpha\left(i,z,w\right)\leq\left(c+b_2\right)n^*+T^R z+\dfrac{\alpha}{1-\alpha}\left(m+T^R\left(\bar{x}+n^*+a^*\right)\right)$.

Condition C3 is guaranteed by Property 5-11 with K = -M. Condition C4 is a consequence of Property 5-11 with W(*i*, *z*, *w*) = M + *rz*. Q.E.D.

5.4.3 Computation and implementation of the optimal control policies

As proved in Theorem 5-3 and 5-4, there exit the same optimal control policies for average-cost MDP and discounted-cost MDP. The related optimal control policies$\pi(\mathbf{n})$, $\mu^1(\mathbf{n})$, and $\mu^2(\mathbf{n})$ can be determined by solving the following LP model:

$$J\left(\mathbf{n}\right)\equiv\text{maximize }\lambda$$

Subject to

$$\psi\left(i,z,w_{i+1}^2\right)\leq c\left(n_i-z\right)^+ + x+T^R\left(z-n_i-x\right)^+ +\psi^1\left(i,x,w_{i+1}^2\right),\forall i=1,...,7$$

$$\psi^1\left(i,x+w_{i+1}^2\right)\leq\psi^2\left(i,y\right)+b_1\left(y-x-w_{i+1}^2\right),\forall i=1,...,7$$

$$\lambda+\psi^2\left(i,y\right)\leq b_2 w_{i+2}^2+\sum_a P_{i+1,a}\psi\left(i+1,y+a,w_{i+2}^2\right),\forall i=1,...,7$$

$$\forall x_i\leq\left(z_i-n_i\right)^+\wedge\bar{x},x+w_{i+1}^2\leq y\leq\max\left\{x+w_{i+1}^2,n_{i+1}\right\},0\leq w_{i+2}^2\leq n_{i+2},\forall i=1,...,7$$

Where J(**n**) is optimal average cost for problem (5.1) under contract **n**. The optimal controls are given respectively by *x* *y* and w^2 reaching equality in the above relations. Further, optimal RTS assignment and one-day advance cancellation are characterized by two control threshold vectors **L** and $\mathbf{S^1}$. However, optimal two-day advance cancellation $\mathbf{S^2(y)}$ is not always a simple threshold policy. From relations x_i=min $(L_i, -w^2_i, (z_i - n_i)^+)$, and w^1_{i+1}=max $((S^1_{i+1} - x_i - w^2_{i+1})^+, 0)$, $w^2_{i+2} = S^2_{i+2}(y)$, the optimal controls **L**, $\mathbf{S^1}$ and $\mathbf{S^2(y)}$ can be easily determined.

The existence of optimal control policies makes the implementation easy. At the end of day t, the implementation of the optimal patient assignment policy first determines the CTS queue length x_t which depends on state variable z_t and w^2_t. The next step is to determine the number of CTS cancelled for day t+1, i.e., w^1_{t+1}, which depend on w^2_{t+1} and x_t. The final step is to make the two-day advance cancellation decision. The number of CTS cancelled for day t+2 depends on state variable $y_t = x_t + w^1_{t+1} + w^2_{t+1}$.

Step 1: The implementation of the optimal patient assignment control policy can be divided into three cases:

Case 1: As shown in Fig. 5.1, if state variable z_t is smaller than n_t, then there exists the number $n_t - z_t$ of unused CTS, and no patients waiting for the incoming time slots.

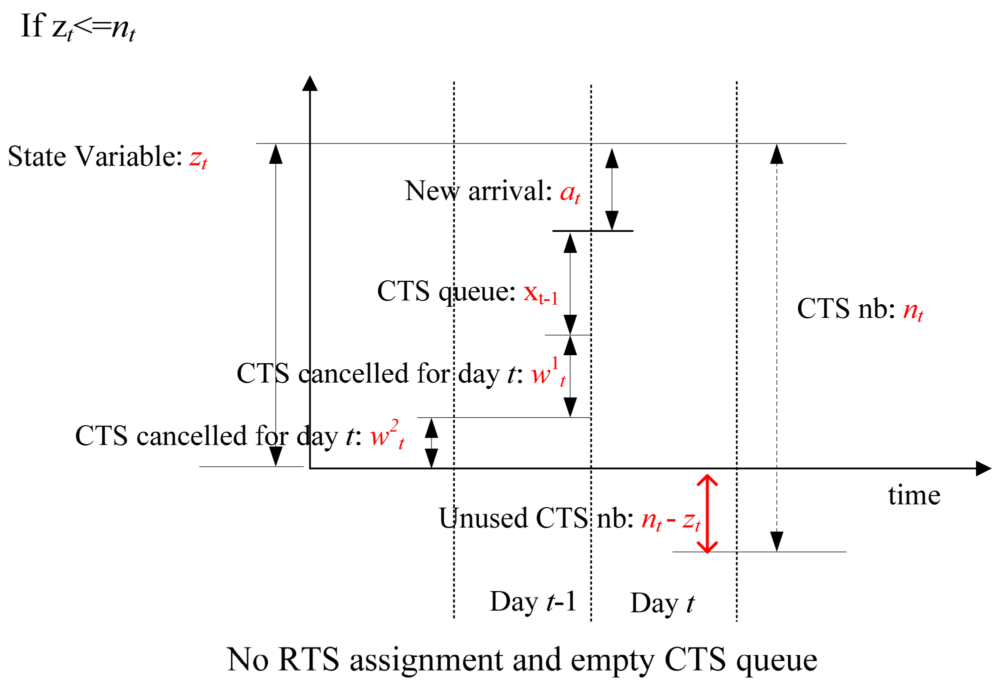


FIG. 5.1 The optimal patient assignment control if z_t <=n_t

Case 2: As shown in Fig. 5.2, if state variable z_t is greater than n_t but smaller than $L_t + n_t - w^2_t$, then all the remaining patients are kept in the CTS queue and no patients are assigned to RTS.

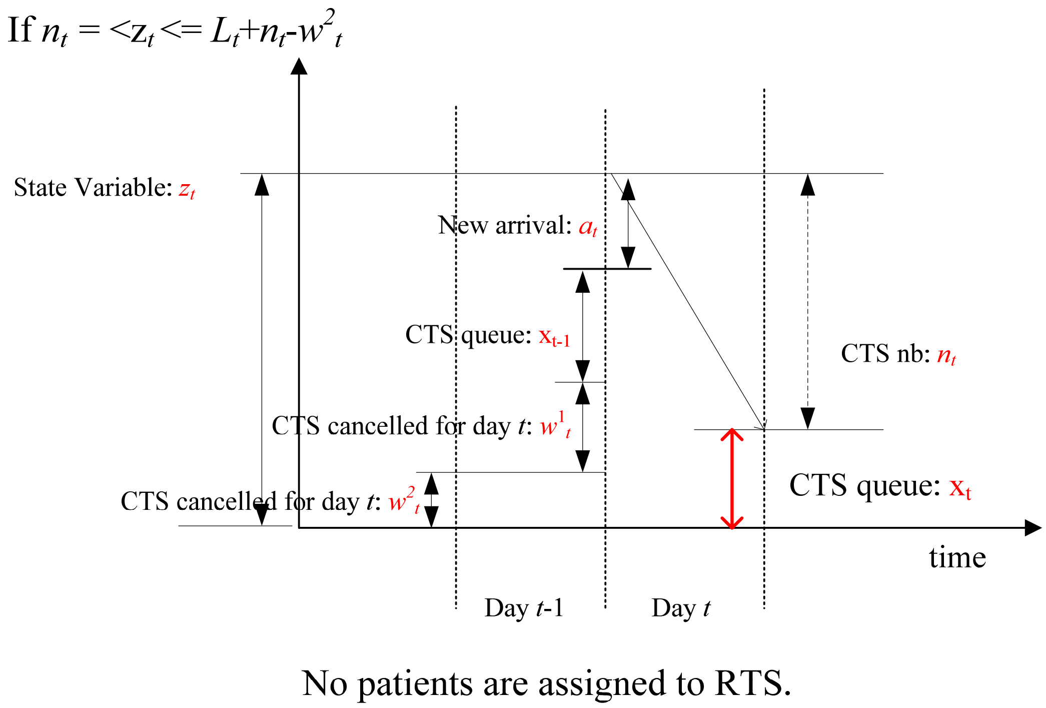


FIG. 5.2 The optimal patient assignment control if $n_t =< z_t <= L_t + n_t - w^2_t$

Case 3: As shown in Fig. 5.3, if state variable z_t is greater than $L_t + n_t - w^2_t$, then the number of patients assigned to CTS is kept at $L_t - w^2_t$, and the other remaining patients are assigned to RTS.

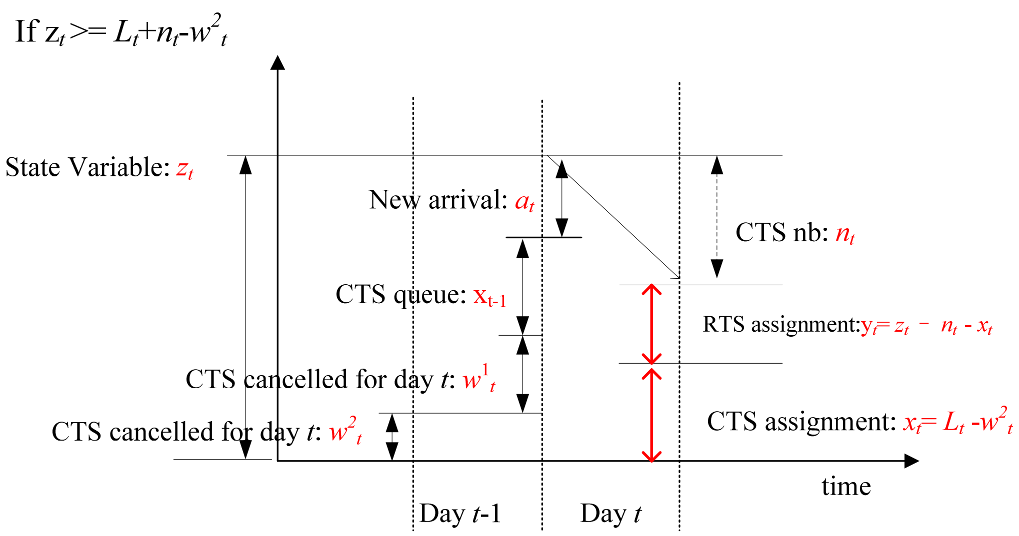


FIG. 5.3 The optimal patient assignment control if $z_t >= L_t + n_t - w^2_t$

Step 2: The implementation of one-day advance cancellation control can be divided into two cases:

Case 1: As shown in Fig. 5.4, if the ending CTS queue at day t plus two day advance cancellation for day t+1, w^2_{t+1},is smaller than S^1_{t+1}, then the number of CTS cancelled for day t+1 is $w^1_{t+1} = S^1_{t+1} - x_t - w^2_{t+1}$.

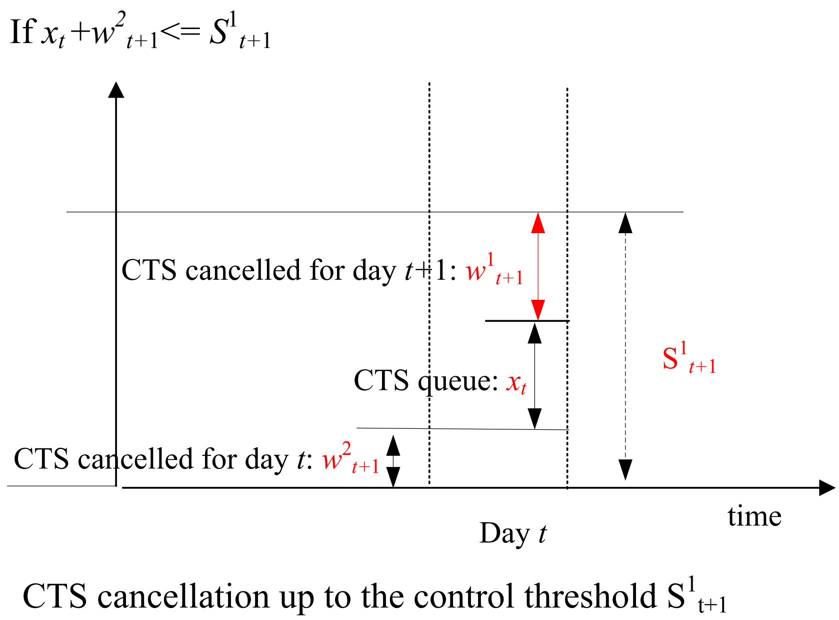


FIG. 5.4 The optimal one-day advance CTS cancellation control if $x_t + w^2_{t+1} <= S^1_{t+1}$

Case 2: If $x_t + w^2_{t+1} >= S^1_{t+1}$, no CTS is cancelled for day t+1.

Step 3: The implementation of two-day advance cancellation depends on state variable $y_t = x_t + w^1_{t+1} + w^2_{t+1}$ which becomes known now. So the number of CTS cancelled for day t+2 is $\mathbf{S^2(y)}$.

As stated in Section 3.3.3 and 4.3.3, the implementation of the contract-based MRI examination reservation process needs the aids of patients scheduling policy. To reduce the variance of patients' waiting times, more work is needed about patients scheduling.

5.5 Local Optimization of Contract

Starting from a given initial contract, this section presents a local search for improving the contract by considering patient assignment, one day advance CTS cancellation, and two-day advance CTS cancellation policies. This local search relies on the structure properties of the previous section especially the optimality equations (5.47)-(5.49) for contract evaluation.

The local search starts from an initial contract $\mathbf{n}^0$. It then iteratively improves this contract. At each iteration, it determines the best neighbor solution among the set of contracts: $\mathbf{n} + \mathbf{e}_k$ (increasing one time slot in day k), $\mathbf{n} - \mathbf{e}_k$ (reducing one time slot in day k), $\mathbf{n} - \mathbf{e}_k + \mathbf{e}_j$ (move one time slot from day k to day j). This process repeats until no improvement can be found.

The overall algorithm for the contract optimization is summarized as follows:

Algorithm (Contract optimization)

1. Select an initial contract $\mathbf{n}^0$, determine the optimal control policies $\pi(\mathbf{n}^0)$, $\mu^1(\mathbf{n}^0)$, $\mu^2(\mathbf{n}^0)$and the optimal average cost $J(\mathbf{n}^0)$ under contract $\mathbf{n}^0$ by solving LP model;

2. Let $\mathbf{n}^* = \mathbf{n}^0$; $J(\mathbf{n}^*) = J(\mathbf{n}^0)$;

3. Determine the neighbor solution $\mathbf{n}'$ with the smallest average cost as follows:

$$\mathbf{n}' = \underset{\mathbf{n}\in\{\mathbf{n}+e_k;\mathbf{n}-e_k;\mathbf{n}-e_k+e_j:1\le k,j\le 7,k\ne j\}\cap IN^7}{\arg\min} J(\mathbf{n})$$

4. If $J(\mathbf{n}') < J(\mathbf{n}^*)$, set $\mathbf{n}^* = \mathbf{n}'$ and go to step 3;

5. The final contract is $\mathbf{n}^*$ and the final control policy is $\pi(\mathbf{n}^*)$ and $\mu(\mathbf{n}^*)$.

5.6 Computational Results

This section presents numerical results to show the benefit of two-day advance CTS cancellation. All numerical experiments are performed on a Intel(R) Core (TM)2 Duo CPU T7250 based PC running at 2.00 GHz with 3.0 GB of Memory. The optimal control policies for the MDP formulation (5.1) are obtained by solving LP model with CPLEX 11 solver.
The numerical experiments are all derived from the base case corresponding to our real case study. From the data collected from the neurovascular department of our study, the average numbers of patient arrivals during the week are as follows: {1, 0.89, 0.95, 1.16, 1.53, 0.16, 0.05}. The number of patients arrived each day is assumed to follow a Poisson distribution truncated at A = 20 which is large enough such that the probability of $a_i > A$ can be neglected. The waiting time for RTS varies from 30 to 40 days with an average of $T^R = 35$ days. The weighting factor of unused CTS, c, is set to 15. CTS one-day advance cancellation cost, b_1, is taken as half of c, i.e. 7.5. Two-day advance cancellation cost, b_2, is assumed as half of b_1, i.e. 3.75.

In the following, the impact of CTS cancellation is analyzed with respect to the cancellation cost b_1, b_2, unused CTS cost c, delay T^R of regular reservation, patient arrival pattern, and the patient arrival rate. The initial contract $\mathbf{n}^0$ is obtained by the method proposed in chapter 4 for optimizing the contract without two-day advance CTS cancellation.

For each case, three solutions are considered: (i) the case with only patient assignment and one-day advance cancellation considered for the given contract $\mathbf{n^0}$. This solution will be denoted **"One-day-cancel"**; (ii) the case by considering one-day, two-day advance cancellation and patient assignment control policies for the given contract $\mathbf{n^0}$. This solution will be denoted **"Two-day-cancel"**; (iii) the contract obtained with local search starting from $\mathbf{n^0}$. This solution will be denoted "**LocalOpt**".

The three solutions are further compared with respect to different performance criteria including the average delay, the unused CTS ratio, the percentage of patients using RTS, the percentage of CTS cancelled.

Note that two-day advance cancellation control policy can be written in the following form for all the instances except for the high demand instance, shown in Fig. 5.5:

$$S_{i+2}^{2}\left(y_i\right)=\begin{cases} X_{i+2} & \text{if } y_i \le Y_{i+2} \\ X_{i+2}+Y_{i+2}-y_i & \text{if } Y_{i+2} \le y_i \le X_{i+2}+Y_{i+2}, \forall i=1,...,7 \\ 0 & \text{else} \end{cases} \tag{5.62}$$

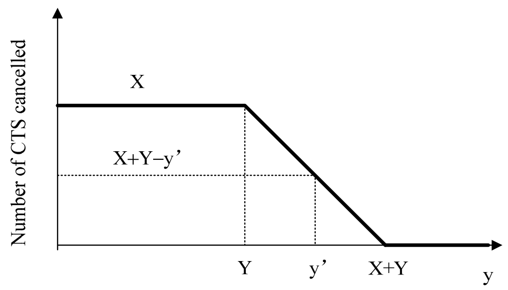


FIG. 5.5 Approximated two-day advance CTS cancellation control policy

5.6.1 Impact of two-day advance cancellation cost

This subsection considers the impacts of the two-day advance CTS cancellation cost b_2 by varying b_2 from 0 to 0.9 b_1.

Fig. 5.6 compares the average cost of the three solutions "One-day-cancel', "Two-day-cancel", and "LocalOpt". From this figure, we note that the two-day advance cancellation cost b_2 has a great impact on the benefit of two-day advance CTS cancellation. Compared with "One-day-cancel", the gain of two-day advance cancellation decreases with the increase of b_2, for example, the gain is 11.95% when b_2=0.1b_1, 3.65% when b_2=0.5b_1, and 0 when b_2=0.8b_1 and b_2=0.9b_1. Similarly, the local optimization further improves both the contract and control policies. The gain of local optimization with respect to the solution of

"Two-day-cancel" also decreases with the increase of b_2 with a gain of 30.97% when b_2=0.1b_1, 7.28% when b_2=0.5b_1, and 0 when b_2=0.8b_1 and b_2=0.9b_1. The total improvement of the two-day advance CTS cancellation and local optimization with respect to the solution "One-day-cancel" decreases from 39.22% when b_2=0.1b_1 to 0% when b_2=0.9 b_1.

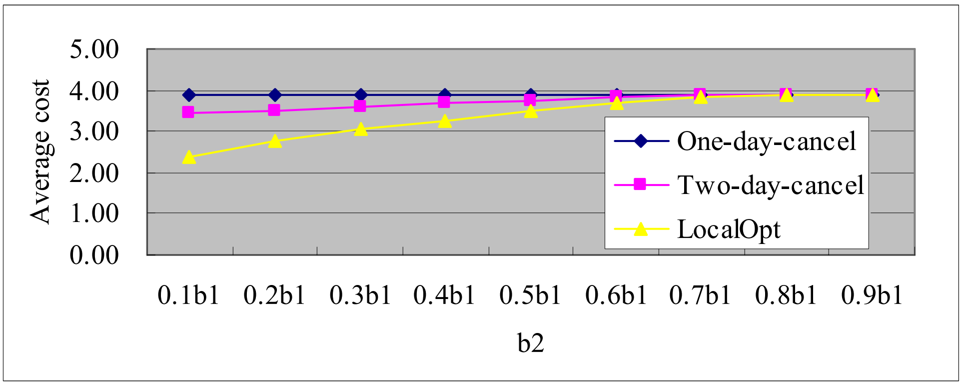


FIG. 5.6 Average costs vs two-day advance cancellation costs b_2

b_2	One-day-cancel				Two-day-cancel				LocalOpt			
	Delay (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)	*Delay* (days)	*Unused* (%)	*RTS* (%)	*Cancel* (%)
0.1b1	2.97	0.98	0.44	17.38	3.19	2.47	0.63	16.05	2.21	0.45	0.00	42.15
0.2b1	2.97	0.98	0.44	17.38	3.05	3.48	0.50	14.94	2.34	0.49	0.00	35.74
0.3b1	2.97	0.98	0.44	17.38	3.05	3.48	0.50	14.94	2.25	3.00	0.01	25.26
0.4b1	2.97	0.98	0.44	17.38	3.05	3.48	0.50	14.94	2.25	3.00	0.01	25.26
0.5b1	2.97	0.98	0.44	17.38	3.05	3.48	0.50	14.94	2.25	3.00	0.00	25.26
0.6b1	2.97	0.98	0.44	17.38	3.06	2.85	0.47	15.54	2.25	3.00	0.00	25.26
0.7b1	2.97	0.98	0.44	17.38	2.95	1.67	0.44	16.70	3.06	1.67	0.47	16.72
0.8b1	2.97	0.98	0.44	17.38	2.97	0.98	0.44	17.38	2.97	0.98	0.44	17.38
0.9b1	2.97	0.98	0.44	17.38	2.97	0.98	0.44	17.38	2.97	0.98	0.44	17.38

TAB. 5.1 Performance comparison for different two-day advance cancellation cost b_2

Table 5.1 summarizes the performance measures for the three different solution strategies. In the table, "*Delay*", "*Unused*", "*RTS*", and "*Cancel*" separately denote the average delay, unused CTS ratio, the percentage of patients assigned to RTS, and the percentage of CTS cancelled. With respect to the strategy "One-day-cancel", the two-day advance cancellation policy slightly increases *RTS* percentage and decreases CTS Cancellation ratio with more unused CTS and longer delay for most instances. It is interesting that the CTS cancellation ratio does not increase under the contract obtained from Chapter 4. The contract "LocalOpt" further takes advantage of two-day advance cancellation. With respect to the solution of

"Two-day-cancel", "LocalOpt" benefits from the two-day advance cancellation by increasing the CTS cancellation ratio. The solution of "LocalOpt" also reduces RTS percentage, with a little decrease in average delay and unused CTS ratio.

Table 5.2 summarizes the contracts, patient assignment, one-day cancellation, and two-day cancellation control policies for the solution strategies of "Two-day-cancel" and "LocalOpt". The optimal contract and the optimal control policies for "One-day-cancel" solution are **n**={0, 1,1, 1, 2, 2, 0}, **L**={10,10,10,11,10,8,9} **S**={0,1,1,1,1,2,0}, which are to the same with the control for "Two-day-cancel" and "LocalOpt" strategies with b_2=0.8b_1 and b_2=0.9b_1. For the solution of "Two-day-cancel", the patient assignment control policy is nearly the same for different b_2. With the increase of b_2, the control policy for one-day advance cancellation increases leading to more one-day cancellation, whereas the control policy for two-day advance cancellation decreases leading to less two-day cancellation. For the optimal contract of "LocalOpt" strategy, more CTS are planned when b_2 is small. Due to the possibility of the two-day advance CTS cancellation, CTS are now planned for Sunday. When b_2 is small, larger patient assignment threshold is used because of more CTS introduced in order to avoid assigning patients to RTS.

b_2	Two-day-cancel	LocalOpt
0.1b_1	**n**={0, 1,1, 1, 2, 2, 0}7CTS $\mathbf{S^1}$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0 } **Y**={0,0,0,0,1,0,0 } **L**={10,10,10,10,9,8,8}	**n**={1,0,1,2,2,1,3}10CTS $\mathbf{S^1}$={0,0,0,1,1,1,3} **X**={1,0,1,2,2,1,3 } **Y**={3,0,0,1,1,2,1 } **L**={22,23,24,24,24,23,21}
0.2 b_1	**n**={0, 1,1, 1, 2, 2, 0}7CTS $\mathbf{S^1}$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,9,8,9}	**n**={1,0,1,2,2,1,2}9CTS $\mathbf{S^1}$={0,0,0,1,1,1,2 } **X**={1,0,1,2,1,1,2 } **Y**={2,0,0,1,2,1,1} **L**={18,19,20,19,19,18,17}
0.3 b_1	**n**={0, 1,1, 1, 2, 2, 0}7CTS $\mathbf{S^1}$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S^1}$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,14,14,15,14,14,12}
0.4 b_1	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S^1}$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S^1}$={0,0,0,0,1,1,2} **X**={0,01,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,14,15,15,14,14,13}

b_2	Two-day-cancel	LocalOpt
0.5 b_1	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S^1}$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S^1}$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,14,15,15,15,14,13}
0.6 b_1	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S^1}$={0,0,0,0,1,2,0} **X**={0,0,1,1,1,1,0} **Y**={0,0,0,0,0,1,0} **L**={10,10,10,11,10,8,9}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S^1}$={0,0,0,0,1,1,2} **X**={0,01,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,15,15,15,15,14,13}
0.7 b_1	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S^1}$={0,1,1,0,1,2,0} **X**={0,0,0,0,1,0,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,11,10,8,9}	**n**={0,1,1,1,2,1,1}7CTS $\mathbf{S^1}$={0,1,1,0,1,1,1} **X**={0,0,0,0,1,0,1} **Y**={0,0,0,0,0,0,1} **L**={10,10,10,10,10,9,9}
0.8 b_1	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S^1}$={0,1,1,1,1,2,0} **X**={0,0,0,0,0,0,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,11,10,8,9}	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S^1}$={0,1,1,1,1,2,0} **X**={0,0,0,0,0,0,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,11,10,8,9}
0.9 b_1	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S^1}$={0,1,1,1,1,2,0} **X**={0,0,0,0,0,0,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,11,10,8,9}	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S^1}$={0,1,1,1,1,2,0} **X**={0,0,0,0,0,0,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,11,10,8,9}

TAB. 5.2 Contracts and control policies vs two-day advance cancellation costs b_2

5.6.2 Impact of one-day advance cancellation cost

This subsection considers the impact of the one-day advance cancellation cost b_1 by varying b_1 from 0.1c to 0.9c by taking b_2 equal 0.5b_1.

Fig. 5.7 compares the average costs of the three solution strategies for different b_1. It is clear that one-day cancellation cost leads to the reduction of average costs, although there is no obvious trends. When b_1 is smaller, two-day advance cancellation only provides little improvement. The greatest improvement of “Two-day-cancel” over “One-day-cancel” is 9.45 when b_1=0.4c. Local optimization further improves the contract and control policies. The combined improvement over “One-day-cancel” reaches 10.67% when b_1=0.5c.

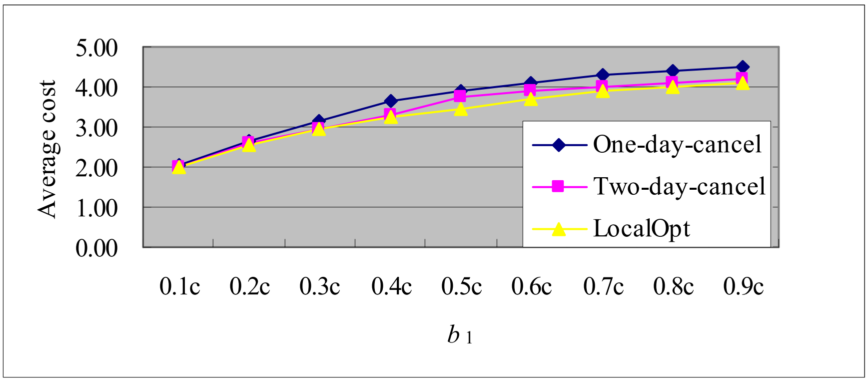


FIG. 5.7 Average costs vs different two-day advance cancellation costs b_1

Table 5.3 summarizes the performance measures of the three different solution strategies for different b_1. With respect to the solution of "One-day-cancel", the "Two-day-cancel" slightly decreases the cancellation ratio when $b_1 \leq 0.6c$, but greatly increases it when $b_1 > 0.6c$. Correspondingly, the unused CTS ratio increases when $b_1 \leq 0.6c$, and decreases when $b_1 > 0.6c$. This means that the two-day advance cancellation plays more important role when $b_1 > 0.6c$. The contract "LocalOpt" further improves the benefits of the two-day advance. With respect to the solution of "Two-day-cancel", "LocalOpt" increases the cancellation ratio, at the same time reduces the unused CTS ratio. With the increase of b_1, cancellation ratio decreases, whereas unused CTS ratio and average delay increase.

b_1	One-day-cancel				Two-day-cancel				LocalOpt			
	Delay (days)	Unused (%)	RTS (%)	Cancel (%)	Delay (days)	Unused (%)	RTS (%)	Cancel (%)	Delay (days)	Unused (%)	RTS (%)	Cancel (%)
0.1c	1.63	0.00	0.00	36.23	1.70	0.00	0.00	36.23	1.59	0.00	0.00	42.60
0.2c	2.06	0.00	0.01	28.26	2.15	0.00	0.01	28.26	1.87	0.00	0.00	36.23
0.3c	2.06	0.00	0.01	28.26	2.23	0.58	0.01	27.69	2.35	0.40	0.01	27.86
0.4c	2.06	0.00	0.00	28.26	2.36	1.82	0.01	26.45	2.27	1.52	0.01	26.75
0.5c	2.97	0.98	0.44	17.38	3.05	3.48	0.50	14.94	2.25	3.00	0.00	25.26
0.6c	2.82	2.66	0.39	15.67	3.11	3.18	0.47	15.21	2.25	3.00	0.00	25.26
0.7c	2.40	9.58	0.30	8.67	3.30	3.00	0.47	15.39	3.11	3.94	0.46	14.45
0.8c	2.26	12.64	0.28	5.60	3.17	3.91	0.45	14.46	3.11	3.94	0.46	14.45
0.9c	2.16	18.22	0.26	0.00	2.87	7.63	0.39	10.69	3.00	5.03	0.41	13.32

TAB. 5.3 Performance comparison for different one-day advance cancellation costs b_1

Table 5.4 summarizes the contracts and the control policies for different b_1. With respect to the control policies in "One-day-cancel", the patient assignment policy in "Two-day-cancel" keeps nearly the same. The one day cancellation control policy $\mathbf{S}^1$ in "Two-day-cancel" is the same when $b_1 \leq 0.3c$ and $b_1 \geq 0.7c$, and smaller in the rest instances. For most cases the same number of CTS is planned in "LocalOpt" except for $b_1=0.1c$, $0.5c$, and $0.6c$. Higher patient assignment thresholds are used when more CTS is planned. "LocalOpt" usually moves one CTS from weekday to Sunday in order to benefit from the two-day advance CTS cancellation. When $b_1 \geq 0.7c$, it moves all CTS in Saturday to Sunday. This is because one-day advance cancellation policy does not work at this time.

b_1	One-day-cancel	Two-day-cancel	LocalOpt
0.1c	**n**={0, 1, 2, 1, 2, 2, 1}; 9CTS **S**={0, 1, 2, 1, 2, 2, 1}; **L**={18, 18, 18, 19, 18, 17, 17}.	**n**={0,1,2,1,2,2,1}; 9CTS **S**^{1}={0,1,2,1,2,2,1} **X**={0,0,0,0,0,0,1} **Y**={0,0,0,0,0,0,2} **L**={18,18,18,19,18,17,17}	**n**={1,1,1,2,2,2,1}10CTS **S**^{1}={1,1,1,2,2,2,1} **X**={1,0,0,1,0,0,1} **Y**={1,0,0,1,0,0,2} **L**={22,23,24,23,23,22,22}
0.2c	**n**={0, 1, 1, 1, 2, 2, 1};8CTS **S**={0, 1, 1, 1, 2, 2, 1}; **L**={14, 14, 14, 14, 14, 13, 12}.	**n**={0,1,1,1,2,2,1};8CTS **S**^{1}={0,1,1,1,2,2,1} **X**={0,0,0,0,0,0,1} **Y**={0,0,0,0,0,0,2} **L**={13,14,14,14,14,12,12}	**n**={1,1,1,1,2,2,1}9CTS **S**^{1}={1,1,1,1,2,2,1} **X**={1,0,0,0,1,0,1} **Y**={1,0,0,0,1,0,2} **L**={18,18,19,19,19,18,17}
0.3c	**n**={0, 1, 1, 1, 2, 2, 1};8CTS **S**={0, 1, 1, 1, 2, 2, 1}; **L**={14, 14, 15, 15, 14, 13, 13}.	**n**={0,1,1,1,2,2,1};8CTS **S**^{1}={0,1,1,1,1,2,1} **X**={0,0,0,0,1,1,1} **Y**={0,0,0,0,1,1,2} **L**={14,14,14,15,14,13,13}	**n**={0,1,1,1,1,2,2}8CTS **S**^{1}={0,1,1,1,0,2,2} **X**={0,0,0,0,0,2,2} **Y**={0,0,0,0,0,0,2} **L**={14,14,14,14,15,14,12}
0.4c	**n**={0, 1, 1, 1, 2, 2, 1};8CTS **S**={0, 1, 1, 1, 2, 2, 1}; **L**={14, 15, 15, 15, 15, 13, 13}.	**n**={0,1,1,1,2,2,1};8CTS **S**^{1}={0,1,0,0,1,2,1} **X**={0,0,0,1,2,2,1} **Y**={0,0,0,0,0,1,2} **L**={14,14,15,15,15,13,13}	**n**={0,1,1,1,2,1,2}8CTS **S**^{1}={0,1,1,0,1,1,2} **X**={0,0,0,0,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,14,14,15,14,14,13}
0.5c	**n**={0, 1,1, 1, 2, 2, 0};7CTS **S**={0, 1, 1, 1, 1, 2, 0}; **L**={10, 10, 10, 11, 10, 8, 9}.	**n**={0,1,1,1,2,2,0};7CTS **S**^{1}={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,1,1,1,2,1,2}8CTS **S**^{1}={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,14,15,15,15,14,13}
0.6c	**n**={0, 1, 1, 2, 1, 2, 0};7CTS **S**={0, 1, 1, 1, 0, 2, 0}; **L**={10, 11, 11, 10, 10, 9, 9}.	**n**={0,1,1,2,1,2,0};7CTS **S**^{1}={0,0,0,1,0,2,0} **X**={0,0,1,1,0,1,0} **Y**={0,0,0,1,0,0,0} **L**={10,10,11,10,10,9,9}	**n**={0,1,1,1,2,1,2}8CTS **S**^{1}={0,0,0,0,0,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,15,15,15,15,14,13}

b_1	One-day-cancel	Two-day-cancel	LocalOpt
0.7*c*	**n**={1, 1, 1, 1, 1, 2, 0};7CTS **S**={0, 0, 0, 0, 0, 2, 0}; **L**={10, 10, 10, 11, 11, 9, 10}.	**n**={1,1,1,1,1,2,0};7CTS $\mathbf{S}^1$={0,0,0,0,0,2,0} **X**={1,1,1,0,0,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,11,9,10}	**n**={0,1,1,1,2,0,2}7CTS $\mathbf{S}^1$={0,0,0,0,0,0,2} **X**={0,0,1,0,1,0,2} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,10,9}
0.8*c*	**n**={1,1,1,1,2,1,0};7CTS **S**={0, 0, 0, 0, 0, 1, 0}; **L**={10, 10, 11, 11, 10, 10, 10}.	**n**={1,1,1,1,2,1,0};7CTS $\mathbf{S}^1$={0,0,0,0,0,1,0} **X**={1,1,1,0,1,1,0} **Y**={0,0,0,0,0,1,0 } **L**={10,10,10,11,10,9,10}	**n**={0,1,1,1,2,0,2}7CTS $\mathbf{S}^1$={0,0,0,0,0,0,2} **X**={0,0,1,0,1,0,2} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,10,9}
0.9*c*	**n**={1,1,1,1,3,0,0};7CTS **S**={0, 0, 0, 0, 0, 0, 0}; **L**={11, 11, 11, 11, 9, 10, 10}.	**n**={1,1,1,1,3,0,0};7CTS $\mathbf{S}^1$={0,0,0,0,0,0,0} **X**={1,1,1,0,1,0,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,11,11,9,10,10}	**n**={0,1,1,1,2,0,2}7CTS $\mathbf{S}^1$={0,0,0,0,0,0,2} **X**={0,0,0,0,1,0,2} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,11,10,10,9}

TAB. 5.4 Contracts and control policies vs one-day advance cancellation costs b_1

5.6.3 Impact of unused CTS cost

This subsection explores the impact of unused CTS cost, *c*, by varying *c* from 1 to 20 with $b_1 = 0.5c$ and $b_2 = 0.5b_1$.

Fig. 5.8 compares the average costs for different unused CTS cost *c*. When *c* is smaller than 5, two-day advance cancellation only provides marginal improvement. Larger improvement is gained when *c* is larger, for example, 9.48% of improvement of “Two-day-cancel” over “One-day-cancel” when *c*=10; 3.65% when *c*=15; and 5.21% when *c*=20. Local optimization further improves the contracts and control policies, with 7.28% of the greatest improvement over “Two-day-cancel” when *c*=15.

Table 5.5 summarizes the performance measures for different solution strategies. Two-day advance cancellation causes slightly longer delay and more RTS assignment. For “LocalOpt”, the average delay and the CTS cancellation ratio increase, while the unused CTS ratio decreases.

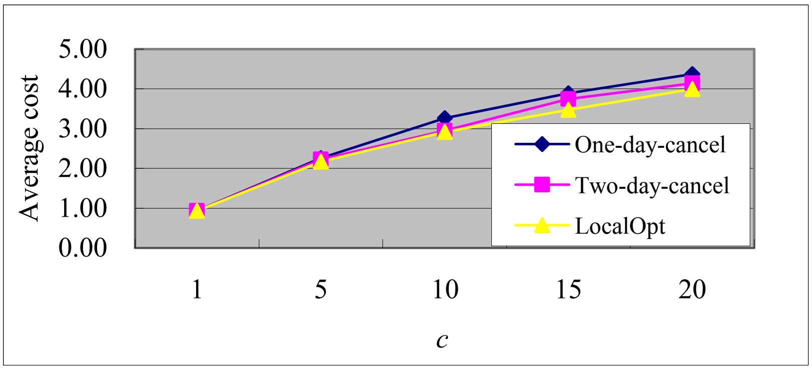


FIG. 5.8 Average costs vs different unused CTS costs *c*

c	One-day-cancel				Two-day-cancel				LocalOpt			
	Delay (days)	Unused (%)	RTS (%)	Cancel (%)	Delay (days)	Unused (%)	RTS (%)	Cancel (%)	Delay (days)	Unused (%)	RTS (%)	Cancel (%)
1	0.52	28.44	0.00	14.17	0.52	28.44	0.00	14.17	0.52	28.44	0.00	14.17
5	1.37	11.45	0.01	16.81	1.58	10.60	0.01	17.66	1.40	7.44	0.00	28.79
10	1.76	3.56	0.00	24.70	2.01	4.77	0.01	23.49	2.15	3.38	0.01	24.88
15	2.97	0.98	0.44	17.38	3.05	3.48	0.50	14.94	2.25	3.00	0.00	25.26
20	2.98	0.98	0.38	17.34	3.20	2.45	0.46	15.93	2.59	1.03	0.00	27.23

TAB. 5.5 Performance comparison for different CTS costs *c*

Table 5.6 compares the contracts and control policies for different *c*. For all solution strategies, the number of CTS decreases with the increase of *c*. “Two-day-cancel” solution has slightly lower RTS thresholds than “One-day-cancel” solution. “LocalOpt” differs with the other strategies in the CTS planned for the Sunday even when *c* is great. “LocalOpt” also allows more CTS cancellation than “Two-day-cancel” solution. Higher RTS assignment thresholds are used when more CTS is planned.

c	One-day-cancel	Two-day-cancel	LocalOpt
1	**n**={1, 2, 1, 2, 2, 1, 1}; 10CTS **S**={0, 0, 0, 0, 0, 1, 1}; **L**={22, 21, 22, 22, 22, 21, 21}	**n**={1,2,1,2,2,1,1};10CTS $\mathbf{S}^1$={0,0,0,0,0,1,1} **X**={0,0,0,0,0,0,0} **Y**={0,0,0,0,0,0,0} **L**={22,21,22,22,22,21,21}	**n**={1,2,1,2,2,1,1}10CTS $\mathbf{S}^1$={0,0,0,0,0,1,1} **X**={0,0,0,0,0,0,0} **Y**={0,0,0,0,0,0,0} **L**={22,21,22,22,22,21,21}
5	**n**={1, 1, 1, 1, 2, 2, 0}; 8CTS **S**={0, 0, 0, 0, 1, 2, 0}; **L**={13, 13, 13, 14, 13, 12, 13}	**n**={1,1,1,1,2,2,0}8CTS $\mathbf{S}^1$={0,0,0,0,0,2,0} **X**={0,0,0,0,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={13,13,13,14,13,12,13}	**n**={1,1,2,1,2,1,1}9CTS $\mathbf{S}^1$={0,0,1,0,0,1,1} **X**={1,0,1,0,1,1,1} **Y**={1,0,0,0,0,0,1} **L**={18,18,18,18,18,18,18}

c	One-day-cancel	Two-day-cancel	LocalOpt
10	**n**={0, 1, 1, 2, 2, 1, 1}; 8CTS **S**={0, 1, 1, 1, 1, 1, 1}; **L**={14, 15, 15, 14, 14, 13, 13}	**n**={0,1,1,2,2,1,1}8CTS $\mathbf{S}^1$={0,0,0,1,0,1,1} **X**={0,0,1,1,1,1,1} **Y**={0,0,0,1,1,0,1} **L**={14,14,15,14,13,13,13}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S}^1$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,0,1,1} **L**={14,14,14,15,14,14,12}
15	**n**={0, 1,1, 1, 2, 2, 0};7CTS **S**={0, 1, 1, 1, 1, 2, 0}; **L**={10, 10, 10, 11, 10, 8, 9}.	**n**={0,1,1,1,2,2,0};7CTS $\mathbf{S}^1$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S}^1$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,14,15,15,15,14,13}
20	**n**={ 0, 1, 1, 1, 2, 2, 0}; 7CTS **S**={ 0, 1, 1, 1, 1, 2, 0}; **L**={11, 11, 11, 11, 10, 9, 9}	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S}^1$={0,0,0,0,1,2,0} **X**={0,0,1,1,1,1,0} **Y**={0,0,0,0,1,1,0} **L**={10,10,11,11,10,9,9}	**n**={1,0,1,2,1,2,1}8CTS $\mathbf{S}^1$={0,0,0,1,0,2,1} **X**={1,0,1,2,1,2,1} **Y**={1,0,0,0,1,0,2} **L**={14,15,16,15,16,14,14}

TAB. 5.6 Contracts and control policies vs unused CTS costs *c*

5.6.4 Impact of other parameters

This subsection considers the relationship between the influence of two-day advance cancellation control and other parameters including (i) the average RTS delay, T^R; (ii) the patient arrival pattern, and (iii) the patient arrival rate.

To investigate the impact of the average RTS delay T^R, numerical experiments are performed for the base case by varying T^R from 25 to 45. For all the strategies, the average cost (see Fig. 5.9), the performance measures (see Table 5.7), and the contracts and the control policies (see Table 5.8) are fairly insensitive to the change of T^R, except that RTS threshold increases with the increase of T^R for all the solutions.

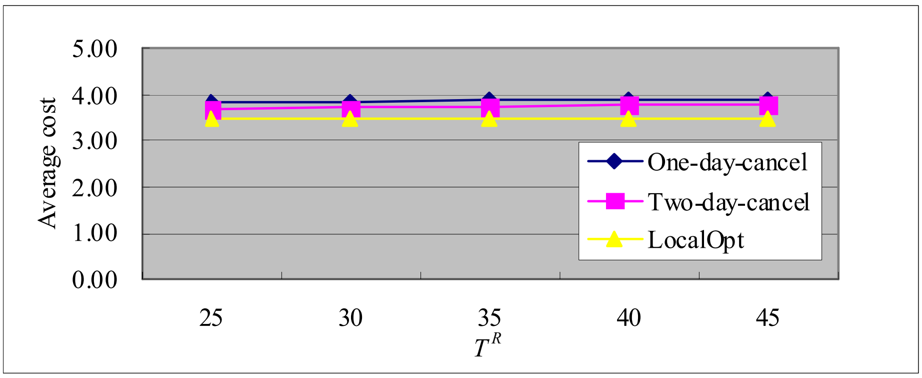


FIG. 5.9 Average costs vs different RTS delays T^R

T^R	One-day-cancel				Two-day-cancel				LocalOpt			
	Delay (days)	Unused (%)	RTS (%)	Cancel (%)	Delay (days)	Unused (%)	RTS (%)	Cancel (%)	Delay (days)	Unused (%)	RTS (%)	Cancel (%)
25	2.87	1.01	0.94	17.77	2.94	3.56	1.03	15.29	2.25	3.00	0.05	25.28
30	2.93	0.99	0.65	17.55	3.01	3.51	0.69	15.07	2.25	3.00	0.01	25.26
35	2.97	0.98	0.44	17.38	3.05	3.48	0.50	14.94	2.25	3.00	0.00	25.26
40	2.85	2.65	0.30	15.60	3.14	3.17	0.35	15.13	2.25	3.00	0.00	25.25
45	2.87	2.64	0.21	15.53	3.16	3.15	0.25	15.06	2.25	3.00	0.00	25.25

TAB. 5.7 Performance comparison for different RTS delays T^R

T^R	One-day-cancel	Two-day-cancel	LocalOpt
25	**n**={ 0, 1, 1, 1, 2, 2, 0};7CTS **S**={ 0, 1, 1, 1, 1, 2, 0}; **L**={8, 8, 8, 9, 8, 7, 7}.	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S^1}$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={8,8,8,9,8,6,7}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S^1}$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={11,11,11,12,11,11,10}
30	**n**={ 0, 1, 1, 1, 2, 2, 0};7CTS **S**={ 0, 1, 1, 1, 1, 2, 0}; **L**={9, 9, 9, 10, 9, 7, 8}.	**n**={0,1,1,1,2,2,0}7CTS $\mathbf{S^1}$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={9,9,9,10,9,7,8}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S^1}$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={13,13,13,14,13,13,11}
35	**n**={0, 1,1, 1, 2, 2, 0};7CTS **S**={0, 1, 1, 1, 1, 2, 0}; **L**={10, 10, 10, 11, 10, 8, 9}.	**n**={0,1,1,1,2,2,0};7CTS $\mathbf{S^1}$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S^1}$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,14,15,15,15,14,13}
40	**n**={ 0, 1, 1, 2, 1, 2, 0};7CTS **S**={ 0, 1, 1, 1, 0, 2, 0}; **L**={11, 11, 11, 11, 11, 9, 10}.	**n**={0,1,1,2,1,2,0}7CTS $\mathbf{S^1}$={0,0,0,1,0,2,0} **X**={0,0,1,1,0,1,0} **Y**={0,0,0,1,0,0,0} **L**={11,11,11,11,11,9,10}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S^1}$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={16,16,16,17,16,16,15}
45	**n**={ 0, 1, 1, 2, 1, 2, 0};7CTS **S**={ 0, 1, 1, 1, 0, 2, 0}; **L**={12, 12, 12, 12, 12, 10, 11}.	**n**={0,1,1,2,1,2,0}7CTS $\mathbf{S^1}$={0,0,0,1,0,2,0} **X**={0,0,1,1,0,1,0} **Y**={0,0,0,1,0,0,0} **L**={12,12,12,11,12,10,11}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S^1}$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={17,18,18,18,18,17,16}

TAB. 5.8 Contracts and control policies vs RTS delays T^R

We now perform the sensitivity analysis with respect to the patient arrival patterns by exchanging the current peak arrival rate of Friday with the arrival rates of any other weekday. Another arrival pattern (Ave) with equal weekday arrival is also considered. The average cost (see Fig. 5.10) and the performance measures (See Table 5.9) are also insensitive to the change of patient arrival patterns. "LocalOpt" usually moves one CTS from peak arrival date to one or two days later and one CTS from Saturday to Sunday, at the same time, one CTS is added in Sunday. Therefore, more CTS cancellation and more patient assignment threshold are expected.

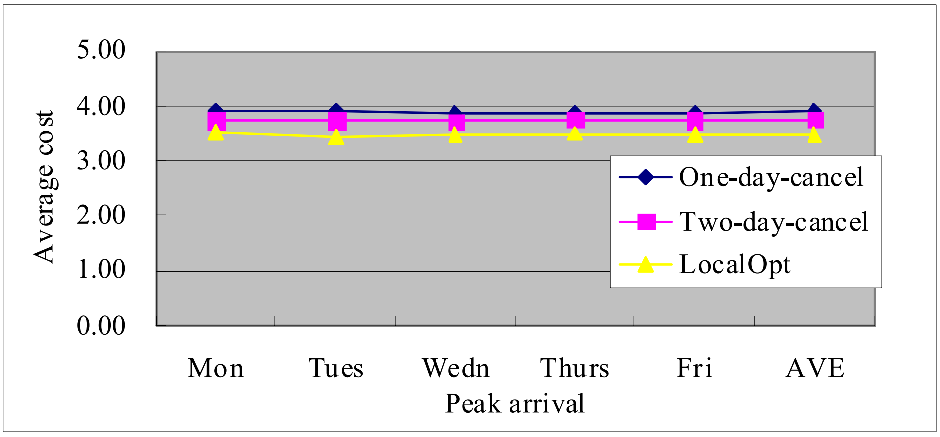


FIG. 5.10 Average costs vs different patient arrival patterns

Peak arrival	One-day-cancel				Two-day-cancel				LocalOpt			
	Delay (days)	Unused (%)	RTS (%)	Cancel (%)	Delay (days)	Unused (%)	RTS (%)	Cancel (%)	Delay (days)	Unused (%)	RTS (%)	Cancel (%)
Mon	2.90	2.08	0.44	16.27	3.07	3.94	0.48	14.43	2.40	2.66	0.01	25.58
Tues	2.90	2.05	0.44	16.30	3.06	3.38	0.50	15.02	2.39	2.19	0.01	26.06
Wed	2.85	2.41	0.44	15.90	3.08	3.33	0.51	15.04	2.54	1.57	0.01	26.64
Thurs	2.95	1.15	0.46	17.21	2.89	3.73	0.46	14.62	2.29	2.90	0.01	25.33
Fri	2.97	0.98	0.44	17.38	3.05	3.48	0.50	14.94	2.25	3.00	0.00	25.26
AVE	2.80	3.09	0.37	15.62	3.33	2.35	0.49	16.46	2.32	2.44	0.01	26.17

TAB. 5.9 Performance comparison for different patient arrival patterns

Peak arrival	One-day-cancel	Two-day-cancel	LocalOpt
Mon	**n**={1, 1, 1, 1, 1, 2, 0};7CTS **S**={0, 1, 1, 0, 1, 2, 0}; **L**={10,10,10,10,10,9,9}.	**n**={1,1,1,1,1,2,0}7CTS $\mathbf{S}^1$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,9,9}	**n**={0,1,2,1,1,1,2}8CTS $\mathbf{S}^1$={0,0,1,0,0,1,2} **X**={0,0,2,1,1,1,2} **Y**={0,0,0,1,0,1,1} **L**={15,15,14,15,15,14,13}

Peak arrival	One-day-cancel	Two-day-cancel	LocalOpt
Tues	**n**={0, 2, 1, 1, 1, 2, 0};7CTS **S**={0, 1, 1, 0, 1, 2, 0}; **L**={10,10,10,10,10,9,9}.	**n**={0,2,1,1,1,2,0}7CTS $\mathbf{S}^1$={0,1,0,0,0,2,0} **X**={0,0,0,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,1,1,2,1,1,2}8CTS $\mathbf{S}^1$={0,0,0,1,0,1,2} **X**={0,0,1,2,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,15,15,14,14,14,13}
Wed	**n**={0, 1, 2, 1, 1, 2, 0};7CTS **S**={0, 1, 1, 0, 1, 2, 0}; **L**={10,10,10,10,10,9,9}.	**n**={0,1,2,1,1,2,0}7CTS $\mathbf{S}^1$={0,0,1,0,0,2,0} **X**={0,0,1,0,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,0,2,1,2,1,2}8CTS $\mathbf{S}^1$={0,0,1,0,1,1,2} **X**={0,0,1,1,2,1,2} **Y**={0,0,0,1,0,2,1} **L**={14,15,15,15,14,14,13}
Thurs	**n**={0, 1, 1, 2, 1, 2, 0};7CTS **S**={0, 1,1, 1, 1, 2, 0}; **L**={10,10,10,10,10,8,9}.	**n**={0,1,1,2,1,2,0}7CTS $\mathbf{S}^1$={0,0,0,1,0,2,0} **X**={0,0,1,1,0,1,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S}^1$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,14,15,15,15,14,13}
Fri	**n**={0, 1,1, 1, 2, 2, 0};7CTS **S**={0, 1, 1, 1, 1, 2, 0}; **L**={10, 10, 10, 11, 10, 8, 9}.	**n**={0,1,1,1,2,2,0};7CTS $\mathbf{S}^1$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S}^1$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,14,15,15,15,14,13}
AVE	**n**={0, 1, 2, 1, 1, 2, 0};7CTS **S**={0, 1, 1, 0, 1, 2, 0}; **L**={11,11,10,10,10,9,9}.	**n**={0,1,2,1,1,2,0}7CTS $\mathbf{S}^1$={0,0,1,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,1,1,0,0,0} **L**={10,11,10,10,10,9,9}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S}^1$={0,0,0,0,1,1,2} **X**={0,0,1,1,2,1,2} **Y**={0,0,0,0,0,1,1} **L**={14,15,15,15,15,14,13}

TAB. 5.10 Contracts and Control policies vs patient arrival patterns

We now perform the sensitivity analysis with respect to the patient arrival rate. Three scenarios are considered "Low" (base case), "Medium" (patient arrival rates 5 times larger), "High" (patient arrival rates 10 times larger). Fig. 5.11 compares the average costs of the three solution strategies. Two-day advance cancellation brings 3.65% improvement over "One-day-cancel" strategy for low demand instance, 5.18% for medium demand, and 5.33% for high demand instance. Local optimization further brings 7.28%, 3.53%, and 3.22% improvement over "Two-day-cancel" strategy, and leads to a combined improvement of 10.67%, 8.53%, and 8.38% respectively for low, medium, high demand instances. We can see that the two-day advance cancellation work well for both high and low demand instances.

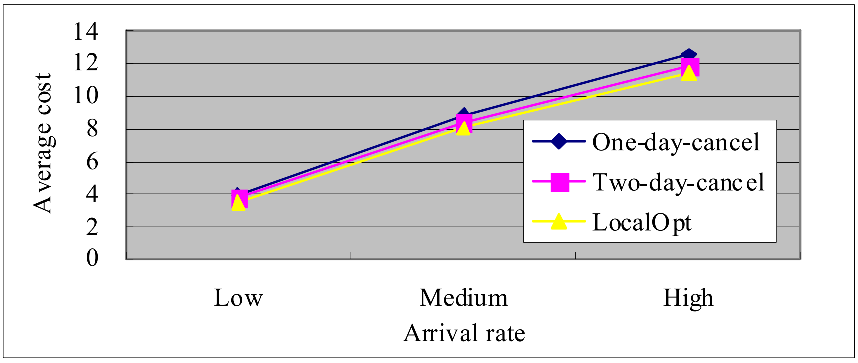


FIG. 5.11 Average costs vs different patient arrival rates

Table 5.11 summarizes the performance measures for different arrival rates. For all the solutions, the increase of demand rate results in shorter delay, less RTS assignment and less CTS cancellation ratio. Unused CTS ratio increases for "One-day-cancel" policy but decreases for the other solution strategies.

Arrival rates	One-day-cancel				Two-day-cancel				LocalOpt			
	Delay (days)	Unused (%)	RTS (%)	Cancel (%)	Delay (days)	Unused (%)	RTS (%)	Cancel (%)	Delay (days)	Unused (%)	RTS (%)	Cancel (%)
Low	2.97	0.98	0.44	17.38	3.05	3.48	0.50	14.94	2.25	3.00	0.00	25.26
Medium	1.14	1.61	0.06	8.78	1.30	2.06	0.08	8.35	1.11	2.07	0.01	10.99
High	0.77	1.85	0.04	5.61	0.90	1.51	0.05	5.95	0.81	1.59	0.01	7.30

TAB. 5.11 Performance comparison for different patient arrival rates

Table 5.12 summarizes the contracts and control policies for different rates. One additional CTS is planned for "LocalOpt" and more CTS are planned in Sunday. Therefore, more CTS cancellation and higher RTS assignment are expected.

Fig. 5.12 and 5.13 show the two-day advance cancellation control policy for "Two-day-cancel" and "LocalOpt" solutions. These figures show the reason why these policies cannot write in the form of relation (5.62), i.e., the control policies for Wednesday and Friday in "Two-day-cancel" solution and for Thursday and Friday in "LocalOpt" solution. However, it can be approximated as the policy in the form of (5.62), as shown in Table 5.13. In this table, "obj_appr" and "obj" are separately the criterion value with the approximated policy and the real optimal policy. "*Gap*" is (1-obj/obj_appr)*100%. From this table, we can see that the *Gap* is very small, no greater than 0.25%, which means that the approximated control policies perform well.

Arrival rates	One-day-cancel	Two-day-cancel	LocalOpt
Low	**n**={0, 1,1, 1, 2, 2, 0};7CTS **S**={0, 1, 1, 1, 1, 2, 0}; **L**={10, 10, 10, 11, 10, 8, 9}.	**n**={0,1,1,1,2,2,0};7CTS $\mathbf{S}^1$={0,0,0,0,0,2,0} **X**={0,0,1,1,1,2,0} **Y**={0,0,0,0,0,0,0} **L**={10,10,10,10,10,8,9}	**n**={0,1,1,1,2,1,2}8CTS $\mathbf{S}^1$={0,0,0,0,1,1,2} **X**={0,0,1,1,1,1,2} **Y**={0,0,0,0,1,1,1} **L**={14,14,15,15,15,14,13}
Medium	**n**={3,5,5,6,8,4,1};32CTS **S**={0,2,2,2,2,4,1}; **L**={25,25,25,25,24,22,23}.	**n**={3,5,5,6,8,4,1}32CTS $\mathbf{S}^1$={0,1,1,1,0,3,1} **X**={0,1,2,2,2,4,1} **Y**={0,0,1,1,1,0,4} **L**={24,24,25,25,24,22,22}	**n**={4,4,5,6,8,3,3}33CTS $\mathbf{S}^1$={0,0,1,1,1,2,3} **X**={1,1,2,2,2,3,3} **Y**={3,0,0,1,1,1,3} **L**={28,29,29,30,29,28,26}
High	**n**={8,9,10,12,15,7,1};62CTS **L**={34,35,35,35,35,31,32}. **S**={0,2,2,2,2,6,1};	**n**={8,9,10,12,15,7,1}62CTS **L**={34,34,35,35,34,30,31} $\mathbf{S}^1$={0,1,1,1,0,6,1} $\mathbf{S}^2$**(0)**={1,1,2,2,2,5,1} $\mathbf{S}^2$**(1)**={1,0,2,2,2,4,1} $\mathbf{S}^2$**(2)**={0,0,1,2,1,3,1} $\mathbf{S}^2$**(3)**={0,0,1,1,1,2,1} $\mathbf{S}^2$**(4)**={0,0,0,0,0,1,1} $\mathbf{S}^2$**(5)**={0,0,0,0,0,0,1} $\mathbf{S}^2$**(6)**={0,0,0,0,0,0,1} $\mathbf{S}^2$**(7)**={0,0,0,0,0,0,0} $\mathbf{S}^2$**(8)**={0,0,0,0,0,0,0}	**n**={8,8,10,12,16,4,5}63CTS **L**={37,39,39,40,38,37,35} $\mathbf{S}^1$={0,0,1,1,1,3,5} $\mathbf{S}^2$**(0)**={0,0,2,3,3,3,5} $\mathbf{S}^2$**(1)**={0,0,2,3,3,3,5} $\mathbf{S}^2$**(2)**={0,0,1,2,2,3,5} $\mathbf{S}^2$**(3)**={0,0,0,1,2,2,5} $\mathbf{S}^2$**(4)**={0,0,0,1,1,1,4} $\mathbf{S}^2$**(5)**={0,0,0,0,0,0,3} $\mathbf{S}^2$**(6)**={0,0,0,0,0,0,2} $\mathbf{S}^2$**(7)**={0,0,0,0,0,0,1} $\mathbf{S}^2$**(8)**={0,0,0,0,0,0,0}

TAB. 5.12 Contracts and control policies vs patient arrival rates

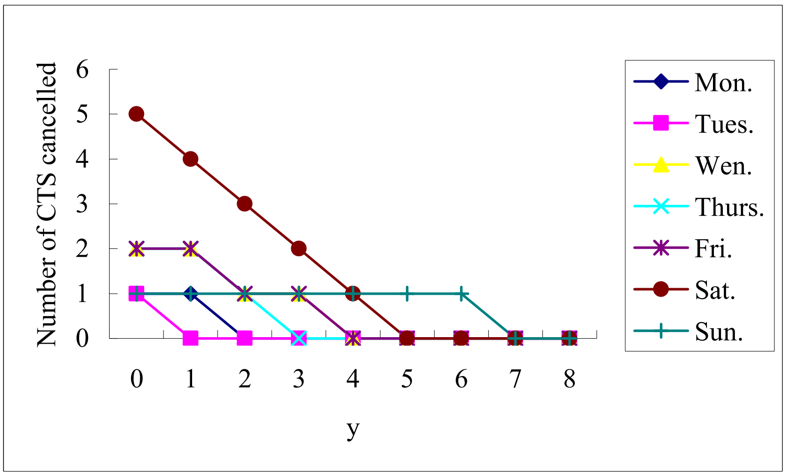


FIG. 5.12 Two-day advance CTS cancellation policy for “Two-day-cancel” solution

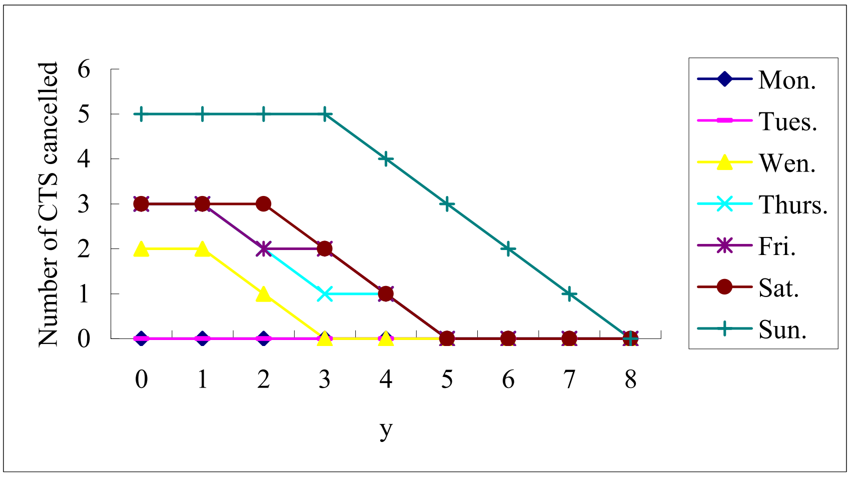


FIG. 5.13 Two-day advance CTS cancellation policy for "LocalOpt" solution

Two-day-cancel				LocalOpt			
Approximated policy	obj_appr.	obj	Gap(%)	Approximated policy	obj_appr.	obj	Gap(%)
X={1,1,2,2,2,5,1} **Y**={1,0,1,2,1,0,6}	11.84	11.84	0.02	X={0,0,2,3,3,3,5} Y={0,0,1,1,1,2,3}	11.48	11.46	0.13
X={1,1,2,2,2,5,1} **Y**={1,0,2,2,1,0,6}	11.84	11.84	0.02	X={0,0,2,3,3,3,5} Y={0,0,1,2,1,2,3} ;	11.48	11.46	0.21
X={1,1,2,2,2,5,1} **Y**={1,0,2,2,2,0,6}	11.84	11.84	0.02	X={0,0,2,3,3,3,5} Y={0,0,1,1,2,2,3}	11.47	11.46	0.10
X={1,1,2,2,2,5,1} **Y**={1,0,1,2,2,0,6}	11.84	11.84	0.03	X={0,0,2,3,3,3,5} Y={0,0,1,2,2,2,3}	11.48	11.46	0.16

TAB. 5.13 The approximated policies and the performance measures

CPU time for local search is less than 11 minutes for all the low demand instances, 5227s for medium demand instance, and 25057s for high demand instance.

5.7 Conclusion

This chapter proposes an average cost MDP formulation to establish the structure properties of patient assignment, one-day advance cancellation, and two-day advance cancellation control policies. Local optimization is used to improve a given initial contract. Numerical results show that the two-day advance cancellation and local optimization can greatly reduce the criterion values and the unused CTS ratio.

Future research can be pursued in several directions. An extension research is needed by assuming that CTS can be cancelled in advance of arbitrary days. Multiple priorities patients and several imaging examinations are the other possible directions.

Chapter 6
Conclusion

Stroke patients need quick diagnosis. However, significant delays are observed as many key examinations depend on expensive and heavily used imaging facilities such as MRI scanners. The objective of this thesis aims to reduce the waiting time of stroke patients for MRI examination without degrading the utilization of MRI scanner.

Based on the field observation of stable weekly patient arrivals, we have proposed a new contract-based MRI examination reservation process for stroke patients, i.e., neurovascular department reserves a certain number of contracted time slots for stroke patients every week. Except for these contracted ones, the time slots by regular request are still possible for stroke patients. To implement this new reservation process, three decisions need to be determined:

Contract decisions, i.e. the number of CTS and its distribution over time.

Patient assignment control policy, i.e., the real time control for assigning the incoming patients to CTS or RTS.

Advance CTS cancellation control policy, i.e., the real time control for cancelling the CTS in advance.

To determine the above decisions, we make the following contributions:

Chapter 3 proposes an integrated stochastic programming model to simultaneously determine the optimal contract and the optimal patient assignment control policy. This problem is difficult as it involves simultaneously two decisions at different levels, the contract at the tactical level and the optimal control policy at the real-time level. In order to solve this model, we first consider the optimal control problem for a given contract. An average cost MDP approach is used to establish structure properties of the optimal control policy. In particular, we show that there exists a threshold L_i for each day i of the week, 1) if the ending CTS queue is shorter than L_i, the optimal control consists in keeping all the remaining patients waiting for CTS; 2) if the ending CTS queue is longer than L_i, the optimal control consists in sending patients to RTS by keeping the number L_i of patients in the CTS queue. The contract optimization problem is solved by a two-step approach. First, the long term average cost is approximated by the average cost over a finite horizon and according to a given sample path of patient arrivals. This Monte Carlo approximation of the contract optimization problem is further simplified by relaxing the non-anticipativeness of the feasible control policy. The relaxed Monte Carlo approximation problem is equivalently transformed into a linear program with seven integer-valued variables corresponding to the contract, which can be efficiently solved by any LP solver. The resulting contract is further improved by some local search procedures. Numerical results show that the proposed approach is very efficient and provides solutions very close to real optimum. Sensitivity analysis is performed to show the impact of different problem data on the contract and control policy. Numerical results also show that the relaxed Monte Carlo approximation always leads to a contract which is at most two local moves away from the best contract identified by exhaustive search for small size problems and by multiple runs of our approach for large size problems. Further, except for one instance, the best contract is reached if the exact criterion values of local solutions are used in local search. Experimental results also show that this contract-based MRI examination reservation process can greatly reduce the average delay of stroke patients, but it also leads to some unused CTS and there still exists some “unlucky” patients, who are assigned to RTS and have to wait much longer time than the other stroke patients.

To avoid “unlucky” patients, this chapter also proposes a new reservation process which still makes use of the contract and **L**-policy. Here, **L**-policy is used to reserve some RTS for the neurovascular department, rather than some particular patients. Experiments show that this new method can better the distribution of stroke patients’ waiting times.

Chapter 4 considers the possibility of one-day advance CTS cancellation in order to reduce the unused CTS ratio. For each given contract, an average-cost MDP approach is proposed to simultaneously optimize patient assignment and CTS cancellation policies. The structure properties of the optimal control policies are established via discounted cost problem. The implementation of the control policies are as follows:

First, we show that there exists a threshold L_i for each day i of the week, and 1) if the ending CTS queue is shorter than L_i, the optimal control consists in keeping all the remaining patients waiting for CTS; 2) if the ending CTS queue is longer than L_i, the optimal control consists in sending patients to RTS by keeping the number L_i of patients in the CTS queue.

Second, there exists another threshold S_{i+1} for each day i of the week, and 1) when CTS queue is below S_{i+1}, the optimal CTS cancellation control consists in making the number of CTS cancelled for day i+1 plus CTS queue length of day i at the threshold S_{i+1}; 2) otherwise, no CTS is cancelled.

Based on the optimal control policies, a local search algorithm is proposed to improve the contract. Numerical results show that the proposed approach is very efficient and advance CTS cancellation allows significant reduction of unused CTS ratio with slight increase of waiting time. Sensitivity analysis is performed to show the impact of different problem data on the contract and control policies.

Chapter 5 explores the possibility of two-day advance CTS cancellation to help the imaging department better the arrangement of other patients to the time slots released from contract. An average cost MDP is formulated to simultaneously explore the nature of patient assignment, one-day advance cancellation, and the two-day advance CTS cancellation control policies. In particular, we show that the implementation of control policies for day *i* are as follows:

1) Patient assignment control policy: there exists a threshold L_i for each day *i* of the week and the optimal patient assignment control consists in sending patients to RTS by keeping the CTS queue length plus the two-day advance CTS cancellation for day *i* the same with L_i; otherwise no patients are assigned to RTS.

2) One-day advance CTS cancellation control policy: there exists a threshold S_{i+1} and the optimal one-day advance CTS cancellation control consists in making the number of one-day advance CTS cancellation for day $i+1$ plus the number of two-day advance CTS cancellation for day $i+1$ plus CTS queue length for day *i* the same with the threshold S_{i+1} if the number of two-day advance CTS cancellation for day $i+1$ plus CTS queue for day *i* is below S_{i+1}. Otherwise no CTS is cancelled.

3) There exists a pair of parameters (X_{i+2}, Y_{i+2}) for an approximated two-day advance control. When *y*, i.e., the number of CTS queue length for day *i* plus one-day advance CTS cancellation for day $i+1$ plus two-day advance CTS cancellation for day $i+1$, is smaller than a certain value Y_{i+2}, the two-day advance cancellation control for day $i+2$ is the same, i.e., X_{i+2}; When *y* is greater than Y_{i+2}, and less than $X_{i+2}+ Y_{i+2}$, the two-day advance cancellation control for day $i+2$ consisting in making *y* plus the number of two-day advance cancellation for day $i+2$ the same with $X_{i+2}+ Y_{i+2}$; when *y* is greater, the two-day advance cancellation is zero.

Local search is proposed to improve the contract decision. Numerical experiments are performed to compare the performance of one-day advance CTS cancellation, two-day advance CTS cancellation, and local optimization. Results show that two-day advance CTS cancellation and Local optimization can reduce the objective criteria and improve the performance.

Starting from the perspectives of neurovascular department, this thesis proposes a contract-based MRI examination reservation process for stroke patients. Contracts and control policies are determined in this thesis. The use of contract gives a long term view of diagnostic capacity available and the neurovascular department can better manage the priority of its stroke patients and reduce the waiting times for examinations. From the perspective of the imaging department, although the use of contract potentially leads to unused time slots, it also gives the imaging department stable demands which can be used to improve staff scheduling and diagnostic facility scheduling. Another advantage of contract-based approach is the possibility for the neurovascular department to better match different diagnosis examinations of the same patient and available contracted CTS for different facilities. The control policies can help in reducing patients' average delay without degrading the utilization of MRI scanners.

This method can be directly applied to the hospital with the existing patient scheduling method, such as FIFO. However, there are still a lot of works to do if we want to better the distribution of patients' waiting times and reduce the variances. In addition, we do not recommend the contract-based approach for all departments but only for critical diagnostic facilities and for some major consumers with stable demands. Results of this paper can be directly used to design separately the contract for each department for each critical facility. If this leads to the over-usage of diagnosis facility, the contract of each department can be refined by limiting the number of time slots to contract each day and all results of this thesis still hold. The relation between different medical departments is not addressed in this thesis but is crucial for implementing the contract-based approach in a hospital. The joint design of contract-based solutions of several departments raises some fundamental questions such as (i) how many time slots of a diagnostic facility to contract and (ii) how to share these time slots among different departments. Results about the optimal control policies of this paper can be extended to evaluate a contract solution. However new approaches are needed to coordinate the contracts for different departments.

In addition, the form of the optimal contract is still an open issue even for purely stationary demand. It is unclear how to determine the optimal contract if Assumption 4 is relaxed. Extension to non stationary patient arrival case is one interesting research avenue. Another immediate extension is the development of real time control strategies for advance cancellation of CTS in case of short CTS queue. Management of multiple classes of patients and multiple imaging examinations is a natural but challenging research direction.

From the service provider perspective, how to optimize the operational efficiency of the imaging department by taking into account different quality requirements of medical units is a rich research area.

Reference

Adan, I.J.B.F. and J.M.H. Vissers (2002). Patient mix optimisation in hospital admission planning: a case study. International Journal of Operations and Production Management. 22(4), 445-461.

Altman, E., T. Jimenez, and G. Koole (2001). On optimal call admission control in a resource-sharing system. Ieee Transactions on Communications. 49(9), 1659-1668.

Arenas, M., et al. (2002). Analysis via goal programming of the minimum achievable stay in surgical waiting lists. Journal of the Operational Research Society. 53, 387-396.

Augusto, V. and X. Xie(2009) Redesigning the neurovascular unit of a health care complex, in Proceedings of the International Conference on Industrial Engineering and Systems Management: Montreal, Canada.

Augusto, V., X. Xie, and V. Perdomo (2008). Operating theatre scheduling using Lagrangian relaxation. European Journal of Industrial Engineering. 2(2), 172-189.

Babes, M. and G.V. Sarma (1991). Out-patient queues at the Ibn-Rochd health centre. The Journal of the Operational Research Society. 42(10), 845-855.

Barut, M. and V. Sridharan (2004). Design and evaluation of a dynamic capacity apportionment procedure. European Journal of Operational Research. 155(1), 112-133.

Belien, J. and E. Demeulemeester (2007). Building cyclic master surgery schedules with leveled resulting bed occupancy. European Journal of Operational Research. 176(2), 1185-1204.

Belien, J. and E. Demeulemeester (2008). A branch-and-price approach for integrating nurse and surgery scheduling. European Journal of Operational Research 189(3), 652-668

Belien, J., E. Demeulemeester, and B. Cardoen (2006). Visualizing the demand for various resources as a function of the master surgery schedule: A case study. Journal of Medical Systems. 30(5), 343-350.

Belobaba, P.P. (1989). Application of a probabilistic decision model to airline seat inventory control. Operations Research. 37(2), 183-197.

Bertsekas, D.P. (1996). Dynamic Programming And Optimal Control. Vol. II: Athena Scientific.

Bharadwaj, A., A. Sen, and A. Vinze (1999). Scheduling cardiac procedures: A knowledge-based approach. Ieee Transactions on Engineering Management. 46(3), 322-334.

Bitran, G.R. and S.M. Gilbert (1996). Managing hotel reservations with uncertain arrival. Operations Research. 44(1), 35-49.

Blake, J.T. and M.W. Carter (1997). Surgical process scheduling: a structured review. Journal of the Society for Health Systems. 5(3), 17-30.

Blake, J.T., F. dexter, and J. Donald (2002). Operating room manager's use of integer programming for assigning block time to surgical groups: A case study. Anesthesia and Analgesia. 94, 143-148.

Blake, J.T. and J. donald (2002). Mount sinai hospital uses integer programming to allocate operating room time. Interfaces. 32, 63-73.

Calichman, M.V. (2005). Creating an optimal operating room schedule. AORN Journal. 81, 580-588.

Cardoen, B., E. Demeulemeester, and J. Belien (2009). Optimizing a multiple objective surgical case sequencing problem. International Journal of Production Economics. 119(2), 354-366.

Cardoen, B., E. Demeulemeester, and J. Belien (2010). Operating room planning and scheduling: A literature review. European Journal of Operational Research. 201(3), 921-932.

Carroll, W.J. and R.C. Grimes (1995). Evolutionary change in product management experiences in the car rental industry. Interfaces. 25(5), 84-104.

Cayirli, T. and E. Veral (2003). Outpatient scheduling in health care: A review of literature. Production and Operations Management. 12(4), 519-549.

Cayirli, T., E. Veral, and H. Rosen (2008). Assessment of patient classification in appointment system design. Production and Operations Management. 17(3), 338-353.

Cayirli, T., Veral, E., and Rosen, H. (2006). Designing appointment scheduling systems for ambulatory care services. Health Care Management Science. 9, 47-58.

Cheng, M.G., et al. (2008). Simulated Annealing Algorithm for Scheduling Problem in Daily Nursing Cares. 2008 Ieee International Conference on Systems, Man and Cybernetics (Smc), Vols 1-6. 1680-1686, 3725.

Conforti, D., F. Guerriero, and R. Guido (2008). Optimization models for radiotherapy patient scheduling. 4or-a Quarterly Journal of Operations Research. 6(3), 263-278.

Conforti, D., F. Guerriero, and R. Guido (2010). Non-block scheduling with priority for radiotherapy treatments. European Journal of Operational Research. 201(1), 289-296.

Denton, B. and D. Gupta (2003). A sequential bounding approach for optimal appointment scheduling. Iie Transactions. 35(11), 1003-1016.

Denton, B., J. Viapiano, and A. Vogl (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. Health Care Management Science. 10(1), 13-24.

Donnan, G.A., et al. (2008). Stroke. Lancet 371 (9624), 1612-1623.

Erdelyi, A. and H. Topaloglu (2009). Computing protection level policies for dynamic capacity allocation problems by using stochastic approximation methods. IIE Transactions. 41(6), 498 - 510.

Erdelyi, A. and H. Topaloglu (2009). Computing protection level policies for dynamic capacity allocation problems by using stochastic approximation methods. Iie Transactions. 41(6), 498-510.

Gans, N., G. Koole, and A. Mandelbaum (2003). Telephone call centers: Tutorial, review, and research prospects. Manufacturing Service Operation Management. 5, 79-141.

Geng, N., et al.(2009a) MRI reservation for neurovascular patients, in Proceedings of the fifth annual IEEE international conference on Automation science and engineering. IEEE Press: Bangalore, India, 391-396

Geng, N., et al.(2009b) Time slots allocation and real-time control of MRI examinations of stroke patients, in 35th Conference on Operational Research Applied to Health Services: Leuven, Belgium.

Geng, N., Xie, X.,and Jiang, Z. (2009c). Contracting MRI time slots and advance cancellation for stroke patients. IEEE Transactions on Automation and Science Engineering (submitted).

Geng, N., et al. (2010a). A Monte Carlo Optimization and Dynamic Programming Approach for Managing MRI Examinations of Stroke Patients. IEEE Transactions on Automatic Control(Conditionally accepted).

Geng, N., et al.(2010b) Planning and Advance Cancellation of MRI Examinations of Stroke Patients, in 2010 IEEE Workshop on Health Care Management: Venice, Italy.

Geraghty, M. and E. Johnson (1997). Revenue management saves national car rental. Interfaces. 27, 107-207.

Gerchak, Y., D. Gupta, and M. Henig (1996). Reservation planning for elective surgery under uncertain demand for emergency surgery. Management Science. 42(3), 321-334.

Ghouila-Houri, A. (1962). Caractérisation des matrices totalement unimodulaires. Comptes

Rendus Hebdomadaires des Séances de l'Académie des Science (Paris). 254, 1192-1194.

Green, L.V. and S. Savin (2008). Reducing Delays for Medical Appointments: A Queueing Approach. Operations Research. 56(6), 1526-1538.

Green, L.V., S. Savin, and B. Wang (2006). Managing patient service in a diagnostic medical facility. Operations Research. 54(1), 11-25.

Guinet, A. and S. Chaabane (2003). Operating theatre planning. International Journal of Production Economics. 85(1), 69-81.

Gunes, E.D. and H. Yaman (2010). Health network mergers and hospital re-planning. Journal of the Operational Research Society. 61(2), 275-283.

Gupta, D. and B. Denton (2008). Appointment scheduling in health care: Challenges and opportunities. Iie Transactions. 40(9), 800-819.

Gupta, D. and L. Wang (2008). Revenue management for a primary-care clinic in the presence of patient choice. Operations Research. 56(3), 576-592.

Haerian, L., T. Homem-De-Mello, and C.A. Mount-Campbell (2006). Modeling revenue yield of reservation systems that use nested capacity protection strategies. International Journal of Production Economics. 104(2), 340-353.

Harper, P.R. (2002). A framework for operational modelling of hospital resources. Health Care Management Science. 5, 165-173.

Harper, P.R. and H.M. Gamlin (2003). Reduced outpatient waiting times with improved appointment scheduling: a simulation modelling approach. Or Spectrum. 25(2), 207-222.

Ho, C.-J. and H.-S. Lau (1992). Minimizing total cost in scheduling outpatient appointments. Management Science 38(12), 1750 - 1764

Ho, C.J. and H.S. Lau (1999). Evaluating the impact of operating conditions on the performance of appointment scheduling rules in service systems. European Journal of Operational Research. 112(3), 542-553.

Hsu, V.N., R. de Matta, and C.Y. Lee (2003). Scheduling patients in an ambulatory surgical center. Naval Research Logistics. 50(3), 218-238.

Jun, J.B., S.H. Jacobson, and J.R. Swisher (1999). Application of discrete-event simulation in health care clinics: A survey. Journal of the Operational Research Society. 50(2), 109-123.

Kidwell, C.S. and S. Warach (2003). Acute ischemic cerebrovascular syndrome: diagnostic criteria. Stroke 34(12), 2995–2998.

Kim, S.C., et al. (1999). Analysis of capacity management of the intensive care unit in a hospital. European Journal of Operational Research. 115(1), 36-46.

Kim, S.C., et al. (2000). Flexible bed allocation and performance in the intensive care unit. Journal of Operations Management. 18(4), 427-443.

Kolish, R., and Sickinger, S. (2008). Providing radiology health care services to stochastic demand of different customer classes. OR Spectrum. 30(2), 375-395.

Koole, G. (1998). Structural results for the control of queueing systems using event-based dynamic programming. Queueing Systems. 30(3-4), 323-339.

Lamiri, M., et al. (2008a). A stochastic model for operating room planning with elective and emergency surgery demand. European Journal of Operational Research. 185(3), 1026-1037.

Lamiri, M., V. Augusto, and X.L. Xie (2008b). Patients Scheduling in a Hospital Operating Theatre. 2008 Ieee International Conference on Automation Science and Engineering, Vols 1 and 2. 627-632, 1030.

Lee, D.K.K. and S.A. Zenios (2009). Optimal Capacity Overbooking for the Regular Treatment of Chronic Conditions. Operations Research. 57(4), 852-865.

Lev, B., et al. (1976). Patient flow analysis and the delivery of radiology service. Socio-Economic Planning Sciences. 10, 159-166.

Liberman, V. and U. Yechiali (1977). Hotel Overbooking Problem - Inventory System with Stochastic Cancellations. Advances in Applied Probability. 9(2), 220-220.

Magerlein, J.M. and J.B. Martin (1978). Surgical demand scheduling: A review. Health Service Research. 13, 418-433.

Marcon, E., S. Kharraja, and G. Simonnet (2003). The operating theatre planning by follow-up of the risk of no realization. International Journal of Production Economics. 85, 83-90.

Min, D.K. and Y. Yih (2010). An elective surgery scheduling problem considering patient priority. Computers & Operations Research. 37(6), 1091-1099.

Mondschein, S.V. and G.Y. Weintraub (2003). Appointment policies in service operations: A critical analysis of the economic framework. Production and Operations Management. 12(2), 266-286.

Muthuraman, K. and M. Lawley (2008). A stochastic overbooking model for outpatient clinical scheduling with no-shows. Iie Transactions. 40(9), 820-837.

Niu, Q., et al.(2007) Performance analysis of the operating room using simulation, in Proceeding of the 2007 CDEN and CCEE Conference.

Olivares, M., C. Terwiesch, and L. Cassorla (2008). Structural estimation of the newsvendor model: An application to reserving operating room time. Management Science. 54(1), 41-55.

Ormeci, E.L. and J. van der Wal (2006). Admission policies for a two class loss system with general interarrival times. Stochastic Models. 22(1), 37-53.

Patrick, J. and M.L. Puterman (2007). Improving resource utilization for diagnostic services through flexible inpatient scheduling: A method for improving resource utilization. Journal of the Operational Research Society. 58(2), 235-245.

Patrick, J., M.L. Puterman, and M. Queyranne (2008). Dynamic Multipriority Patient Scheduling for a Diagnostic Resource. Operations Research. 56(6), 1507-1525.

Perdomo, V., V. Augusto, and X.L. Xie (2006). Operating theatre scheduling using Lagrangian relaxation. 2006 International Conference on Service Systems and Service Management, Vols 1 and 2, Proceedings. 1234-1239, 1659.

Perros, H.G. and K.M. Elsayed (1996). Call admission control schemes: A review. Ieee Communications Magazine. 34(11), 82-91.

Persson, M. and J. Persson (2009). Analysing management policies for operating room planning using simulation. Health Care Management Science.

Persson, M. and J.A. Persson(2007) Analysing management policies for operating room planning using simulation. Working paper, Blekinge Institute of Technology, Sweden.

Puterman, M.L. (1994). Markov Decision Processes - Discrete Stochastic Dynamic Programming. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistic Section: A wiley-Interscience Publication.

Ridge, J.C., et al. (1998). Capacity planning for intensive care units. European Journal of Operational Research. 105(2), 346-355.

Robinson, L.W. and R.R. Chen (2003). Scheduling doctors' appointments: optimal and empirically-based heuristic policies. Iie Transactions. 35(3), 295-307.

Rohleder, T.R. and K.J. Klassen (2002). Rolling horizon appointment scheduling: A simulation study. Health Care Management Science. 5(3), 201-209.

Ross, K.W. and D.H.K. Tsang (1989). The Stochastic Knapsack-Problem. Ieee Transactions on Communications. 37(7), 740-747.

Schutz, H. and R. Kolisch(2009) Capacity allocation for magnetic resonance imaging scanners, in Working paper.

Shmueli, A., C.L. Sprung, and E.H. Kaplan (2003). Optimizing Admissions to an Intensive Care Unit. Health Care Management Science. 6(3), 131-136.

Shumsky, R.A. and F.Q. Zhang (2009). Dynamic Capacity Management with Substitution. Operations Research. 57(3), 671-684.

Sickinger, S. and R. Kolisch(2008) A generalized bailey-welch rule and simple tabu search procedure for outpatient appointment scheduling, in Technical Report. TUM Business School, Technishce Universitat Munchen.

Thompson, S., et al. (2009). Efficient Short-Term Allocation and Reallocation of Patients to Floors of a Hospital During Demand Surges. Operations Research. 57(2), 261-273.

Topkis, D. (1979). Equilibrium Points in Nonzero-sum n-person Submodular Games. SIAM Journal of Control Optimization 17(6), 773-787.

Vanden Bosch, P.M. and D.C. Dietz (2000). Minimizing expected waiting in a medical appointment system. Iie Transactions. 32(9), 841-848.

Vanden Bosch, P.M., D.C. Dietz, and J.R. Simeoni (1999). Scheduling customer arrivals to a stochastic service system. Naval Research Logistics. 46(5), 549-559.

Vasanawala, S.S. and T.S. Desser (2005). Accommodation of requests for emergency US and CT: Applications of queueing theory to scheduling of urgent studies. Radiology. 235(1), 244-249.

Vissers, J.M.H.(1994) Patient Flow based Allocation of Hospital Resources. Thesis, Eindhoven University of Technology, Eindhoven.

Vissers, J.M.H. (1998). Patient flow-based allocation of inpatient resources: A case study. European Journal of Operational Research. 105(2), 356-370.

Walter, S.D. (1973). A comparison of appointment schedules in a hospital radiology department. British Journal of Preventive and Social Medicine. 27, 160-167.

Wang, D. and J.P. Xu (2008). A Fuzzy Multi-objective Optimizing Scheduling for Operation Room in Hospital. Ieem: 2008 International Conference on Industrial Engineering and Engineering Management, Vols 1-3. 614-618,2200.

Wijewickrama, A. and S. Takakuwa (2005). Simulation analysis of appointment scheduling in an outpatient department of internal medicine. Proceedings of the 2005 Winter Simulation Conference, Vols 1-4. 2264-2273, 2717.

Wijewickrama, A.K.A. (2006). Simulation analysis for reducing queues in mixed-patients' outpatient department. International journal of simulation modelling. 5(2), 56-68.

Zhang, B., et al. (2009). A mixed integer programming approach for allocating operating room capacity. Journal of the Operational Research Society. 60(5), 663-673.

List of publications

Health care:

1. Na GENG, Vincent AUGUSTO, Xiaolan XIE, Zhibin JIANG, MRI Reservation for Neurovascular Patients, 5th Annual IEEE Conference on Automation Science and Engineering (CASE 2009), 391-396.

2. Na GENG, Vincent AUGUSTO, Xiaolan XIE, Zhibin JIANG, Time slots allocation and real-time control of MRI examination of stroke patients, Operational Research Applied to Health Services 2009.

3. Na GENG, Vincent AUGUSTO, Xiaolan XIE, Zhibin JIANG, Planning and advance cancellation of MRI examinations of stroke patients, IEEE Workshop on Health Care Management, Venice, Italy, 2010.

4 Na GENG, Xiaolan XIE, Vincent AUGUSTO, Zhibin JIANG, A Monte Carlo optimization and dynamic programming approach for managing MRI examinations of stroke patients. IEEE Transactions on Automatic Control (Conditional acceptance), 2010.

5 Na GENG, Xiaolan XIE, and Zhibin JIANG, Implementation of MRI Examination Reservation Strategies for Stroke Patients, IEEE Conference on Automation Science and Engineering, (acceptance), 2010.

6 Na GENG, Xiaolan XIE, and Zhibin JIANG, Contracting MRI time slots and advance cancellation for stroke patients. IEEE Transactions on Automation and Science Engineering (submission), 2009.

Semiconductor wafer fabrication:

7. Na Geng, Zhibin Jiang and Feng Chen, Stochastic programming based capacity planning for semiconductor wafer fab with uncertain demand and capacity, European Journal of Operational Research, Volume 198, Issue 3, 1 November 2009, Pages 899-908.

8. Na Geng and Zhibin Jiang, A review on strategic capacity planning for semiconductor manufacturing industry, International Journal of Production Research, Volume 47, Issue 13 January 2009 , pages 3639 - 3655.

9. Na Geng and Zhibin Jiang, Stochastic-programming-based capacity planning for wafer fabrication, Journal of Shanghai Jiao Tong University, 40(11), 2006: 1935-1938 (EI No.:070310372749) (in Chinese);

10. Na Geng and Zhibin Jiang, The capacity planning for wafer fabrication based on an improved stochastic-programming, Journal of Shanghai Jiao Tong University, 41(5), 2007: 825-829. (EI No.: 072710691929) (In Chinese);

11. Na Geng and Zhibin Jiang, Capacity planning for semiconductor wafer fabrication with uncertain demand and capacity, Proceedings of the 3rd Annual IEEE Conference on Automation Science and Engineering, Scottsdale, AZ, USA, Sept 22-25, 2007: P100-105;

12. Na Geng and Zhibin Jiang, A review on medium term capacity planning for semiconductor wafer fabrication, IEEE International Conference on Systems, Man and Cybernetics, Montreal, QC, Canada, 7-10 Oct. 2007: 680-684.

École Nationale Supérieure des Mines
de Saint-Étienne

N° d'ordre : 2010 EMSE 0572

耿 娜

基于组合优化与马尔科夫决策支持的核磁共振检查计划研究

专业：工业工程

关键词：计划，核磁共振（MRI）检查，合同，提前取消，马尔科夫决策支持，随机规划，最优策略

摘要：

本研究是在与一家法国医院合作中受到启发，目的是减少在神经内科进行治疗的中风病人的住院时间。中风病人需要得到及时的检查、诊断与治疗，这些检查又依赖于昂贵且使用量极高的图像仪器，如核磁共振仪（MRI）。因此，对神经内科来说，通过减少病人等待检查的时间来减少病人的住院时间是非常重要的。

为了在不降低MRI利用率的前提下，减少中风病人等待MRI检查的时间，本文从神经内科着手，提出了一个新的MRI检查预约过程。为保证中风病人尽快进行检查，图像科每周预留一部分合同时间槽（CTS）给中风病人；除了这部分CTS外，中风病人也可以通过预约常规的时间槽（RTS）进行检查。

本文首先提出了一个随机规划模型来同时确定合同决策（即CTS的数量及其在时间轴上的分布）和病人分派策略（即指定病人等待CTS或RTS），目的是在病人等待时间与闲置的CTS数量之间达到最好的权衡。为了求解该模型，首先，在给定合同策略的前提下，用平均成本马尔科夫决策支持（MDP）的方法对最优控制策略的结构性质进行了研究和证明。然后通过蒙特卡罗模拟和局部优化确定合同决策。试验结果发现所提的算法能很有效地求解模型。新的预约过程能大大减少病人等待时间，但也会存在一定数量的闲置CTS。

为了减少闲置的CTS，本文进一步研究了CTS提前一天取消和提前两天取消的控制策略。用MDP方法对最优提前取消策略进行了研究。数值试验发现，考虑CTS的提前取消策略能大大减少目标值和CTS的闲置率。

École Nationale Supérieure des Mines
de Saint-Étienne

N° d'ordre : 2010 EMSE 0572

Na GENG

Combinatorial optimization and Markov decision process for planning MRI examinations

Specialty: Industrial Engineering

Keywords: Planning, MRI exams (Magnetic Resonance Imaging), Contract, Advance cancellation, Markov decision process, Stochastic programming, Optimal strategies

Abstract:

This research is motivated by our collaborations with a large French university teaching hospital in order to reduce the Length of Stay (LoS) of stroke patients treated in the neurovascular department. Quick diagnosis is critical for stroke patients but relies on expensive and heavily used imaging facilities such as MRI (Magnetic Resonance Imaging) scanners. Therefore, it is very important for the neurovascular department to reduce the patient LoS by reducing their waiting time of imaging examinations.

From the neurovascular department perspective, this thesis proposes a new MRI examinations reservation process in order to reduce patient waiting times without degrading the utilization of MRI. The service provider, i.e., the imaging department, reserves each week a certain number of appropriately distributed contracted time slots (CTS) for the neurovascular department to ensure quick MRI examination of stroke patients. In addition to CTS, it is still possible for stroke patients to get MRI time slots through regular reservation (RTS).

This thesis first proposes a stochastic programming model to simultaneously determine the contract decision, i.e., the number of CTS and its distribution, and the patient assignment policy to assign patients to either CTS or RTS. To solve this problem, structure properties of the optimal patient assignment policy for a given contract are proved by an average cost Markov decision process (MDP) approach. The contract is determined by a Monte Carlo approximation approach and then improved by local search. Computational experiments show that the proposed algorithms can efficiently solve the model. The new reservation process greatly reduces the average waiting time of stroke patients. At the same time, some CTS cannot be used for the lack of patients.

To reduce the unused CTS, we further explore the possibility of the advance cancellation of CTS. Structure properties of optimal control policies for one-day and two-day advance cancellation are established separately via an average-cost MDP approach with appropriate modeling and advanced convexity concepts used in control of queueing systems. Computational experiments show that appropriate advance cancellations of CTS greatly reduce the unused CTS with nearly the same waiting times.

École Nationale Supérieure des Mines
de Saint-Étienne

N° d'ordre : 2010 EMSE 0572

Na GENG

Planification des examens IRM à l'aide de processus de décision markovien et optimisation combinatoire

Spécialité: Génie Industriel

Mots clefs: Planification, examens IRM, Contrat, Annulation, Processus de décision markovien, Programmation stochastique, politiques optimales

Résumé :

Cette thèse propose un nouveau processus de réservation d'examens IRM (Imagerie par Résonance Magnétique) afin de réduire les temps d'attente d'examens d'imagerie des patients atteint d'un AVC (Accident Vasculaire Cérébral) soignés dans une unité neurovasculaire. Le service d'imagerie réserve chaque semaine pour l'unité neurovasculaire un nombre donné de créneaux d'examens IRM appelés CTS afin d'assurer un diagnostic rapide aux patients. L'unité neurovasculaire garde la possibilité de réservations régulières appelées RTS pour pallier les variations des flux de patients.

Nous donnons d'abord une formulation mathématique du problème d'optimisation pour déterminer le nombre et la répartition des créneaux CTS appelée contrat et une politique d'affectation des patients entre les créneaux CTS ou les réservations RTS. L'objectif est de trouver le meilleur compromis entre le délai d'examens et le nombre de créneaux CTS non utilisés. Pour un contrat donné, nous avons mis en évidence les propriétés et la forme des politiques d'affectation optimales à l'aide d'une approche de processus de décision markovien à coût moyen et coût actualisé. Le contrat est ensuite déterminé par une approche d'approximation Monté Carlo et amélioré par des recherches locales. Les expérimentations numériques montrent que la nouvelle méthode de réservation permet de réduire de manière importante les délais d'examens au prix des créneaux inutilisés.

Afin de réduire le nombre de CTS inutilisé, nous explorons ensuite la possibilité d'annuler des créneaux CTS un ou deux jours en avance. Une approche de processus de décision markovien est de nouveau utilisée pour prouver les propriétés et la forme de la politique optimale d'annulation. Les expérimentations numériques montrent que l'annulation avancée des créneaux CTS permet de réduire de manière importante les créneaux CTS inutilisés avec une augmentation légère des délais d'attente.